The Institute for Earth and Environmental Software (ISEES): A Strategic Plan

Version 1

March 1, 2015

Steering Committee:

Matthew B. Jones <jones@nceas.ucsb.edu> (Principal Investigator)
Lee Allison <lee.allison@azgs.az.gov>
Dan Ames <dan.ames@byu.edu>
Bruce Caron <bruce@tnms.org>
Scott Collins <scollins@sevilleta.unm.edu>
Patricia Cruse <cruse.patricia@gmail.com>
Peter Fox <pfox@cs.rpi.edu>
Stephanie Hampton <s.hampton@wsu.edu>
Chris Mattmann <chris.a.mattmann@jpl.nasa.gov>
Carol Meyer <carolbmeyer@esipfed.org>
William Michener <william.michener@gmail.com>
Erin Robinson <erinrobinson@esipfed.org>
Mark Schildhauer <schild@nceas.ucsb.edu>

This material is based upon work supported by the National Science Foundation under Grant No. OCI-1216894.

Executive Summary

The need for deeply informed understanding and stewarding of the earth system has never been greater, in the face of rapid climate change and increasing human impacts on the land, sea, and atmosphere. Fortunately, the potential for science to meet these challenges also has never been greater, due to recent *technological advances* affording researchers access to an unprecedented flood of observational earth science data, from local to global scales, and across a multitude of thematic and spectral dimensions. In parallel, exponential increases in computational power, storage, and bandwidth are enabling ever more powerful modeling and analyses of those data, to accelerate discovery and insights into complex earth science phenomena. These technology advances have catalyzed fundamental changes in the earth system researchers' analytical program over the past several decades, placing us in the "Fourth Paradigm" of scientific research, "eScience"-- driven by a merging of *computational methods* with *data science*, carried out collaboratively by researchers communicating over high-speed networks, and working across distributed data and analytical resources. The lone, localized "desktop" researcher is rapidly becoming a vestige of the past.

Ironically, as the earth and environmental sciences become increasingly reliant on large-scale, Big Data-driven, computationally demanding, multidisciplinary, and integrative methods-- aside from a small elite cadre, much of the research community is falling behind in their ability to capitalize on the possibilities of eScience. We believe the core problem arises from a disconnect beween the science community and the software and data engineering communities that leads to inefficiencies in how critical technology services are used by our nation's researchers. We outline a strategic plan for the Institute for Sustainable Earth and Environmental Software (ISEES) in which we develop new models for how scientific software can be better developed, adopted, and supported on behalf of our nation's researchers.

ISEES' mission is to make fundamental advances in software and data science and education that accelerate and transform earth and environmental science.

To accomplish this mission, ISEES will pursue four main goals. First, transform the culture of environmental science to embrace software and data science. Second, enable new partnerships between environmental scientists, data scientists, and software engineers that are mutually beneficial. Third, stimulate and support community-driven software advances that benefit environmental science. And fourth, transform education in the US environmental science community to include data and software development techniques as core components of earth and environmental science disciplines.

ISEES will operate as virtual organization with bi-coastal meeting centers that will meet these goals via five principal strategic areas of activity:

- Strategy 1: Stimulate collaboration in environmental science and research software engineering

    - Activity: Support Data Science Partnerships
- Strategy 2: Coordinate and provide education and training
    - Activity: Coordination of training activities across partners
    - Activity: Open Science for Synthesis Training
- Strategy 3: Provide Support, Infrastructure and Consulting Services
    - Activity: Software support and consulting
    - Activity: Science software infrastructure
- Strategy 4: Build a vibrant science and software community
    - Activity: Coordination of science/software communities
- Strategy 5: Create a sustainable organization
    - Activity: Operate under a Participatory Governance regime
    - Activity: Build diverse funding streams

Table of Contents

Software and data challenges in environmental science

Challenged with pressing questions about the impacts of global climate change on environmental systems, effects of biodiversity loss on ecosystem service provisioning, implications of soil carbon sequestration for food security, and other issues requiring improved understanding of both natural and human systems, the earth and environmental science community increasingly recognizes a dire need for integrated, transdisciplinary approaches to science. This applies not only to the research process itself, but also to the computational infrastructure that supports it. Particularly in light of the expanding role of sophisticated quantitative methods, applied to ever-increasing amounts of data over a remarkable range of temporal and spatial scales, it is clear that continued progress toward meeting these scientific challenges will depend upon the availability of robust scientific software components and data that can interoperate as never before.

Unfortunately, science-relevant software development efforts frequently suffer from ad hoc development processes, use of stove-piped and proprietary systems, undesirable code complexity and opacity, lack of software testing, lack of scalability, lack of openness and interoperability, and lack of formal versioning and management of software evolution and maintenance. This prevents the research community from realizing the full benefits that modern scientific software could offer. Overcoming these challenges will form the core mission for the Institute for Sustainable Earth and Environmental Software. ISEES will work to integrate, mature, and sustain software used throughout the scientific lifecycle, from initial data acquisition in observational systems, to data management, quality assurance, data integration, analysis, and synthetic modeling.

The ISEES Vision

Technological advances are creating outstanding new opportunities for investigating challenging science research questions at levels of detail and generality that were impossible in the past. Full realization of technological potential, however, requires that the scientific community far more effectively develop, reuse, extend, and share software and analyses than is currently the case.

We envision an Institute for Sustainable Earth and Environmental Software (ISEES) that coordinates development and sustainable support of innovative and interoperable scientific software tools that can transform science at the intersection of earth, environmental, and life sciences. ISEES will advance the state of science software by engaging earth and environmental research communities to address the software barriers that most impede grand challenge earth science. At ISEES, researchers will collaboratively address the entire software lifecycle, from product conceptualization, to requirements analysis, design, development, testing, deployment, long-term support, and decommissioning. A robust workforce development program will sustain these software advances made through ISEES.

ISEES will address many of the barriers to effective software use in earth and environmental science, including:

- Isolated development
    - The development of science software typically is done in isolation, within individual investigator labs and coded for single use.
- Need for training
    - Most scientists are not trained in software development, but need to be.
- Code complexity
    - Scientific codes can become massively complex, and would benefit from the use of common design patterns from the software world
- Lack of formal development processes (e.g., testing, code review, version control)

Results from multiple community workshops indicate that fundamental change in all of these areas would significantly increase the pace of science. Notable advances in sharing and preserving scientific research data should now be accompanied by similar attention to the codes, programs, models, and applications that we use to collect, organize, simulate, and interpret scientific phenomena. While data provide the evidence for understanding natural phenomena, it is the software used to analyze and interpret the data that enables scientific insight.

This strategic plan represents the collective vision of a coalition of computational earth system and environmental scientists, working together to identify and propose solutions to the critical software challenges that currently impede more rapid, robust, relevant, and reproducible scientific insights into the Earth System. Software usage patterns today largely reflect familiarity, affordability, (potentially outdated) disciplinary traditions, and ease-of-use, rather than the more pertinent dimensions such as portability, scalability, reusability, and robustness of output, that are needed for Fourth Paradigm science. It is imperative that traditional approaches in the use and development scientific software must undergo serious review and potentially dramatic revision, to more fully benefit from the opportunities that technological advances are providing.

The vision for an improved software ecosystem that we put forth for our nation's earth science researchers prescribes a balance between pure and applied considerations, and between general and specialized software tools and applications, accompanied by strong commitments to support and training programs in those tools and applications. Our vision is informed by the requirements of a disparate community of researchers with varying levels of computational resources and skills. We propose mechanisms to enable rapid, iterative dialogue that can efficiently clarify and prioritize evolving domain user communities' software needs, and the engineering efforts required to develop and sustain the software solutions to resolve these needs.

