EarthCube

2015 Highlights
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-2015-1

First released: May 27, 2015

Picture credits – On cover from top to bottom: NASA International Space Station Science (CC BY-NC 2.0). The rest of the pictures in this document were taken by members of the EarthCube community at various workshops, field trips, and meetings. Artwork and visualizations were provided by EarthCube members.

This document is released under a Creative Commons CC BY 4.0 license. This license allows 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. This document should be cited as:


It is available online at:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EarthCube in brief</td>
<td>5</td>
</tr>
<tr>
<td>EarthCube in 2015</td>
<td>6</td>
</tr>
<tr>
<td>Research Coordination Networks: Science communities</td>
<td>6</td>
</tr>
<tr>
<td>Building Blocks: Technology demonstrations</td>
<td>12</td>
</tr>
<tr>
<td>Conceptual Designs: Towards an EarthCube architecture</td>
<td>31</td>
</tr>
<tr>
<td>Test Governance: Standing up the EarthCube organization</td>
<td>32</td>
</tr>
<tr>
<td>Community activities</td>
<td>34</td>
</tr>
<tr>
<td>Strategic planning for EarthCube</td>
<td>34</td>
</tr>
<tr>
<td>Roadmaps for science activities</td>
<td>35</td>
</tr>
<tr>
<td>A roadmap for technology activities</td>
<td>36</td>
</tr>
<tr>
<td>Developing use cases</td>
<td>37</td>
</tr>
<tr>
<td>Identifying technology gaps</td>
<td>37</td>
</tr>
<tr>
<td>Adopting standards</td>
<td>37</td>
</tr>
<tr>
<td>Standing up a testbed</td>
<td>37</td>
</tr>
<tr>
<td>Converging on architecture</td>
<td>37</td>
</tr>
<tr>
<td>The Council of Data Facilities</td>
<td>38</td>
</tr>
<tr>
<td>Engaging the community</td>
<td>39</td>
</tr>
<tr>
<td>Involvement and adoption in the community</td>
<td>40</td>
</tr>
<tr>
<td>Faces of EarthCube</td>
<td>41</td>
</tr>
</tbody>
</table>
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.

EarthCube’s long-term vision is

a community-driven dynamic cyberinfrastructure that supports

standards for interoperability,

infuses advanced technologies to improve and facilitate interdisciplinary research, and

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

An extensive overview of EarthCube activities can be found in the 2014 EarthCube report “EarthCube: Past, Present, and Future”, also available in an abridged version from http://www.earthcube.org
CReSCyNT is a multi- tiered and multidisciplinary network that identifies needs, best practices, bottlenecks, and approaches to advance the design and development of data management, visualization, and image processing capacity for the coral reef science community.

PIs: Ruth Gates, Megan Donahue, Erik Franklin, Judy Lemus (Hawaii Institute of Marine Biology, University of Hawaii at Manoa), Gwen Jacobs (Information Technology Services, University of Hawaii at Manoa).

"Developing a toolbox of methods to convert the breadth in the coral reef science domain to image data is the next major challenge in the coral reef science domain."

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.

"Survey current data sharing practices of CReSCyNT participants and identify areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

CReSCyNT building blocks: Identify current data sharing practices of CReSCyNT participants and areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.

"Survey current data sharing practices of CReSCyNT participants and identify areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

CReSCyNT building blocks: Identify current data sharing practices of CReSCyNT participants and areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.

"Survey current data sharing practices of CReSCyNT participants and identify areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

CReSCyNT building blocks: Identify current data sharing practices of CReSCyNT participants and areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.

"Survey current data sharing practices of CReSCyNT participants and identify areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

CReSCyNT building blocks: Identify current data sharing practices of CReSCyNT participants and areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.

"Survey current data sharing practices of CReSCyNT participants and identify areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

CReSCyNT building blocks: Identify current data sharing practices of CReSCyNT participants and areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.

"Survey current data sharing practices of CReSCyNT participants and identify areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

CReSCyNT building blocks: Identify current data sharing practices of CReSCyNT participants and areas of redundancy and barriers to sharing their study sites, linking disciplinary expertise and coral reef data resources.

Create a geospatial map of researchers and their study sites, linking disciplinary expertise to place-based research and sharing current barriers to cyberinfrastructure, and computer scientists.

Participation in Brown Dog Workshop and Establishment of initial CReSCyNT nodes

The coral reef science domain produces and uses image data across spatial scales from intracellular structures to remotely sensed landscapes. CReSCyNT will identify cyberinfrastructure solutions to meet current and future challenges in image analysis across domains and scales.
### Goals of the RCN
- Foster collaboration among paleogeoscientists, paleobiologists, geo- and bio-informaticists, stratigraphers, geochronologists, geographers, data scientists, and computer scientists.
- Provide a curated catalog of paleogeoscience resources.
- Discuss and promote best practices for data, metadata, and sample management and curation.

