



EarthCube



Past, Present, and Future





EarthCube

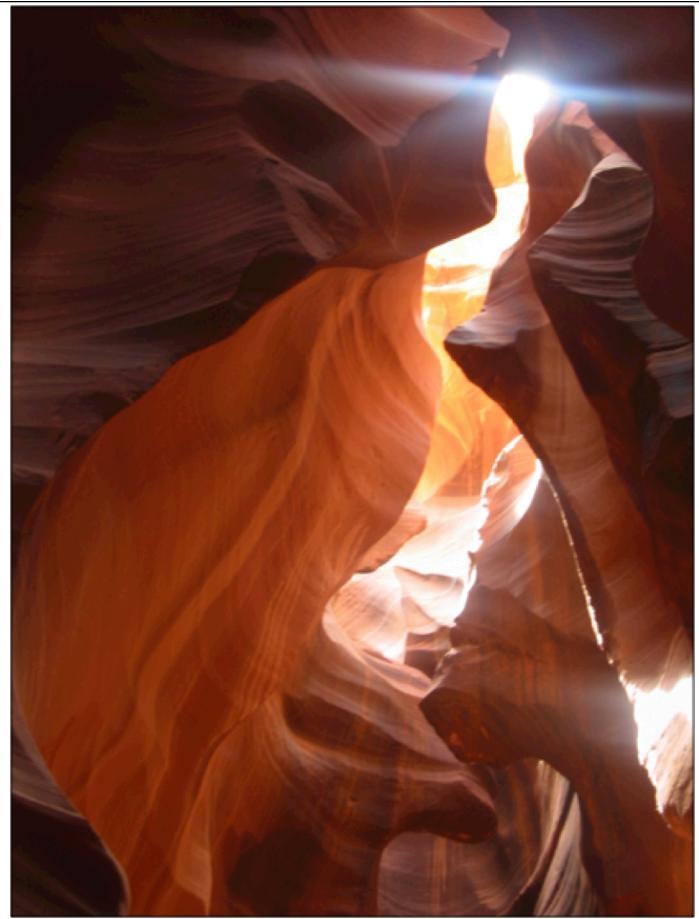


[EarthCube](#) is an initiative of the [National Science Foundation](#).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

EarthCube Project Report EC-2014-3

First released: December 12, 2014



Picture credits – On cover from top to bottom: NOAA Okeanos Explorer Program (CC BY-SA 2.0), NASA/Kathryn Hansen (CC BY 2.0), and Canyonlands National Park/Neal Herbert (CC BY-NC-SA 2.0). On back page from top to bottom: Hurricane Danielle, NASA (CC BY-NC 2.0), Other Worlds, NASA (CC BY-NC-ND 2.0). On this page: Jurassic Navajo Sandstone, Antelope Canyon, Northern Arizona/Marjorie Chan (CC BY 4.0). The rest of the pictures in this document were taken by members of the EarthCube community at various workshops, field trips, and meetings.



This document is released under a [Creative Commons CC BY 4.0 license](#). This license lets others distribute, remix, tweak, and build upon the work, even commercially, as long as they credit this document. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials. It should be cited as:

“EarthCube: Past, Present, and Future.” Yolanda Gil, Marjorie Chan, Basil Gomez and Bruce Caron (Eds). EarthCube Project Report EC-2014-3, December 2014.

It is available online at:

[*http://earthcube.org/document/2014/earthcube-past-present-future*](http://earthcube.org/document/2014/earthcube-past-present-future)



Table of Contents

Executive Summary	5
Message from NSF: Transforming Geosciences through EarthCube	7
EarthCube Vision and Mission	9
A Timeline for EarthCube	11
End User Workshops: Gathering Requirements from Scientists	13
EarthCube Funded Projects	23
Research Coordination Networks: Engaging Science Communities	25
Building Blocks: Exploring Solutions and Demonstrating Utility	41
Conceptual Designs: Initial Planning for Enterprise Architecture	75
Test Governance: Formal Mechanisms to Involve the Community	85
Getting Involved	97
The Science Committee	100
The Technical and Architecture Committee	100
The Engagement Team	100
The Liaison Team	100
The Council of Data Facilities	101
The Leadership Council	102
Other Opportunities	103



About the Editors

Yolanda Gil is Director of Knowledge Technologies at the Information Sciences Institute and Research Professor in the Computer Science Department of the University of Southern California. She received her M.S. and Ph. D. degrees in Computer Science from Carnegie Mellon University. Her expertise is in artificial intelligence, in particular intelligent systems to support scientific discovery. She chaired the World Wide Web Consortium (W3C) Provenance Group that led to the W3C PROV community standard, and is now engaged with the Open Geospatial Consortium (OGC) to investigate provenance challenges of geospatial data. Dr. Gil has served in the Advisory Committee of the Computer Science and Engineering Directorate of the National Science Foundation. She was elected Fellow of the American Association of Artificial Intelligence (AAAI) in 2012. She led the EarthCube Roadmap for Workflows in Geosciences, and is Co-PI of the EarthCube Building Block on "GeoSoft: Software Stewardship for Geosciences." She Co-Chairs the EarthCube Technology and Architecture Committee, and was elected as its representative to the EarthCube Leadership Council.

Marjorie Chan is a Professor in the Department of Geology and Geophysics of the University of Utah. She received her Ph.D. in geology from the University of Wisconsin-Madison, and her B. Sc. from the University of California-Davis. Her research interests are in clastic sedimentary geology, and multidisciplinary studies using aspects of facies, basin analysis, fluid flow, and modeling (with applications toward environmental and predictive problems). Recent work examines sandstone color, iron oxide concretions, and weathering patterns as terrestrial analogs for the red planet Mars. She is an elected fellow of the Geological Society of America (GSA) and was the 2014 GSA Distinguished International Lecturer delivering 54 lectures in 6 countries. She was co-convener of the Sedimentary Geology EarthCube End-User Workshop, and is a Co-PI of the EarthCube RCN for "EC3 Earth Centered Cyberinfrastructure Collaboration". She is Co-Chair of the EarthCube Engagement Team, and was elected as its representative to the EarthCube Leadership Council.

Basil Gomez is an Adjunct Professor in the Department of Geography at the University of Hawai'i at Mānoa. He received a D.Sc. and Ph.D. from the University of Southampton, and his Bachelor's degree from Plymouth Polytechnic. His research focuses on interrelations between landscape, water and the materials it transports, which can shed light on such diverse issues as: paleoearthquake recurrence on the Hikurangi margin; Holocene climate variability across the mid-latitude South Pacific; the potential rate of bed-load transport, and the source provenance of Cypriot Bronze Age and Roman pottery. He is Co-Chair of the EarthCube Science Committee, and was elected as its representative to the EarthCube Leadership Council.

Bruce Caron is an active online-community architect, and is looking to help virtual science organizations build community governance and achieve their promise. Bruce was trained as a social anthropologist and an urban cultural geographer. He recently led the NASA Science on Drupal Central Project, which built collaboration and collective intelligence capabilities for NASA earth science Drupal developers. Bruce is the founder of the New Media Studio and the New Media Research Institute in Santa Barbara. He has served as the president of the Federation of Earth Science Information Partners and is a founding and current board member of the Foundation for Earth Science in Washington, DC. In 2010 he was awarded the Martha Maiden Lifetime Achievement Award from the ESIP Federation. He is currently the Program Office Manager for EarthCube.



Executive Summary

EarthCube began in 2011 as joint initiative between the National Science Foundation (NSF) Directorate for Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI). This evolving, dynamic community effort is not only a new way for the NSF to partner with the scientific community, but also a challenge for the many academic, agency and industry stakeholders in the geo-, cyberinfrastructure, computer and social sciences to create new capabilities for sharing data and knowledge and conducting research.

EarthCube's vision is a community-driven dynamic cyberinfrastructure that will support standards for interoperability, infuse advanced technologies to improve and facilitate interdisciplinary research, and help educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science. EarthCube's mission is to enable geoscientists to address the challenges of understanding and predicting a complex and evolving Earth system by fostering a community-governed effort to develop a common cyberinfrastructure to collect, access, analyze, share and visualize all forms of data and resources, using advanced technological and computational capabilities.

Achieving these objectives requires a long-term effort, which the NSF anticipates supporting until at least 2022. It also requires a community desire to identify common solutions and best practices, adapt and respond to change as cyberinfrastructure evolves, and adopt new technologies and approaches.

Between 2012 and 2014, the NSF funded twenty-four EarthCube domain end-user workshops. Their purpose was to allow the constituent geoscience communities to articulate their cyberinfrastructure needs and science goals, particularly in relation to the accessibility of data and information both within their disciplines and from other fields. In 2013 and 2014, the NSF funded several dozen projects focused on software component development, architecture design, efforts to advance community-building, and governance.

An important result of these activities was the development of an initial EarthCube governance through several community events. This governance structure was put in place in the Fall of 2014 and will be operating in the next few months to develop processes to coordinate community input and ongoing work. EarthCube's Leadership Council oversees the activities of the five components of EarthCube governance: the Science Committee, the Technology and Architecture Committee, the Council of Data Facilities, the Liaison Team, and the Engagement Team. The Council is comprised of community-elected representatives drawn from each of these components, and at-large members of the Atmosphere/Space; Oceans, Earth Sciences, Polar and Cyberinfrastructure communities.

Above all, EarthCube is an evolving, dynamic community effort that actively seeks to engage individuals and partners from across the geosciences and cyberinfrastructure. The more scientists and technologists are engaged in future EarthCube activities, the more EarthCube can and will achieve.





EarthCube

Message from NSF: Transforming Geosciences through EarthCube

Since the beginning of EarthCube in 2011, we at the National Science Foundation have appreciated anew the frontier science challenges undertaken by the academic geosciences community and their partners. We know the community will seek even greater challenges in the future to understand the fundamental processes of the Earth system, within the atmospheric, earth, geospace, ocean and polar sciences and across those boundaries. EarthCube is intended to support this endeavor and transform the conduct of geosciences research by creating a more productive research environment, with new capabilities for sharing data and knowledge by and beyond the geosciences.

As a joint effort of the NSF Directorate for Geosciences and the Division of Advanced Cyberinfrastructure, EarthCube is a new approach for NSF to partner with the scientific community. It envisions an iterative process for creating community-driven and governed cyberinfrastructure, and requires collaboration among the many stakeholders in the geosciences, cyberinfrastructure and computer sciences, social sciences, as well as agency and international partners that share these goals. It is critical to this partnership to have a staged and deliberate approach to EarthCube, allowing time for broad and open involvement, as well as assessment and responsiveness from both NSF and the community.

We are now in a new phase for EarthCube where the responsibilities for and drivers of EarthCube come from the scientific community. NSF welcomes this transition and the wide adoption of EarthCube's vision and process, as reflected by this document. It presents the community's view of where we have been and how EarthCube will be formed through collaborative activities, from the strategic visioning of EarthCube to conceiving and standing up a functioning community governance through Test Enterprise Governance; from the vital input and guidance of the geosciences domains workshops and Research Coordination Networks to the initial elements of EarthCube found within the Building Blocks and Conceptual Designs.





EarthCube

We know this will be a long-term effort with many changes over time as cyberinfrastructure evolves to accommodate changing user needs and emerging technologies and services. EarthCube will be supported by the existing foundation of cyberinfrastructure investments, including databases, software services and community facilities that have been created by the geosciences and cyberinfrastructure communities over the past two decades. Success in serving the entire geosciences community will depend in part on identification of common solutions and best practices, and strategic adoption of new technologies and approaches. It will also depend on the continued participation and work of the community, and we are grateful for the many contributions that have already been made on behalf of EarthCube.



NSF commends the community for their whole-hearted spirit of collaboration, the bright vision they see for geosciences in the future and our ability to work in partnership to achieve this vision. We look forward to working closely with the entire EarthCube community in the years to come.



Eva Zanzerkia
Program Director, EarthCube
Directorate for Geosciences
National Science Foundation



EarthCube



EarthCube Vision and Mission

EarthCube's long-term vision is

a community-driven dynamic cyberinfrastructure that will support

standards for interoperability,

infuse advanced technologies to improve and facilitate interdisciplinary research, and

help educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science.

EarthCube's mission is

to enable geoscientists to address the challenges of understanding and predicting a complex and evolving Earth system

by fostering a community-governed effort to develop a common cyberinfrastructure

to collect, access, analyze, share and visualize all forms of data and resources,

using advanced technological and computational capabilities.

(From the EarthCube Leadership Council, December 2014)

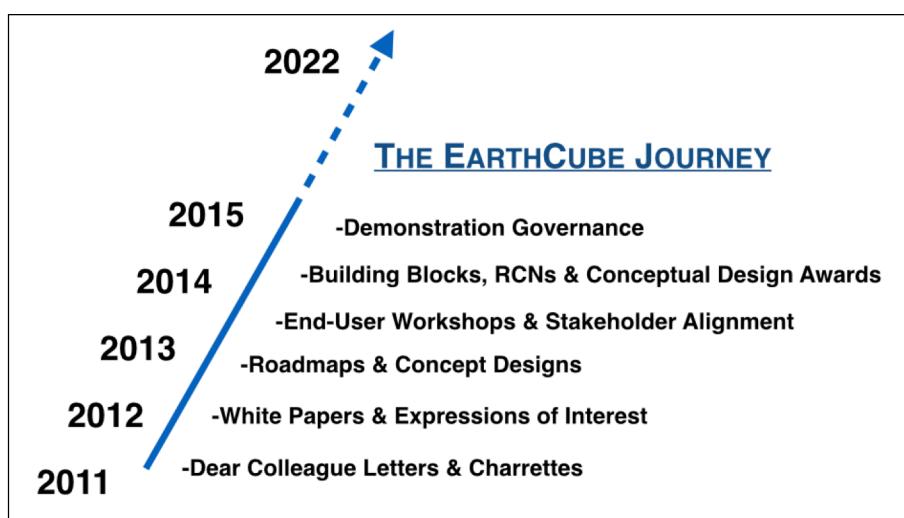


EarthCube



A Timeline for EarthCube

EarthCube was established in 2011 via a collaborative partnership between NSF's Directorate for Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI). An NSF Dear Colleague Letter¹ was released in June 2011 announcing the partnership and initial goals for EarthCube. Several webinars followed, and an additional document, EarthCube Guidance for the Community², gave more detailed guidance. These announcements launched the first conversations about the future of EarthCube. Over the next two years, a series of webinars, community meetings (Charrettes), and White Paper and Roadmap solicitations provided a forum for potential participants to propose what EarthCube should look like in terms of science requirements, technology solutions, designs, and governance.³



Between March and August 2012, each of the Community Groups and Concept Teams was tasked with preparing Roadmaps to show how to move their area of EarthCube forward. These roadmaps were the culmination of months of research,

community outreach, and deliberations in virtual and physical meetings, and they identified the initial stakeholders and cyberinfrastructure components. Collectively they served to provide NSF and other interested parties with a cross spectrum of ideas and concepts from the Earth, computer, information science and other stakeholder communities regarding key elements needed to build EarthCube. They were presented to NSF and the EarthCube community at the second EarthCube charrette.

The second EarthCube charrette took place in Roslyn, VA, in June 2012. This event engaged 190 physical and 60 remote attendees and focused on moving EarthCube forward. A goal of the charrette was to review and integrate the Community Group and Concept Team draft roadmaps, to forge a common vision, and create a cohesive set of milestones to move

¹ Dear Colleague Letter: <http://1.usa.gov/1BtXS7V>

² EarthCube Guidance for the Community: <http://1.usa.gov/1wrkkxo>

³ These documents are available in the EarthCube Document Repository: <http://workspace.earthcube.org/document-repository>



EarthCube

EarthCube forward. Activities and discussions focused on the identification of common themes, challenges, and synergies; that could be merged into one common roadmap for EarthCube.

Beginning in summer 2012, NSF funded a series of 24 EarthCube domain end-user workshops. These workshops targeted a broad spectrum of Earth, atmosphere, ocean, and allied senior, mid- and early-career scientists. The purpose of these workshops was to allow the constituent geoscience communities to articulate their cyberinfrastructure needs and science goals, particularly in relation to the accessibility of data and information both within their disciplines and from other fields.



Meetings of the EarthCube Community Groups and Concept Award Teams Principal Investigators were held in June and October 2012. The first meeting fostered the discussions and roadmap integration efforts, which began at the June 2012 charrette, and generated ideas for use cases, reference architecture, governance and timelines. The follow-up meeting furthered integration of the roadmaps, thereby developing a more cohesive vision of how to move EarthCube forward. Significant steps have been made towards achieving this goal, and a comprehensive technical roadmap for EarthCube is being assembled.

In September 2013, the NSF, marking a new phase for EarthCube, announced \$14.5 million in funding for initial software components development for EarthCube ('Building Blocks'), projects to develop broad architecture design white papers ('Conceptual Designs'), and Research Coordination Networks (RCNs) to advance community- building exemplars in several domain science communities, and a project to develop and test a prototype community governance framework. NSF support for EarthCube continued with the announcement of another round of awards in September 2014.

To date, a great deal of collaborative work has been done to further the advancement of EarthCube, but the process is far from complete, and in the coming years the geoscience community will continue to build and expand upon the work already done.



EarthCube

End User Workshops

Gathering Requirements from Scientists

(2012-2014)



<http://www.earthcube.org/page/end-user-workshops>



EarthCube



End User Workshops

NSF funded 24 EarthCube domain end-user workshops targeting a broad spectrum of Earth, atmosphere, ocean, and related senior, mid- and early career scientists. The purpose of these workshops was to allow geoscience communities to articulate and document their cyberinfrastructure needs with the object of improving data and information access within and outside their disciplines. An additional goal of these workshops was to gather information about the science-drivers and data utilities, and the requirements for user-interfaces, models, software, tools, etc. so that EarthCube can be designed to help geoscientists more easily do the science they want and would like to accomplish.

Outcomes of the domain workshops and other community engagement programs continue to actively shape EarthCube's form and function. Additionally, the workshops served to introduce EarthCube to a broad spectrum of end-users, and encouraged them to think about how data-enabled science could help them achieve their scientific goals. This information is currently being analyzed by working groups and teams constituted under EarthCube's Demonstration Governance Structure, with the object of determining how it can best be leveraged to enable and enhance the future development of EarthCube..

Included in the Executive Summaries for each of these workshops are statements about the community's science vision and the challenges it faces, as well as the technical requirements needed to address them, and priorities and recommendations for action.





EarthCube

End User Workshops

Meetings of Young Researchers in Earth Science (MYRES) V: The Sedimentary Record of Landscape Dynamics

August 8, 2012, Salt Lake City, UT

Organizer: Elizabeth Hajek, Pennsylvania State University

Envisioning Success - A Workshop for Next Generation EarthCube Scholars and Scientists

October 16-17, 2012, Carnegie Institute, Washington, DC



Organizers: Joel Cutcher-Gershenfeld (PI), University of Illinois; Steve Diggs, Scripps Institute of Oceanography (UCSD); Yolanda Gil, Information Sciences Institute (USC); Robert Hazen, The Carnegie Institute of Washington; and Danie Kinkade, Woods Hole Oceanographic Institution

Structural Geology and Tectonics

October, 20-21 2012, Chicago, IL

Organizer: J. Douglas Walker

EarthScope

October 29-30, 2012, Arizona State University, Tempe, AZ

Organizers: Ramon Arrowsmith, Arizona State University and the EarthScope Cyberinfrastructure Subcommittee (ECISC)

Experimental Stratigraphy

December 11-12, 2012, University of Texas at Austin

Organizers: Wonsuck Kim, University of Texas at Austin; Leslie Hsu, LDEO; Raleigh Martin, University of Pennsylvania; and Brandon McElroy, University of Wyoming



EarthCube

Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction

December 17-18, 2012 UCAR, Boulder, CO

Organizers: Mohan Ramamurthy, UCAR; Fuqing Zhang, Penn State University; Russ Schumacher, Colorado State University



Engaging the Critical Zone Community to Bridge Long Tail Science with Big Data

January 21-23, 2013, University of Delaware

Organizers: Anthony Aufdenkampe, Stroud Water Research Center; Christopher Duffy, Penn State University; and Gregory Tucker, Stroud Water Research Center

Envisioning a Digital Crust for Simulating Continental Scale Subsurface Fluid Flow in Earth System Models

January 29-31, 2013, USGS Powell Center, Fort Collins, CO

Organizers: Jennifer Arrigo, CUAHSI; Norman Jones, Brigham Young University; and Ying Fan Reinfelder, Rutgers University

Cyberinfrastructure for Paleogeoscience

February 4-6, 2013, University of Minnesota

Organizers: Anders Noren (University of Minnesota), Jack Williams, Kerstin Lehnert, Shanan Peters, Eric Grimm, Julie Brigham-Grette, Emi Ito, Dave Anderson, Lisa Boush.





EarthCube

Education

March 4-5, 2013, Scripps Institution of Oceanography, La Jolla, CA

Organizers: Kim Kastens (Education Development Center), Ruth Krumhansl (Education Development Center), Chery Peach (Scripps Institute of Oceanography)

Petrology and Geochemistry

March 6-7, 2013, National Museum of Natural History, Washington, DC

Organizers: Chuck Connor (University of South Florida), Elizabeth Cottrell (National Museum of Natural History, Smithsonian Institution), Radjeep Dasgupta (Rice University), Kerstin Lehnert (Lamont-Doherty Earth Observatory, Columbia University), Abani Patra (SUNY Buffalo).



Sedimentary Geology

March 25-26, 2013, University of Utah

Organizers: Marjorie Chan (University of Utah) and David Budd (University of Colorado Boulder)





Modeling for the Geosciences

April 22-23, 2013, Boulder, CO

Organizers: Louise Kellogg (University of California, Davis), Jennifer Arrigo (Consortium of Universities for the Advancement of Hydrologic Science), Lorraine Hwang (Computational Infrastructure Dynamics), Scott Peckham (Community Surface Dynamics Modeling System), David Tarboton (Utah State University)

Integrating Inland Waters, Geochemistry Biogeochemistry and Fluvial Sedimentology Communities

April 24-26, 2013, Boulder, CO

Organizers: Emilio Mayorga (University of Washington) and Albert Kettner (University of Colorado)

Deep Seafloor Processes and Dynamics

June 5-7, 2013, University of Rhode Island, Graduate School of Oceanography, Narragansett, RI

Organizers: Vicki Ferrini, Lamont-Doherty Earth Observatory, Karyn Rogers, Carnegie Institution of Washington, Annette DeSilva, ex officio member, UNOLS

Integrating Real-time Data into the EarthCube Framework

June 17-18, 2013, National Center for Atmospheric Research (NCAR), Boulder, CO

Organizers: Mike Daniels (NCAR, Chair), Frank Vernon (Scripps Institution of Oceanography), Sara Graves/Sandra Harper (University of Alabama Huntsville), Branko Kerkez (University of Michigan) and Chandra Chandrasekar (Colorado State University)

Ocean 'Omics

**August 21-23, 2013,
University of Southern
California, Wrigley
Marine Institute,
Catalina Island, CA**

Organizers: Katrina Edwards (USC), Ed DeLong (MIT), John Heidelberg (USC), and Ginger Armbrust (U of Washington)





EarthCube

Developing a Community Vision of Cyberinfrastructure Needs for Coral Reef Systems Science

**September 18-19, 2013, Hawai'i Institute of Marine Biology, University of Hawai'i
October 23-24, 2013, National Center of Ecological Analysis and Synthesis (NCEAS),
University of California at Santa Barbara**

Organizers: Ruth Gates (Hawai'i Institute of Marine Biology, University of Hawai'i) and Mark Schildhauer (National Center of Ecological Analysis and Synthesis (NCEAS), University of California Santa Barbara)

Bringing Geochronology into the EarthCube Framework

**October 1-3, 2013,
University of Wisconsin-Madison, Pyle Center**

Organizers: Brad S. Singer (University of Wisconsin-Madison), Douglas J. Walker (University of Kansas), and Shanan Peters (University of Wisconsin-Madison)

Articulating Cyberinfrastructure Needs of the Ocean Ecosystem Dynamics Community

October 7-8, 2013, Woods Hole Oceanographic Institution, Woods Hole, MA
Organizers: Danie Kinkade (WHOI) and Peter Wiebe (WHOI)





EarthCube

Engaging the Atmospheric Cloud/ Aerosol/ Composition Community

October 21-22, 2013, George Mason University, Fairfax, VA

Organizers: Liping Di (George Mason University), Athanasios Nenes (Georgia Institute of Technology), Akua Asa Awuku (University of California-Riverside), Stefan Falke (George Mason University)

Rock Deformation and Mineral Physics Research

November 12-14, 2013, Alexandria, VA

Organizers: Chris

Marone (Penn State)
and Jay Bass
(University of
Illinois)



Science-Driven Cyberinfrastructure Needs in Solar- Terrestrial Research

**August 13-15,
2014, Newark, NJ**

Organizers: Gelu M. Nita, Dale E. Gary, Andrew J. Gerrard, Gregory D. Fleishman, Alexander G. Kosovichev, Vincent Oria, Marek Rusinkiewicz.