The "Institute for Sustainable Earth and Environmental Software" (ISEES) will not be just another facility; rather it will be a set of services and capabilities, distributed and community-driven, informed by domain scientists working closely with computer scientists and software engineers, to collectively guide, advance, and sustain a more efficient and effective software ecosystem to facilitate scientific inquiry. While our focus is on the earth and environmental sciences, our proposed framework should have far broader implications for how scientific software can evolve and grow, to promote more effective, accelerated, reproducible eScience research in the future.

The ISEES Mission and Goals

ISEES' mission is to make fundamental advances in software and data science and education that accelerate and transform earth and environmental science.

To accomplish this mission, ISEES will pursue four main goals through the strategies and activities described in the remainder of this plan.

- Goal 1: Transform the culture of environmental science to embrace software and data science
- Goal 2: Enable new partnerships between environmental scientists, data scientists, and software engineers that are mutually beneficial
- Goal 3: Stimulate and support community-driven software advances that benefit environmental science
- Goal 4: Transform education in the US environmental science community to include data and software development techniques as core components of earth and environmental science disciplines

Strategic Planning Process

The community-driven ISEES conceptualization process engaged a broad interdisciplinary group consisting of domain scientists (e.g., environmental, earth, life, and social scientists), computer and informatics scientists, digital librarians, and experts in governance and sustainability (see Appendix C). Six working groups organized into 3 clusters were charged with exploring and proposing alternative designs for a software sustainability institute for the environmental and earth sciences (see Appendix A). The Science Cluster articulated grand challenges within earth observational sciences that will drive ISEES' software activities and defined exemplary collaborative science activities that supported detailed requirements analysis (see Appendix B). The Software Cluster analyzed requirements for scientific software and proposed approaches for ISEES to address these via improvements across the full science software lifecycle. The Sustainability and Adoption Cluster examined sustainability and governance challenges, and proposed models for engaging the research community, governing ISEES, and developing an effective workforce that can sustain the portfolio of science software. Community experts led each working group and collectively comprised a 12- person Steering Committee that synthesized recommendations, presented results, and gathered feedback at the Earth Science Information Partnership (ESIP) summer and winter meetings, as well as at the American Geophysical Union meeting in December 2014.

In addition to these planning activities, we conducted two intensive pilot activities to better understand the role that ISEES should play specifically in workforce development and community building. The first was our Open Science for Synthesis training, an experiment in training using virtual teleconferencing to conduct a 3-week, project-based, bi-coastal training on data science and software skills for environmental scientists (See Appendix A). We collaborated on OSS with the Water Science Software Institute, enabling us to train 45 scientists divided between locations at NCEAS and RENCI. Second, seeing the need for more integrated collaboration between science researchers and the research software community, we organized the Open Science Codefest (http://nceas.github.io/open-science-codefest/), attracting over 100 people from the environmental science and software communities to collaborate on research software to benefit science (see Appendix A). The huge success of these two pilot activities gave us valuable insights into the software training and community engagement needs of the earth and environmental science communities, and drove the strategic plan presented in this report.

Strategies to transform earth and environmental software

To realize its mission and goals, ISEES should be created to pursue five fundamental strategic objectives: Strategy 1, change the culture between science and software communities; Strategy 2) alter the education and training landscape; Strategy 3) provide support, consulting, and infrastructure services to the community; Strategy 4) build a vibrant and integrated science and software community; and, Strategy 5), create a sustainable organization. Each of these will be executed through one or more associated activity areas that can grow and expand over time as community needs evolve.

Strategy 1: Stimulate collaboration in environmental science and research software engineering

The advent of NSF-funded synthesis centers such as NCEAS, NESCent, and NIMBios brought a new emphasis on collaboration, data re-use, and interdisciplinary synthesis that has extensively impacted the environmental sciences (Hampton and Parker, 2011; Hackett et al. 2008). Similarly, the ISEES workshops have articulated the need for effective integration between environmental science and computing disciplines which is sorely lacking today (see Appendix A). Such collaboration is critical both to improve the efficiency and capabilities of environmental scientists and to advance the reproducibility and utility of scientific software tools. And yet, despite this well-recognized need, computing and science practitioners truly come from two distinct cultures that seldom meet, much less inform one another. Bridging this cultural divide will be among the primary goals of ISEES, and will be accomplished by providing opportunities for environmental scientists, software engineers, computer scientists, and cyberinfrastructure specialists to regularly collaborate on problems of mutual interest.

Activity: Support Data Science Partnerships

ISEES will fund collaborative working groups with the specific intent to convene interdisciplinary expertise needed to solve pressing scientific problems that are currently impeded by a lack of appropriate models, software tools, cyberinfrastructure, and data. The goal of these partnerships will be to:

- Enable the research community to work together and create one culture around scientific computing
- Prioritize and accelerate software development needed for specific science research problems that will have a major impact on society (see Appendix B)
- Support the full software lifecycle for software that the research community deems as critical for science

To accomplish this, ISEES will call for the broader research community to propose specific partnerships once or twice a year, and groups will be evaluated based on 1) the science and engineering impact of the proposed collaboration, 2) the degree of integration of the two cultures, and 3) the tractability of the proposed work within the confines of a working group model. Each group that is selected would be provided with resources for collaborative meetings over a two to three year period to work on the problem, software development support and consulting, computing infrastructure for software development, hardening, and execution, and training to enable effective use of these computing environments.

Expected products from the partnerships will include a set of science outcomes such as publications, as well as a set of software and infrastructure outcomes, such as newly integrated or hardened software products, data systems, or infrastructure services.

Strategy 2: Coordinate and provide education and training

Knowledge and skills needed by people that develop scientific software are diverse and decidedly interdisciplinary, spanning many skills in software and data engineering, as well as quantitative modeling, statistics, and analysis (see Table 1 in Appendix A). Individuals coming from computing backgrounds are best equipped to do environmental science software development when they have significant training in a relevant domain science. And earth and environmental science researchers would be much better equipped to create and mature scientific codes after training in software design, coding practices, and testing. In general areas for training span 1) computing, 2) mathematics and statistics, and 3) the natural and physical sciences. Some knowledge from each of these three areas is optimal for workers at any level, from basic to advanced, but required for the most advanced worker.

ISEES' goal in education and training is to transform education in the US environmental science community to include data and software development techniques as a part of the core curriculum taught to all researchers in the field. To accomplish this, training and education activities would include coordination across training programs, providing open science and synthesis training courses, and helping to change the nature of graduate programs in earth and environmental science.

Activity: Coordination of training activities across partners

A wide variety of training activities in software and data science are underway in the community, but have been uncoordinated and have yet to reach a critical mass. For example, NSF funded NCEAS, the Ecological Society of America, the Long-term Ecological Research Network (LTER), and NimBIOS, among others, to conduct training in data science techniques since the late 1990's through today. These have generally been short, small workshops ranging from a few hours to two weeks in duration, and generally reach 20 to 40 people at a time. In the last two years, these evolved at NCEAS into the 3-week Open Science for Synthesis course, which focused on data science techniques within a curriculum focused around real-world scientific synthesis projects. More recently, the Software Carpentry two-day workshop has been gaining tremendous momentum, and the Sloan and Moore funded data science institutes have announced new degree programs. Meanwhile, graduate departments around the country continue to offer courses in statistics, modeling, simulation, and data mining, and some have begun to link these into new degree programs. Because these efforts have been disconnected, they will benefit tremendously from coordination through ISEES. In this activity, ISEES will convene leaders from these various initiatives (starting with a workshop in April, 2015) to develop a shared vision of training content, online materials, and course progression, with the goal that participants will be able to participate in training from multiple institutions and be able to use courses from one organization to fulfill prerequisites from another.

Data and software in science has traditionally fallen between the cracks, and is not fully covered in domain science curricula, nor in computer science nor software engineering. ISEES will help coordinate curriculum efforts by providing infrastructure for curriculum development and sharing, provide opportunities for various degree and training programs to collaborate, and provide opportunities for multi-institutional, distributed data science seminars.