### Communities involved
- Paleogeosciences, computer sciences, bioinformatics communities
- Grad students, postdocs, faculty, database managers, and museum curators

### Activities to date
- Workshop focused on Paleobiology
- Workshop focused on Computing with Time
- Bi-weekly webinar series of relevant systems and projects for paleogeoscientists, cyberinfrastructure researchers, and the wider EarthCube community
- Creation of a catalog of software, databases, and physical repositories
- Collaboration with the CINERGI EarthCube Building Block to expose paleogeoscience resources
- AGU and GSA Town Halls and workshops
- Early career event joint with the iSamples RCN at AGU 2014 about management and curation of physical samples

### Plans
- **Continue fostering the C4P Community of practice** through identification for networking, coordination, and collaboration of the many existing and nascent C4P-related groups.
- **Advance adoption and implementation of leading practices emerging from C4P events.**
- **Continue to emphasize engagement of early career researchers in C4P activities.**

### Results
- **Result 1:** Catalog of Paleogeoscience resources
- **Result 2:** Workshop reports with recommendations for future directions in paleobiology and computing with time
- **Result 3:** A resource of archived webinar talks at www.youtube.com/cyber4paleo
Communities involved
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 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.

Activities to date
- Second EC3 field trip (August 2nd – 7th, 2015)
- Town hall meeting at GSA, Baltimore
- Town hall meeting at AGU, San Francisco

Results of 2014 EC3 Field Trip
- After the field trip, both geologists and computer scientists reported that they had a better understanding about the current challenges facing the field geoscientists with respect to getting their data fully incorporated into the appropriate data infrastructure.
- An important result of the EC3 field trip was the consensus of the need for developing an open source field data collection application that adopts community-developed data/metadata standards.

"We are now at a critical junction to make field-based geologic data an integral part of EarthCube."
Mookerjee et al., 2015, GSA Today (accepted)
From environment, to base pairs, to ocean science & geobiology: Defining tools & technologies for an interoperable environmental ‘omics cyberinfrastructure

Activities to date
- Conducted a community-wide survey to identify needs and challenges in ‘omics research
- Continued development of resource viewer, in collaboration with CINERGI
- Participation in EarthCube governance and formation of the Science Strategic Plan

Upcoming: First ECOGEO workshop this Fall!
27-28 August, University of Hawai‘i

Goals of the RCN
- Create a strategic network and community of field and cyber scientists to explore new facets of ‘omics data
- Articulate needs, challenges, and solutions for cyberinfrastructure, workflows, and database and resource sustainability
- Develop best practices for curation and analysis of ‘omics data/metadata that also facilitates collaboration and training

Communities involved
- Environmental microbiology
- Oceanography
- Geobiology
- Bioinformatics
- Computer science
- NSF Science and Technology Centers

Lead PI: Edward F. DeLong, Department of Oceanography, University of Hawai‘i
Department of Civil and Environmental Engineering, MIT

Email - ecogeo.rcn@gmail.com
Website - http://earthcube.org/group/ecogeo
Listserv - http://earthcube.org/mailman/listinfo/ecogeo_earthcube.org

Funded September 2014
- Stage 1 – Assess community needs
- Workshop 1 – Identify practical solutions and key resources
- Stage 2 – Develop framework for integration of best practices
- Workshop 2 – Implementation and training within EarthCube

Results
- Survey to assess community needs
  - Expanded functionality for data queries
  - Stable platform for data/metadata/tools
  - Best practices and workflows
  - Case-studies and training webinars
- Resource Viewer
  - Contains data sets, databases, tools, people
  - Interactive and adaptable
- Use-cases: developed using EarthCube template
Goals of the RCN
• Develop a shared vision for an Internet of Samples in the Earth Science.
• Build consensus for best practices and standards (sample identification, citation, and metadata) that can be adopted across the diverse stakeholder community.
• Enhance awareness of and access to existing resources that will advance preservation, access, and management of samples.

Communities involved
• Domain Scientists
• Data Facilities
• Curators
• Computer & Cyberinfrastructure Scientists
• Software Engineers
• Large-scale science programs, sampling campaigns, & observatories
• Publishers and Professional Societies
• Agencies

Activities to date
• Early Career Forum: 60+ early career scientists discussed sample management practices at the 2014 AGU FM.
• Stakeholder Alignment Survey: 280+ respondents (see http://earthcube.org/workspace/isamples/isamples-survey).
• Stakeholder Alignment Workshop: 50+ participants met to converge on a shared vision of success, determined appropriate governance structure, and formed 5 working groups.