Increasing the Access to and Relevance of Marine Seismic Data

December 10-12, 2014, San Francisco, CA

Organizers: Jamie Austin (PI), University of Texas at Austin Institute for Geophysics (UTIG); Nathan Bangs, UTIG; Suzanne Carbotte, Lamont-Doherty Earth Observatory (LDEO); Jon Childs, US Geological Survey; Adrian McGrail, ION; John Snedden, UTIG; David Arctur, UT Austin.



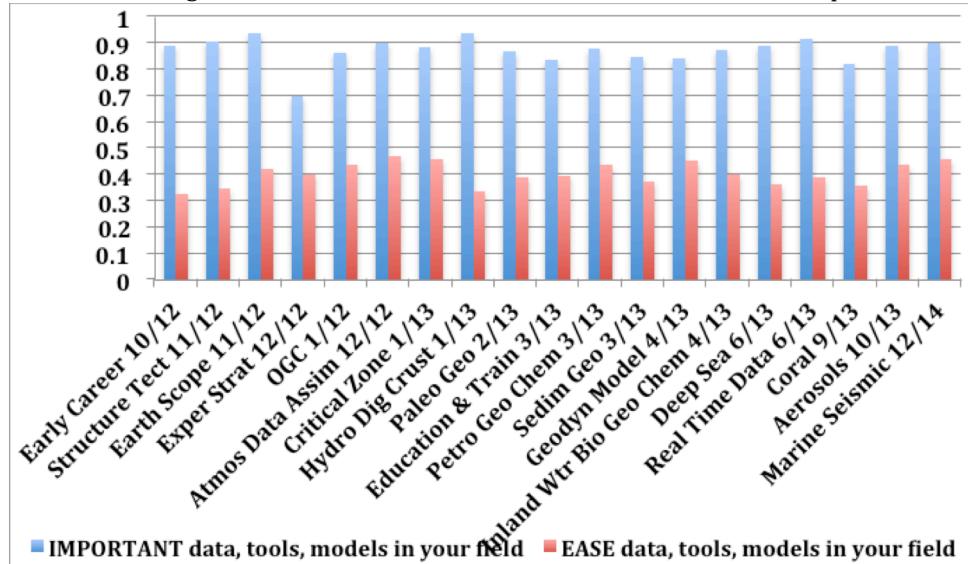


Reports of End User Workshops

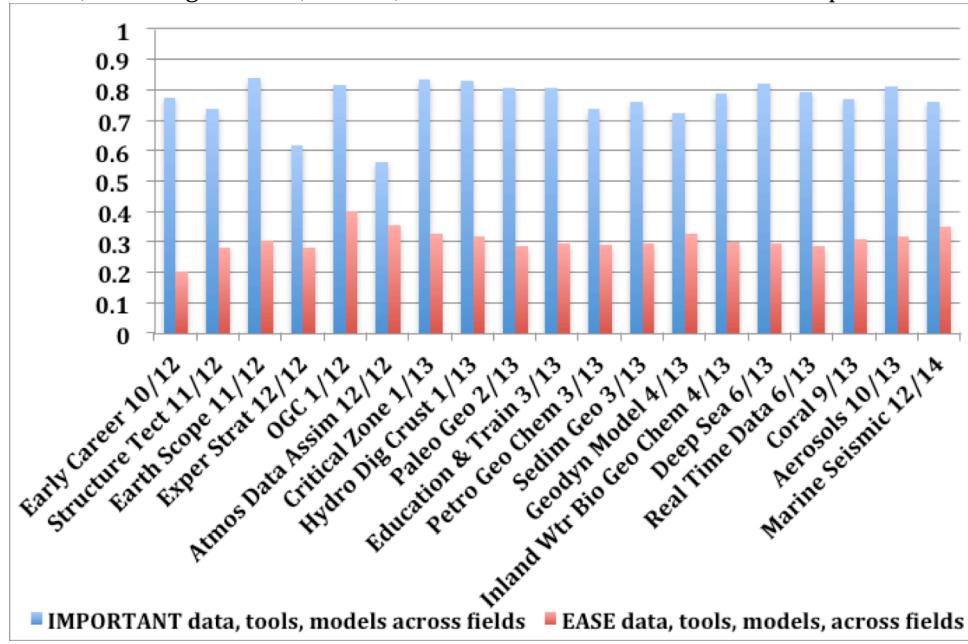
End user workshop reports and other materials are available from <http://www.earthcube.org/page/end-user-workshops>, including a compilation of workshop executive summaries.

Data from Participants of End User Workshops

Workshop participant views on how IMPORTANT it is and how EASY it is to find, access, and integrate data, models, and software WITHIN fields and disciplines:



Workshop participant views on how IMPORTANT it is and how EASY it is to find, access, and integrate data, models, and software ACROSS fields and disciplines:

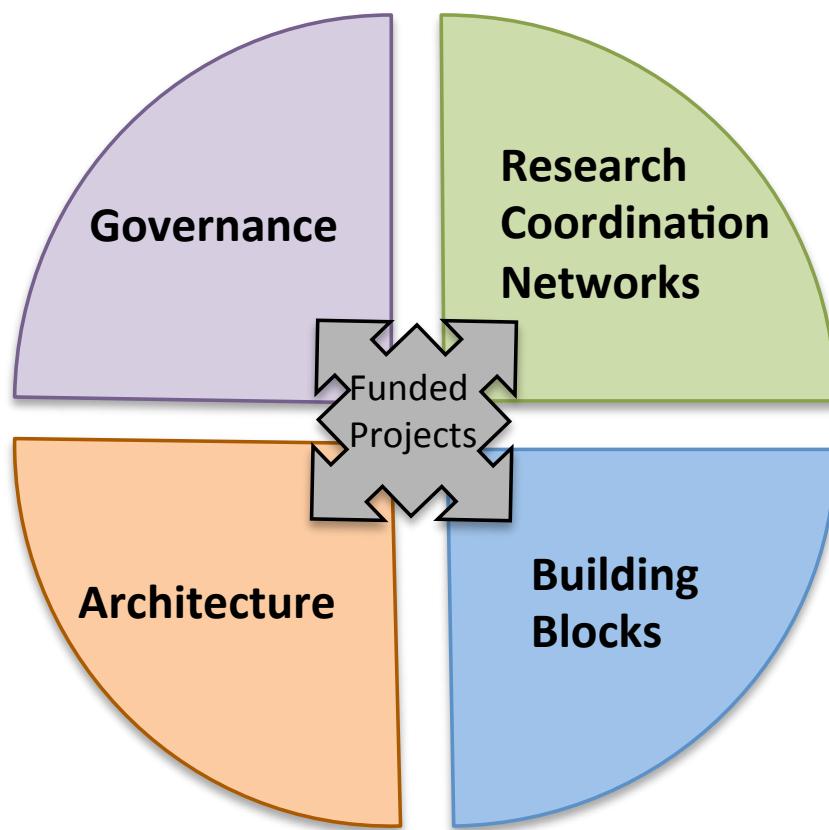




EarthCube

EarthCube Funded Projects

(2013 and 2014 Awards)

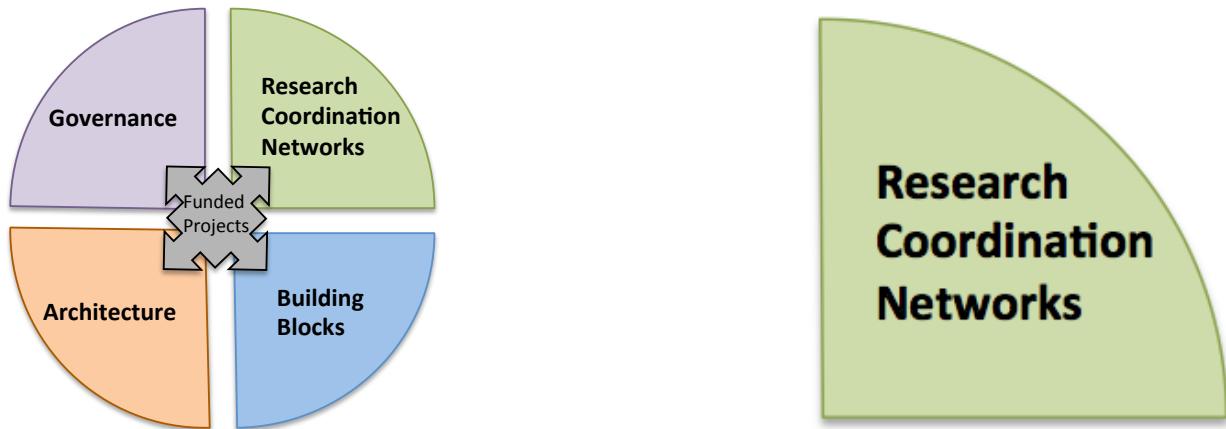






EarthCube

Research Coordination
Networks (RCNs)



EarthCube Research Coordination Networks

Engaging Science Communities

<http://workspace.earthcube.org/rcns>



EarthCube

Research Coordination
Networks (RCNs)



What Are Research Coordination Networks?

EarthCube Research Coordination Networks (RCNs) are intended to advance geosciences cyberinfrastructure by building and strengthening partnerships between geo- and cyber/computer / information scientists.

Examples of RCN outcomes include:

- Development of community standards, data citation or other community plans for data management in one or more field of the geosciences.
- Articulation of common cyberinfrastructure and technology grand challenges across different geosciences disciplines, including dialog towards designing potential solutions for data integration, computation, modeling, software and/or visualization needed to meet future scientific and education goals.
- Agreements on data and/or cyberinfrastructure issues involving multiple geosciences fields that will result in improved interdisciplinary access to products of scientific work or training and education.

Each RCN has a steering committee primarily composed of academic geoscientists. Cyber and/or computer / information scientists have key roles within the network. Network participants typically involve investigators at diverse organizations, including new researchers, post-docs, graduate students, and undergraduates. The RCNs include mechanisms to maintain openness, ensure access, and actively promote participation by interested parties outside of that initial list of participants.

Results from these projects will influence the direction of EarthCube, including architecture and geosciences-wide cyberinfrastructure developments.





EarthCube

Research Coordination
Networks (RCNs)

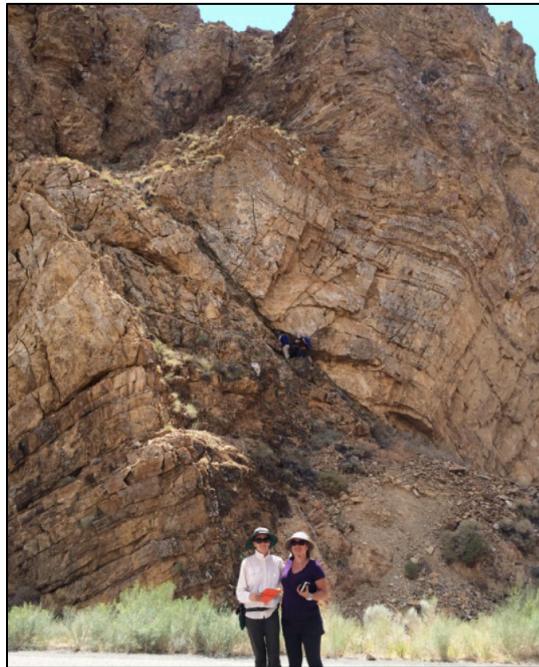
Portfolio of Research Coordination Networks

2013 Research Coordination Networks Awards

Name & Web Site	Project	Page
C4P	Collaboration and Cyberinfrastructure for Paleogeosciences	29
EC3	Earth-Centered Communication for Cyberinfrastructure: Challenges of Field Data Collection, Management and Integration	33
SEN	Sediment Experimentalist Network	39

2014 Research Coordination Networks Awards

Name & Web Site	Project	Page
CReSCyNT	Coral REef Science and CYberinfrastructure NeTwork	31
ECOGEO	EarthCube's Oceanography and Geobiology Environmental 'Omics	35
iSamples	The Internet of Samples in the Earth Sciences	37



C4P

Collaboration and Cyberinfrastructure for Paleogeosciences

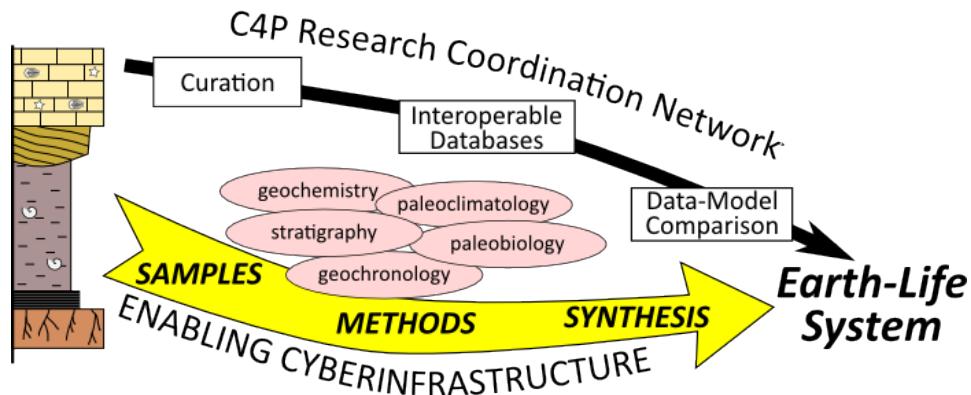
The community

The C4P project establishes and operates the EarthCube Research Coordination Network (RCN) *“Collaboration and Cyberinfrastructure for Paleogeosciences”* to advance the role of cyberinfrastructure in unraveling the large-scale, long-term evolution of the Earth-Life system through the study of the geological record.

This RCN intends to foster collaboration among **paleogeoscientists, paleobiologists, geo-and bioinformaticists, stratigraphers, geochronologists, geographers, data scientists, and computer scientists** with an aim to dramatically improve the application of modern data management approaches, data mining technologies, and computational methods to analyze and mine data within the paleogeosciences and other domains and disciplines.

Challenges

One of the primary challenges is to build new partnerships between domain scientists and cyber-and computer scientists that are essential for the development of useful, meaningful, and sustainable CI resources that will be adopted and embraced by the users and that can be maintained and evolve as part of a long-term infrastructure.



Approach

C4P is engaging a diverse community by facilitating communication and interaction between domain scientists and cyber/computer scientists through webinars, workshops, and web-enabled networking, and creating a descriptive catalog of existing CI resources that was initiated at the EarthCube Paleogeoscience End-User workshop and integrating it with evolving EarthCube inventories like CINERGI.

- 17 webinars are archived at the C4P YouTube channel (youtube.com/cyber4paleo)
- C4P workshop reports are available at workspace.earthcube.org/c4p



Participation

C4P's highest priority is to achieve broad network participation with inclusion of early-career through more senior members who can represent the paleogeosciences, computer sciences and bioinformatics communities. C4P conducts and participates in outreach events such as symposia, town hall meetings, and forums at major conferences (for example: GSA, AGU, ESIP), and coordinates with other CI developments by inviting active members of other RCNs and EarthCube projects to present at C4P webinars and participate in C4P workshops.

2014 Webinar participants included: CINERGI, NeotomaDB, Geosamples and IGSN, Encyclopedia of Life, PaleoPortal, Scientific Drilling Cyberinfrastructure, Paleobiology Database, iDigBio, GeoDeepDive, Specify Software, Lifemapper Computational Platform, NOW database, dbSEABED, GPlates, iSamples, and more.

The 2015 Webinar Series will resume in early February 2015.

Benefits to the broader geoscience community – next steps

The C4P RCN focuses on the development and deployment of standards for aggregation and dissemination of paleogeoscience data, with the goal of facilitating research focused on Earth-Life history. Increasing accessibility to “dark data” and hidden data and specimen collections are important goals of this proposed project. All activities of the RCN aim to facilitate community interaction across disciplines and to allow for broader participation amongst all participants.

PIs and Steering Committee Members



Kerstin Lehnert (PI)
Lamont-Doherty,
Columbia U.



Jack Williams
U. Of Wisconsin,
Madison



Mark D. Uhen
George Mason
U.



Chris Jenkins
U. Colorado at
Boulder

Douglas Fils (Ocean Leadership)

W Christopher Lenhardt (RENCI)

Anders Noren (U. Minnesota)

Thomas Olszewski (Texas A&M)

Lisa Park-Boush (U. Connecticut)

Dena Smith (U. Colorado at Boulder)

David Anderson (NOAA) (2013-2014)

Leslie Hsu (Lamont-Doherty, Columbia U.)

This project is funded by the National Science Foundation under the EarthCube initiative through grant 1340301. For more information, contact Kerstin Lehnert at lehnert@ldeo.columbia.edu or visit the project web site at <http://workspace.earthcube.org/c4p>.

The community

The Coral Reef Science & Cyberinfrastructure Network (CReSCyNT) is a multi-tiered and multidisciplinary EarthCube Research Coordination Network comprised of a dynamic, diverse, durable, and creative community of coral reef researchers, ocean scientists, cyberinfrastructure specialists, and computer scientists. CReSCyNT is an organically assembled network of interacting geoscience and technology nodes comprised of scientists across the boundaries of biological, physical and chemical oceanography, climate science, remote sensing, modeling and engineering working in the domain of coral reef science.

Challenges

Information generated by the multidisciplinary research of coral reefs crosses biological scales from genes to ecosystems, ranges in temporal scale from nanoseconds to millennia, and spans spatial scales from nanometers to kilometers. The data represent both static and continuous measures and qualitative and quantitative traits that are very different in nature. Adding to the challenge is the fact that individuals or small groups collect most coral reef data: large collaborative endeavors are rare and the field is highly competitive and has a culture reticent to share data. Given the complex array of data types generated by investigators a range of expertise who are working in locations that are biologically, socially, and culturally distinct, the task of managing and integrating data within the coral reef domain arguably encompasses many of the challenges implicit in the broader effort to create integrated data management infrastructures across the Geosciences.

Approach

CReSCyNT nurtures the growth of core activities in a network of geoscience nodes, where each node represents a sub-field or discipline within the coral reef or computer science domains. Each of these nodes will be led by one or two recognized members of that subdiscipline who embrace the values of the facilitating committee and EarthCube in advocating sharing, transparency, open dialog, creativity and innovation to promote the agenda of integrating data across the geosciences. These node coordinators (NCs) will capitalize on the positions of trust and respect they hold in their respective disciplines, as well as their sensitivity to their specific subdisciplinary culture, to



Coral reef science generates image data from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image data handling and analysis.



EarthCube

Research Coordination Networks (RCNs)

assemble a network of researchers in that disciplinary node that encompasses a diversity of career stages, geographic distribution, ethnicity, and gender to capture the complete range of activities conducted within that node's discipline. Participation will be facilitated by CReSCyNT meetings (in person and web based), social media, and virtual workspace.

Participation

CReSCyNT welcomes participation from all interested coral reef scientists. CReSCyNT is building on two NSF Earthcube-supported workshops that engaged more than 50 U.S. academic coral reef researchers on the data and cyberinfrastructure needs of the coral reef science community. The facilitating committee is identifying individuals to serve as node coordinators for each subdiscipline: node coordinators identify key players in their subdiscipline and recruit them to the node network for their subdiscipline. Each recruited network member will be asked for suggestions of others who might be interested, and the group will grow accordingly. These nodes will be allowed to expand, coalesce, or divide to meet the needs and interests of the subdisciplinary communities, while maintaining a connection to CReSCyNT through the node coordinators and ongoing network activities.

Benefits to the broader geoscience community – next steps

Coral reef scientists have an urgent need to address the management and analysis of diverse image data that spans scales from Transmission Electron Micrographs to satellite derived images of whole reef tracts. Collection of image data is rapidly expanding, and there is a deep legacy of image data not yet digitized, catalogued, or analyzed. Through the EarthCube network, CReSCyNT is developing partnerships with disciplines that are further advanced in image data handling and analysis to identify best practices. CReSCyNT will translate these best practices for the coral reef science domain and to other geosciences facing these same challenges.

While domain-centered nodes represent “horizontal”, question-driven science, we are also facilitating a system-centered “vertical” integration through place-based collaborations. As part of our CReSCyNT activities, we are overlaying the domain-centered network with a geospatial network to identify untapped collaborations between individuals working at the same location, gaps in expertise at particular locations, and areas of weak network connectivity that may result in redundancy and data silo-ing. Our experience integrating both horizontal and vertical expertise will translate to other geoscience disciplines.



Ruth Gates
Hawai'i Institute
of Marine Biology,
University of
Hawai'i at Mānoa



Megan Donahue
Hawai'i Institute of
Marine Biology,
University of
Hawai'i at Mānoa



Erik Franklin
Hawai'i Institute of
Marine Biology,
University of
Hawai'i at Mānoa



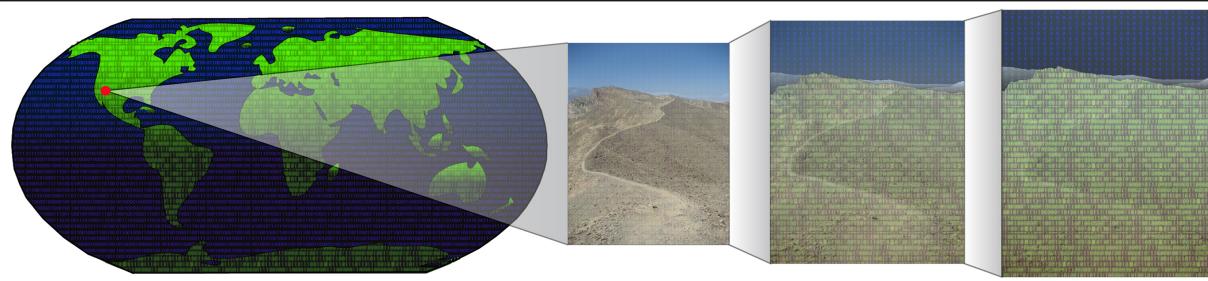
Gwen Jacobs
Information
Technology,
University of
Hawai'i at Mānoa



Judy Lemus
Hawai'i Institute of
Marine Biology,
University of
Hawai'i at Mānoa

EC3

**Earth-Centered Communication for Cyberinfrastructure:
Challenges of Field Data Collection, Management and
Integration**



The community

EC3 is attempting to engage all geoscientists who collect field data. We are building connections between this field-based geoscience community and members of the cyberinfrastructure communities.