Activity: Open Science for Synthesis Training

Because graduate students already have extremely full schedules, ISEES will continue to provide the Open Science for Synthesis (OSS) training annually. OSS will fill a gap in training because it is more in-depth than the typical two day Software Carpentry workshop, but not as much of a time commitment as a semester long data science course, or a full data science degree. The OSS course is specifically designed as a survey course that builds upon the short introductory workshops offered by Software Carpentry, extending the material into more detailed treatments of engineering techniques for scientific software development, including design patterns, scientific workflows, and testing. These engineering skills are complemented by exposure to new techniques in analysis, modeling, and data mining that are intended to broaden knowledge of the quantitative techniques available to scientists. And finally, all of these skill-oriented course segments are glued together with 3-week long small group synthesis projects modeled after the synthesis working group model pioneered at NCEAS. This allows researchers to use their new data science skills in real-world science scenarios that typically results in both science publications and software products as outcomes.

Strategy 3: Provide Support, Infrastructure and Consulting Services

Many of the recommendations for services that were needed by the science community were for support and consulting services that are typically difficult to provide under a traditional research grant. Participants strongly recommended that ISEES provide these services on a fee-for-service basis with the intent of strengthening the science community.

Activity: Software support and consulting

ISSES should provide an evolving set of support and consulting services on a fee-for-service basis. These would include services in support of the development of scientific software, such as:

- Code porting service
- Code development and design service
- Code optimization services
- Code review prior to release/publication? (e.g. many researchers primarily want to avoid embarrassment in releasing code) (Institute's Stamp of approval?)
- Security consulting

Participants also envisioned support services that directly target end users of scientific software. Many scientific software projects are critical to science, but too small and targeted at particular academic communities to be viable to have their own support infrastructure, especially if the market is too small to support commercialization. For example, the AD Model Builder environmental modeling software has been adopted by NCEAS and the science community for support and maintenance, and hundreds of other packages could use similar support. Participants envisioned a strong role for ISEES to provide support for software identified by the community as critical to science. This support could come in several forms, including:

- Development and maintenance of online documentation
- Creation of tutorials and help systems
- Operation of shared support help-desk
- Coordination of community members that can provide support and help

Some of these services would be subsidized, and some would be provided on a fee basis through the sale of support contracts.

Activity: Science software infrastructure

In addition to support and consulting, infrastructure for scientific software was identified as a critical gap that ISEES could provide. For example, while it is well-recognized that scientific software should be published, there are few venues that are both accessible and focused on long-term preservation. On the one hand, scientists can use version control systems such as GitHub, but as commercial ventures, these do not have preservation as a core part of their mission. If past repositories such as Source Forge are an indicator, the community is quite fickle with respect to which of these systems receive support. In addition, version control systems are not tailored to work with the academic publishing and citation model. On the other hand, there are language-specific package repositories (e.g., CRAN, PyPI, Debian) that can be used for distributing scientific software packages. These, however, lack cross-language/cross-operating system support, and tend to be restricted to re-usable software, as opposed to capturing software as used in particular scientific projects and publications. Consequently, ISEES could and should play a role in coordinating among these various software archiving activities in order to provide interoperable infrastructure for depositoing and preserving scientific software. This would obviously need to be collaborative, working in conjunction with other groups such as Zenodo and the Mozilla Science Lab that are working towards solutions in these areas.

Strategy 4: Build a vibrant science and software community

ISEES success will be measured through the extent to which it succeeds in aligning the interests of the earth and environmental science community with the research software and data science communities. While these latter communities are both focused on serving science, they are frequently disconnected. ISEES' goal to substantially change the culture of these communities so that they fully embrace one another will be both critical and difficult.

Activity: Coordination of science/software communities

Many of the problems we face at the intersection of science and software are driven primarily by social and cultural forces, rather than by technology. ISEES should play a strong role in connecting these communities in a way that natural builds positive connections between researchers, increases understanding of the challenges faced by each, and promotes collaboration between interested groups. Towards that end, ISEES will support a variety of community building activities. For example, ISEES will continue to organize and support the Open Science Codefest (Appendix A), which was hugely successful in connecting science and software practitioners and building working relationships between them. In addition to specific activities such as OS Codefest, ISEES will also help promote community activities that are already underway, such as the ESIP Software cluster, the Mozilla Science Lab, the Research Data Alliance, Earth Cube, DataONE, and ROpenSci. These activities span the continuum from science to software to infrastructure work, and it is only through effectively connecting these communities and activities that we will see cultural change that values science, data, and software synergies.

Strategy 5: Create a sustainable organization

A fundamental challenge for ISEES is how to create a sustainable organization that can support the scientific software community and advance science. The funding model at NSF has not been conducive to sustainability, in that the typical 3-5 year grant cycle creates a highly fragmented and competitive funding landscape. Few efforts at NSF last beyond a decade, which is significantly due to the funding model. For ISEES to be sustainable, participants envisioned an institute that would be valuable for a diverse group of participants and stakeholders, including:

- Researchers from academia, government, and corporate sectors;
- Meta-organizations including professional societies, other networks and NGOs;
- Institutions that support scientific software development such as academic libraries and academic computing;
- Funding sources from public, corporate sector and foundations;
- Data repositories and networks that hold much of the science output; and
- Software community including developers, users, and code repositories

Consensus during our strategic planning was that sustainability for ISEES would be tied to 1) the value that ISEES brings to the science community, and 2) a diverse funding model that supports many kinds of activities.

Activity: Operate under a Participatory Governance regime

To provide value for the science and software communities, ISEES must be responsive to community needs. And therefore, it must be governed by the community itself. We envision a participatory governance structure in which a non-profit ISEES is governed by a community-led Board of Directors and executed by a Director and staff that are located at university partner organizations. Although the governance model is not yet finalized, we envision several key features.

First, ISEES would be operated by a Board of Directors drawn from our stakeholder community. The Board would be responsible for all activities of the institute, would control its funding, and would delegate operations to the ISEES Director. The Board would be actively engaged in fundraising for ISEES, and would assist the DIrector in preparing proposals and reviewing the fiscal operations of the Institute. The Board would conduct or commission annual audits to ensure that ISEES is operating to the benefit of the science software community.

Second, ISEES would be a collaborative, virtual organization operated by a partnership between a non-profit organization and university partners. Currently, we imagine that the Director for the institute would be from one of the participating universities, but that the governance would be operated by the Foundation for Earth Science, a non-profit organization that currently supports the Earth Science Information Partners (ESIP). Operating as part of the Foundation would allow ISEES to utilize a diverse array of funding streams and minimize adminsitrative overhead while staying clearly focused on the needs of the community. The Foundation would also support any staff positions that are necessary and that are not directly affiliated with a university location.

Third, day to day operations of ISEES would be run by a Director located at one of the university partners or at the Foundation for Earth Science. The Director would be responsible for overseeing and implementing ISEES' activities and would report to its board.

Fourth, although ISEES is a virtual organization, we see a need for a physical location to host meetings, training events, and the envisioned Data Science Partnerships. Initially we envision ISEES as having a location at UC Santa Barbara, co-located with the complementary synthesis center at NCEAS. We also have been developing a collaboration with the Renaissance Computing Institute at the University of North Carolina, as we imagine that having the ability to meet at bi-coastal locations will best serve the community.

Finally, the Institute will convene a 12 member Advisory Board that will participate in strategic recommendations about the activities of the Institute. This advisory board will operate on a two year cycle, with half of the members rotating each year. It will consist of individuals from both the science and research software communities with expertise that can be applied to focusing and selecting Institute activities that will maximize its value to the community. For example, the Advisory Board will be tasked with reviewing and recommending activities to support under the Data Science Partnerships program, as well as prioritizing and recommending software systems to be supported under the ISEES support and consulting services program.