Ongoing Themes and Activities
• Working groups (allied with other efforts):
  • Metadata & identifiers;
  • User stories & science requirements;
  • Communication, training, & education;
  • Architecture & workflows;
  • Physical infrastructure, sample rescue, & digitization of collections.
• Education: Development of educational modules for sample management (B. Hallett)
• Knowledge Creation: development of the iSamples Knowledge Hub, a crowd-sourced semantic wiki about sample collections and curation procedures (Y. Gil).

Upcoming iSamples Events:
• Workshop at the Joint Conference on Digital Libraries (JCDL), June 21-25th, 2015
• ESIP, Monterey, CA, July 14-17th, 2015
Goals of the RCN

- Improve the efficiency and transparency of sediment experiment research.
- Build a knowledge base for guidance on best practices for data collection and management.
- Facilitate cross-institutional collaborative experiments.

Communities involved

- Earth Surface Process community – experimentalists, modelers, and field geologists, and others who use SEN data.
- Grad students, postdocs, faculty, technicians, educators

Activities to date

- SEN Workshops at Nagasaki University and Utrecht University
- AGU Scientific sessions and Town Halls
- Establishment of working groups on SEN knowledge base, education & standards, and experimental collaboratories
- Establishment of the SEN knowledge base at sedexp.net
- Lectures at Summer Institute for Earth-Surface Dynamics
- Papers in *The Sedimentary Record* and *Geomorphology*
- Collaborations with CINERGI, Geosoft, and Geosemantic Framework Building Blocks
- Live-broadcasting of in-progress experiments from collabortory members
- Work with SEN postdoc Kim Miller in developing and applying data management practices while addressing sedimentology grand challenges through experimental research and building of collaboratories.

Plans

- A third workshop in 2016 joint with CSDMS focused on data and model connections.
- Continued collaboration with EarthCube Building Blocks to supply necessary use cases for long-tail research and learn more about useful cyberinfrastructure.
- Continued population, development, and promotion of the sedexp.net knowledge base, with application and refinement of the experimental data template.

Results

- **Result 1:** Established and tested a knowledge base at sedexp.net.
- **Result 2:** Published a recommended experimental data template for community comment in *Geomorphology*.
- **Result 3:** Established three working groups for knowledge base, education & standards, and experimental collaboratories.

"The sediments are a sort of epic poem of the Earth." — Rachel Carson

PIs: Wonsuck Kim (UT Austin); Leslie Hsu (LDEO Columbia U); Brandon McElroy (U Wyoming); Raleigh Martin (UCLA)

earthcube.org/group/sen

2015 RESEARCH COORDINATION NETWORK UPDATES

Sediment Experimentalist Network (SEN)
Motivation

- Many search portals exist, yet geoscientists state that data discovery is a key challenge
- The goal is cross-disciplinary discovery of geoscience information resources, over high quality community-curated metadata, with transparent resource registration and search

Problems

- It is not clear what resources exist in EarthCube
- Metadata development and curation is time-consuming and manual
- Domain communities have limited ways to organize their information resources
- Inconsistent and incomplete descriptions, semantic and other differences make data discovery queries challenging

Three scenarios are demonstrated:

- Data facilities spend significant manual effort to ensure high quality and completeness of their metadata. We demonstrate how the CINERGI metadata curation pipeline improves existing metadata descriptions using a geoscience-wide ontology, a range of metadata enhancement components, and provenance and validation services.
- Several EarthCube Research Coordination Network projects assemble information resources for their domains. We demonstrate several CINERGI Community Resource Viewers developed to support such domain communities.
- Geoscientists in several disciplinary domains including hydrology, geochemistry, metagenomics, sedimentology, and critical zone science, are interested in finding information for their research. We demonstrate how the CINERGI inventory can be enhanced to support their use cases.

Novel contributions:

- Community Resource Viewers: online applications for crowdsourcing community resource registry development, curation and search
- Metadata enhancement pipeline and a set of document enhancers to automatically improve various aspects of metadata descriptions
- User interfaces for advanced resource discovery
- Geoscience-wide ontology to support semantic tagging and faceted search across domains
- Metadata provenance, validation and annotation services

Benefits

- Improved metadata for information discovery across domains
- Ability to organize and jointly curate domain-specific resources
- Transparent metadata enhancement and improved search in complex research scenarios
More queries at github.com/CINERGI/UseCases/issues

Find data from bedload tracer experiments in the field, and other nearby studies.

The CINERGI catalog returns results for user search queries, using an ontology built from domain knowledge, that helps compile data from a large pool of resources for research.