Challenges

The main challenge for the field-based geoscience community is that our data are most often collected in an analog fashion (e.g., analog compass and field notebook). The onerous nature of transferring data from a field notebook into a digital format has prevented many researchers from sharing their complete field datasets with the larger scientific audience.

Approach

In order to come up with solutions that address these challenges and are technically feasible and sync up well with our existing/developing cyberinfrastructure, we are getting field scientists and computer scientist together in the field, on a four-day field trip to discuss these topics. There is no better place to have these conversations than the field.

Participation

Anyone can apply to participate in one of these field trips. See the list below of the 2014 summer field trip participants. There will be another field trip in the summer of 2015. Interested individuals can download the application from our web page (the above link).

Benefits to the broader geoscience community – next steps

In addition to the 2015 field trip, we will be hosting town hall meetings at both the AGU and GSA national meetings (2014 and 2015).

EC3 Steering Committee Membership: Richard Allmendinger, Cornell U; Jim Bowring, College of Charleston; Marjorie Chan, U of Utah; Amy Ellwein, Rocky Mountain Bio Lab; Yolanda Gil, U of Southern CA; Paul Harnik, Franklin and Marshall College; Eric Kirby, Penn State U; Ali Kooshesh, Sonoma State U; Matty Mookerjee, Sonoma State U; Rick Morrison, Comprehend Systems Inc; Terry Pavlis, U of Texas, El Paso; Shanan Peters, U of Wisc, Madison; Bala Ravikumar, Sonoma State U; Paul Selden, U of Kansas; Thomas Shipley, Temple U; Frank Spear, Rensselaer Poly. Inst; Basil Tikoff, U of Wisc, Madison; Douglas Walker, U of Kansas; Mike Williams, U of Mass., Amherst

2014 field trip mix: 10 geologists, 12 computer scientists, 4 graduate students, 2 undergraduates, 1 high-school student, 1 cognitive scientist, 2 applied linguists.



2014 field trip participants: Geology: Joe Andrew (UofKS), Forest Fortescue (NCRWQCB), Stephen Dornbos (UofW, Milwaukee), Mike Williams (Amherst), Basil Tikoff (Madison), Frank Spear (RPI), Terry Pavlis (UTEP), Matty Mookerjee (SSU), Marjorie Chan (UofUT), Steven Whitmeyer (James Madison); Comp Sci: Ilkay Altintas (SDSC), Michael McClenen (Madison), Jesse Eickholt (Central MI), Jessica Good (AZGS), Adrian Sonnenschein (AZGS), Yolanda Gil (USC), Susan Winters (UofMD), Sibel Adali (RPI), Michael Haberman (UofIL), Ken Lawrie (BGS), Nichola Smith (BGS), George Percivall (OGC); Grad Students: Sarah Ramdeen (Chapel Hill), John Donahue (NM Tech), Magdalena Donahue (NM Tech), Zachary Michels (Madison); Undergrads: Daniel Vieira (SSU), Anna Zeng (USC), Kevin Zeng (USC); Cognitive Sci: Thomas Shipley (Temple U.); Applied Linguistics: Mick Smith (UCLA), Charles Goodwin (UCLA).

This project is funded by the National Science Foundation under the EarthCube initiative through grant award number 1340265. For more information, contact Matty Mookerjee at matty.mookekerjee@sonoma.edu or visit <http://www.sonoma.edu/users/m/mookerje/EC3>.



ECOGEO

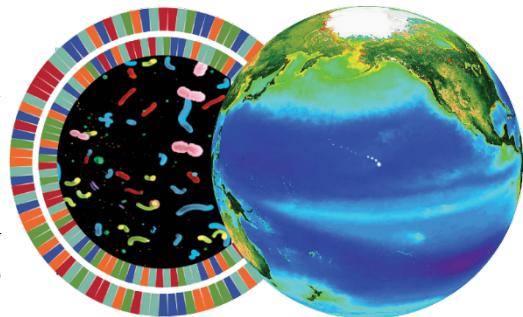
EarthCube's Oceanography and Geobiology Environmental 'Omics

The community

The ECOGEO community connects researchers from an array of disciplines within oceanography and geobiology that employ 'omics techniques to answer a variety of scientific questions. We are united by our investigations into high-throughput data of DNA (genomics, metagenomics) and RNA (metatranscriptomics), as well as proteins (proteomics), sugars (glycomics), and even lipids (lipidomics). This vast array of complex cellular component data represents the core architecture underpinning biological processes, community dynamics, and interactions between life and its environment.

Challenges

One of the main challenges in 'omics research is the standardization of procedures – from sample and metadata collection, to data formatting and analysis – that allow for integration across individual studies, 'omics domains (listed above), and spatiotemporal scales. The community is also currently limited by storage and large-scale, high-performance computational resources capable of archiving, sharing, processing, and analyzing large 'omics data sets. Finally, we lack visualization and statistical tools that function across a heterogeneous data landscape (e.g. connecting environmental parameters to complex meta-'omics sequence data to physical and biogeochemical models).



These challenges are not limited to the 'omics community. Any field that utilizes "big data" to answer complex science questions will have to address similar issues with data integration, storage, and communication. We have joined forces with EarthCube so that we can address, and solve, these challenges as a larger geoscience community.

Approach

The ECOGEO RCN has three main goals to address these challenges and provide results:

- Create a strategic network and community of field and cyber scientists to explore new facets of 'omics data.
- Articulate needs, challenges, and practical solutions that address: 1) development of infrastructure, 2) integration and implementation of workflows, and 3) database and resource sustainability to support ocean and geobiology environmental 'omics research.
- Develop a community-based framework that integrates best practices for curation and analysis of 'omics data and metadata, and facilitates collaboration and training among environmental microbiology, geobiology, and computer science disciplines.



EarthCube

Research Coordination Networks (RCNs)



Participation

Participation in ECOGEO is open and available to anyone using, or curious about using, environmental ‘omics data to explore scientific questions. Our initial recruitment strategy has been two-fold: creation of a survey to identify community needs with respect to ‘omics research (distributed throughout our steering committee’s networks, several NSF Science and Technology Centers, EarthCube’s newsletter and social media, several professional societies, and the Gordon and Betty Moore Foundation), and an interactive website at EarthCube that allows community members to join ECOGEO and sign up for our listserv. As this RCN is still under construction, the range of fields and domains we represent is far from complete. However, our initial efforts are primarily targeting the ocean science and geobiology communities.

Benefits to the broader geoscience community – next steps

Outcomes from the ECOGEO RCN will provide a much needed, and community-designed, federated framework for ‘omics research. We will have highlighted key areas for development and resource allocation so that further ‘omics community efforts can be focused and efficient. This federated framework will also allow the ocean/geoscience and cyberinfrastructure communities to conduct higher-level analyses, such as integration across ‘omics domains and environmental data, provide training for future ‘omics scientists through dynamic and practical use-case studies, and better adapt to the rapidly changing technologies for data acquisition and analysis.



Edward F. DeLong – Lead PI

Professor

Department of Oceanography, University of Hawai'i, Mānoa
Department of Civil and Environmental Engineering, and
Department of Biological Engineering, MIT

This project is funded by the National Science Foundation under the EarthCube initiative through grant 1440066. For more information, contact Ed DeLong at edelong@hawaii.edu or visit the project web site at <http://workspace.earthcube.org/ecogeo>.

iSamples

The Internet of Samples in the Earth Sciences

The community

The iSamples RCN builds, grows, and fosters a community of practice, in which domain scientists, curators and managers of sample repositories and collections, computer and information scientists, tool developers and technology innovators engage in and collaborate on defining, articulating, and addressing the needs and challenges of physical samples as part of digital data and information infrastructures. The ultimate goal of the RCN is to radically improve the discovery, access, sharing, analysis, and curation of physical samples and the data generated by their study for the benefit of science and society as part of the EarthCube program.

Challenges

The volume and diversity of samples that are collected across the Earth Sciences is huge. Samples have been acquired over decades or even centuries and are stored in a large number and variety of institutions including museums, universities and colleges, state geological surveys, federal agencies, and industry. Many samples are not curated at all, remain unknown and inaccessible to the broader science community due to the lack of a digital online presence that makes them discoverable and accessible, and are often at risk of being discarded due to lack of space and resources for sample curation. The RCN will address the wide range of issues to ensure proper digital documentation, identification, and accessibility of samples, as well as policies and practices for physical preservation.

Approach

The RCN will achieve its goals by focusing on three themes: *socialization*, *knowledge creation*, and *best practices*. *Socialization* will help build, grow, and foster the iSamples community, its structure, communication, and collaboration to converge on a vision of a shared cyberinfrastructure and related physical infrastructure for samples. *Knowledge creation* will help identify and articulate needs and challenges of physical samples in cyberinfrastructure, allow evaluation of existing practices, tools, and technologies, and support implementation of *Best Practices* and standards for sample identification, documentation, citation, curation, and sharing across the entire Earth Science community that the RCN aims to develop, implement, and promote.





Research Coordination Networks (RCNs)

Participation

Activities of the iSamples RCN include surveys and questionnaires (the first survey is going to be released before the AGU Fall Meeting 2014), workshops (first workshop to be held at the University of Texas as Austin in early 2015), and "Town Hall" events at various meetings including AGU, ESIP, and GSA. The RCN will specifically target early career scientists (an Early Career Forum will take place at the AGU Fall Meeting 2014) and engage the international community. More than 80 stakeholders from the domain, computer, and information sciences, including repository and collection curators, have already asserted their participation in the iSamples RCN. The iSamples Wiki (to be released in early 2015) will allow broad participation in the gathering and sharing of sample curation practices and resources and provide a registry of existing sample collections.

Benefits to the broader geoscience community – next steps

Samples in the Earth & Space Sciences are key to understanding the Earth's dynamical systems, its history and evolution, and the history and evolution of our solar system. They provide a basis for progress across many disciplines, from the study of global climate change now and over the Earth's history, to present and past biogeochemical cycles, to anthropogenic impact on soils and water quality. This RCN aims to better integrate samples into cyberinfrastructure, thereby creating exciting new opportunities to discover, share, and re-use physical samples, and allow previously impossible linking and integration of sample-based observations across data systems and within the scientific literature. Ultimately, improved digital management and access to samples will advance scientific discovery, productivity, effectiveness, and efficiency.

PI and Steering Committee Members



Kerstin Lehnert (PI)
Lamont-Doherty,
Columbia U.

Betty Adrian (USGS, NGDC)
Samuel Bowring (MIT)
Michael Denslow (NEON, Inc.)
Jane Greenberg (U. North Carolina)
Leslie Hale (Smithsonian Institution)
Benjamin Hallett (U. Wisconsin, Oshkosh)
Allegra Hosford Scheirer (Stanford U.)
W Christopher Lenhardt (RENCI)

Ann Molineux (U. Texas, Austin)
Carla Moore (NOAA, NGDC)
Anders Noren (U. Minnesota)
Lindsay Powers (NEON, Inc.)
Steve Richard (AZGS)
Ramona Walls (U. Arizona)
Joel Cutcher-Gershenfeld (U. Illinois, Urbana-Champaign)

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1440351. For more information, contact Kerstin Lehnert at lehner@ldeo.columbia.edu or visit the project web site at <http://workspace.earthcube.org/isamples>.

SEN

Sediment Experimentalist Network

The community

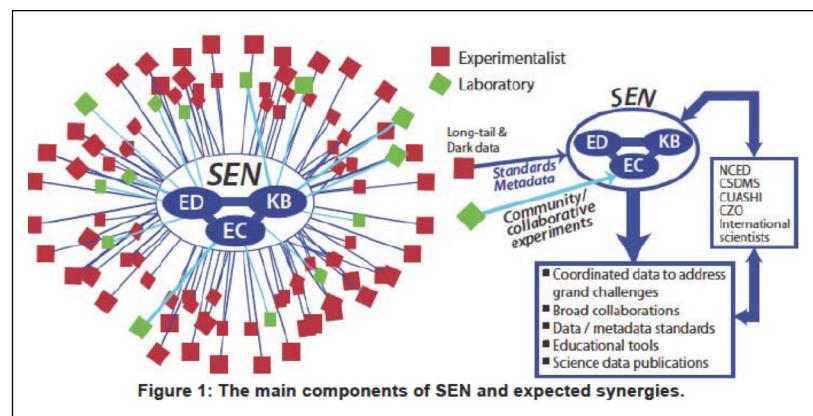
The **Sediment Experimentalist Network (SEN)** integrates the efforts of sediment experimentalists to build a knowledge base for guidance on best practices for data collection and management. The network facilitates cross-institutional collaborative experiments and communicates with the research community about data and metadata guidelines for sediment-based experiments. **This effort aims to improve the efficiency and transparency of sedimentary research for field geologists and modelers as well as experimentalists.**

Challenges

Research in experimental geomorphology is in a data-rich era with a level of high-resolution, digital data sets that was not available even a few years ago. Facilities are evolving to allow a greater number of more sophisticated experiments at ever larger scales and finer resolutions. However, these extremely large volumes of laboratory data are not easily accessible, often stored in individual hard drives or other digital media and gradually abandoned. This lack of organization comes at substantial cost to the experimental geomorphology community through lost opportunities for coordinating experimental data with models and field data. Building a global research network of these scientists, facilities, and data is key to addressing grand scientific challenges faced by the sediment experimentalist community.

Approach

There are three major components of SEN. The **Knowledge Base (SEN-KB)** provides a centralized place for people to post and request experimental data and method descriptions. The **Education and Standards group (SEN-ED)** provides training, especially for early career scientists, in order to promote a culture of data sharing and data stewardship. The **Experimental Collaboratories (SEN-EC)** provides a testbed for collaborative work involving data and tools sharing among many investigators, spanning a range of career stages and geographic locations, including educators and international scientists.





EarthCube

Research Coordination Networks (RCNs)

Participation

Participation in SEN is built through mailing lists, annual workshops, town halls, scientific sessions at the AGU meeting, twitter @sedimentexp, monthly newsletters, community experiments that are broadcast and tweeted in real time, and lectures at the National Center for Earth-surface Dynamics Summer Institute.

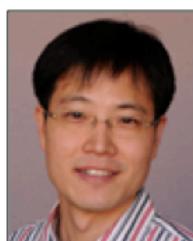
SEN is working with EarthCube through projects like CINERGI, GeoSoft, and Geosemantic Framework to bring the latest cyberinfrastructure to practicing domain scientists.

Benefits to the broader geoscience community – next steps

SEN-KB promotes research community progress through new experimental and educational tools and methods with example data and user needs for EarthCube cyberinfrastructure researchers. SEN-ED outreach and annual workshops train students and young geoscientists for best practices of data collection and sharing. The last two workshops in Asia and Europe built international ties, and a future joint meeting with CSDMS will integrate with efforts to provide well-documented and easily accessed experimental data as model benchmarks.

SEN serves the voice of the experimental geomorphology community to the cyberinfrastructure field through EarthCube projects, and brings cyberinfrastructure technologies back to the domain scientists for consideration, comment, and use.

PIs and Steering Committee Members



Wonsuck Kim (PI)
U. Texas
at Austin



Leslie Hsu
Lamont-Doherty,
Columbia U.



**Brandon
McElroy**
U. Wyoming



Raleigh Martin
U. California
Los Angeles

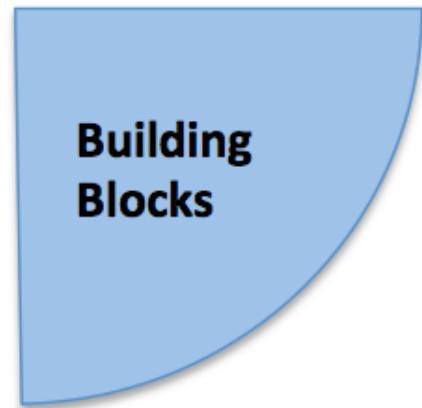
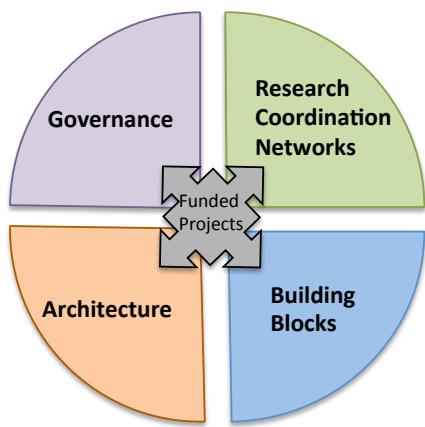
Kimberly Litwin Miller (U. Texas, Austin)
Kealie Goodwin (U. Texas, Austin)
Lisa Johnston (U. Minnesota)
Albert Kettner (U. Colorado, Boulder)
Francois Metivier (IPGP, France)
Tetsuji Muto (U. Nagasaki, Japan)
Charles Nguyen (U. Minnesota)
Aaron Packman (Northwestern U.)
Chris Paola (U. Minnesota)
Leonard Sklar (San Francisco State U.)

This project is funded by the National Science Foundation under the EarthCube initiative through grant EAR-1324760. For more information, contact sedimentexp@gmail.com or visit <http://workspace.earthcube.org/sen>.



EarthCube

Building Blocks



EarthCube Building Blocks

Exploring Solutions
and Demonstrating Utility

<http://workspace.earthcube.org/building-blocks>



EarthCube

Building Blocks



What Are Building Blocks?

EarthCube Building Blocks are created to leverage existing resources that have resulted from investments to date on cyberinfrastructure for geosciences and other sciences.

Building Block projects aim to:

- 1) **Integrate existing technology components** to extend capabilities to a broader set of geoscientists than are currently served,
- 2) **Create or modify cyberinfrastructure to overcome shortcomings** identified by the geosciences community, or
- 3) **Introduce novel cyberinfrastructure** into the geosciences

They are designed to demonstrate utility to geosciences communities within 24 months.

The Building Blocks must articulate how they extend and fit into an overall cyberinfrastructure ecosystem, and how the solution might be broadly applied across all geosciences.

The Building Blocks involve both computer scientists and geoscientists, and often include social and library scientists.



Portfolio of Building Blocks

2013 Building Block Awards

Name & Web Site	Project	Page
BCube	A Broker Framework for Next Generation Geoscience	45
CINERGI	Community Inventory of EarthCube Resources for Geosciences Interoperability	49
DisConBB	Integrating Discrete and Continuous Data	55
Earth System Bridge	Earth System Bridge: Spanning Scientific Communities with Interoperable Modeling Frameworks	59
GeoDeepDive	A Cognitive Computer Infrastructure for Geoscience	63
GeoSoft	Collaborative Open Source Software Sharing for the Geosciences	69
GeoWS	Geoscience Web Services	71
ODSIP	An Open Data-Services-Invocation Protocol (ODSIP)	73

2014 Building Block Awards

Name & Web Site	Project	Page
CHORDS	Cloud-Hosted Real-Time Data Services for the Geosciences	47
CyberConnector	Bridging the Earth Observations and Earth Science Modeling for Supporting Model Validation, Verification, and Inter-comparison	51
Digital Crust	An Exploratory Environment for Earth Science Research and Learning	53
EarthCollab	Enabling Scientific Collaboration and Discovery through Semantic Connections	57
GeoDataspace	Simplifying Data Management for Geoscience Models	61
GeoLink	Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences	65
GeoSemantics	A Geo-Semantic Framework for Integrating Long-Tail Data and Models	67



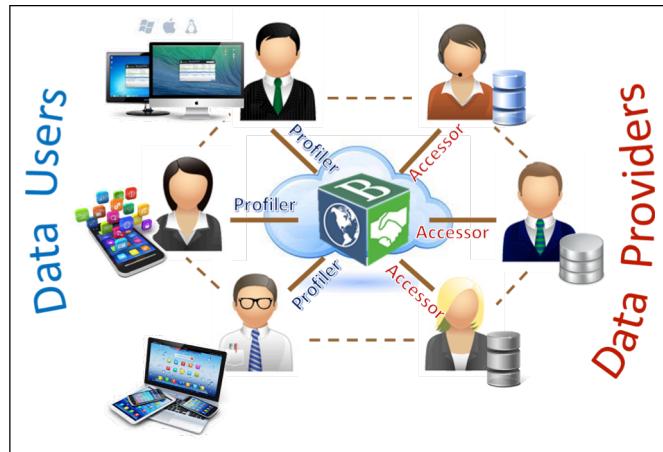
BCube: A Broker Framework for Next Generation Geoscience

Science Challenges

New advances in the geosciences require working across traditional disciplinary boundaries. This need has grown as scientists address the complex challenges of understanding and predicting changes in the Earth's environment. It is becoming ever more necessary to find, access, and use diverse and increasingly large volumes of data. The technology for managing and accessing scientific data is advancing rapidly, but leveraging these developments for interdisciplinary research is in itself a major challenge.

Technical Approach

The BCube team is exploring the use of brokering technologies, i.e. software components that mediate interactions between existing and emerging information systems, to demonstrate how brokering can make it easier for scientists to discover, share and access data, and thus become more effective and productive. Another focus of the project is improving access to difficult to find, yet valuable data by developing web crawlers to discover descriptions of data and services that are ready to be brokered.



A testbed, using Amazon cloud services, has been established and is being used for prototyping.

Science Drivers

Geoscientists on the BCube team, from the fields of Hydrology, Oceanography, Cryosphere and Climate, have defined detailed science scenarios that are guiding technology developments. A multi-disciplinary team, including educators and social scientists, is convened for each scenario to investigate the ability of the BCube infrastructure to meet the identified needs to the satisfaction of the scientists.



EarthCube

Building Blocks

Major Repositories Integration

Scenarios often require access to a wide variety of resources in forms other than their native formats. This leads to interconnecting into the testbed major repositories such as OOI (Ocean), USGS (Hydrology, NCAR/RAL (weather), and NSIDC (cryosphere).

Benefits to Scientists

Brokering and crawling are both intended to become part of a greater cyberinfrastructure aimed at serving all of the geosciences. The specific scientific benefits of the BCube project are being demonstrated and documented in the context of the science scenarios. The benefits are expected to propagate through the respective communities as the functionality is shared with colleagues and the capabilities of the BCube framework are expanded and integrated with other Building Blocks

The BCube Team

BCube Principle Investigator



From bottom right to top left: S. Browdy, R. Duerr, F. Pearlman, O. Schofield, D. Fuka, J. Pearlman, S. Nativi, S. Slota, E. Boldrini, M. Santoro, SJS Khalsa, Z. Easton

Siri Jodha Singh Khalsa, *University of Colorado, NSIDC*

Co-Principle Investigators

Ruth Duerr, *Crawler Design and Development, University of Colorado, NSIDC*

Stefano Nativi, *Broker Design and Development, CNR*

Jay Pearlman, *Cross-Domain Team, J and F Enterprise*

Francoise Pearlman, *Outreach and Communications, J and F Enterprise*

CyberInfrastructure Scientists

Steve Browdy, *OMS Tech, Enrico Boldrini, CNR, Mattia Santoro, CNR*

Geoscientists

Ocean: Oscar Schofield, Rutgers; Scott Doney, WHOI, and Wally Fulweiler, Boston U

Hydrology: Zach Easton and Dan Fuka, Vtech

Weather: Aaron Braeckel, NCAR REL

Cryosphere: Kevin Schaefer, NSIDC

Social Scientists

Geof Bowker and Steve Slota, UC Irvine

Education

Cheryl Peach, UCSD

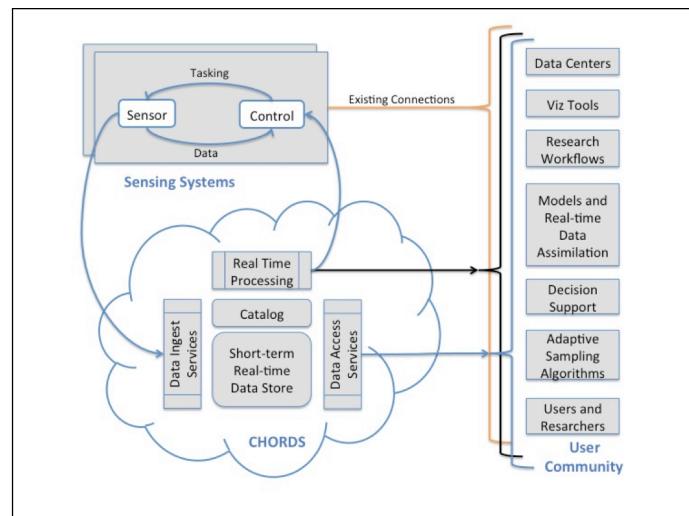
This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER 1343802. For more information, contact Siri Jodha Singh Khalsa via email: khalsa@colorado.edu.