Activity: Build diverse funding streams

Diverse funding will be key to the success of ISEES, as the community changes stimulated by ISEES will require shepherding over many years. We envision funding ISEES in multiple phases. Startup funding would be secured from the NSF S2I2 program as solicitations become available, and this would support the creation of the institute and its operation during the first 3-5 years.

During this period, ISEES Board and Director would seek additional funding from a diverse group of foundations and institutions. We see the strong potential for funding from private foundations that have a vested interest in open science principles and practice, as well as foundations focused on the earth and the environment.

In addition, during the initial startup phase, we will add fee-based services in serveral areas as described above. These would potentially include tuition for training courses (but with a scholarship model to support students in need), and fees for both software and systems support and for software consulting services.

As ISEES matures, we expect funding to transition from its initial heavy reliance on one or a few grants and contracts, to a sustainable model where fee-based operations operate continuously and are supplemented by smaller grants and contracts for advances in particular areas of research and development.

Synopsis and Expected Impact of ISEES

Recent advances in scientific software have revolutionized discovery and invention throughout the sciences. However there are many missed opportunities for advance due to a lack of coordination and failure to follow best practices in scientific software development [Wilson et al.](http://doi.org/doi:10.1371/journal.pbio.1001745). Efforts such as the [Science Code Manifesto](http://sciencecodemanifesto.org/) show that software is widely reconized as a cornerstone for science, and yet most scientists rarely consider software issues in their daily practices. We believe that open software is critical to the transparency and reproducibility of science; that open data and open science efforts require open software; that engaging scientific societies and user groups throughout the software lifecycle is critical to scientific advance; that software reuse is relatively rare but should not be; that sustainable scientific software requires investment from diverse sectors; and that research software organizations that effectively coordinate and focus scientific software must be community driven via participatory governance models. WIth these principles in mind, we have envisioned an Institute for Sustainable Earth and Environmental Software that will significantly accelerate science by supporting a strong partnership between the research engineering and science communities.

Appendix A: The ISEES Strategic Planning process

The plan for ISEES was developed over the course of six strategic planning workshops and two pilot activities that jointly involved over 198 participants from diverse disciplines (Appendix C). We describe each of these planning activities to provide more detail about the genesis of the ideas for ISEES.

Science Drivers

A principal goal of the ISEES design process was to identify key areas where sustainable software could significantly improve the conduct and pace of science. More than two-dozen scientists representing the breadth of Earth and environmental science domains participated in a two-day workshop at the National Center for Ecological Analysis and Synthesis. The express goals of the workshop were to: (1) propose a range of multi- and interdisciplinary "grand challenge" Earth and environmental science questions, the answers to which could meaningfully impact science and society; (2) provide examples of the types of data and software needed to address a diverse subset of the questions by analyzing workflows that could be used to derive answers; and (3) prioritize the functions of a software sustainability institute that could accelerate interdisciplinary research in the Earth and environmental sciences. Appendix B outlines the science questions that drive the vision and mission of ISEES.

ESIP Town Hall

Early in the planning process we held a town-hall meeting at the Earth Science Information Partners summer meeting in 2013. We convened leaders from the ISEES initiative, as well as from the Water Science Software Institute and the Empowering Long Tail Research (IELTR) project to present ideas about the need for an institute to the earth sciences community. A panel discussion with many members of the ESIP Federation stimulated a wide variety of of the community building and engagement ideas presented in this plan, and led to specific objectives to coordinate ISEES activities with ESIP and related organizations.

Software Lifecycle

The software lifecycle workshop convened community-wide experts in software development and software frameworks to examine the role of a software institute. Workshop participants recognized academic, open source, and commercial models as three distinct scientific software lifecycle models that have distinct characteristics but also have linkages among them. Academic software was characterized as arising from university, agency, and NGO research groups, often reaching a prototype level of functionality, and is often developed by lone individuals or small groups with science background but often little formal training in software engineering. Open source scientific software is developed collaboratively by communities that share an interest in a science software platform, and often grows out of successful academic software projects. Finally, commercial scientific software provides large-scale frameworks for computation and data management, frequently focusing on disciplines with extensive resources, and at times used in science as a side effect of its development for more voluminous consumer markets. Each of these models was considered to bring strengths to the development of scientific software, but there was consensus that the open source model had the most to offer to reproducible, and efficient scientific software community.

Workshop participants brainstormed, refined, and prioritized the functions that should be performed by an institute. These included four main categories of functions. 1) Community building functions, meant to fuse the science and software research communities into a better functioning partnership. 2) Training and Advocacy functions focused on promulgating needed skills in software and data science into the earth and environmental sciences, as well as advocating for and sustaining open source software that is widely useful to the research community. 3) Consulting Services focused on assisting science groups in software processes, including software design and architecture, product hardening, maintenance, and preservation, as well as consulting on licensing, testing, and cost modeling. 4) Infrastructural Services to support the software lifecycle within the earth and environmental sciences, including software discovery services, a software review and certification program, and the provision of software use and quality metrics.

Software Components

Workforce Development

There was strong consensus in the workforce development workshop that raising the basic skill level of domain scientists, broadly, is a critical need. Areas of emphasis include training earth and environmental scientists in software and data engineering practices, and engineers in science practices (Table 1). The group recommended that this is where ISEES should focus its efforts. ISEES should also strategically support innovators who are working at intermediate levels and help to advance their skills to expert levels.

Table 1: Knowledge and skills needed in the workforce, in order to advance environmental science through data-enabled research approaches and to create software that is sustainable within this field.

Basic	Intermediate	Advanced
data documentation	version control	intellectual property
command line programming	data enabled science	scalability of computation
how to cite software	data archiving	model interoperability
directories and files	basic command line programming	numerical analysis
units and dimensional analysis	data structures	parallelization - code, hardware
how to archive data	diversity of algorithms	numerical stability

computational literacy	uncertainty, analysis/model assumptions / error	verification (code)
visual literacy	spatial analysis	cloud computing
units and dimension analysis	exploratory data analysis	object-oriented design
data wrangling – handling diverse “messy” data	issue tracking	code that interacts with the web
inter-disciplinary thinking	open science	algorithms - code for big data (attached to scalability)
data-enabled science skills	standards and tools	interoperability (package API)
collaboration skills	knowledge of public repositories	unit testing
fundamental computing architecture	workflows	advanced metadata
conceptual modeling	semantics, ontologies, and taxonomies, vocabularies	high performance computing
pseudocode	software life cycle	software licensing for coders
data management best practices		hardware knowledge

Methods to foster the growth of these areas of knowledge and skill among this workforce are equally diverse, and ultimately there was no recommendation made for the areas that ISEES should pursue. Participants agreed that a more thorough scan than could be provided within the time allowed would be necessary in order to 1) scan the current activities such that ISEES does not duplicate other efforts, 2) review pedagogical literature available on the efficacy of various approaches, and 3) scope the cost effectiveness of each approach.

However there was general agreement that many of the basic skills and knowledge areas lend themselves to web dissemination both because they are relatively simple and because the audience is comparatively large and highly distributed. Online modes range from webpages and aggregation of curated web resources (e.g. how to cite software) to hosting distributed online courses. For intermediate to advanced topics, interpersonal interactions were emphasized, including face-to-face working groups and workshops that may be either hosted in a central location or distributed, visiting scholar programs that bring together learners with more advanced mentors, as well as supporting virtual communities in advancing their skills. Again targeting universities was highlighted as a major opportunity, recognizing and leveraging the peer-to-peer, mentor-trainee and teacher-student relationships that are already structurally reinforced within university communities. For example, powerful approaches might include providing materials to university activities (e.g. suitable for classrooms or for lab meetings), creating “train the trainer” workshops in which trainers take lessons back to university communities, and taking actions to influence practices or attitudes within universities.