A key challenge is “…lack of a central repository for finding, accessing, and using data and software…” “There is no one-stop shop for even knowing what is available, let alone accessing it…”

– Participants in the Critical Zone and Data Prediction and Ensemble Assimilation workshops
Motivation
- Advance our ability to study and relate data and models of discrete and continuous phenomena, applied to...
- Improve emergency response and community resilience to flooding events.
- Achieve sustained innovation in the research-to-operations process.

Problem:
- Carry out near-real-time hydrologic simulations at high spatial resolution, nationally, communicated to local level.
- Manage large volumes of time series data.
- Provide adequate data density throughout.
- Generate high-resolution flood inundation maps nationwide.

Demonstration scenario:
National Flood Interoperability Experiment
Every 3 hours, the US National Weather Service updates their national 15-hr rainfall forecast, putting this workflow in motion:
1. Accessing the national high-resolution weather forecast for precipitation rates;
2. Land surface modeling of runoff, downscaled to 3km catchment scale;
3. Streamflow routing simulation across the US at scale of 3km stream reaches;
4. Flood inundation predictions at local scales;
5. National alerting and support for locally anticipated flood events and emergency response.

Note: results of this project are considered research to inform National Weather Service; not yet public forecasts.

Novel contributions:
- First nationwide simulation of streamflow in a single model.
- Helping harmonize coordination of weather, streamflow, and flooding data across federal agencies, vendors, and research communities.
- Developed common information model and tools for converting between WaterML and NetCDF.
- Helped advance Tethys hydrologic modeling platform (open source GIS, cloud, Python framework).

Benefits
- Better & more timely information: applies the benefits of sophisticated forecasting methods to all communities.
- Improved emergency operations: this approach will improve the information available, through addition of streamflow and hydraulics mapping to traditional weather maps.
1. Weather model and forecasts (High Resolution Rapid Refresh, NWS)

2. Land-Atmosphere Model (WRF-Hydro, NCAR)

Forecasts and Model Workflow for National Flood Interoperability Experiment

1. National Weather Service posts 15-hour precipitation forecast every 3 hours.
2. WRF-Hydro models runoff, downscaled to 3km catchments.
3. RAPID models streamflow for 2.7 million stream reaches (segments) in continental US.
4. Probabilistic flood height is estimated from streamflow & used to determine inundation.
5. Steps 1-3 complete in less than 1 hour; keeps up with NWS weather forecasts; step 4 varies.

3. Streamflow routing (CONUS)

Routing Application for Parallel computation of Discharge (RAPID) at Texas Advanced Computing Center (TACC)

4. Probabilistic flood forecasts

3km average Catchment-level forecasts

"Interoperability seems to be about the integration of information. What it’s really about is the coordination of organizational behavior.”
-- David Schell
Motivation

- Scientists have developed a large number of numerical models across the geosciences.
- Sophisticated modeling frameworks have emerged that allow models to be reused by coupling them as plug-and-play components.
- Capitalize on prior investments in modeling.

Problem: Powerful modeling frameworks from the federal/operational and academic communities exist, but are not interoperable.

- It should be easy to use any geoscience model in any modeling framework, so that modelers have maximum flexibility.

Demonstration Scenarios:

- Show that a model that has been augmented with a Basic Model Interface (BMI) can be plugged into a framework-specific adapter that allows it to be used in that modeling framework with very little effort. Target frameworks include: CSDMS (Community Surface Dynamics Modeling System), ESMF/NUOPC (Earth System Modeling Framework), OMS (Object Modeling System), OpenMI (Open Modeling Interface) and Pyre.
- Couple the MPIPOM-Tropical Cyclones ocean model to a local-scale hydrologic or inundation model. (MPIPOM-T = Message Passing Interface Princeton Ocean Model for Tropical Cyclones)
- Complete implementation and testing of a Basic Model Interface (BMI) for the WRF model (Weather Research and Forecasting) and link it within the CSDMS framework to the local-scale, spatial hydrologic TopoFlow model.

Novel Contributions:

- Model Coupling Metadata (MCM): standardized metadata schema for describing models
- Browser-based, graphical MCM Tool for entering model metadata via MCM.
- Earth System Framework Description Language (ES-FDL) for describing modeling frameworks.
- Basic Model Interface (BMI) adapters allow BMI-enabled models to be used in multiple plug-and-play modeling frameworks (CSDMS, ESMF/NUOPC, OMS, OpenMI, Pyre).
- Ontology and extensions to the CSDMS Standard Names (CSN) naming conventions
- Crosswalk between CSN Variable Names and CF Standard Names.
- Crosswalk between CSN Variable Names and CUAHSI VariableName CV.
Case Study: Hydrologic Impacts of Hurricane Landfall

- 140 mph top sustained wind; Min. pressure: 918 mbar
- 15 fatalities; Cost: $1.22 Billion USD
- 6th costliest in Mexican History. High rainrates caused severe flooding.
- Need to couple a regional-scale hurricane model (HWRF) and a city-scale hydrologic model, (WRF-Hydro). (Note: HWRF runs provided by M. Biswas and J. Vigh)

“Never doubt that a small group of thoughtful, committed citizens can change the world; indeed, it’s the only thing that ever has.”