**CHORDS
pilot****Cloud-Hosted Real-time Data Services for the Geosciences: a pilot project****Science Challenges**

The importance of real-time scientific data is ever increasing, particularly in mission critical scenarios, where informed decisions must be made rapidly. Advances in the distribution of real-time data are leading many new transient phenomena in space-time to be observed, however, real-time decision-making is infeasible in many cases that require streaming scientific data to be coupled with complex models. While EarthCube will provide an unprecedented framework for disseminating historical data sources, the use of real-time data raises an additional set of complex challenges. Our pilot project seeks to begin to investigate the role of real-time data within the geosciences to enable adaptive experimentation and real-time hypothesis testing.

Technical Approach

Real-time dissemination of research observations is often uncommon, which constrains science and limits the use of data to specific domains. The proprietary data distribution systems directed only toward displays and tools used by mission scientists, are usually expensive to develop, brittle, and painful to maintain. A user-friendly, simple and inexpensive web-based infrastructure will address these challenges, lower the barriers to entry, and enable broader real-time observational data collection, management and distribution.



Data and Service Interoperability will be important issues of the CHORDS project in an effort to perpetuate the use and incorporation of standards that will serve to improve the seamless interchange of sensor data. Community efforts, such as the Open Geospatial Consortium's Sensor Web Enablement (SWE) specifications can be instrumental in building interoperable infrastructure frameworks supporting improved sensor data capabilities. SWE, as well as other data and service specifications and standards relevant to sensor data and communications, will be evaluated by the project team as to applicability to the EarthCube sensor community and for potential recommendations as best practices.



Science Drivers

The CHORDS team is diverse and made up of technical and scientific representatives with connections to Hydrology, Computer Science, Solid Earth, Oceanography, Space and Atmospheric Sciences. Our team members serve as representatives of these science domains and will work with others to shape the CHORDS project as it develops. Our initial work will consist of defining requirements, designing specifications and ingesting a small subset of geosciences data streams into a CHORDS instance built in the cloud. We will also document the system sufficiently to allow it to work with new data sources in the future. The team will also participate in activities that strengthen the integration of real-time data being ingested via CHORDS into other EarthCube Building Block initiatives that are under development in order to advance science. At the conclusion of this pilot project, we will explore expanding the work of CHORDS into a more complete building block.

Benefits to Scientists

CHORDS stakeholders form an amalgam of scientists and university researchers who need to manage and distribute real-time data streams, but who do not have the time or resources to develop networking, software and computational infrastructure. The current CHORDS project is a prototype from which a full implementation can be built upon in the future. With minimal cost and effort, an instance of CHORDS will be created on a cloud server, and simply configured through a web interface. The researcher can then add simple functionality to their distributed data systems, which transmit observations (or any other real-time artifact) to CHORDS using straightforward web transactions. At the completion of a full CHORDS implementation, these workflows could include systems ranging from highly integrated command and control systems, data assimilation into models, field project control centers, standalone applications, or even web browsers and spreadsheets.

Project Principal Investigators (left to right)

PI: Mike Daniels, Manager of Earth Observing Laboratory's Computing Data and Software (CDS) Facility, National Center for Atmospheric Research

Co-PI: Dr. V. Chandrasekar, Professor of Electrical and Computer Engineering, Colorado State University

Co-PI: Dr. Sara Graves, Director, Information Technology and Systems Center, University of Alabama Huntsville

Co-PI: Dr. Branko Kerkez, Assistant Professor, Civil and Environmental Engineering, University of Michigan

Co-PI: Dr. Frank Vernon, Research Geophysicist at University of California at San Diego and Scripps Institute of Oceanography



This project is funded by the National Science Foundation under the EarthCube initiative through grant #1440133. For more information, contact Mike Daniels at daniels@ucar.edu.



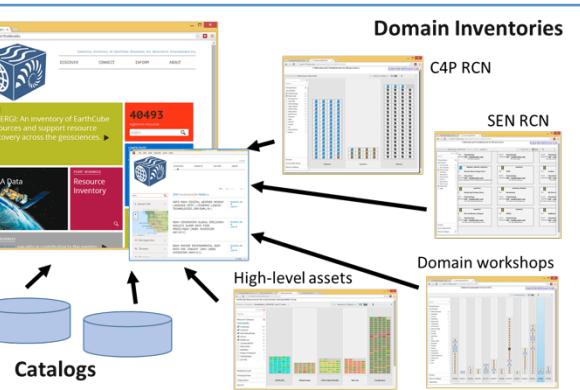
Science Challenges

Finding appropriate data is a difficulty that has been most often articulated by geoscientists during EarthCube end-user workshops. It becomes especially challenging when researchers work on interdisciplinary problems. Researchers finding and interpreting data across domain boundaries have to deal with unfamiliar terminology and research designs, implicit measurement assumptions, and disparate metadata and data formats. Despite the wealth of geoscience information available in digital form, and a plethora of databases, services and data portals already developed, there is no single inventory of available information across domains. The goal of CINERGI is to compile such an inventory, developing mechanisms to ensure that different resources have consistent and easy-to-interpret descriptions, traceable origins, and documentation that is as complete as possible. The scope includes datasets commonly catalogued by many organizations, as well as documentation for catalogs, vocabularies, data services, process models, repositories, etc. This inventory will help researchers answer both relatively simple and complex queries - in the latter case possibly requiring several iterations and a link to a domain data catalog for additional search options.

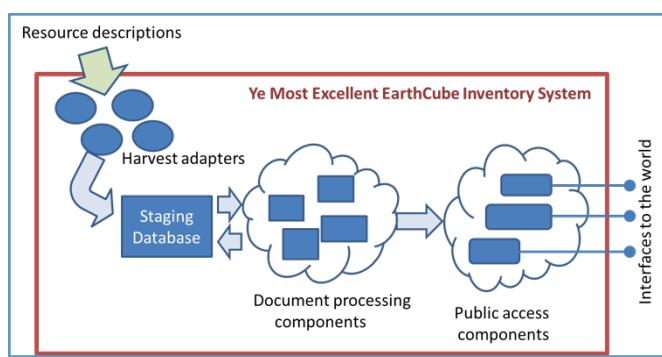
Technical Approach

Compiling and curating a large inventory of geoscience information resources requires integration of metadata records from standards-compliant catalogs maintained by domain data facilities and large projects, and information about data sources that are used and/or generated by multitude of smaller research projects, typically referred to as the “long tail of science”. While there are a relatively limited number of protocols for harvesting metadata from such catalogs, we found little consistency in metadata content.

To address this challenge, we are developing a CINERGI metadata processing pipeline: metadata documents are harvested using a number of adapters, loaded into a staging database, validated against content standards, and then processed to improve metadata content before being republished via a standard interface.



Metadata enhancements include checking and validating spatial extent, assigning an extent based on available information if applicable; analyzing and adding keywords to make the metadata easier to discover across domains; making the dataset title more descriptive; correcting temporal extent as needed; validating organization names against standard vocabularies; and adding standard thematic category and





resource type classification terms. As the enhancers change the content of the record, a corresponding provenance record is being created and made accessible via CINERGI search interface. Project information and links to inventories are at <http://workspace.earthcube.org/cinergi>

Science Drivers

Assembling and validating a large collection of geoscience metadata cannot be done without direct involvement of many groups of geoscientists, who specify which data resources are important for members of their domain, which metadata elements are important to expose for cross-disciplinary search, and validate assembled metadata and query responses. This engagement comes in several forms: (1) working with EarthCube Research Coordination Network (RCN) projects to jointly assemble resources used in their domains and make them searchable through the CINERGI system, (2) describing and registering resources mentioned by geoscientists in the course of EarthCube end-user workshops, in responses to EarthCube member surveys, and appearing in similar inventories; (3) interacting with managers of domain data facilities, (4) registering resources developed by EarthCube partners, and (5) exploring more complex query scenarios through collaboration with several geoscience researchers – in paleogeology, hydrology, and critical zone science.

Benefits to Scientists

CINERGI will reduce the burden of finding, interpreting and evaluating fitness-for-use of different types of information resources, across geoscience domains. A number of geoscience data facilities and projects - in geochemistry, hydrology, ocean sciences, ecology and other fields - have developed excellent data repositories and metadata catalogs: CINERGI will enable accessing them via a single standards-based catalog interface, and improve metadata descriptions to make data discovery more uniform and less time consuming.

Ilya Zaslavsky San Diego Supercomputer Center, UCSD	Stephen Richard Arizona Geological Survey	Tanu Malik Computation Institute, University of Chicago and Argonne National Lab	Amarnath Gupta San Diego Supercomputer Center, UCSD
Kerstin Lehnert Lamont Doherty Earth Observatory, Columbia University	Luis Bermudez Open Geospatial Consortium	Jeffrey Grethe Center for Research in Biological Systems, UCSD	

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1343816. For more information, contact Ilya Zaslavsky at zaslavsk@sdsc.edu.



CyberConnector

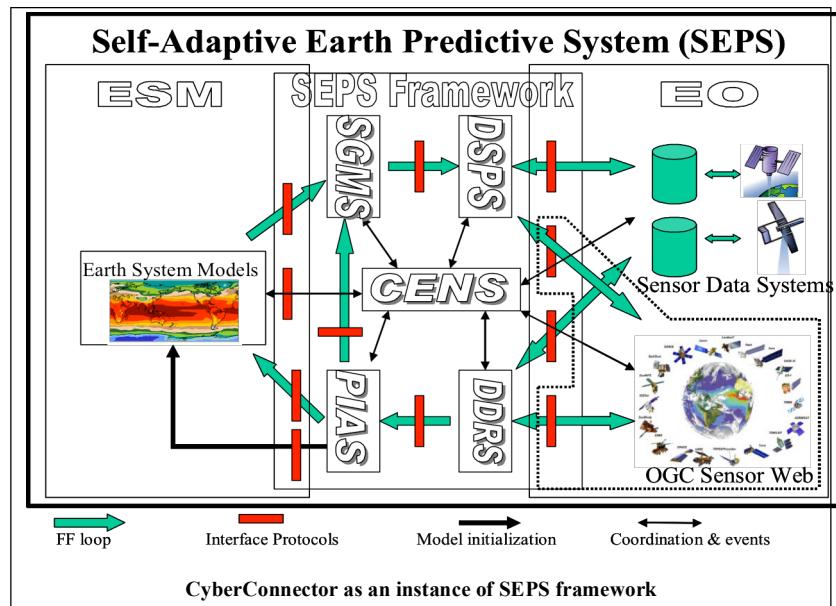
Bridging the Earth Observations and Earth Science Modeling for Supporting Model Validation, Verification, and Inter-comparison

Science Challenges

Modeling is one of major approaches for understanding the Earth system (ES) and predicting the future state of the system. Meanwhile, Earth observation (EO) via remote sensing is the most important method for measuring the current state of the Earth system. Products derived from EO data are widely used in model initialization, validation, verification, and inter-comparison. Currently, the process to prepare EO observation data for use in the modeling, as well as to release model outputs for public use, is done in ad hoc way. Such process is very laborious and time consumptive. This project will build the CyberConnector to automatically process the EO data into the right products in the right form needed for initialization, validation, and inter-comparison of ES models and to facilitate the publication of model outputs.

Technical Approach

This project leverages the online availability of huge volume of EO data at EO cyberinfrastructure of federal agencies (e.g., NSF, NASA, NOAA, USGS), and international EO organizations (e.g., CEOS members), advancements in standard-based geospatial web service and sensor web technologies, ISO and OGC standards on geospatial data and services, and successful cyber-systems built by the PI and others. The CyberConnector automates the process from EO data to products needed by ES models with the geospatial processing model (GPM) mechanism and workflow approach. The *CyberConnector* approach is a general one that can support many different ES models with the same cyberinfrastructure.



Science Drivers

The project includes the cloud, aerosol, and ocean modelers as the co-PIs. They will provide the science requirements to the project and ensure the results of this project are ready useful to the multiple modeling communities. The science Co-PIs will also be responsible to outreach to their community so that multiple communities can benefit from this project.



EarthCube Building Blocks

Benefits to Scientists

We expect the *CyberConnector* can reduce at least one order of magnitude in time needed for modelers to prepare model inputs from EO data and model outputs for data-model intercomparison and dissemination. The *CyberConnector* provides a new way for Earth science models to utilize the huge volumes of Earth observation data quickly and effectively.



PI: Dr. Liping Di, Center for Spatial Information Science and System (CSISS), George Mason University (GMU)



Co-PI: Dr. Ben Domenico, Unidata, UCAR



Co-PI: Dr. Haosheng Huang, Department of Oceanography and Coastal Science, Louisiana State University



Co-PI: Dr. Daniel Tong, CSISS, GMU



Co-PI: Dr. Xiaqing Wu, Department of Geological and Atmospheric Sciences, Iowa State University

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1440294. For more information, contact Prof. Liping Di at ldi@gmu.edu or visit <http://www.csiss.gmu.edu/cyberconnector>



Digital Crust

A 4D Data System on Crustal Material Properties and an Exploratory Environment for Geoscience Research and Learning

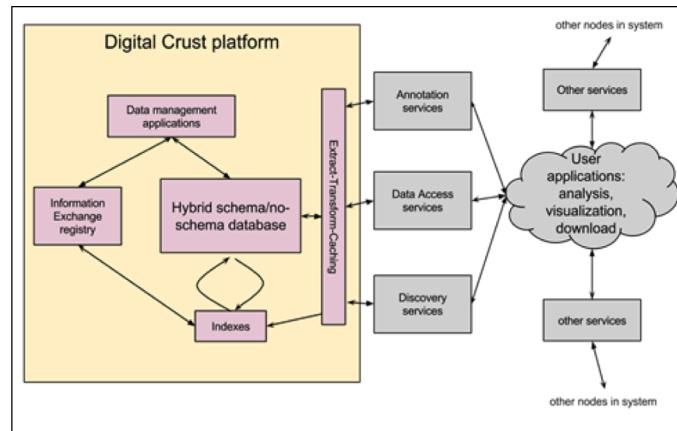
Science Challenges

Although the earth's upper crust has been mapped extensively, the resulting data have not been integrated into a single framework. Through the Digital Crust Building Block, we intend to build a 4D data system that integrates the existing data, and allows scientists to visualize and contribute to its content, conduct syntheses for detecting emerging patterns (e.g., Macrostrat), and derive 3D gridded data products such as crustal permeability for simulating fluid flow.

Technical Approach

The architecture is based on document-based data store (e.g. MongoDB) to allow an open-world information model providing flexibility/extensibility while promoting standardization. Information model registry provides platform for documenting content models and mapping between them.

A key function is providing services for mapping between spatial representations from field obs, syntheses, or models geolocated using vector features (points, lines, polygons), grids with various geometry, and hybrid schemes (e.g. Macrostrat). Web services provide loose coupling between data and user applications.



Science Drivers

The project is led by geoscientists and intended to serve any disciplinary scientist with interests in the earth's upper crust. It will be built on Macrostrat which has integrated many stratigraphic units and allows users to edit the columns. As a demo, we will produce a 3D permeability dataset for N. America for modeling groundwater and geochemical stores/fluxes in large-scale models.

Benefits to Scientists

Digital Crust can be a place for a scientist to find out what data/model already exist on the structure and composition of the earth's upper crust, visualize the geologic environment, contribute data and models, and derive data products for domain-specific applications.



Sky Bristol
USGS
Denver



Rick Hooper
CUAHSI
Boston



Steve Ingebritsen
USGS
Menlo Park



Nils Moosdorf
ZMT Bremen
Germany



Shanan Peters
Geoscience
U. Wis-Madison



Y Fan Reinfelder
Earth-Planet. Sci.
Rutgers U.



Steve Richard
AZGS
AZ



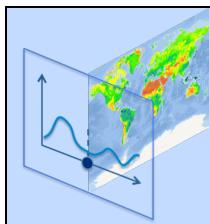
Ilya Zaslavsky
Spatial Info Lab
SDSC

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1440288. For more information, contact Ying Fan Reinfelder at yingfan@eps.rutgers.edu.



EarthCube

Building Blocks



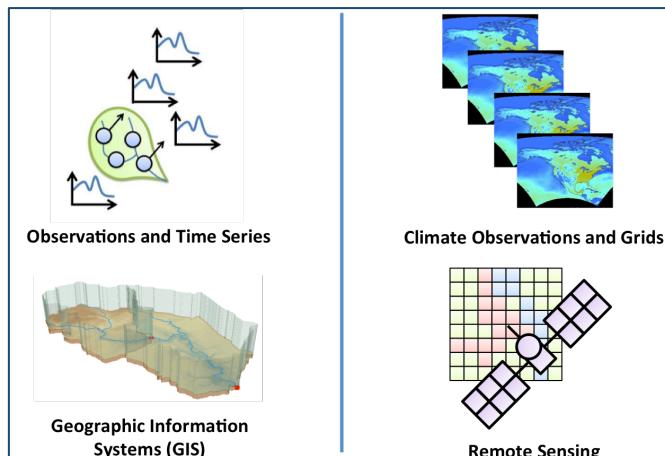
DisconBB: Integrating Discrete and Continuous Data

Bridging the “digital divide” between hydrologic and atmospheric sciences

Science Challenges

While we seek to understand similar phenomena in atmospheric data and water resource data, such as precipitation and evaporation, the information and tools for studying these phenomena are quite different. In the **hydrological science** studied at CUAHSI, there is an emphasis on time series of data collected at point locations, and a study typically involves a narrow region of space, a deep horizon of time, and synthesis of observed and modeled information for a few variables over this domain of space and time. In the **atmospheric science** information conveyed by Unidata, the focus is on near-term synoptic scale observations and modeling of a large number of interrelated variables conveyed as multidimensional arrays covering a large region of space for a short interval of time.

Fundamentally, these various classes of information can be described as spatially **discrete** for collections of points, lines and areas representing geographical features and their location over a study domain; and spatially **continuous** for raster grids and multidimensional arrays that are spatially distributed in a regular fashion through a study domain. Time series are more readily shared in hydrology, while gridded arrays are more readily shared in atmospheric science. There are also **semantic differences** in the way the same variables are described in hydrologic and atmospheric sciences – precipitation can have different names and meanings depending on the information source. We are seeking to develop a common information model and tools for unambiguous, interoperable exchange of data across these domains.



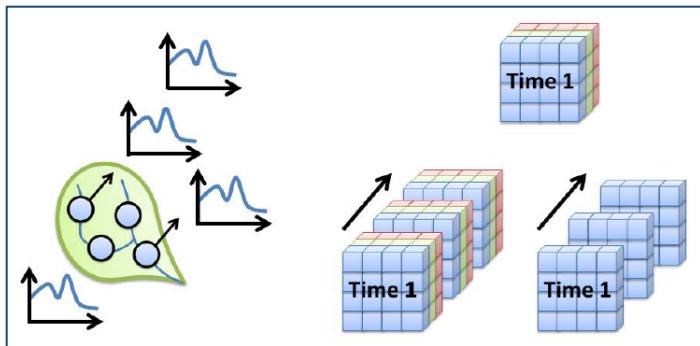
Technical Approach

The first step is to develop a **common information model** for representing both discrete and continuous data, having various temporal scales. Whereas one time series represents one variable for hydrology, a single gridded array can describe many variables and time sampling schemes. CUAHSI has developed an XML-based data structure called WaterML, which describes the type of observation and then presents the time series data for one variable. Unidata works primarily with netCDF, which implements the multidimensional



EarthCube

Building Blocks



array structure. Because this multidimensional array approach is the more flexible, we are investigating best ways of putting WaterML content into netCDF's information model, using the Climate & Forecasting (CF) Conventions version 1.6 with Discrete Sampling Geometries.

As we progress with this, we are also

developing tools to better visualize these relationships in space & time, which should help enable other domains of geosciences to take advantage of this approach.

Science Drivers

Scientists in hydrology, hydraulics, atmospherics, meteorology and informatics are involved in this project, making sure we're solving the right problems and coming up with correct results. Our demonstration project is a National Flood Interoperability Experiment, sponsored by USGS, NOAA National Weather Service, US Corps of Engineers, and FEMA. We have state-level engagement with the Iowa Flood Center, the Texas Division of Emergency Management, and engagement with 20 participating institutions.

Benefits to Scientists

Besides the direct benefits to climate, weather, and hydrology scientists, this work can be applied in other domains working with multidimensional spatial and temporal data, such as for seismic analysis, soil chemistry changes over deep time, oceanographic data, and many others.



PI: David Maidment, University of Texas at Austin

Co-PI: Ethan Davis, Unidata

Co-PI: Alva Couch, Tufts University

Co-PI: Dan Ames, Brigham Young University

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1343785. For more information, contact David Maidment at maidment@utexas.edu.



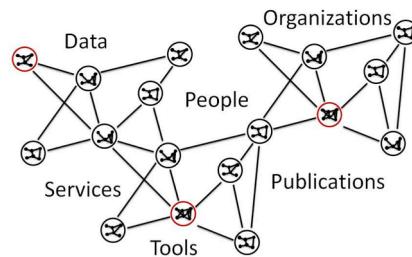
Science Challenges

Scientific projects are increasingly organized as “virtual organizations,” with many distributed participants, digital communications, information sharing, and heterogeneous products. This challenges the ability of researchers to bring information together from multiple sources, to identify which data sets, people, or tools, might be relevant to their research, and to understand how certain data, instruments, or methods were used to produce particular results.

Technical Approach

The EarthCollab project will demonstrate the benefits of linked open data for managing scientific information and data to improve the finding and sharing of information to advance research. The VIVO software suite will be the technology base for the project. Developed by Cornell University Library in collaboration with many partners, VIVO is used by over 100 organizations to manage information related to researchers, research projects, organizations, publications, data, and instruments.

Networked Science



We will leverage the VIVO software suite to create structured, interpretable linked data representations of our two case studies, permitting direct interlinking of information and data across platforms and projects. Our efforts will structure data into an ontology-based, standard data format (i.e., RDF) for re-use, leveraging web identifier and vocabulary structures that have been widely adopted in the geoscience and cyberinfrastructure communities. In addition, we will develop new capabilities within VIVO to facilitate exchange of information across linked data stores, thereby allowing research projects to leverage representations of researchers, data sets, tools, and organizations in linked data form.

Science Drivers

Our project focuses on two geoscience-based case studies: 1) An interdisciplinary field program, the Bering Sea Project, whose data archive is hosted by NCAR’s Earth Observing Laboratory (EOL), and 2) Diverse research projects informed by geodetic tools, such as GPS networks and LiDAR imagery, which are operated and/or maintained by UNAVCO. We have organized one workshop with researchers from these case studies, and plan to continue engaging with those communities as our project develops through additional workshops, targeted surveys, and via our advisory board.