Several comments from participants were discussed at length and need further consideration by ISEES. 1) There was strong consensus that ISEES should help tie together the broader community of organizations with aligned interests and activities in workforce development, to make it stronger, and not duplicate or compete with these other efforts, including those that may not be focused on software but which foster diversity in the sciences, such as the SEEDS program in ecology. 2) There was concern that NSF may or may not be the organization to fund the level of workforce development that is now necessary – it is a large effort, and NSF does not have a history of funding such large educational efforts. 3) Metrics of success, modes of evaluation and gathering feedback (e.g. from professionals in educational assessment) should be established early and taken seriously at the highest level of organization.

Sustainability and Governance, and Community Engagement

The ISEES Community Engagement workshop focused on identifying the stakeholder community, strategies for engagement, and incentives for participating in the envisioned institute. Building off past workshop recommendations on the potential functions of ISEES - Training, Infrastructure Services, Advocacy and Consultancy - workshop participants identified the stakeholder community that ISEES would serve:

- Researchers from academia, government, and corporate sectors;
- Meta-organizations including professional societies, other networks and NGOs;
- Institutions that support scientific software development such as academic libraries and academic computing;
- Funding sources from public, corporate sector and foundations;
- Data repositories and networks that hold much of the science output; and
- Software community including developers, users, and code repositories

The workshop offered a diverse set of strategies for engagement of the stakeholder community by both ISEES-initiated contact and leveraging existing channels for engagement. Ideas for engagement included:

- Utilizing professional associations as a communication channel, point of engagement for meetings (e.g. workshops, hackathons, meet ups) and source of community-identified challenges.
- Social media for communication, identification of trends and challenges and to identify thought leaders for emerging software developments and practices.
- Leveraging online classes and academic technology centers
- Tracking scientific literature for trends in academic software
- Serving as an ‘Expertise Brokerage’ for scientific projects
- Training of early career researchers
- Promoting new collaborations through exposing ideas and people

Engaging early career scientists was viewed as having the biggest impact factor, as they are viewed as early adopters of new approaches. Post-docs were characterized as especially eager to soak up new ideas and opportunities.Targeting new academic staff during faculty orientations or similar activities would help engage them at critical moments as they start their professional careers. Other engagement and training opportunities include offering advanced short courses with compelling training materials, particularly at professional meetings. Development of a “receptivity coefficient” would be useful in assessing the effectiveness of different engagement tools and approaches.

In addition to the more traditional concepts of community building, there was animated discussion about a more crowd-sourcing, viral approach to engaging colleagues to build a community. Proponents (“evangelists”) in universities would build one-to-one connections. But building community requires a place to meet, whether in person or virtually. This approach to community engagement leads to an evolutionary process for organizational structure and decision making. The idea is to start with a community networking process and see where it goes. This mirrors the open source development processes – see what people rally around and then put support into those areas.

Incentives for participating in ISEES were highlighted by a discussion of the dual nature of activities - both volunteer and supported. There was consensus that the institute would be driven principally by community contributions (voluntary) but that some incentives for participation would be required. Among the noted incentives were:

- Reputation building and rewards for software contributions
- Cross-community pollination
- Access to existing & reusable resources - human and technical
- The promise of faster science results
- Efficiency and economies of scale
- Extending reach
- Potential for greater interoperability of resources

Training Pilot: Open Science for Synthesis

In order to gain understanding of the workforce development needs in the environmental sciences, ISEES investigators Hampton, Jones, Schildhauer, and Regetz organized and conducted the 2013 NCEAS Summer Institute focusing on teaching critical data science, analysis, and modeling skills to early career researchers (see http://www.nceas.ucsb.edu/outreach/summer-institute/2013/summer-institute-2013). Demand for the institute was high, with over 400 applicants for 22 spots in a short application period. Participants received hands-on guided experience using best practices in the technical aspects that underlie successful synthesis – from data discovery and integration to analysis and visualization, and special techniques for collaborative scientific research. Special attention was paid to successful software development for synthetic science, including introductions to version control, workflow management, and data management. A series of surveys were conducted before, during, and after the institute to assess student expectations, knowledge, and learning outcomes.

A second pilot was conducted in 2014 when we organized and ran the Open Science for Synthesis (OSS) training for 45 early career researchers. The goal of this pilot was to experiment with virtual delivery of data and software science training in an intensive three week course, but one in which participants are distributed bi-coastally at NCEAS and our partner RENCI. Similarly to 2013, participants received training in data science, programming, engineering techniques, scientific collaboration, and scientific synthesis, but in this case some of the instructional material was delivered from a remote campus. Participants still had access to local experts, and participated in intensive, real-world synthesis activities, but we were able to double the number of participants by holding two events simultaneously. Like the first pilot, we conducted extensive assessments during the pilot, which allowed us to determine the impact that local versus remote delivery of materials had on the participant experience. Overall, both pilots highlighted the need to these types of meso-scale training events that allow more in-depth exploration than activities such as Software Carpentry but are less intensive than a semester long course. Participant comments about the trainings included:

- "... a really amazing experience at OSS. I really value the knowledge and experience I gained and can't put to words how useful it is going to be for me."
- "What we learned really ought to be taught to every incoming PhD student in an ecology program, but it just isn't."
- "It was full on, eye opening, and an incredible experience."
- OSS "really transformed how I think of open science and working across groups."

Community Pilot: Open Science Codefest

Over September 2 to 4, 2014, more than 70 scientists from all over the world gathered in Santa Barbara for the first-ever Open Science Codefest (OSCodefest, http://nceas.github.io/open-science-codefest/). OSCodefest brought together computer programmers and environmental scientists to collaborate, problem solve, code, and share skills. The unconventional, unstructured "unconference" format of the event allowed flexibility for organic work flows and synergies between attendees, many of whom had not worked together prior to OSCodefest.

This conference was organized to stimulate productivity and community building, while providing ample opportunities for collaborative coding and design sessions. By the conclusion of the meeting, over 20 breakout sessions had been completed, and these new collaborations for skill-sharing and product generation will continue long beyond OS Codefest.

Participants themselves proposed and lead the breakout sessions, each aimed at producing a specific desired result. For example, one session focused on generating cool, impactful visualizations for large, multifaceted datasets. Another group of coders worked together to develop an interactive citizen science smartphone ecology gaming application featuring amphibians, called "EcoFrog." Another breakout session included a team of over 20 scientists co-authoring a peer-review journal article describing how researchers can make their science open access, currently available as a preprint in PeerJ (Hampton et al., 2014).

We strongly encouraged the "Law of two feet" coined by Harrison Owen, originator of the open space approach — if you are neither learning nor contributing, use your two feet and move another group that better suits your needs. Overall, Codefest was a great success and everyone was exposed to new ideas and resources, generated new code for their science projects, made new connections, and had lots of fun along the way.

Open Science Codefest (http://nceas.github.io/open-science-codefest/) was co-sponsored by ISEES, the National Center for Ecological Analysis and Synthesis (NCEAS), the National Science Foundation (NSF), Renaissance Computing Institute (RENCI), the Water Science Software Institute, rOpenSci, DataONE, and the Mozilla Science Lab.

Appendix B: Science Challenges Enabled by Sustainable Earth and Environmental Software

A principal goal of the ISEES design process was to identify key areas where sustainable software could significantly improve the conduct and pace of science. More than two-dozen scientists representing the breadth of Earth and environmental science domains participated in a two-day workshop at the National Center for Ecological Analysis and Synthesis. The express goals of the workshop were to: (1) propose a range of multi- and interdisciplinary "grand challenge" Earth and environmental science questions, the answers to which could meaningfully impact science and society; (2) provide examples of the types of data and software needed to address a diverse subset of the questions by analyzing workflows that could be used to derive answers; and (3) prioritize the functions of a software sustainability institute that could accelerate interdisciplinary research in the Earth and environmental sciences.