--Margaret Mead

Column Land Surface Models:
Noah/NoahMP/SAC-HTET

Evapotranspiration, Soil moisture/Soil Ice, Snowpack/snowmelt, Runoff, Radiation Exchange, Energy Fluxes, Plant Water Stress

Terrain Routing Models:
Overland & Subsurface Flow

Stream Inflow, Inundation Depth, Groundwater Depth, Soil Moisture

Channel & Reservoir Routing Models:
Hydrologic & Hydraulic

Streamflow, River Stage, Flow Velocity, Reservoir Storage & Discharge
Large amounts of geoscience data currently reside in the text, tables and figures of scientific publications.

- Many important science questions require synthesizing data from widely scattered and heterogeneous published sources.
- Efforts to date to synthesize literature-derived data involve slow and costly manual extraction and then entry into a pre-defined database structure.

**Problem: difficult to find and retrieve data in the published literature.**

- Legacy data in all publications need to be made openly accessible to automated text and data mining.
- There is currently no digital library resource that contains all of the necessary published data, despite the fact that many published sources are available online in open access.
- Static databases produced from literature cannot be readily updated, assessed, or augmented with new data.

**Demonstration scenario: automated space-time indexing of published literature in Elsevier holdings via Macrostrat and our new text and data mining (TDM)-ready digital library resource.**

- Negotiated contractual agreement with Elsevier to develop TDM-ready library that can be used by EarthCube (and other) communities. Other publisher negotiations ongoing.
- Developed secure and high-throughput computing infrastructure to support automated document retrieval and pre-processing with common TDM tools (e.g., natural language parsing, document layout and font recognition)
- Working API leveraged by DigitalCrust project

**Novel research contributions:**

- A text and data mining-ready digital library and scalable high-throughput computing infrastructure capable of supporting the development of sophisticated literature data extraction applications.
- The capacity to quickly and accurately identify documents with potentially relevant data using simple or complex criteria that leverage full document contents.
- A supporting framework for space-time integration of geoscience resources; initially rock-record focused but extensible to geo- and bio-relevant areas.

**Benefits**

- Reduced time and cost of literature-based data compilation
- Greater return on current and past investments in data production in the “long tail” of geoscience
- General platform applicable to many areas of research in Earth and Life sciences
Results from the PaleoDeepDive (PDD) machine reading system compared to synthetic results achieved by humans reading and entering data from the published literature. Two different machine compilations are shown: PDD overlapping corpus, a set of documents also read by humans and entered into the Paleobiology Database (http://paleobiodb.org) and a larger set of documents (whole corpus). Human compendium shown here assembled by Jack Sepkoski over the course of more than 10 years. Also shown is the machine-generated genus diversity history for select class-level taxa. See Peters et al. (2014) PLoS One (doi:10.1371/journal.pone.0113523).

"Countless person-hours and dollars have been spent on characterizing and analyzing the composition, structure, and history of the geologic record. The resultant literature is vast, and much of it is underutilized. Leveraging machine-learning techniques for extracting relevant data from the literature will allow broad research questions to be more rapidly answered.”
— paleobiologist Noel Heim
Motivation
Geoscientists need workspaces that
• Self-curate
• Track cause and effect between data and models (code)
• Are reproducible
• Support services: discovery, sharing, publication

Problem
Lack of workspaces for geoscientists leads to
• Tedium in compiling, installing, and reproducing models in different environments using different data

Demonstration Scenario
The GeoDataspace Workspace and its use in our partner geoscience domains (Seismology, Hydrology, Space Science, CSDMS Model Repository). We will showcase the following features:
• **Self-curation:** Annotate model code and data during development and its use in Seismology
• **Provenance-tracking:** Track a Hydrology data processing pipeline for a model
• **Reproducibility & Preservation:** Reproduce a downloaded CDMS model, extend it and preserve it
• **Service-integration:** Integration with discovery, sharing and publication services
The workspace captures the science activity in a geounit, which is created, published, shared, reproduced, and preserved.

Novel Contributions
A workspace creates geounits and supports
• Command-line annotation management for self-curation
• Data-flow tracking integrated with version control
• Reproducibility with automated dependency management and containerization
• Access to highly-available data management services through Globus, PROVaaS, and Docker
• Integration with data facilities and enterprise science systems
The workspace has minimal learning curve, no change to development tools and improves with time.