Benefits to Researchers



EarthCube

Building Blocks

This project will provide insight into how the geosciences can leverage linked data to produce more coherent methods of information and data discovery for large multi-disciplinary projects and virtual organizations. The systems we build will enable researchers to more easily find people, organizations, and research resources that are relevant to their work. This in turn can help advance research and education endeavors within the geosciences and with other fields of study.

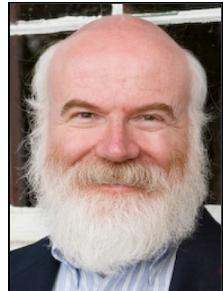
Project Principal Investigators (left to right)

PI: Matthew S. Mayernik, Project Scientist and Research Data Services Specialist, NCAR Library, National Center for Atmospheric Research (NCAR)

Co-PI: Michael D. Daniels, Manager of Earth Observing Laboratory's Computing, Data and Software (CDS) Facility, National Center for Atmospheric Research (NCAR)

Co-PI: Dean B. Krafft, Chief Technology Strategist and Director of Information Technology, Cornell University Libraries

Co-PI: Linda R. Rowan, Director of External Affairs, UNAVCO



This project is funded by the National Science Foundation under the EarthCube initiative through NSF grant #1440293. For more information, contact Matthew Mayernik at mayernik@ucar.edu.



Science Challenges

Scientific modeling helps us understand our environment. Diverse scientific communities team up to predict the environmental hazards of weather/climate, ecology, and deep Earth processes. However, technological and logistical gaps impede streamlined communication between academics whose research advances these predictions, and mission-driven federal agencies that use their tools for the benefit of society. There is also a need to integrate these modeling tools within and across academic communities. This project is developing innovative approaches that allow different Earth System modeling frameworks to work together or “interoperate,” thereby helping bridge these gaps, and enabling geoscience communities to collaborate and advance earth system science.

Technical Approach

The Earth System Bridge (ESB) team includes leaders from several major modeling frameworks and projects - both federal and academic - including: **CSDMS** (Community Surface Dynamics Modeling System), **ESMF** (Earth System Modeling Framework), **CIG/Pyre** (Computational Infrastructure for Geodynamics), **CUAHSI** (Consortium of Universities for the Advancement of Hydrologic Science) and **UCAR** (University Consortium for Atmospheric Research). Leaders from the **OMS** (Object Modeling System) and **OpenMI** (Open Modeling Interface) modeling frameworks are also contributing to the project.

Designing the Bridge: Framework Definition Language (ES-FDL)

The Earth System Framework Description Language (ES-FDL, pronounced ES-fiddle) is a metadata schema that is being developed to describe modeling frameworks, much like a blueprint. It will serve as a theoretical basis for understanding and connecting frameworks such as CSDMS, ESMF, the Model Coupling Toolkit (MCT), OMS, OpenMI and the OASIS coupler. The ES-FDL activity builds on and extend a series of international workshops that brought together developers of Earth system model coupling technologies for collaborative comparison and analysis (Dunlap et al., 2013). These workshops, coordinated by the International Working Committee on Coupling Technologies (IWCCT) resulted in improved understanding of individual coupling technologies and published reviews of the state of the art (Valcke et al., 2013). An anticipated next step is cross-framework analysis of features and development of connections and collaborations. The FDL will accelerate this goal by enabling more precise classification of frameworks, and by informing the nature of the bridges that can be built between them.

Building the Bridge, Span by Span: Sharing Resources Between Frameworks

One of the technologies being used and further developed for this project is the *Basic Model Interface* (BMI), developed for the CSDMS modeling framework. This is a *framework-agnostic* model interface that is easy to implement in different computer languages but provides the caller (any modeling framework) with fine-grained control via initialize,



EarthCube

Building Blocks

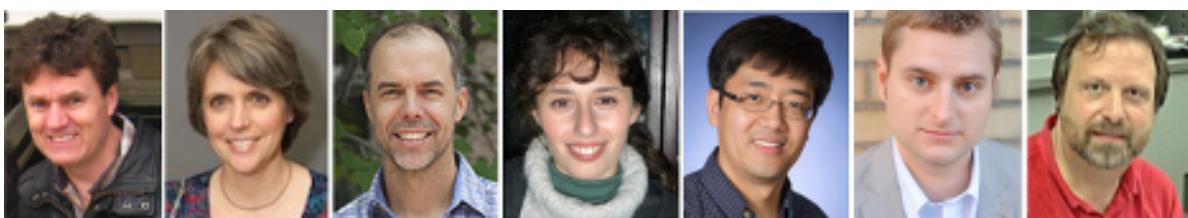
update and finalize methods, as well as a *standardized* description of its input and output variables, its grid(s), time-stepping scheme, and other information that is required by the framework to enable plug-and-play model coupling. The fact that it is framework agnostic means that developers don't need to adapt their model for any particular framework, and also means that adapters can be written that allow a BMI-augmented model to be used in multiple modeling frameworks. The CSDMS framework already supports BMI-compliant models and the Bridge team is working on adapters that connect BMI-compliant models to the ESMF/NUOPC , Pyre/St. Germain, OMS and OpenMI frameworks. A crosswalk between the CF (Climate and Forecasting) Standard Names and the CSDMS Standard Names is also being developed in support of this effort.



Crossing the Bridge: Linking Federal and Academic Models
One of the goals of the Bridge project is to use tools like ES-FDL, BMI and standardized metadata to link together multiple frameworks in order to create new modeling and predictive capabilities. The Bridge Team is partnering with the National Weather Service, CSDMS, and the WRF-Hydro project to link new models to the NOAA operational Environmental Modeling System (NEMS). There are two main parts to this effort:

- A regional atmospheric model called Nonhydrostatic Multiscale Model on the B grid (NMM-B) will be coupled to the regional Princeton Ocean Model (POM) for hurricane prediction using ESMF. This will involve testing the use of BMI within ESMF.
- The new NMMB-POM system in NEMS will be coupled to a land model that uses the hydrological components from WRF-Hydro. This effort links NOAA operational models, the NOAA Water Center, and the hydrological research community.

To demonstrate improved interoperability between modeling frameworks, the team is also linking a deep Earth process model called SNAC (a geodynamics / tectonics model, used by the CIG (Computational Infrastructure for Geodynamics) community with a surface process model called CHILD (a landscape evolution model), available within the CSDMS framework. SNAC is being augmented with a BMI interface and new CSDMS Standard Names are being developed for variables used in geodynamics models, with help from domain experts.



Scott Peckham
CU Boulder Cecelia DeLuca
CU Boulder/NOAA David Gochis
UCAR Anna Kelbert
OSU Eunseo Choi
Univ. of Memphis Rocky Dunlap
CU Boulder/NOAA Rick Hooper
CUAHSI

This project is funded by the National Science Foundation's EarthCube initiative, grant ICER-1343811. For more information, please contact Dr. Scott Peckham, Scott.Peckham@colorado.edu.



GeoDataspace: Simplifying Data Management for Geoscience Models

Science Challenges

Sharing and repeating geoscience applications is crucial for verifying claims, reproducing model/simulation runs, and promoting reuse of complex geoscience model applications. However, geoscientists lack effective mechanisms that enable easy sharing and efficient repeatability. It is not unusual for geoscientists to spend vast amounts of time and effort to capture, manage and organize the various data elements that a typical geoscience application or model requires to operate: the input files, processing and manipulation scripts, manifests, or databases that must be assembled and organized appropriately for the model to function correctly. When developing, testing, validating, and comparing models, particularly coupled models, the number of such data elements and the complexity associated with their management can soon outgrow, forcing researchers to narrow the scope of their application, compromise the quality of research, or conduct analysis within restricted teams. The goal of the GeoDataspace project is to make it easy for geoscientists to share and repeat their geoscience model applications in an efficient and effective way. The GeoDataspace system captures models and data in an integrated way, encapsulates them as a single shareable package, and allows the user to share/publish the package for wider community use or self-preserve it for further analysis.

Technical Approach

We are establishing the **GeoDataspace** framework, a cyberinfrastructure that will assist scientists and communities to create and maintain collections of *geounits* that pertain to a specific research project. For example, a GeoDataspace for solid earth might accumulate a set of geounits representing different simulation runs performed by different people, with different models, and with different inputs and parameters. Once created, geounits will provide a single handle to various model-related data items and source codes, offering benefits of shareability, reusability, and reproducibility during model development, testing, and validation.

Creating and establishing a geounit involves: (i) capturing, (ii) encapsulating, and (iii) cataloging. In the capture phase all the required data, processes, and environment variables are tracked and recorded comprehensively. In the encapsulate phase, the system creates a shareable package of the various recorded entities along with reference executions of how the data was consumed by different process and which data was generated. In the catalog phase, sufficient metadata from the data files is added to the shareable package for publishing the package and disseminating it to the wider community.

GeoDataspace is powered by Globus services, which provides a set of cloud-based services for scientific data management. Globus endpoints exist on more than 8000 nodes, and Globus drives the operation of end system software from the cloud, avoiding many end-user complexities of configuration and installation. Project information and links to source code, hosted geounits for various communities will soon be available from <http://workspace.earthcube.org/geodataspace>.



EarthCube

Building Blocks

Science Drivers

Sharing, repeating, and reproducing geoscience applications and models is inherent to every geoscience domain. Currently, we have engaged with geoscientists in four domains: (i) Solid Earth, (ii) Hydrology, (iii) Space Science, and (iv) Surface Dynamics, who have a critical need for using the GeoDataspace system to create shareable geounits of their respective modeling software. In Solid Earth, geounits are being created of the GPlates software, which ingests mantle seismic images obtained from the EarthScope project, stored in the IRIS Data Management Center, and merges the mantle convection modeling code (from the Computational Infrastructure for Geodynamics, CitcomS) with data assimilation software. In Hydrology, geounits are being established for the data-processing pipeline of the iRODS-enabled VIC model. In Surface Dynamics we are engaging with CSDMS to create shareable packages of coupled models, and in Space Science, we plan to track data flows in models developed by the CCMC Center at NASA Goddard Space Flight Center. In addition we are engaging with geoscientists at the governance level, the “big science” partners at NCAR and with single investigators at the “long tail of science” to understand their needs for sharing, repeating and reproducing scientific modeling software.

Benefits to Scientists

GeoDataspace will reduce the time to capture all data dependencies associated with a geoscience model/application, whether local or distributed. It will ease the process of curation by automating collection of some of the vital metadata associated with the data and source code files. It will provide mechanisms for them to easily and efficiently share their created software for repeatability and reproducibility, and publish their application into community-specific repositories. Created using robust, time-tested, and widely-adopted Globus services, the GeoDataspace framework aims to redefine sharing and reproducibility of geoscience models.

Tanu Malik Computation Institute, University of Chicago	Ian Foster Computation Institute, University of Chicago	Kyle Chard Computation Institute, University of Chicago	Mike Gurnis Seismological Laboratory, California Inst of Technology
Jonathan Goodall Dept. of Civil and Environmental Engg, Univ. of Virginia.	Scott Peckham INSTAAR, University of Colorado, Boulder		Joseph Baker Electrical and Computer Engineering, Virginia Tech.

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1440327. For more information, contact Tanu Malik at tanum@ci.uchicago.edu.



Science Challenges

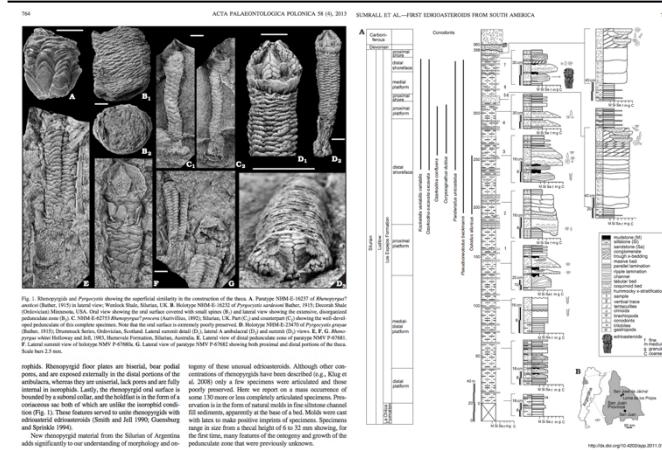
Text, tables, and illustrations in scientific publications are the primary means of distributing data and information in many domains of Earth and life science, particularly those based on physical samples. Although efforts are underway to integrate data archiving protocols into existing scientific workflows, a vast amount of valuable “dark” legacy data remains sequestered in publications. Manually assembling the synthetic databases necessary to test a wide range of hypotheses is labor intensive and results in static lists of predetermined facts that are separated from their original contexts.

Technical Approach

Here, we address problems in geoscience and computer science by developing a machine reading and learning system that meets or exceeds human quality in complex multi-lingual text- and image-based data extraction tasks. Our system builds upon the DeepDive machine reading and learning system and leverages the computational infrastructure of Condor and the US Open Science Grid. The input to the system is a set of documents (e.g., PDFs or HTML) and a database structure that defines the entities and relationships of interest. The first step in is to execute OCR, NLP, and other document processing tasks, which yield data that can be used to define features which in turn relate entities (e.g., by part of speech, or occurrence of entities in the same row of a table).

After extracting features, the next step is to generate a factor graph. Evidence variables derive from existing dictionaries and existing knowledge bases can be used in distant supervision. One key challenge is that factor graphs can be large. DeepDive uses recent research in theory and systems to overcome this computational challenge.

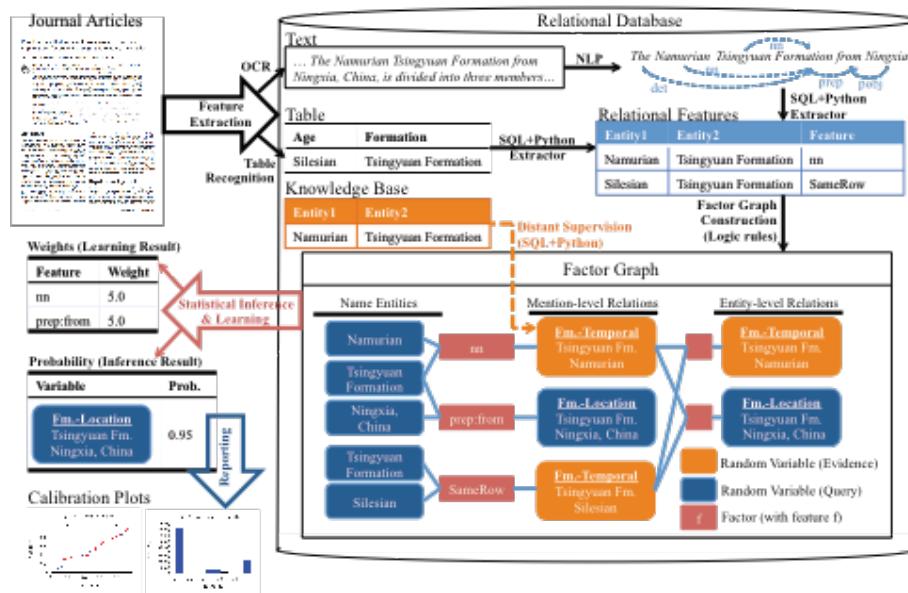
Given a factor graph, our system learns the weight for each factor and runs inference tasks to estimate the probability of each random variable. The output is a probabilistic database, in which each fact is associated with an estimated probability of being correct, and a set of statistics that describe the performance of the system. Improving quality is an iterative process.



Many important data are contained in publications, but manually locating, extracting and synthesizing them can be prohibitively time consuming.

Fig. 1. Shallowly and poorly showing the superficial similarity in the construction of the bivalve shells. A. Photograph NHM 5-1827 of *Brachiodontes*? sp. (Bivalvia: Carditidae) from the Lower Cretaceous of the Deltoid Group, St. Paul Island, Minnesota, USA. Out view showing the outer surface covered with small apertures (D) and lateral view showing the extremely disorganized punctation of the outer surface. B. Stratigraphic column from South America showing the distribution of different dolomite facies and the typical dolomitization of the complex sequence. Note that the out surface is extremely poorly pectinated. B. Holotype NHM 5-2349 of *Propecaena* sp. (Bivalvia: Carditidae) from the Lower Cretaceous of the Deltoid Group, St. Paul Island, Minnesota, USA. Lateral view of dorsal pectinate zone of pecten NHM P.7.981. E. Lateral view of dorsal pectinate zone of pecten NHM P.7.981. F. Lateral view of dorsal pectinate zone of pecten NHM P.7.981. G. Lateral view of dorsal pectinate zone of pecten NHM P.7.981. Scale bars: 2.5 mm.

morphs. Rhomboidal floor plates are vertical, but pedal pores are horizontal. In the third portion of the shell periphery, where they are oriented, lack pores and are fully rounded. The fourth portion of the shell periphery is bounded by a robust collar, and the bedding is in the form of a continuous, slightly irregular, horizontal layer about 1.30 m more or less completely encircled specimens. Pedal eversion is in the form of numerous lobes and ridges that will be used to make precise imprints of 1 to 2 mm showing, for the first time, many features of the ontogeny and growth of the peduncular area that were previously unknown.



Science Drivers

To validate our system and test the hypothesis that a machine can extract structured data from publications, we are attempting three test cases. The first, now completed and published (DOI: 10.1371/journal.pone.0113523), is to recreate and extend the human-constructed Paleobiology Database (<http://paleobiodb.org>). The second is to extract geochemical measurements, such as stable carbon isotopes, and integrate them with the Macrostrat database (<http://macrostrat.org>). The third is to extract data, such structural orientation data, from published geological maps.

Benefits to Scientists

In addition to developing general capabilities in machine reading that can benefit scientists in all domains, we are creating infrastructure to lower the barrier to text and data mining (TDM) activities. Our infrastructure involves the creation of a next-generation, TDM-ready digital library that is well curated, constructed in collaboration with publishers and library staff, and built directly on top of a high throughput computing capability.



Chris Ré
Dept. Computer Science
Stanford University



Shanan Peters
Dept. Geoscience
U. of Wisconsin-Madison



Miron Livny
Computer Sciences Dept.
U. of Wisconsin-Madison

This project is funded by the National Science Foundation under the EarthCube initiative through NSF grant ACI-1343760. For more information, contact Shanan Peters at peters@geology.wisc.edu.

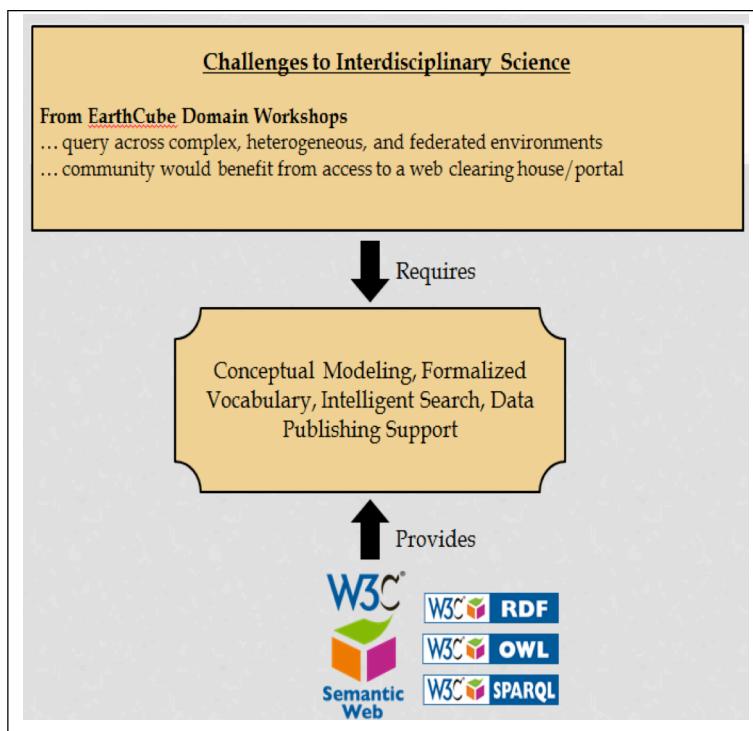


Science Challenges

Today's geoscientists need to discover, integrate, and synthesize data from multiple sources. Often, these data cross interdisciplinary boundaries or are being used for new purposes. Yet, geoscientists are often unaware of all of the data available to them. In addition, data are often heterogeneous and exist in federated environments. Discovering, sharing, and integrating data are currently major hurdles within, and across, geoscience domains.

Technical Approach

A wide spectrum of maturing methods and tools, collectively characterized as the Semantic Web, is helping to vastly improve the dissemination of scientific research. Creating semantic integration requires input from both domain and cyberinfrastructure scientists. The GeoLink project will demonstrate semantic technologies through the integration of geoscience science data repositories, library holdings, conference abstracts, and funded research awards.



GeoLink's outcomes will include a set of reusable ontology design patterns (ODPs) that describe core geoscience concepts, a network of Linked Data published by participating repositories using those ODPs, and tools to facilitate discovery of related content in multiple repositories..

Science Drivers

The challenges being addressed in GeoLink are scientific, technical, and social. As a result, this project involves experienced researchers and students from geoscience, computer science, and library and information science. Seeded with domain examples from the ocean and ecological sciences, our inter-disciplinary team is addressing this project from multiple angles.



EarthCube Building Blocks

Benefits to Scientists

GeoLink will greatly enhance the capabilities for scientists to discover and interpret relevant geoscience data and knowledge. It will lower barriers to cross-repository data discovery and access, while respecting and preserving repository autonomy and heterogeneity. The cyberinfrastructure underlying the approach is extendable, sustainable, and affordable - leveraging state of the art developments in Linked Open Data and formal semantics, grounded through shared ontology design patterns. The value of the approach will finally be showcased by developing a Web portal that enables searching and browsing of integrated content from multiple repositories, serving a diverse collection of geoscience communities.

Project Investigators:



Top left to right: Robert Arko, Lamont-Doherty Earth Observatory; Cyndy Chandler Woods Hole Oceanographic Institution; Doug Fils, Consortium for Ocean Leadership

Bottom left to right: Pascal Hitzler, Wright State University; Krzysztof Janowicz, University of California Santa Barbara; Tom Narock, Marymount University; Mark Schildhauer, National Center for Ecological Analysis & Synthesis

Additional Information

This project is funded by the National Science Foundation under the EarthCube initiative through grant 1440221. For more information, contact Robert Arko at arko@ldeo.columbia.edu or visit us on the web: <http://workspace.earthcube.org/geolink/>

GeoLink is a continuation of the EarthCube project OceanLink. For details on OceanLink and the transition to GeoLink see: http://workspace.earthcube.org/geolink/oceanlink_ieee_big_data.pdf

Or our technical report: http://workspace.earthcube.org/geolink/oceanlink_technical_report.pdf

The GeoLink team is engaged with other closely-related EarthCube efforts including the Community Inventory (CINERGI), the VIVO/Semantic Connections Building Block, and the Council of Data Facilities.



Science Challenges

Our goal is to enable the integration of long-tail data, i.e. data collected by individual researchers or small research groups, and long-tail models, i.e. models developed by individuals or small modeling communities, using a framework rooted in semantic techniques. We focus on these long-tail resources because despite their often-narrow scope, they have significant impacts in scientific studies and present an opportunity for addressing critical gaps through automated integration. We aim is to develop a decentralized knowledge-based platform that can be easily adopted across geoscience communities comprising of individual and small group researchers, to allow semantically heterogeneous system to interact with minimum human intervention.