Challenging Multi- and Interdisciplinary Earth and Environmental Science Questions

Workshop participants first brainstormed criteria that would be used to identify grand challenge questions in the Earth and environmental sciences. Seven high-priority criteria were established:

1. Tractability. There should already be sufficient investment in data, tools, and conceptual development so that concerted effort could lead to an answer within a reasonable time frame (e.g., 10 years).
2. Generalizability. The answer(s) to the question should be broadly applicable across space, time, and culture.
3. Multi- / inter-disciplinarity. Teams of researchers working across and at the intersection of different disciplines are necessary to resolve the complexities of the problem.
4. Forward-looking. The question should lead to substantive innovation and scientific progress.
5. Societal relevance. Addressing the question will save money and improve lives.
6. Transformative potential. Research into the issue can potentially shift scientific paradigms.
7. Impact. Answers to the question can have real and meaningful impact.

Workshop participants individually identified one or two challenging scientific questions in the Earth and environmental sciences that met the seven criteria outlined above. Collectively, participants then vetted the questions and, where appropriate, combined and re-phrased similar questions. Eighteen questions, grouped into six categories, resulted from this effort.

Sustainability and extraction of natural resources

Question 1: How is the Earth's metabolism responding to anthropogenic global change, and what surprises are in store?

Question 2: How will projected patterns in global energy use over the next three decades impact the sustainability of the world's ecosystems?

Question 8: How can cities be redesigned based upon principles of environmental quality, social equity, and economic feasibility, and the best available science, so that they can persist into the future?

Question 16: How can the negative impacts on ecosystems and the services they provide be minimized given growth in the global trade of natural resource commodities like food and biofuels?

Question 18: How will coastal human and ecological communities adapt at local scales to global climate changes?

Global pathways of pollution

Question 3: What are the safe limits or critical thresholds for atmospheric pollutants within and among ecosystems?

Question 7: What are the controls, impacts, and societal responses to atmosphere–land–water transfer of pollutants, and how will they change under multiple, global-change stressors?

Global availability of water

Question 4: How will coupled human and biophysical systems shape and be shaped by water availability?

Question 14: What is the past, present, and future state of flux of all water everywhere?

Question 17: What are the major quantifiable feedback loops among human water use, landscape change, and global water cycles that drive availability of water from local to global scales?

Biodiversity

Question 5: What is the mean and uncertainty for global extinction of species critical to ecosystem services, and what drives shifts in these means and uncertainties under expected/possible future anthropogenic habitat and climate change?

Question 11:What are the bioservices of organisms that contribute to human well-being? Where and when in the history of life did they arise? How valuable are they in economic terms?

Refining global models for better forecasting and decision-making

Question 6: How can we better forecast ecosystem responses, feedbacks, and services, for a rapidly emerging new state of the Earth system, one with no analogy or baseline in societal and scientific experience?

Question 12: What will be the impact on services provided by the earth system based on policy decisions made today?

Question 13: How does the understanding of flow of material and organisms across scale and systems allow us to predict responses to disturbances?

Crossing and integrating scales of understanding

Question 9: How do spatial and temporal patterns of microclimates affect plant species resistance and resilience to regional and range-wide climate change?

Question 10: How is the complexity of interactions at multiple scales between different trophic levels and abiotic factors influencing ecosystem response to global climate change?

Question 15: Society and wildlife depend on vascular plants, so how do various factors influence soil nutrient/water availability and biotic uptake of these under future global changes and across spatial and temporal scales?

Table 1. Challenge multi- and interdisciplinary scientific questions in the Earth and environmental sciences.

Data and Software Needed to Address Grand Challenge Earth and Environmental Science Questions

Once the questions were agreed upon, small teams worked to flesh out the individual questions by providing context, outlining the necessary research and highlighting the importance of the research to science and society. Box 1 provides an example for one of the three questions.

Box 1. Description of one of the challenging Earth and environmental science questions that informed the identification of data and tools needed by researchers.

Question 4: How will coupled human and biophysical systems shape and be shaped by water availability?

The water cycle is an integral component of the climate system, intersects all major biogeochemical cycles, and provisions freshwater resources that are essential to human society and most terrestrial ecosystems. Physical and biological drivers of the water cycle are varied, including planetary energy balance, internal climate system variability, soil and aquifer properties, plant water use, and direct human intervention. Cycling of water also involves processes acting across an enormous range of temporal and spatial scales, from water's molecular role as the electron donor in photosynthesis to global-scale transport by atmospheric and ocean circulation. This incredible complexity makes the water cycle a highly non-linear system, and our current ability to make precise predictions of water cycle response to perturbations at local, regional and global scales is limited. For example, intercomparison of climate and land surface models indicates widespread disagreement in not only the magnitude but also the sign of future regional precipitation changes associated with future climate warming (IPCC Working Group I 2007) and recent evapotranspiration changes due to simple land-cover shifts (Pitman et al. 2009). The timing, amount, and form of water availability also constrain and shape human- and eco-systems, with potential feedbacks on water cycling. For example local water scarcity necessitates widespread water diversion for crop irrigation in California's Central Valley, which may alter summer monsoons in Arizona (Lo & Famiglietti 2013), and interactions between atmospheric CO_2 levels and water availability will drive future changes in plant water use that influence river runoff and continental evapotranspiration (Betts et al. 2007). Developing a predictive understanding of the water cycle and its 2-way interaction with human and biophysical systems will require the tight integration of data, theories, and algorithms from a large number of disparate domains, including climate science, hydrology, hydroecology, remote sensing, agricultural science, civil and environmental engineering, water management and planning and economics.

A comprehensive, predictive understanding of the water cycling and its 2-way interaction with human and biophysical systems will support a vast portfolio of research in related scientific domains. With this information researchers will be able to more accurately predict changes in global and regional climate change, ecosystem structure, and geomorphic and biogeochemical processes. This information would also advance our retrospective understating of events in Earth history such as major biogeochemical perturbations, evolutionary transitions, and collapse of ancient human societies. The strategies and transdisciplinary collaborative infrastructure developed in pursuit of predictive understanding of water availability and its connectivity to human and biophysical systems would benefit efforts to develop understanding of other major Earth cycles and systems.

Water availability already constrains human development in many regions on Earth (Vörösmarty et al. 2010), and sustaining and improving human well-being will require continued access to water. Prediction of the amount, form and timing of water availability in the environment is necessary to inform development of built infrastructure and management practices that ensure stable supplies of water for human use, sustain availability of water resources that support critical ecosystem services, and protect human and biophysical systems from hydrological extremes.

Next workshop participants broke into three breakout groups that diagrammed the workflow necessary to address each of three different questions, and identified related data and software needs. Figure 1 illustrates many of the high-level scientific steps that would be necessary to understand "how will coupled human and biophysical systems shape and be shaped by water availability?".