Benefits
**Computational/Data Geoscientists:** Simplifies model and data management
**Research Groups (Faculty & Students):** Faster hand-shakes among students with reproducible research
**Modeling Centers:** Improved support for modeling community for impactful science.

“About 85 percent of my “thinking” time was spent getting into a position to think, to make a decision, to learn something I needed to know” – J.C.R. Licklider (American computer scientist)
Civilization advances by extending the number of important operations which we can perform without thinking them.

-Alfred North Whitehead (English mathematician and philosopher)
Demonstration scenario: Semantic interoperability of long-tail resources in modeling the Critical Zone

To close the long-tail resource discovery and use gap, standard information in terms of semantic annotation (semantic metadata) is required to improve the search and integration process.

- Kim is a researcher, who collected data from the IML CZO and shared this data on clowder.
- She annotates the data using standard vocabularies; the annotation service guides her.
- Data Network extractor retrieves the spatial temporal contextual relationships with similar datasets.
- Through semantic annotation, Kim finds a model that can process this data.
- Kim creates a model using Model-as-a-Service consisting of semantically enabled models with their dependencies identified using Data Alignment Service.
- The simulation results are shared in clowder but Kim didn’t find the exact Standard Name to annotate the new dataset.
- Kim uses the GeoSemantics Wiki to define the new standard name given available names.

Novel contributions:

- **Graph Knowledge base** for storing linked resources.
- **GeoSemantics Wiki** for geoscience communities to annotate their standard names.
- **Data Networks Service** for inferring the contextual relationships between given resources.
- **Semantically enabled models** as a foundation for advancing Model-as-a-Service.
- **Data Alignment service** for handling the semantic mediation between model and data resources.
- **Semantic tagging service** for resources with standard names and incorporating semantics in the development of models.

Geosemantics framework:

- Advances the interoperability of model and data resources using semantic annotations.
- Allows cross walks between standard names using a collaborative knowledge management system.
- Augments the semantic mediation and matching between models and data with minimum human intervention.

Reusability and interoperability of long-tail models and data are limited.

- Long-tail resources are often stored in less structured databases.
- Resource synthesis is complex due to the diversity of their attributes.
- Long-tail resources are important and can become powerful resource when they become interoperable across scientific disciplines.

Resource discovery gap over the web is increasing tremendously

- Resources should be semantically enabled by using the Semantic Web standards and Linked Standard Names.

Elaborate, precise automated searches will be possible when semantics are widespread on the Web.

-- Tim Berners-Lee, James Hendler, Ora Lassila

http://geosemantics.hydrocomplexity.net

Pics: Praveen Kumar, Mostafa Elag, Luigi Marini, U. affinois, Urbana, IL
Leslie Hsu, Lamont-Doherty Earth Observatory, Columbia U., Palisades, NY

A Geo-Semantic Framework for Integrating Long-Tail Data and Models
“As the volume, complexity, and heterogeneity of data resources grow, scientists increasingly need new capabilities that rely on new “semantic” approaches (e.g., in the form of ontologies—machine encodings of terms, concepts, and relations among them).”

-- Peter Fox, James Handler, The Fourth Paradigm
Today, most software developed by scientists is never shared
- There are repositories of model software (e.g., CSDMS), also code repositories (e.g., GitHub, CRAN)
- However, software is rarely shared, particularly all the data pre-processing and visualization software.

Problem: Hard to build on other work, especially across disciplines
- Software captures valuable geoscience knowledge and should be shared

Demonstration scenario: Cross-disciplinary modeling in the Critical Zone Observatories
- To test that groundwater and forest life-cycle are important agents in the shaping of landforms over time scales of $10^3$ to $10^7$ years, a student wants to create an integrated model for landscape evolution
- He looks for a hydrologic model and GIS tool in OntoSoft, decides to use PIHM and QGIS
- He first builds a prototype in MATLAB, and gets advice from OntoSoft to share and describe the code.
- He ports the code to C and integrates it with PIHM to develop a coupled landscape evolution-hydrologic model (LE-PIHM)
- He shares results in OntoSoft including Computable Document Format files with data, workflow, visualizations online.
- Others can then find the software and easily reuse it (especially scientists in other disciplines or who do not program).!
- Other CZO scientists use LE-PIHM for new questions, e.g., how do evolving landscapes affect the water cycle and geochemical weathering for different geologic settings.