Technical Approach

The framework consists of three layers: knowledge base, knowledge management, and Web application (Figure 1). The knowledge base defines upper level ontologies to describe spatial and temporal relationships among long-tail resources, and mini-ontologies that describe the scientific content of each geoscience discipline. Ontology Web Language (OWL) is used for coding the ontologies. The knowledge management layer implements JENA to ingest and harmonize the mini-ontologies. In addition, this layer includes packages for semantic mediation and matching between models and data. This tier turns the natural language arguments of a query statement to standard arguments that can be interpreted by models and/or data. The top layer of the framework represents the Web application, which is coded using Scala and Play to expose six services (Figure 2). These services will support and enhance the semantic integration between models, and model and data. Geo-Semantic framework source code is on Stash <https://opensource.ncsa.illinois.edu/stash/projects/ECGS>.

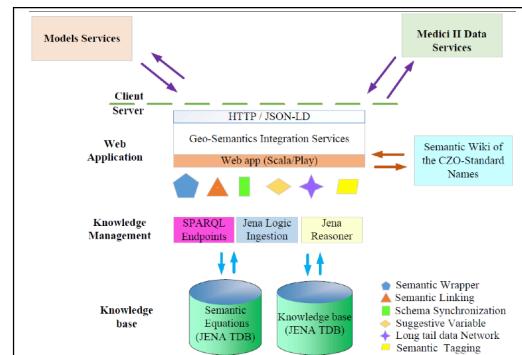


Figure 1 GeoSemantics Framework Architecture

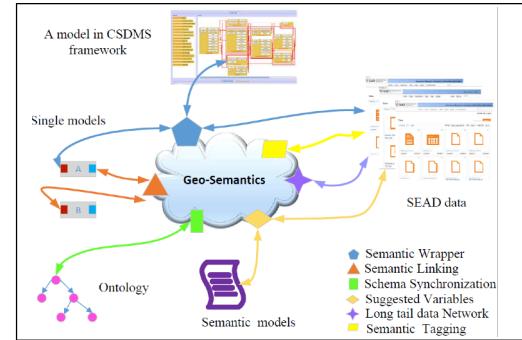


Figure 2 The Proposed Web Services

Science Drivers

To achieve our goal we will build on two existing technologies: (1) SEAD (Sustainable Environmental Actionable Data) funded by NSF DataNet Program since 2011 for supporting the full life-cycle of long-tail data, which includes collection, curation, discovery,



EarthCube

Building Blocks

sharing, and preservation; and (2) CSDMS (Community Surface Dynamics Modeling System) funded by NSF since 2007 which has been developing technologies to convert existing models into a plug and play system for interoperable integration. Both these technologies are broadly applicable to the broad field of geoscience and, therefore, serve as compelling prototypes. We will also build on ongoing EarthCube initiatives including GeoSoft, Earth System Bridge, and SEN (Sediment Experimentalist Network), and eWELL (Workforce Education and Learning Library) efforts in EarthCube. The outreach activates of the project includes collecting the community feedback through surveys, questionnaires, and informal communications. Currently, we are developing a Semantic Wiki, combination of wiki and Semantic Web technology, for CZO communities to share mini-ontologies about their geoscience process http://ecgs-dev.ncsa.illinois.edu/mediawiki/index.php/Main_Page. Finally, we invite input, feedback, and suggestions from the community at <http://workspace.earthcube.org/geo-semantic>.

Benefits to Scientists

The Geo-Semantic knowledge framework increases the semantic interoperability among geoscience communities by introducing a flexible software infrastructure to close the loop from models queries back to data sources in a seamless paradigm. The proposed framework will pioneer the development of a robust knowledge framework that will: (i) minimize human intervention in semantic mediation; (ii) minimize the context ambiguity of data that advances a machine physical interpretation of metadata physical meaning; and (iii) automate the crosswalks between geoscience standard names. This framework will directly impact research questions that require multidisciplinary interaction between different geoscience communities, where integrating models and data is a daunting process. We believe that advancing semantic interoperability among long-tail resources will enable the creation of a sophisticated loosely coupled system that enables the exploitation of models and data for scientific exploration in ways that have not been possible before.

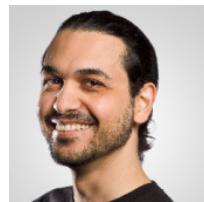
GeoSemantics Team



Praveen Kumar
Department of Civil
and Environmental
Engineering
University of Illinois
at Urbana-
Champaign (UIUC).



Mostafa M. Elag
Department of Civil
and Environmental
Engineering
University of Illinois
at Urbana-
Champaign (UIUC).



Luigi Marini
National Center for
Supercomputing
Applications (NCSA)
at the University of
Illinois at Urbana-
Champaign (UIUC).



Scott D. Peckham
Institute of Arctic
and Alpine
Research
University of
Colorado, Boulder,
CO.



Leslie Hsu
Lamont-Doherty
Earth Observatory,
Columbia
University,
Palisades, NY.

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER 1440315. For more information, contact Professor Praveen Kumar at kumar1@illinois.edu.



GeoSoft: Collaborative Open Source Software Sharing for the Geosciences

Science Challenges

Geosciences software embodies important scientific knowledge that should be explicitly captured, curated, managed, and disseminated. Recurring issues of provenance and uncertainty of data could be better addressed with improved treatment of geoscience software. Scientists recognize the value of sharing software to avoid replicating effort or to reproduce results from others. However, the stewardship of software in geosciences must be greatly improved. First, while modeling frameworks have dramatically improved software sharing, there are orders of magnitude more codes devoted to preparing data for input to a model (what we could call “pre-model software”) and preparing data that results from a model (“post-model” software). Studies show that scientists spend between 60%-80% of a project’s effort collecting and preparing data before doing new science. This software is only rarely shared (e.g., through libraries such as NetCDF-Java and NumPy) and rarely reused, particularly across disciplines. Second, the inaccessibility of software as an explicit science product that, like data, should be shared and reused leads to barriers for those who cannot afford such investments, and to barriers for software experts that could otherwise become involved in supporting software efforts. Finally, the education of future scientists crucially depends on the ability to actively explore problems by experimenting with science-grade data and software.

Technical Approach

The goal of GeoSoft is to enable the creation of a germinal ecosystem for software stewardship in geosciences that will empower scientists to manage their software as valuable scientific assets in an open transparent mode that enables broader access to that software by other scientists, students, and decision makers. Scientific software stewardship requires a combination of cyberinfrastructure, social



infrastructure, and professional development infrastructure. Our cross-disciplinary team has had direct experience with a variety of aspects of the scientific software lifecycle, from conception to development, deployment, characterization, integration, composition, and dissemination through open source communities and geosciences modeling frameworks.

The project focuses on:

1. **Facilitating software publication** through TurboSoft, a personal assistant that guides a user through best practices to componentize, describe, license, and maintain their software. It will assist in metadata capture for open source publication, the formation of



EarthCube

Building Blocks

communities around the software, and set up mechanisms for software citation and credit.

2. **Enabling broad software dissemination** through GeoSoft, a “software commons” for geosciences that supports software contributions (prepared through TurboSoft or otherwise), software discovery through multi-faceted search, and that fosters social interactions through dynamic formation of communities of interest. GeoSoft will interoperate with existing software repositories and modeling frameworks in geosciences. GeoSoft will be based on explicit incentives from social sciences and open source communities.
3. **Providing just-in-time training materials** through SoftCamp, an annotated collection of educational units (videos, screen captures, decision trees, reports) ranging from basic education to professional training on all aspects of software stewardship.

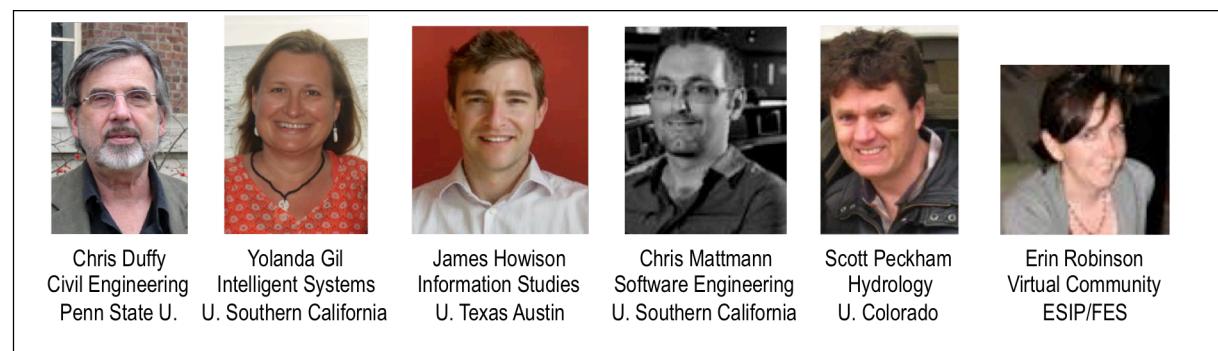
Science Drivers

Our work is driven by demonstration scenarios in the context of the Critical Zone Observatory (CZO) project. Many of the new CZO models are extensions to existing models but their development is often not coordinated with the management and ongoing development of the model they are derived from. For example, during the first years of the Shale Hills CZO funding, several water and energy models were developed and now these models are being extended to model transport processes, stable isotopes, biogeochemistry and the carbon-nitrogen system. Our framework will provide connections that will provide significant benefit to this emerging, cross-disciplinary science.

We also have close collaborations with several EarthCube Research Coordination Networks (RCNs), and actively gather requirements through our Early Career Advisory Committee that is diverse in its composition and cuts across geosciences disciplines.

Benefits to Scientists

The GeoSoft project will result in a germinal social site for the EarthCube, where scientists can discover alternative approaches to release free software, use intelligent interfaces to explain how their software works, and form productive communities around software projects. More broadly, this work will improve our understanding of how to promote software sharing in science, support better software stewardship, and capture metadata for scientific software.



This project is funded by the National Science Foundation under the EarthCube initiative through grants ICER-1343800 and ICER-1440323. For more information, contact Yolanda Gil at gil@isi.edu.



Science Challenges

We have all heard that researchers spend 80 to 90% of their time and effort finding and accessing needed data for their specific research problem and only 10% actually doing science. The fundamental goal of the GeoWS Building Block is to provide more uniform ways to discover, access, and use data from various earth science data centers. The GeoWS strategy is to promote the use of similar web services as methods of interacting with domain data centers. While the standardization of the space-time query for discovery is a high priority, the GeoWS approach is flexible enough for the general approach to be implemented at all of the GeoWS data centers on a variety of domain-specific data sets. Very deliberately, the GeoWS primary partners include centers from solid earth, atmospheric and ocean sciences, which represent the largest sections of the GEO Directorate at NSF, along with secondary partners from other geoscience-related disciplines.

Technical Approach

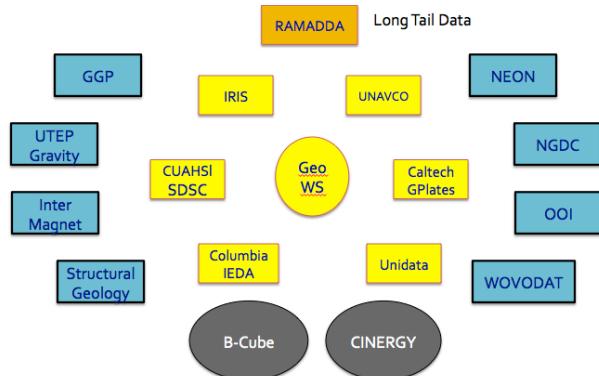
The GeoWS approach is to enable search, discovery, and access by similarly constructed URLs containing a traditional directory structure and query parameters in the URL. Each GeoWS partner has a base URL from which their services can be accessed. For instance, the base for GeoWS compliant services at IRIS is <http://service.iris.edu/fdsnws>. Specific services and versions are available under the base URL such as **/event/1** for the IRIS event service version 1. This can be followed by a series of parameters that provide constraints for the query. By default, information is returned as XML but to aid usability GeoWS services also supports easily understood, simple text content. For example, the following URL represents a query resulting in a text listing of all earthquakes greater than magnitude 6.0 occurring in the last two months of 2014 using the IRIS event service:

```
http://service.iris.edu/fdsnws/event/1/query?starttime=2014-11-01T06:30:00&endtime=2014-12-31T06:30:00&minmag=6&orderby=time&format=text
```

One can paste the above URL into a browser and get a listing of earthquakes matching the constraints provided. Other GeoWS partners have similar capabilities. Such services are designed for easy use in Perl, Java, curl, wget and many other data collection and processing environments. See <http://workspace.earthcube.org/geows/geows-web-services> for information about GeoWS services.

Science Drivers

GeoWS data partners were selected primarily to address a user scenario within geodynamics, augmented slightly to show that the concept could be extended to other disciplines supported by the NSF GEO Directorate. To further demonstrate flexibility of the approach we also have partners supported by the NSF Biological Directorate. Additionally, two partners of GeoWS are either funded by other federal agencies (NGDC) or non-US sources (WOVODAT). GeoWS engages the



The above figure shows the GeoWS partners. They include the 6 funded partners in yellow, the 8 unfunded partners that IRIS will implement similar web services for, in blue. The RAMADDA system (orange) will show examples of how it also can support GeoWS web services. Other Building Block projects that GeoWS is working with include B-Cube



EarthCube

Building Blocks

community through a Science Advisory Committee that will, among other things, act as testers of the GeoWS capabilities when fully deployed.

Benefits to Scientists

GeoWS will benefit a wide range of scientists by reducing the burden currently required to discover, access and use data across multiple scientific domains. While the services can be accessed directly at the providers' service locations, GeoWS is also working closely with two other Building Block proposals: the CINERGI Building Block (<http://workspace.earthcube.org/cinergi>) that serves as an EarthCube wide registry of services as well as the BCube (<http://workspace.earthcube.org/bcube>) building block that offers a brokering capability and a single portal through which the GeoWS data assets can be accessed.

The GeoWS partners include six data centers funded by the NSF GEO Directorate including the GPLates effort at Caltech, the IEDA: Marine Geoscience Data System at Columbia University, the IRIS Seismological Data Center, the CUAHSI effort at San Diego Super Computer, the Geodetic Data Center at UNAVCO, and the Atmospheric Sciences data facility at Unidata. Additionally, GeoWS is directly implementing web service concepts at 8 other locations including 1) NEON (ecological), 2) NGDC (various geophysical data sets), 3) Ocean Observatory Initiative (OOI) (oceanographic data), 4) WOVOdat (volcanological data), 5) GGP (superconducting gravimeter data), 6) InterMagnet (magnetic observatory data), 7) UTEP (gravity and magnetic collections), and 8) U. of Kansas (structural geology data). GeoWS is also supporting the developer of RAMADDA (<http://www.ramadda.org>) to incorporate GeoWS web services into the RAMADDA system where, in theory, any data collection can be managed.



Tim Ahern
Director of Data Services
IRIS



Suzanne Carbotte
Lamont Research Professor
IEDA (Columbia)



Michael Gurnis
Director, Seismological Lab
Caltech



Chuck Meertens
Director of Geodetic
Data Services
UNAVCO



Mohan Ramamurthy
Director
Unidata/UCAR



Ilya Zaslavsky
Director of Spatial Information
Systems Laboratory
SDSC / UCSD

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1343709. For more information, contact Tim Ahern (tim@iris.edu).

**ODSIP****Specifying and Implementing ODSIP, A Data-Service Invocation Protocol****Science Challenges**

The impedance induced by *remote* data usage is growing as geoscience becomes more data-intensive and multidisciplinary. This derives primarily from three factors.

1. Distance between data sources and users introduces latency and Internet reliance. Trans-discipline and international studies often entail large separations.
2. Growing data volumes exacerbate distance factors, even to the extent that data-to-computer latencies can make certain studies impractical.
3. Variations in form and format have long been obstacles data use between scientists, and the multidiscipline character of modern science heightens the impedance. This is compounded by distance, which can create language barriers and/or decrease the likelihood of finding data expertise down the hall

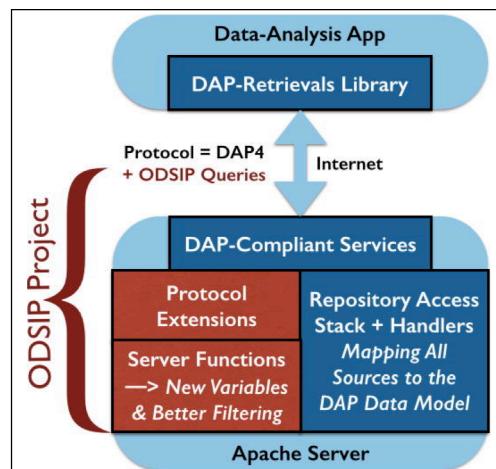
Technologies to surmount the above are themselves limited by the same distance and volume factors: transforming a dataset into something useful (i.e., addressing variations in form) may entail so much processing that it is practical only in close proximity to the data source (where latencies are minimal); even subset-selection (to reduce data-transfer volumes) may entail similar computation. However, data providers often do not consider processing to fall within their missions.

Technical Approach

Trends toward data-access as a (Web) service have already yielded progress in overcoming the above challenges. In particular, use of whole-file data-transfer (FTP, e.g.) is being supplanted by services (such as the Data Access Protocol or DAP) that offer certain filtering (i.e., subset-selection) services as an integral part of data-access.

ODSIP extends this concept to fully embrace the invocation of new server functions within the DAP framework. In other words, this project stretches the notion of data-as-service to prototype several pre-retrieval operations that lower the impedance of remote data usage. Naturally, the focus is on operations that *reduce* the volumes of data transferred to users.

Along with ODSIP capabilities for calculating variables not present in the source data, such as statistical summaries, an important set of server functions will be prototyped to enrich the DAP tools for selecting subsets from the source (or the calculated) data.





The ODSIP outcome will include a nascent algebra of pre-retrieval operations prototyped for use in the scientific contexts discussed below. These prototypes are expected to show major reductions in data-transfer demands and hence significantly increased usability of important remote data.

Science Drivers

Among the many areas of Geoscience study impeded by remote data-usage challenges, three have been selected as drivers for the ODSIP project. The investigator team includes scientists active in these fields, and the project efforts will yield prototype server functions targeted to each of their needs. In turn, the needs of these investigators embody sufficient mathematical breadth to indicate wide applicability of the ODSIP concept, once success has been achieved with the prototypes.

The three prototypes will be developed for:

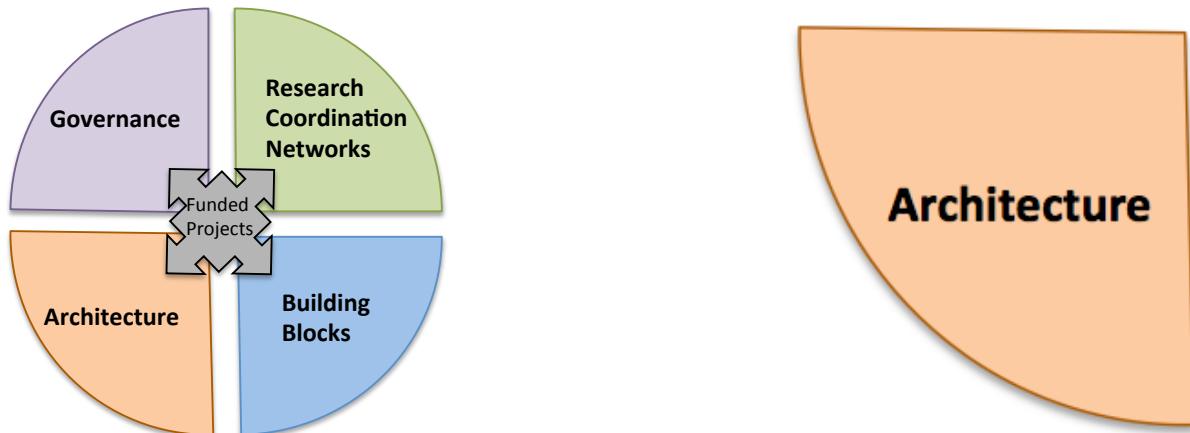
1. Climate-Model Downscaling — Joining climate runs with weather-model ensembles, estimate the probabilities of fine-scale events critical to native-Hawaiian well being.
2. Storm-Surge Prediction — From huge (triangular-mesh) models, give officials in North Carolina simplified info to help anticipate/prepare for coastal emergencies.
3. Analysis/Synthesis of Sea-Surface-Temperature (SST) Fronts — Using data from dissimilar satellites, estimate front locations and retrieve proximate imagery.

Benefits to Scientists

ODSIP is meant to help scientists undertake previously impractical studies, reducing—by orders of magnitude—the volumes of data that must be moved from source repositories to end-user premises. By embedding these new functions in DAP systems well established as secure and effective, an ODSIP premise is that data providers will welcome the prospect of offering pre-retrieval computation in exchange for reduced data transmission.

Dave Fulker OpenDAP	Brian Blanton U-NC	Steven Businger U-HI	Peter Cornillon U-RI	James Gallagher OpenDAP	Mohan Ramamurthy Unidata

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1343761. For more information, contact Dave Fulker at dfulker@opendap.org.



EarthCube Conceptual Designs

Initial Planning for Enterprise Architecture

<http://workspace.earthcube.org/conceptual-designs>



EarthCube

Conceptual Designs



What Are Conceptual Designs?

EarthCube Conceptual Design projects will generate reports describing an initial enterprise architecture design for EarthCube.

These projects interact with the EarthCube community to understand architecture requirements based on the scope of scientific challenges as well as the existing and planned cyberinfrastructure resources.

The architecture must address:

- The diversity of technologies and infrastructure approaches
- The incorporation of existing systems that are used by geoscientists
- The integration of diverse information in an easy-to-use system
- The dynamic nature of architecture requirements as new research opportunities and innovative technologies arise

The strategy for selecting the EarthCube enterprise architecture will be phased. In the first phase, conceptual designs will be reviewed by the community as the basis for the architecture. The second phase will focus on Design Refinement, where the concept designs will be revised to reflect additional requirements gathered through ongoing EarthCube activities.



Portfolio of Conceptual Designs

2013 Conceptual Design Awards

Name & Web Site	Project	Page
DAsHEr	Developing a Data-Oriented Human-Centric Enterprise Architecture for EarthCube	79
GEAR	Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences	81

2014 Conceptual Design Awards

Name & Web Site	Project	Page
SC-DA	A Scalable Community-Driven Architecture	83



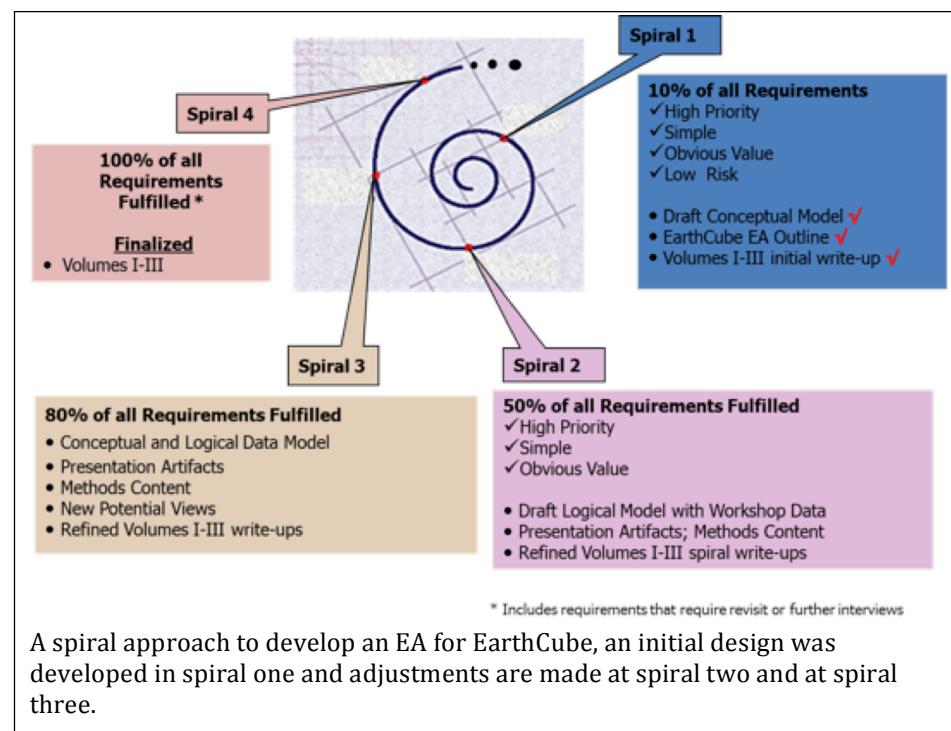
Conceptual Framework

Although various geoscience resources (e.g., data, tools, experts) have become available from cyberinfrastructure, geoscientists and other users still have to spend a significant amount of time to understand and process those resources to meet the requirements of their research. This Conceptual Design project is to design an EarthCube Enterprise Architecture (EA) to support EarthCube for facilitating data communication and human collaboration in pursuit of collaborative geosciences.