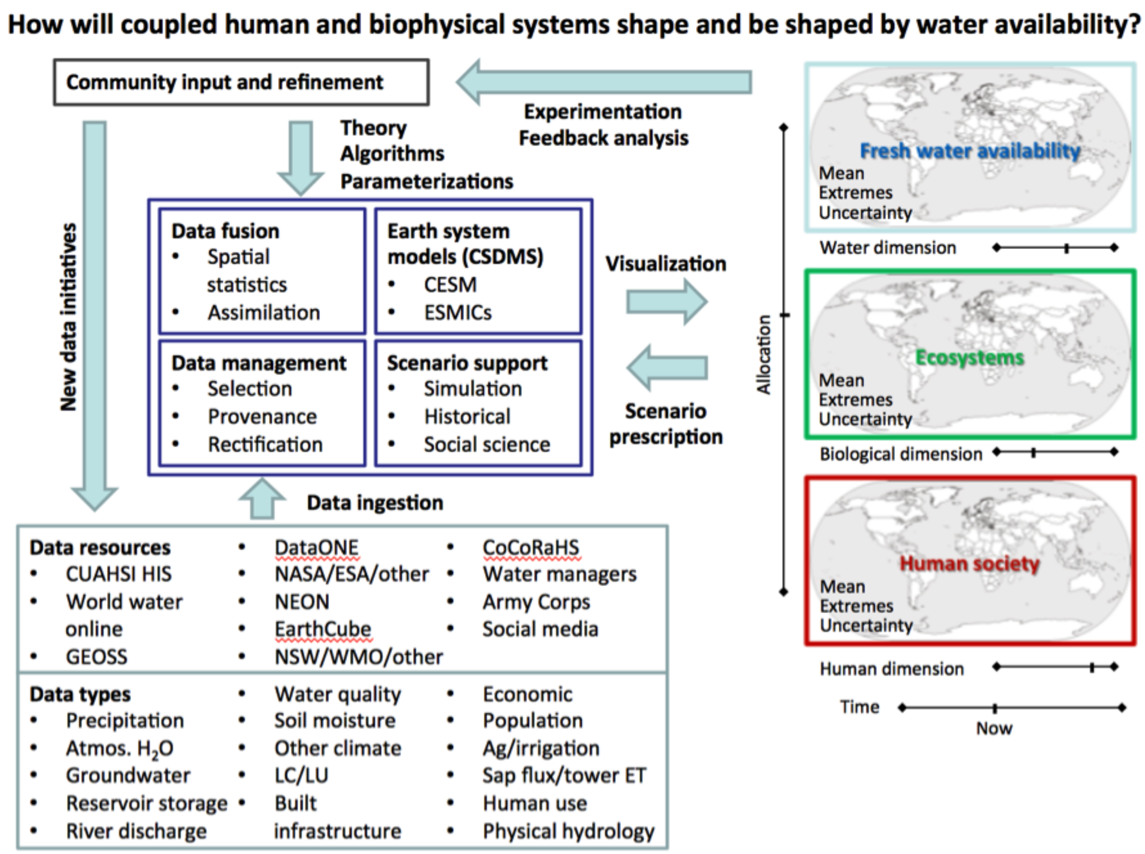

Figure 1. Illustration of the scientific workflow, including high-level analytical processes, data resources and data types that would be necessary to address one of the "grand challenge" Earth and environmental science questions.

Functions of a Software Sustainability Institute Identified by the Scientific Community

After reviewing the three questions and scientific workflow scenarios, as well as the lists of needed data and software, workshop participants then identified and prioritized the functions of a Software Sustainability Institute that they believed could best accelerate interdisciplinary research in the Earth and environmental sciences. The functions, ranked from highest to lowest priority, included:

1. Offer **computation training** for early career and mid and senior scientists
2. Support **assimilation and QA/QC tools** for heterogeneous data
3. Provide a **collaborative environment** for ecologists, computing scientists, social scientists, etc.
4. Develop dynamic, flexible **visualization tools**
5. Provide support for **software maintenance and sustainability**, including software building blocks (e.g., modules)
6. Improve tools for **capturing decisions and workflows** in collaborative research projects
7. Support **software discovery**: One-stop shopping for finding and characterizing software and models -- focus on users
8. Provide **consultants, collaborators** for software, CS, for researchers
9. Develop a **community hub** for standards convergence
10. Facilitate **merging of disparate software** tools
11. Develop **user-friendly interfaces** to existing models
12. Provide a **framework for multiscale, coupled modeling** systems
13. Make **high performance computing available** to the average ecologist and environmental scientist
14. Provide software to **help with uncertainty and error** propagation in spatial models
15. Provide **web-based software services**, i.e. ability to run analyses on ISEES servers via accessible interfaces
16. Support **software vetting** (check software being developed in-house)
17. Help domain researchers **contribute to community** software
18. Provide **taxonomy scrubbing** software
19. Support **improved model intercomparison**

Appendix C: Participants

The conceptualization of ISEES would not have been possible without the dedication and brilliant ideas of the 198 participants in our strategic planning activities. The following individuals contributed ideas, work, and encouragement to bring ISEES into being.

Last	**First**	**Institution**
Abercrombie	Parker	Abercrombie Consulting
Afflerbach	Jamie	University of California, Santa Barbara
Akyildiz	Bugra	New York University
Allison	M. Lee	Arizona Geological Survey
Altintas	Ilkay	University of California, San Diego
Ames	Daniel	Brigham Young University
Anderson	Sean	California State University Channel Islands
Aoki	Betsy	Microsoft
Arko	Robert	Columbia University
Arrigo	Jennifer	Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)
Azzari	George	University of California, Irvine
Bagby	Sarah	University of California, Santa Barbara
Baker	Leanne	Baylor University
Bart	Ryan	University of California, Berkeley
Baskett	Marissa	University of California, Davis
Belnap	Jayne	US Geological Survey (USGS)
Benedict	Karl	University of New Mexico
Benjamin	Alexandra	University of California, Santa Barbara
Best	Benjamin	University of California, Santa Barbara
Boettiger	Carl	University of California, Santa Cruz
Boustani	Maziyar	Jet Propulsion Laboratory of the National Aeronautics and

		Space Administration (NASA)
Bowen	Gabriel	Purdue University
Brand	Nick	University of California, Santa Barbara
Bryant	Annie	University of Utah
Budden	Amber	University of New Mexico
Cabunoc	Abigail	Ontario Institute for Cancer Research
Caron	Bruce	New Media Research Institute
Cavender-Bares	Jeannine	University of Minnesota, Twin Cities
Chamberlain	Scott	University of California, Berkeley
Chandler	Cynthia	Woods Hole Oceanographic Institution
Chaudhary	Aashish	Kitware, Inc.
Chen	Ying-Jung	University of California, Santa Barbara
Choate	Janet	University of California, Santa Barbara
Christopherson	Laura	RENCI
Clavelle	Tyler	University of California, Santa Barbara
Cole	Dave	MapBox
Collins	Scott	University of New Mexico
Cooper	Larry	Southern California Coastal Water Research Project
Cornejo-Donoso	Jorge	University of California, Santa Barbara
Couture	Jessica	University of California, Santa Barbara
Cross	Alex	California Polytechnic State University, San Luis Obispo
Cruse	Patricia	University of California, Office of the President
Czaplewski	John	University of Wisconsin, Madison
Davis	Frank	University of California, Santa Barbara
DeLuca	Cecelia	NOAA, Earth System Research Laboratory (ESRL)
Diggs	Stephen	University of California, San Diego
Dozier	Jeff	University of California, Santa Barbara

Duerr	Ruth	University of Colorado, Boulder
Earl	Stevan	Long Term Ecological Research (LTER)/Central Arizona-Phoenix (CAP)
Ederer	Gregory	NASA Goddard Institute for Space Studies
Englert	Vyki	NationBuilder
Ettinger	Kate	Mural Institute
Fernandez del Viso	Denny	University of Puerto Rico, Humacao
Ferretti	Francesco	Stanford University
Filazzola	Alessandro	York University
Fox	Peter	Rensselaer Polytechnic Institute
Frew	Jim	University of California, Santa Barbara
Fulker	David	Open-source Project for a Network Data Access Protocol (OPeNDAP)
Gamon	John	University of Alberta
Gil	Yolanda	University of Southern California
Glenn	Nancy	Boise State University
Gray	Cameron	Sustain3
Gries	Corinna	University of Wisconsin, Madison
Grimm	Nancy	Arizona State University
Gross	Louis	University of Tennessee, Knoxville
Gunn	William	Mendeley
Hallett	Lauren	University of California, Berkeley
Hampton	Stephanie	Washington State University
Han	Xueying	University of California, Santa Barbara
Harris	David	University of California, Davis
Hart	Edmund	National Ecological Observatory Network, Inc. (NEON)
Hart	Edmund	University of British Columbia