Novel research contributions:
- An intelligent user interface that organizes the interaction in terms of science tasks (e.g., understand assumptions, do research with the software, cite the software, etc)
- An ontology to represent metadata for scientific software
- Automated crawlers that extract metadata for the software from the user’s web site
- Training modules to help scientists learn to share and describe software

Benefits
- **Open science**: Easy to disseminate models and software across disciplines
- **Accelerate research**: By reusing software
- **Reproducibility**: Easy to replicate results
- **Accessibility**: For non-programmers
- **Quality**: Best software that is well tested
Hypothesis: Groundwater flow and the life-cycle of forests introduce Important new feedbacks on landscape co-evolution from tectonic uplift and hydroclimatic formation of soil, regolith and stream networks

Joint work with Yu Zhang and Rudy Slingerland, Penn State University

Redistribution of soil moisture and groundwater (<1-10 yr) affects the life cycle of trees (~100 yrs). Tree-fall introduces bioturbation that effects macropore formation and soil movement (creep) (~100 yrs). Tectonic uplift (10^3-10^6 yrs) exposes country rocks to decay from physical and chemical weathering. Together these processes lead to the formation of regolith/soil (100’s yrs), stream networks and entire watersheds (>10^6 yrs).

Simulated evolution of an initially uniform landscape to a complex terrain and river network over 10^8 years.

"If we shared science software routinely, we would do cross-disciplinary research routinely."
-- CZO scientist

OntoSoft supports software sharing and reuse

Wrote a simplified version in MATLAB
Found a scalable hydrology model: PIHM
Found an efficient differential equation solver: SUNDIALS
Wrote data preparation scripts
Extended scalable model: LE-PIHM
Found an open source GIS software: QGIS

Guidance to describe software with semantic metadata that will allow others to find it
Faceted search allows feature-based comparison of software, customized to science needs
Assistance to choose a license, to get a DOI so others can cite the software, etc
Social recommendations of software and visualizations used by similar users

Simplified model is reused by others for other sites
Extended model is reused by other researchers for new questions
Data preparation scripts are used by others with similar data
Approach:
• Use of brokering technologies, i.e. software that mediates interactions between systems, to make it easier for geoscientists, especially those working across traditional disciplinary boundaries, to discover, access and share data, thereby improving their efficiency and productivity
• Piloting a new paradigm for data advertising and discovery through generation of data ads and specialized, semantically enabled web crawling
• Provide alternative options for interfacing with data facilities, minimizing their need to adopt new protocols and standards

Highlights
• Creation of a cloud-base brokering framework for interfacing with, initially, data resources in the fields of Hydrology Oceanography, Cryosphere and Climate.
• A series of detailed, investigator-defined science scenarios to guide our technology developments
• Community-developed brokering modules that make new data resources accessible through the brokering framework
• A repository of geoscience-related online resources discovered by a custom web crawler and queryable through a SPARQL (semantic query language) endpoint.
• Semantic technologies to discriminate potentially useful data discovered by crawling

Plans
• Complete enactment of the Hydrology Oceanography, Cryosphere and Climate science scenarios
• Develop brokering support for a multi-domain scenario
• Developing a natural language processing model for refining crawling results
• Refine interactions with other selected building blocks
Approach:
- Expose data assets through web services with similar styles.
- Provision of clear documentation for humans and for machines.
- Systems to formulate correct web service links to collections through URL builders.
- Provision of data in standard domain and cross formats as well as well-structured text formats easing interdisciplinary access across domains.

Highlights:
- All six funded partners have exposed data assets through web services with similar styles.
- Developed customized or Swagger interfaces as GUIs to build relevant URLs.
- Supported Domain formats, interdisciplinary formats, or well-structured text formats.
- Unfunded Partners
  - GeoWS services through the IRIS Web Service Shell
  - NGDC
  - UTEP Gravity and Magnetics
  - Intermagnet
  - Ingestion into the IRIS DMC and exposure through IRIS

Plans:
- Continued development of all services.
- Complete services for NEON and WOVDAT.
- Complete Brokering Accessors for all data partners.
- Expose additional data collections managed by GeoWS partner.

Contact GeoWS Services:
- www.geows.org
- http://earthcube.org/group/geows-geoscience-web-services
- Contact GeoWS

Services:
GeoWS: Geoscience Web

2015 Building Block Updates
Vision
- Advance “data-as-service” to embrace a broad set of pre-retrieval operations, beyond extant filtering/subsetting ops.

Goals
- Establish DAP (with rich server functions) as a key EarthCube interoperability tool, lowering data-usage barriers imposed by form, volume & distance.
- 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.

Project outputs:
- Prototype DAP server with extended data-access functions for statistical calculations, subsetting of UGRIDS, & feature extraction.
- Tech challenge: functional richness within simplicity/safety of DAP protocol/data model
- Tech approach: (tentative) embedding user-callable Python modules in the server.

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

Cyberinfrastructure focus areas:
Web services as enablers for:
- Data Access
- Data Interoperability
- Cross-Domain Data Use
The ODSIP outcome will include a nascent algebra of pre-retrieval operations prototyped for use in the scientific contexts below.