Technical Approach

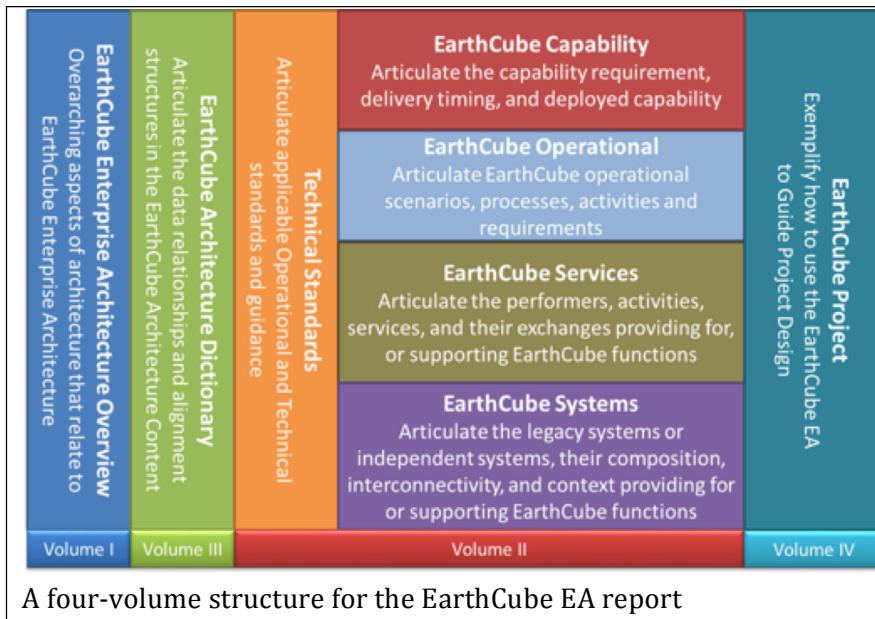
The EarthCube EA framework is developed by leveraging several popular frameworks (i.e. Zachman, Gartner, TOGAF, FEAF, and DoDAF) that have been demonstrated to be efficient in practice. The design of EarthCube EA comprises three major components: 1) the overview of current geoscience resources and architectures, 2) the capabilities of the envisioned EarthCube EA, and 3) the synthesis of proposed changes in the communities that would help reshape geoscience cyberinfrastructure for achieving the goals of EarthCube.

This enterprise architecture conceptual design follows a spiral development approach, as shown in the figure. There are four spirals for the two years, including requirement analysis based on community inputs, initial development of enterprise architecture,





The final design will be in a four-volume report containing different views which describe the EarthCube architecture from different perspective as laid out in the figure.



reports and roadmaps presents the requirements from one community.

2. During drafting of the design report, workshops are organized during ESIP meetings to engage domain experts to discuss EA in general, review the current design, and comment for improving the design. Meanwhile, the architecture will be presented at conferences and meetings to accumulate the ideas from the broader community.
3. Workshops will be organized again during ESIP meetings before finalizing the design. The design results will be discussed with scientist to ensure the consistency between the design and scientists' requirements.

Benefits to Scientists

The design results will provide a reference framework for building EarthCube and other geoscience cyberinfrastructures. The framework could be used as a comprehensive guide on developing geoscience cyberinfrastructure by different stakeholders in the EarthCube community and as a source of priorities that identify topics of highest interest in the community.

Chaowei Yang, NSF Spatiotemporal Innovation Center, George Mason University,	Erin Robinson, Foundation of Earth Science Information Partners	Min Sun, NSF Spatiotemporal Innovation Center, George Mason University,

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1343759. For more information, contact Chaowei Yang at cyang3@gmu.edu.

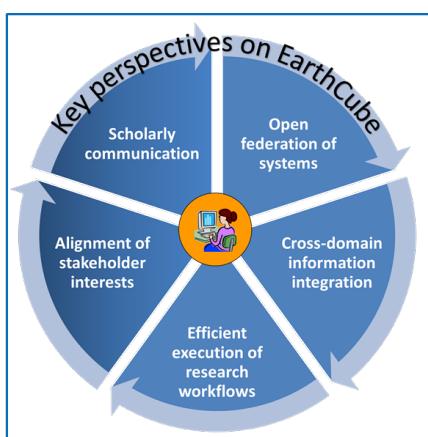
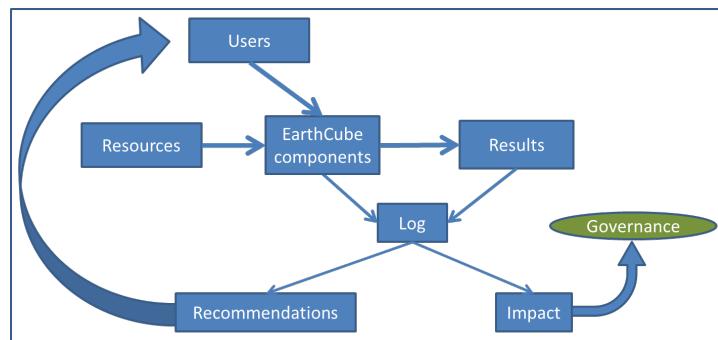


Conceptual Framework

This project seeks to integrate traditional information system components developed in various branches of the geosciences with a number of emerging technologies that support scholarly communication, self-organization and social networking in order to create an enterprise architecture that enables more comprehensive, data-intensive research designs and knowledge sharing within EarthCube. EarthCube is conceptualized as an evolving, community-governed cyberinfrastructure supporting geoscience research and education enterprise. It is considered from five distinct perspectives: as an open federation of systems, as an environment for scholarly communication, as an environment for cross-domain information integration, as a system to enable efficient execution of research workflows, and as an environment for alignment of stakeholder interests.

Technical Approach

Development of the conceptual design for EarthCube involves: (1) analysis of geoscience user needs and research scenarios, to understand the scope and functional requirements of the enterprise system; (2) review of architecture designs and technical approaches adopted in similarly large cross-domain information system projects, in geosciences and elsewhere, at the international (e.g. INSPIRE in Europe, GEOSS) and national (e.g. Federal Enterprise Architecture, NIF, NIEM, iPlant) levels; (3) development of a high-level EarthCube information model to include main functional components and their interactions; (4) analysis of emerging technologies and approaches that support scholarly communication in the science enterprise, such as researcher profile systems, social networks, Q&As, and stakeholder maps; (5) demonstrating how EarthCube will support configuring systems on demand for new research designs; (6) enabling reproducible science; (7) monitoring of different types of feedback in the science system as an instrument for configuring and tuning system to changing research designs and available resources. The design generally follows the Reference Model for Open Distributed Processing (RM-ODP) in the use of various viewpoints to specify large, distributed systems.





The project will demonstrate how the key concepts of the design will be implemented and validated in several use cases at the intersection of different geoscience domains, including critical zone science, hydrology, geochemistry, and geology.

Science Drivers

The project involves both geoscientists and computer scientists. The main drivers are: 1) geoscience grand challenges as articulated in a series of NSF vision documents and refined through the first several years of EarthCube development; and 2) adapting science practice to leverage the accelerating pace of technology innovation. Responding to grand challenges such as climate change prediction, water sustainability, analysis and management of hazards, and CO₂ sequestration, requires new insights into the earth system dynamics, transcending disciplinary, organizational, political and other boundaries. A long-term design for EarthCube must accommodate science drivers and research designs that are not yet fully articulated, using technology that is constantly introducing new possibilities.

Benefits to Scientists

From the perspective of geosciences, the primary contribution of this work is the development of a conceptual framework for technology to support efficient, scientifically-sound and sustainable resource sharing and collaboration across geoscience disciplines to address grand challenge problems. In the operational EarthCube system, researchers will be able to efficiently execute complex research workflows; discover and integrate data from different domains; contribute their data and research products; visualize information in multiple dimensions, efficiently collaborate with other researchers, and jointly participate in advancing the infrastructure to accommodate new data types, models, and research designs. It will become easier to re-generate simulation results, independently validate modeling conclusions, or re-run models with new data obtained from different geoscience domains. From the perspective of cyberinfrastructure research, the main impact of the project is in extending enterprise architecture design methodology for the case of a highly complex, multi-stakeholder federation of systems, and generating new insights into the needs, opportunities and limitations of resource sharing and information integration across different types of geoscience data and models.

			
Ilya Zaslavsky San Diego Supercomputer Center, UCSD	Stephen Richard Arizona Geological Survey	Tanu Malik Computation Institute, University of Chicago and Argonne National Lab	Amarnath Gupta San Diego Supercomputer Center, UCSD

This project is funded by the National Science Foundation under the EarthCube initiative through grant ICER-1343813. For more information, contact Ilya Zaslavsky at zaslavsk@sdsc.edu or visit <http://workspace.earthcube.org/transformative-research-collaboration>.



Conceptual Framework

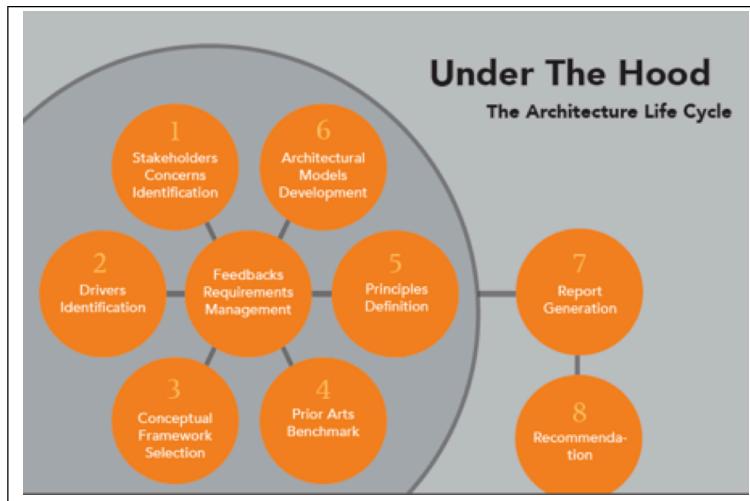
Caltech, Element84 and NASA/JPL are developing a conceptual architecture that aims to serve as the blueprint for the definition, construction, and deployment of both existing and new software components to ensure that they can be unified and integrated into an evolutionary national geosciences infrastructure for data that is part of the EarthCube enterprise.

Technical Approach

This study includes definition of process, technical, and information architectures based on a set of guiding architectural principles that examines the lifecycle of data to ensure that different stakeholder needs are addressed. Caltech, Element84, and NASA/JPL will heavily leverage efforts from multiple disciplines to develop a scalable and extensible approach that can support both a data and computationally intensive environment to enable scientific data management and discovery.

Architecture Principles Examples

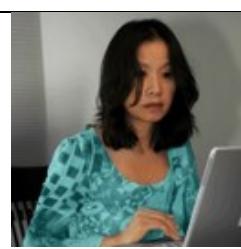
- Location Independent – Allow for the data and services to reside at network accessible locations
- Model Driven – Support by cross cutting information architecture
- Distributed Federation – Allow for integration and distributed services into a coherent cyberinfrastructure
- Standards – Adopt, adapt and develop standard specifications
- Diversity – Support diverse needs of stakeholders
- Scalability – Scale the core functions as the volume of data and services increase
- Explicit Design – Explicitly defined with unambiguous specifications
- Data Integrity – Adheres to rigorous data integrity and provenance policy



George
Djorgovski,
Caltech



Dan Crichton,
NASA/JPL



Emily Law,
NASA/JPL



Dan Pilone,
Element84



Tracey Pilone,
Element 84

This project is funded by the National Science Foundation under the EarthCube initiative through grant 1343661. For more information, contact George Djorgovski at george@astro.caltech.edu.



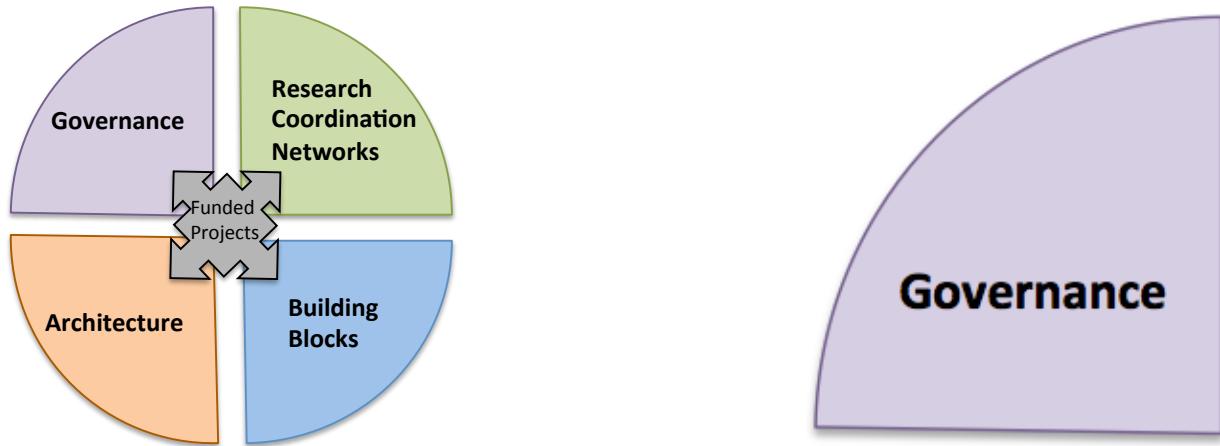
EarthCube

Conceptual Designs



EarthCube

Test Governance



EarthCube Test Governance

**Formal Mechanisms to
Involve the Community**

<http://workspace.earthcube.org/test-governance>



EarthCube

Test Governance



EarthCube Test Enterprise Governance: An Agile Approach

Conceptual Framework

This project for Test Enterprise Governance outlines an agile model to identify, test and evaluate governance models to manage the development of Geosciences cyberinfrastructure. This model seeks broad engagement and participation of the EarthCube stakeholders to define and assess governance models while seeking evaluation and cross-checks from advisory committees and evaluation mechanisms.

This effort employs an iterative deployment across the range of EarthCube stakeholders to encourage transparency, consensus, and inclusiveness. A broad coalition of stakeholder groups comprise the Assembly and served as a preliminary venue for evaluating and testing governance models in 2012-2013. A Secretariat acted as the coordinating body throughout the first phase of the project, carrying out duties such as planning, organizing, communicating. To ensure broader end-user participation in evaluating governance models, a crowdsourcing approach was used for members not involved in the Assembly.

In 2014-2015, a community-led Demo governance is being tested and improved. The organizational structure will be demonstrated and evaluated. The structure and activities of the Demo governance rely heavily on the outcomes of earlier activities, including 26 end-user workshops. The role of the test governance demonstration is to facilitate community convergence on a reference architecture, procedures for standards, and coordination among emerging EarthCube elements.

2012-2013: Set up test governance structure and charter documents



2014-2015: Demonstrate test governance framework





Approach

The agile approach for Test Enterprise Governance has the potential to advance knowledge by providing community-guided solutions to the governance of cyberinfrastructure for the geosciences. It incorporates multiple science disciplines and encourages broader participation of early career scientists. This project encourages cross-disciplinary research and discovery, not only across core science disciplines but also the computer and information sciences.

The screenshot shows the EarthCube website at www.earthcube.org. The header features the EarthCube logo and the tagline "TRANSFORMING GEOSCIENCES RESEARCH". Below the header is a navigation bar with links for HOME, ABOUT, GET INVOLVED, NEWS AND EVENTS, CONTACT, and EARTHCUBE COMMONS. The main content area has three sections: "Welcome to EarthCube" with an icon of the EarthCube logo, "Join Our Community" with a photo of a group of people, and "EarthCube Workspace" with a button labeled "Access the Workspace". A sidebar on the right contains the text: "The EarthCube Commons workspace is where EarthCube project teams and".

By building a collaborative, virtual community of practice, this project hopes to enable distributed knowledge communities to cooperate, pool resources and work together across disciplines, geography, and cultures. Building upon existing NSF investments in cyberinfrastructure, this effort allows for, and enables, a diverse set of end users previously unengaged in discussions within cyberinfrastructure—to define expectations, decision-making and leadership processes, leading to a sustainable and adaptable cyberinfrastructure for the geosciences that is extensible and scalable.

A significant responsibility of the Test Enterprise Governance is facilitation of geosciences community activities related to cyberinfrastructure, data policies and standards, and outreach related to these issues. The project will provide web services, forum and other engagement activities to facilitate this outreach.





Involving the Community

Governance for EarthCube has been a community-driven effort since its inception. Organized by the Enterprise Test Governance project, governance conversations were woven across charrettes (November 2011 and June 2012), workshops (2012-2014), and other virtual and face-to-face meetings. Inputs from the community also came through white papers (November 2011), Technical Roadmaps and Concept Design reports (March-August 2012), and other documents produced by the community.

Outreach to the community has been a priority. The Test Governance project has ensured a presence for EarthCube at numerous conferences and meetings, with Town Halls, presentations, exhibition booths, and advertising materials. EarthCube webinars have been held to inform the community of major stages in the development of EarthCube.

An annual EarthCube All Hands Meeting served as a catalyst for discussions on EarthCube work and the overall governance process.

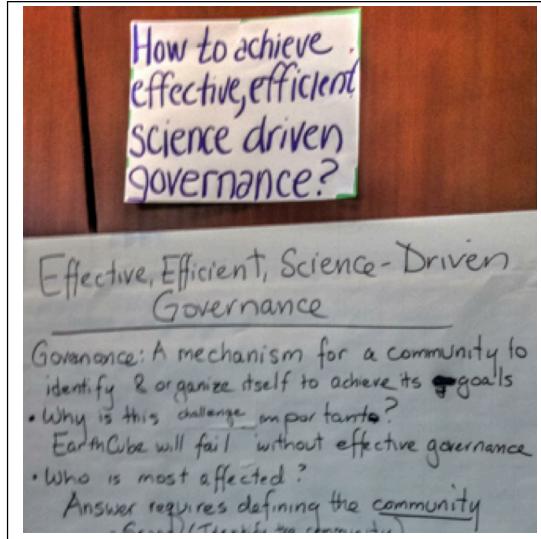
A governance working group was also established and dedicated to this topic. The group has open membership and holds regular virtual meetings.

The conception of governance through this process has spanned from the social/cultural practices of community building and decision policies to notions of shared standards and best-practice open science.

Evaluation and Assessment

Two aspects of developmental evaluation have supported EarthCube — survey data collection and hands-on facilitation. The NSF-funded Stakeholder Alignment Collaboration (see separate entry below) has used survey data as a baseline indication of the views across diverse fields and disciplines. Additional rounds of survey data will track progress and indicate emerging challenges and opportunities. The Spark Policy Institute also synthesized materials from the various governance workshops.

Both the Spark Policy Institute and the Stakeholder Alignment Collaboration have also provided facilitation support for the development of the overall governance architecture, as well as specific session designs, charters, and vision statements.





Designing an Initial Governance Structure and Charter for EarthCube

The governance structure and charter for EarthCube were designed in a three-phase process.

In the initial phase, groups of stakeholders were assembled to identify common goals and design a proposal of how their interests could be represented in the EarthCube governance structure. To this end, several EarthCube Stakeholder Assembly workshops were held in January-April 2014. Stakeholder Assembly workshops were held for data facilities, EarthCube funded projects, IT/computer science/open-source software, science end-users, and professional societies. Hundreds of people participated in these workshops. These workshops generated proposals for the components of the governance structure and associated charters.

In the second phase, a single group representative of all stakeholders was assembled to synthesize an overall governance structure and charter. This was a small group (twenty people) of thought leaders and champions that were identified by the community during the Stakeholder Assembly Workshops held in the prior phase because they took a leadership role (such as helping charter a proposed committee or forming a proposed working group). In addition, participation was balanced with the following criteria: 1) near equal distribution from each Stakeholder Assembly workshop, 2) near equal domain knowledge distribution (Atmosphere, Earth, Ocean, Polar, Cyberinfrastructure), 3) a PI of a Conceptual Design award. There was also an open call for additional representation to ensure transparency. The group met at an EarthCube Assembly Synthesis Workshop held in April 2014. Participants were asked to review the outcomes of all of the Stakeholder Assembly workshops and reach consensus on: 1) clear language defining the EarthCube vision, mission, and goals; 2) primary functions of EarthCube governance; and 3) major structures of governance necessary to support the identified functions. The result was a unified proposal for EarthCube consisting of two documents: a Demonstration Governance Structure and a Demonstration Governance Charter.

The final phase sought community feedback as well as final approval from NSF as the sponsor of EarthCube activities. Community feedback on the governance proposal was sought through two primary venues: a series of online surveys, and sessions at the 2014 EarthCube All Hands Meeting. Comments collected through the surveys were incorporated into the exercises and facilitated discussions at the All Hands Meeting, which served to elicit further feedback on the charter. All the feedback was incorporated into the governance documents. The revised Governance Charter and Governance Structure documents were presented to NSF for final review and approval.

The process by which these founding governance documents were developed recognizes and builds upon the work of many hundreds of people actively involved in the EarthCube initiative since its inception.



Vision: EarthCube enables transformative geoscience by fostering a community committed to providing unprecedented discovery, access, and analysis of geoscience data.

Mission: EarthCube streamlines the path to scientific discovery by overcoming social, institutional, and technical barriers to data sharing and access through a network of interoperable computations resources. These resources will be developed and supported by a community of practice consisting of academic geoscientists and related communities.