Heard	Jefferson	University of North Carolina, Chapel Hill
Heidorn	Bryan	University of Arizona
Hernandez	Rebecca	Stanford University
Hetmank	Thomas	University of California, Santa Barbara
Hills	Andrew	Vizzuality
Holmes	Elizabeth	NOAA, National Marine Fisheries Service (NMFS)
Hooper	Richard	Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)
Howison	James	University of Texas, Austin
Hsu	Leslie	Columbia University
Huesca-Martinez	Margarita	University of California, Davis
Humphries	Grant	University of California, Davis
Hunt	Ryan	MFMP/Hunt Uttilities Group
Idaszak	Ray	University of North Carolina, Chapel Hill
Jaimes	Aline	University of Delaware
Jasiak	Erik	University of Colorado, Boulder
Jones	Matthew	University of California, Santa Barbara
Jones	Sydney	University of New Mexico
Joppa	Lucas	Microsoft Research Ltd.
Joyce	Michael	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Kajokaite	Kotrina	University of California, Los Angeles
Kansa	Sarah	Alexandria Archive Institute
Kennedy	Jessie	Napier University
Khudikyan	Shakeh	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Kishor	Puneet	Creative Commons
Ledley	Tamara	Technical Education Research Centers (TERC)

Leinfelder	Ben	University of California, Santa Barbara
Lenhardt	W. Christopher	University of North Carolina, Chapel Hill
Lin	Jennifer	Public Library of Science (PLoS)
Liu	Sophia	US Geological Survey (USGS)
Longo	Catherine	University of California, Santa Barbara
Lynnes	Chris	National Aeronautics and Space Administration (NASA)
Lyons	Andrew	Stanford University
Ma	Stephanie	University of California, Santa Barbara
MacCuish	John	Mesa Analytics & Computing, Inc.
MacDonald	Andrew	University of British Columbia
Marru	Suresh	Indiana University
Mattmann	Chris	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Mayall	Tess	Science Exchange
McCartney	Peter	National Science Foundation
McClintock	William	University of California, Santa Barbara
McGibbney	Lewis John	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Meyer	Carol	Foundation for Earth Science
Michelou	Vanessa	University of Hawaii, Manoa
Michener	William	University of New Mexico
Miles	Brian	University of North Carolina, Chapel Hill
Mills	Bill	Independent
Mudge	Joseph	Texas Tech University
Muir	Brian	University of California, Riverside
Murphy	Sylvia	NOAA, Earth System Research Laboratory (ESRL)
Novaglio	Camilla	University of Tasmania

Oates	George	Stamen Design
O'Brien	Margaret	University of California, Santa Barbara
Odat	Nidal	Al Hussein Bin Talal University
Palmer	Margaret	National Socio-Environmental Synthesis Center (SESYNC)
Paris-Limouzy	Claire	University of Miami
Peach	Kristen	University of California, Santa Barbara
Pinsky	Malin	Princeton University
Piwowar	Heather	Impact Story
Pollak	Jonathan	Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)
Pollard	Amina	US Environmental Protection Agency (EPA)
Porras-Alfaro	Andrea	Western Illinois University
Pourmokhtarian	Afshin	Boston University
Priem	Jason	Impact Story
Quisel	Tom	Data Science Consultant
Ram	Karthik	University of California, Berkeley
Ramamurthy	Mohan	University Corporation for Atmospheric Research
Ramirez	Paul	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Ratay	Sarah	University of California, Los Angeles
Read	Jordan	US Geological Survey (USGS)
Rebich Hespanha	Stacy	University of California, Santa Barbara
Redding	Jacob	Drupal Association
Reese	Jordan	University of North Carolina, Chapel Hill
Regetz	Jim	University of California, Santa Barbara
Rheinheimer	David	University of California, Merced
Richard	Stephen	US Geoscience Information Network (USGIN)

Ripplinger	Julie	Arizona State University
Robinson	Erin	Foundation for Earth Science
Roth	Keely	University of California, Davis
Scheuerell	Mark	NOAA, National Marine Fisheries Service (NMFS)
Schildhauer	Mark	University of California, Santa Barbara
Shepherd	Adam	Woods Hole Oceanographic Institution
Slaughter	Peter	University of California, Santa Barbara
Smith	Paul	Southern California Coastal Water Research Project
Smith	Rachel	Grove Consultants International
Spahn	Noah	University of California, Santa Barbara
Staples	Thornton	Smithsonian Institution
Starch	Michael	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Stealey	Michael	University of North Carolina, Chapel Hill
Stewart Lowndes	Julia	University of California, Santa Barbara
Stoddart	Nick	Kaizen SCN
Strasser	Carly	University of California, Office of the President
Szuwalski	Cody	University of California, Santa Barbara
Tao	Jing	University of California, Santa Barbara
Tarboton	David	Utah State University
Terrel	Andy	Continuum Analytics
Thaney	Kaitlin	Mozilla Science Lab
Tilmes	Curt	National Aeronautics and Space Administration (NASA)
Tomar	Shivira	Industrial Ecology Research Services (IERS)
Truslove	Ian	University of Colorado
Turner	Mark	California Institute of Technology
Twidale	Michael	University of Illinois

Vargas	Rodrigo	University of Delaware
Verma	Rishi	Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA)
Vieglais	David	University of Kansas
Viñas Garcia	Maria José	National Aeronautics and Space Administration (NASA)
Walker	Lauren	University of California, Santa Barbara
Wasser	Leah	National Ecological Observatory Network, Inc. (NEON)
Weber	Nicholas	University of Illinois, Urbana Champaign
Wee	Brian	National Ecological Observatory Network, Inc. (NEON)
Werner	Chhaya	University of California, Davis
White	Christine	Environmental Systems Research Institute (ESRI)
Whitton	Mary	RENCI
Wilber	Mark	University of California, Santa Barbara
Willcocks	Angela	University of Redlands
Williams	John	University of Wisconsin, Madison
Wilson	Anne	University of Colorado, Boulder
Wilson	Gregory	Software Carpentry
Winslow	Luke	US Geological Survey (USGS)
Wolkovich	Elizabeth	University of British Columbia
Woo	Kara	University of California, Santa Barbara
Wood	Zoe	California Polytechnic State University, San Luis Obispo
Wright	Dawn	Environmental Systems Research Institute (ESRI)
Wu	Kevin	University of California, Santa Barbara
Yi	Hong	University of North Carolina, Chapel Hill
Yocum	Dan	University of California, Santa Barbara
Zimmerman	Naupaka	University of Arizona

Appendix D: References Cited

Betts R. A., et al. 2007. Projected increase in continental runoff due to plant responses to increasing carbon dioxide. Nature 448(7157):1037-1041.

Hackett, E. J., J. N. Parker, D. Conz, D. Rhoten, and A. Parker. 2008. Ecology Transformed: The National Center for Ecological Analysis and Synthesis and the Changing Patterns of Ecological Research. Pages 277–296 in G. M. Olson, A. Zimmerman, and N. Bos, editors. Scientific Collaboration on the Internet. The MIT Press.

Hampton SE, Anderson S, Bagby SC, Gries C, Han X, Hart E, Jones MB, Lenhardt WC, MacDonald A, Michener W, Mudge JF, Pourmokhtarian A, Schildhauer M, Woo KH, Zimmerman N. (2014) The Tao of Open Science for Ecology. PeerJ PrePrints 2:e549v1 http://dx.doi.org/10.7287/peerj.preprints.549v1

Hampton, S. E., and J. N. Parker. 2011. Collaboration and Productivity in Scientific Synthesis. BioScience 61:900–910. doi:10.1525/bio.2011.61.11.9

IPCC Working Group I. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller, editors. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Lo, M. H. and J. S. Famiglietti. 2013. Irrigation in California's Central Valley Strengthens the Southwestern U. S. Water Cycle. Geophysical Research Letters 40(2):301-306.

Pitman A. J., et al. 2009. Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. Geophysical Research Letters 36 (14): L14814.

Vörösmarty C. J., et al. 2010. Global threats to human water security and river biodiversity. *Nature* 467:555-561.