Goals: EarthCube will accomplish its mission by:

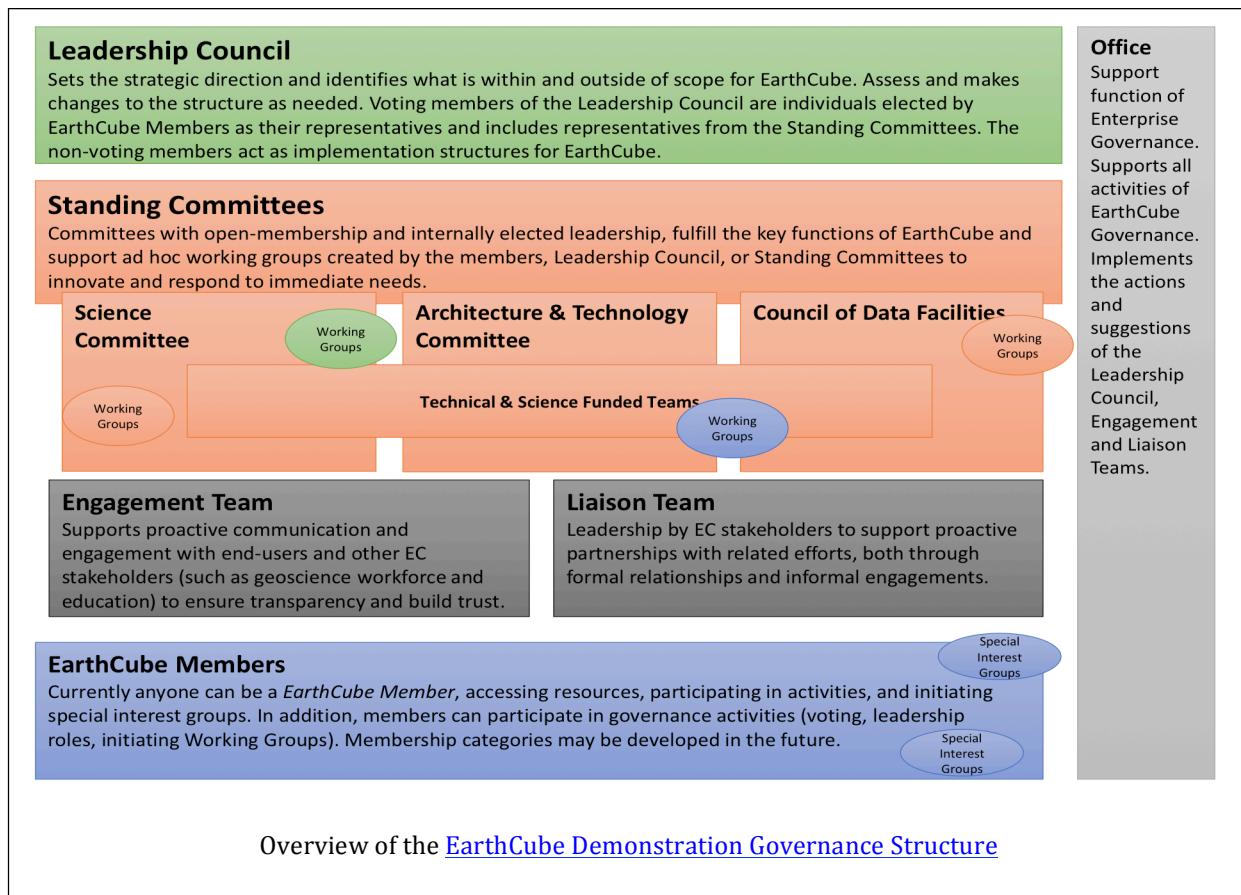
- Maintaining a knowledge base to guide the development and maintenance of the EarthCube system and to assist users in interacting with that system
- Making it easy and more productive to share, identify, access, use and evaluate quality data, streamlining the path to scientific discovery
- Enable data-driven hypothesis testing that would be intractable without EarthCube resources
- Answering questions that require access to multiple types and sources of data
- Enabling trust in data and software
- Developing sustainable and interoperable software and hardware resources
- Seeking opportunities to build technological and social interfaces to other related efforts, and jointly identifying capability gaps that need to be fulfilled with further research and development
- Supporting projects that implement and test EarthCube resources to address grand challenges in the geosciences

Overview of the [EarthCube vision, mission, and goals](#) from the initial governance charter

Initial Governance Structure and Charter for EarthCube

The Demonstration Governance Charter document lays out the vision, mission, and goals for EarthCube. It also describes the specific functions of each of the components of the governance structure.

The initial Demonstration Governance Structure includes: 1) a Leadership Council that oversees the strategic direction and priorities for EarthCube, 2) a Science Committee to identify and prioritize science requirements and connect the geoscience and technology communities, 3) a Technology and Architecture Committee to facilitate technology development, architecture, and a testbed for components, 4) a Council of Data Facilities representing well established institutions that serve data to the geosciences community, 5) an Engagement Team to reach out to the community and seeking involvement, 6) a Liason Team to establish partnerships with existing cyber-initiatives, agencies, associations, and other efforts external to the NSF core constituency, 7) and an EarthCube Office with a support function. The governance structure recognizes open membership for anyone, and the ability of members to participate in the governance structure and voting processes. The structure also includes provisions for chartering Working Groups for specific objectives of high priority for the community, and Special Interest Groups to facilitate community discussions.



Demonstrating Community Governance for EarthCube

October 2014 marked the beginning of a demonstration governance period as described in the Governance Charter. Governance activities for this period include populating the committees and teams in the charter, and establishing leadership and decision making processes to allow EarthCube to move forward on community priorities, such as identifying science drivers and converging on system architecture issues.

The purpose of the demonstration governance period is to facilitate community convergence on a system of governance that can sustainably carry EarthCube into the future.

The effectiveness of the demonstration governance structure will be evaluated throughout the demonstration period, and elements may be adjusted to better meet community needs. A key milestone will be the 6-month point, when a community meeting will be convened for intensive evaluation.



Phase One Project Team

Lee Allison, PI, Arizona Geological Survey

Christopher Keane, Co-PI, American Geosciences Institute

Erin Robinson, Co-PI, Foundation for Earth Science

Advisory Team

David Arctur, University of Texas at Austin

Jim Bowring, College of Charleston

Bruce Caron, The New Media Studio

Gary Crane, Southeastern Universities Research Association

Steve Diggs, Scripps Institution of Oceanography

Cecelia DeLuca, University of Colorado at Boulder/NOAA

Ted Habermann, HDF Group

Kerstin Lehnert, Columbia University and

George Percivall, Open Geospatial Consortium

Steve Richard, Arizona Geological Survey

Brian Wee, NEON, Inc

Evaluation Team:

Joel Cutcher-Gershenfeld, University of Illinois Urbana-Champaign

Jewlynn Lynn, Spark Policy Institute

Operations Center (at Arizona Geological Survey):

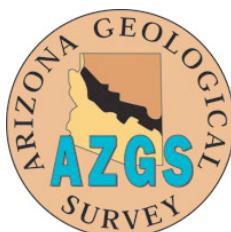
Kim Patten (also Co PI, for yr 1)

Rachael Black

Anna Katz

Kate Kretschmann

Genevieve Pearthree



Lee Allison
AZGS



Kate Kretschmann
AZGS



Anna Katz
AZGS



Rachael Black
AZGS



Bruce Caron
EarthCube PO

This project is funded by the National Science Foundation under the EarthCube initiative through grant 1340233. For more information, contact Lee Allison at lee.allison@azgs.az.gov. More information about the EarthCube Test Enterprise Governance can be found at <http://workspace.earthcube.org/demonstration-governance/>.



Stakeholder Alignment

Enabling Transformation in the Social Sciences, Geosciences, and Cyberinfrastructure

Science Challenges

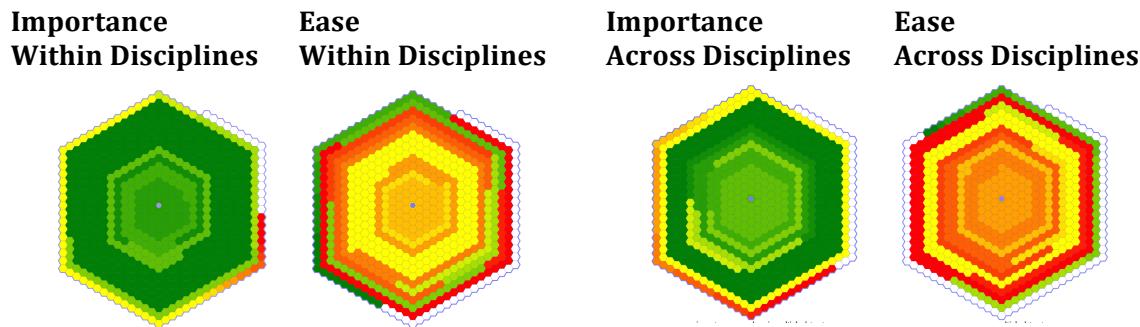
In 2012 the President of the National Academy of Sciences, Ralph Cicerone stated at the 149th Annual Meeting of the Academy that, "Today's most troubling and daunting problems have common features: some of them arise from human numbers and resource exploitation; they require long-term commitments from separate sectors of society and diverse disciplines to solve; simple, unidimensional solutions are unlikely; and failure to solve them can lead to disasters." He went on to observe, "In some ways, the scales and complexities of our current and future problems are unprecedented, and it is likely that solutions will have to be iterative . . . Institutions can enable the ideas and energies of individuals to have more impact and to sustain efforts in ways that individuals cannot."

Our research is designed to enable transformation at institutional and systems levels of analysis so that institutions can better enable the work of scientists, rather than being barriers to progress on the grand challenges Cicerone highlights.

Technical Approach

Stakeholder alignment is the foundation for governance. Stakeholders include the builders and operators of cyberinfrastructure, the diverse fields and disciplines of the geosciences, and others in society impacted by the earth as a system. Alignment is the dynamic process by which sufficient agreement is reached to enable collective action.

With NSF support, we have developed a new visualization tool that enables quick identification of points of alignment or misalignment among stakeholders in complex systems, termed a "z-flower." Here are four z-flowers for key questions about the importance and ease of finding, accessing, and integrating data, models, and software within and across fields and disciplines:

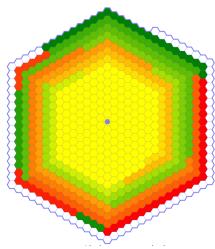


Here is how to read a z-flower: The middle represents the central tendency among stakeholders and the outside represents the outliers. The darker shades of red indicate increasingly negative views; the darker shades of green indicate increasingly positive views; and the yellow represent neutral views. The responses – positive or negative – alternate in a spiral from the middle, which is the mean. Blanks are for non-responses. The

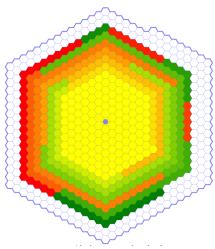


above z-flowers indicate that major gap between the perceived importance of data sharing and the actual ease of doing so. These data represent a baseline for governance of EarthCube and helped to motivate the outreach to domain scientists through the end user workshops. The challenge was reinforced with other data, including what is illustrated with these z-flowers:

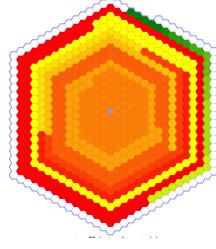
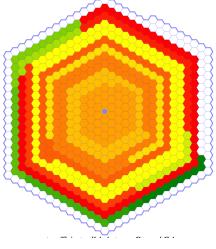
Perceived Sharing Among Geoscientists **Perceived Sharing Among Cyber Developers**



Communication Between Geo and Cyber Communities



End-user Training and Knowledge to Use Cyber Tools



Science Drivers

The visualization of stakeholder alignment and the associated analytics advance the science of large-scale systems change, which integrates institutional theory, systems science, information science, team science, the science of science, and related domains.

Benefits to Scientists

The data and visualizations help to accelerate dialogue and agreement among geoscientists and cyberinfrastructure experts. The stakeholder alignment approach has been joined, in the case of EarthCube, with action research engagement through facilitated support for the development of charters, a shared vision, and other aspects of governance.

Stakeholder Alignment Collaboration

We are a multi-disciplinary team based at the University of Illinois, Urbana-Champaign, including: Karen Baker, UIUC; Nick Berente, University of Georgia; Burcu Bolukbasi, University of Kansas; Dorothy Carter, Georgia Tech University; Noshir Contractor, Northwestern University; Joel Cutcher-Gershenfeld, UIUC; Leslie DeChurch, Georgia Tech University; Courtney Flint, Utah State University; Gabriel Gershenfeld, Cleveland Indians; Michael Haberman, UIUC; John L. King, University of Michigan; Eric Knight, University of Sydney; Barbara Lawrence, UCLA; Spenser Lewis, General Dynamics; Pablo Lopez, UIUC; Ethan Masella, Brandeis University; Matt Mayernik, NCAR/UCAR; Charles McElroy, Case Western Reserve University; Barbara Mittleman, Nodality, Inc.; Victor Nichol, consultant; Mark Nolan, UIUC; Sunjin Pak, UIUC; Melanie Radik, Brandeis University; Dechying Ruengvisesh, UIUC; Namchul Shin, Pace University; Cheryl Thompson, UIUC; Susan Winter, University of Maryland; and Ilya Zaslavsky, UCSD.



This project is funded by the National Science Foundation under NSF OCI RAPID 1229928, "Stakeholder Alignment for EarthCube" and NSF SciSPR-STS-OCI-INSPIRE 1249607, "Enabling Transformation in the Social Sciences, Geosciences, and Cyberinfrastructure." For more information, contact Joel Cutcher-Gershenfeld at joelcg@illinois.edu.



EarthCube

Test Governance



EarthCube

Getting Involved In EarthCube



<http://workspace.earthcube.org/demonstration-governance/>





The Opportunity

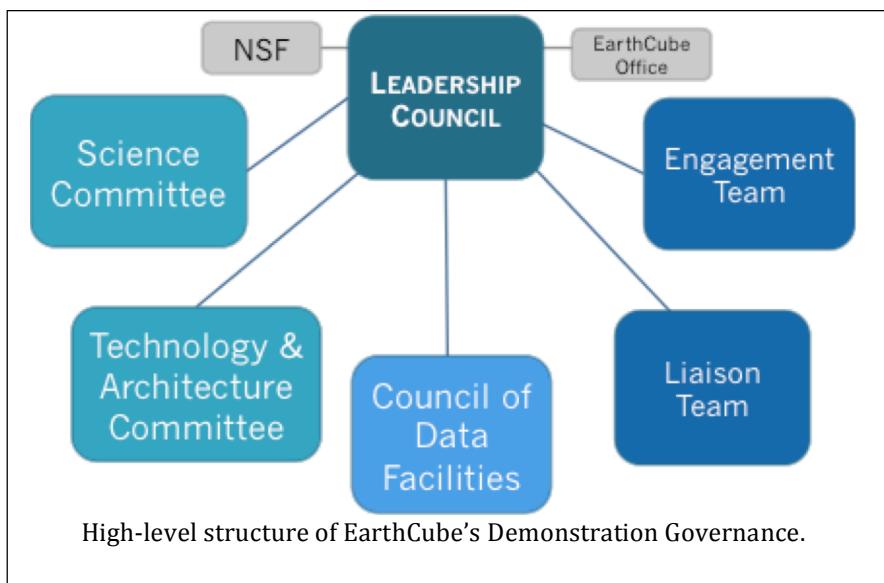
EarthCube is an **evolving, dynamic community effort** that seeks to ensure successful outcomes by actively involving individuals and partners from across the geosciences and cyberinfrastructure sectors. The current phase of EarthCube's development builds upon the outcomes of the 24 end-user workshops, that incorporate input from ~1500 participants. It also seeks to energetically engage stakeholders whose activities will be furthered and enhanced by the improved access to data and resources that EarthCube's emerging, community-governed cyberinfrastructure will facilitate.

You can become involved and positively impact the future of EarthCube. This is an opportunity for you to influence how data will be collected, accessed, analyzed, visualized, shared, and archived; facilitate and participate in interdisciplinary research; and help educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science. Collectively these activities will help foster a sustainable future through a better understanding of our complex and changing planet, and enable the geosciences community to develop a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core⁴.

Participating in EarthCube Governance

The demonstration governance structure that resulted from the activities of the Test Governance project is designed to facilitate individual involvement in committees and working groups, and thereby encourage broad representation from across the geoscience community.

The components of the demonstration governance represent the diversity of EarthCube functions:



⁴ *Earth Cube Guidance for the Community*, NSF11085.



EarthCube

1. **The Science Committee** role is to identify and prioritize end user requirements, and to connect the academic and technology communities. It is co-chaired by Basil Gomez and Emma Aronson and, in the short-term, the work is enabled by three working groups whose tasks are to synthesize: the overarching science drivers identified by the participants of the twenty-four end-user workshops; the funded project's science goals; and novel use cases.
2. **The Technology & Architecture Committee** role is to test and facilitate technology and architecture development. It is co-chaired by Yolanda Gil and Jay Pearlman, and its work is enabled by working groups whose initial focus is on identifying the technical requirements through science use cases; conducting a funded projects gap analysis; developing testbeds for the funded projects; and the identification of appropriate standards for EarthCube.
3. **The Council of Data Facilities** (CDF) is a federation of existing and emerging geoscience data facilities that serve as a foundation for EarthCube and cyberinfrastructure for the geosciences. The interim chair is Mohan Ramamurthy, and it is co-chaired by Kerstin Lehnert and Don Middleton. The CDF Charter was approved in December 2014. The Council is holding a General Assembly Meeting in January 2015 (after the ESIP Winter Meeting), at which time the members of the Executive Committee (the chair, a vice-chair, a secretary, and four other representatives) will be elected.
4. **The Liaison Team** seeks to establish partnerships with existing cyber-initiatives, agencies, associations, and other efforts external to the NSF core constituency, including international activities as well as the private sector. It is co-chaired by Rick Ziegler and Lindsay Powers. The current focus is on mapping the larger geo/CI landscape and community; populating landscape map with organizations, initiatives, agencies, data facilities, etc.; and assessing where EarthCube fits into this landscape. The Liaison Team plans to organize a joint session between EarthCube and COOPEUS - (Cooperation EU/US) at the RDA – Research Data Alliance Fifth Plenary Meeting in March 2015.
5. **The Engagement Team** role is to encourage involvement in EarthCube by proactively reaching out to the geoscience community. The team is chaired by Marjorie Chan. Three working groups are centered around mapping the community engagement scope; conveying new tools and cases; and facilitating internal communication.
6. **The Leadership Council** oversees the activities of the five other components of EarthCube governance, in coordination with NSF as the sponsor and with the EarthCube Office for logistic support. It is presently comprised of community-elected representatives drawn from the Science Committee (Basil Gomez); Technology & Architecture Committee (Yolanda Gil); Council of Data Facilities (Mohan Ramamurthy); Liaison Team (Rick Ziegler); and Engagement Team (Marjorie Chan); as well as members at-large of the Cyberinfrastructure (David Actur); Atmosphere/Space (Farzad Kamalabadi); Oceans (Danie Kinkade); Earth Sciences (Kerstin Lehnert); and Polar (unfilled) communities. Basil Gomez is the Leadership Council's Interim Chair.



The EarthCube Council of Data Facilities

Geoscience data facilities enable transformational science, foster technological development, and promote educational opportunities. They provide direct societal benefits through increased coordination, collaboration, and innovation in the acquisition, curation, preservation, and dissemination of geoscience data, tools, models, software, and services. A federation of existing and emerging geoscience data facilities – through the Council – serve as an effective foundation for EarthCube and related contributors to the cyberinfrastructure for earth system science (lithosphere, hydrosphere, biosphere, atmosphere, and space environment).

The EarthCube Council of Data Facilities (CDF) is a federation of existing and emerging geoscience data facilities that serve as a foundation for EarthCube and cyberinfrastructure for the geosciences.

The Council's mission is to serve in a coordinating and facilitating role that includes advancing the following goals and responsibilities:

- Provide a collective voice on behalf of the member data facilities to the NSF and other foundations and associations, as appropriate.
- Identify, endorse, and promote standards and best or exemplary practices in the organization and operation of a data facility.
- Collaborate with standard-setting bodies with respect to shared feedback on standards for data, models, software sharing interoperability, metadata, and related matters.
- Identify opportunities for and supporting the development and utilization of shared cyberinfrastructure, professional staff development and training, and other related activities.
- Foster innovation through collaborative and interdisciplinary projects.
- Increase understanding and engagement with relevant stakeholders.

Founding members of the council, which already include thirty-four organizations, unanimously approved its initial charter in December 2014.

In advancing this mission, the CDF is committed to working with relevant agencies, professional associations, initiatives, and other complementary efforts. If you work at an Earth data facility, you can become involved in EarthCube by joining the CDF. A data facility is eligible for membership in the Council if it acquires, curates, preserves, and/or disseminates data, software, and/or models for one or more defined communities or disciplines in the EarthCube initiative.



The EarthCube Leadership Council

The Leadership Council is the elected voice of the EarthCube community, setting the strategic direction for EarthCube and making decisions critical to the success of EarthCube with input from the community and in consultation with NSF.

The Leadership Council is formed by representatives of the EarthCube governance components as well as at-large members of the community. Nine voting members include the chair, three standing committee representatives (for the Science Committee, the Technical & Architecture Committee, and the Council of Data Facilities), and five at-large representatives from constituencies of geosciences (one each of Atmosphere, Earth, Oceans, and Polar) and Cyberinfrastructure. Four non-voting members include an Engagement Team Representative, a Liaison Team Representative, a representative from the National Science Foundation, and the Director of the EarthCube Office.

To fulfill this role the Leadership Council will:

- Ensure consistency and transparency in policies, procedures, and decision-making, including providing multiple ways for people to participate in the process of making decisions, and communicating outcomes of decisions to the broad EarthCube community.
- Enable communication between governance organizational units to close gaps, eliminate duplication, and build synergies.
- Establish and manage Standing Committees and Working Groups as needed to perform critical functions.
- Foster business models to sustain and maintain the infrastructure of EarthCube.
- Establish, facilitate, and maintain policies and procedures.
- Provide for public dispute resolution and proactive management of risk and conflicts of interest.
- Act as the single point of communication for coordinating with and making recommendations to the NSF and other funding agencies on behalf of EarthCube.

The EarthCube Office supports the Leadership Council and all the other governance components, and is part of the Test Governance project.

All components of the demonstration governance are open to any and all individuals who wish to participate. **We invite you** to sign up and share in these activities by visiting the *EarthCube Commons* at <http://workspace.earthcube.org/>.



Other Opportunities

Other major arenas that are open to your participation include:

- 1. Professional Meetings:** At the major professional meetings of the geosciences (e.g., AGU - American Geophysical Union, GSA - Geological Society of America, and others), diverse technical sessions and town hall meetings showcase EarthCube activities. Technical sessions permit you to see what projects are in progress and learn about significant results and outcomes. Town hall meetings present thematic information about EarthCube, and offer opportunities for you to raise questions and highlight your concerns. There typically will also be an EarthCube exhibit booth for you to visit, where you can learn about funded EarthCube projects, the new tools that are being developed, and use-case and demonstration science data products.
- 2. Workshops:** Future EarthCube activities will encompass science and technology retreats, workshops, training events, and research opportunities. A focus of these activities will be the fostering of interdisciplinary connections and interactions between scientists and technologists. Announcements will be posted on the *EarthCube Commons*, publicized in the bi-weekly community newsletter and communicated to community members by electronic mail.
- 3. Research Projects:** The NSF periodically announces proposal opportunities with the object of addressing specific aspects of EarthCube to advance its goals. The more that you know about ongoing EarthCube activities, the more you will understand how to pursue these opportunities.



EarthCube is a compelling and evolving vision for the geosciences. The more scientists and technologists are engaged, the more EarthCube can achieve. This is your opportunity to help turn ideas into reality!

To receive EarthCube announcements and other information about the program, you can **subscribe** to the mailing list at <http://workspace.earthcube.org>.

Please **contact us** if you have suggestions or have further questions about how you can be involved: leadership@earthcube.org.



Subscribe to the newsletter:

<http://workspace.earthcube.org>

Join the EarthCube Groups:

<http://workspace.earthcube.org/science-committee>

<http://workspace.earthcube.org/technology-committee>

<http://workspace.earthcube.org/engagement-team>

<http://workspace.earthcube.org/liaison-team>

<http://workspace.earthcube.org/council-data-facilities>

Contact us:

leadership@earthcube.org

<http://workspace.earthcube.org/leadership-council>



EarthCube.org