Geoscience focus areas/drivers:
A geoscience-wide concept, prototyped in 3 domain-specific but illustrative use cases:
- State Climate-model downscaling
- Coastal storm-surge prediction
- SST-front analysis/synthesis
**CHORDS: Cloud-Hosted Real-time Data Services for the Geosciences**

**Approach**
- Develop an easy to use portal for geosciences real-time data providers
- Provide services and tools for real time data consumers using standard protocols

**Highlights**
- Hydro and atmospheric real time streams being accessed via standard OGC protocols
- Prototype portal developed for data providers to easily ingest data into CHORDS

**Plans**
- Define and address requirements
- Test hydro and atmos r/t streams
- Test prototype provisioning tools for r/t data providers
- Collaborate with other EarthCube BBs

---

**Digital Crust – A 4D Exploratory Environment for Earth Science Research and Learning**

**Approach:**
- create space-time (4D) framework that enables geoscientists to register, organize and integrate Earth-system data relevant to the upper crust.

**Highlights**
- leveraging existing database resources developed by USGS and
- general 4D framework for data integration

**Plans**
- generate continent-scale permeability model that accounts for surface and subsurface geology
- develop interactive data contribution and editing environment for crust-related data

---

**EarthCollab: Enabling Scientific Collaboration and Discovery through Semantic Connections**

**Approach**
- Use semantic and linked data technologies to support scientific projects and resources
- Use the VIVO software to show connections between people, projects, data, & documents

**Highlights**
- Science and data center use cases outlined
- Ontology structures defined and modeled

**Plans**
- VIVO deployment and population for geodetic and arctic science (within UNAVCO and NCAR EOL)
- Software dev. to support information exchange between VIVO instances
- Connect with other EC BB semantic web projects
GeoLink: Semantics and Linked Data for the Geosciences

**Approach**
- Bring together experts from the geosciences, computer science, and library science;
- to develop Semantic Web components;
- that support discovery and reuse of data and knowledge.

**Highlights**
- ODPs published; knowledge base deployed and populated.

**Plans**
- Publish set of reusable ontology design patterns (ODPs) that describe datasets, samples, field expeditions, journal articles, funding awards, people, etc;
- Upgrade repositories to publish Linked Data using the ODPs;
- Populate a shared/integrated knowledge base and test science use cases.

geolink.org
**DAsHER CD:**
Developing a Data-Oriented Human-Centric Enterprise Architecture for EarthCube

**Approach:**
1. Designs an Enterprise Architecture (EA) for facilitating data communication and human collaboration
2. Follows a spiral development approach.

**Highlights**
1. Produces a four-volume design
2. Engage EarthCube communities through the development of EA.

---

**GEAR: Geospatial Enterprise Architecture for Research**

**Approach:**
- EarthCube as an evolving standards-based socio-technical federation of systems
- Supporting science enterprise through service integration, usage monitoring, community feedback, transparent scholarly communication

**Highlights**
- Registry of geoscience information resources and resource compositions
- EarthCube information model

---

**A Scalable Community Driven Architecture**

**Approach:**
- Service Oriented architecture framework
- Integration of existing technologies and capabilities
- Iterative design, community engagement and leverage of EarthCube on-going activities

**Highlights**
- Data Life Cycle, Information Model, Technology Mapping
- Active participation in TAC activities

---

**Plans**
- Conceptual Design Report
- Architecture recommendations
- Demonstration of functional components of geoscience cyberinfrastructure

---

**Plans**
1. External review for Volume 1 & 2 in May.
2. Complete first draft for Volume 3 and 4 by the end of Jun.
3. Coordinating with other working groups through the development.

---

**Plans**
1. External review for Volume 1 & 2 in May.
2. Complete first draft for Volume 3 and 4 by the end of Jun.
3. Coordinating with other working groups through the development.
**Goals:**
- Establish a governance structure for EarthCube and associated charters
- Demonstrate that the governance works and successfully engages the community

**The EarthCube Office**
- Supports the Leadership Council and the EarthCube organizational units, members, and funded projects.

**Status**
- Worked with the community to design a governance structure and associated charters
- Supported the community and emerging governance through workshops, web site, and discussion forums
**Overall Goal**
Documenting and explaining the transformational realignment of the institutions of the geosciences (social and technical) to better enable scientific break-through on the great challenges facing science and society.

**Selected Findings**
- **Strategic Alignment**
  - Shared vision of success for EarthCube
- **Structural Alignment**
  - Chartered EarthCube Forums
  - Establishing key standards, such as DOIs for data; IGSNs for physical samples
- **Process Alignment**
  - Focus on integration of funded projects
- **Technology Alignment**
  - Technology and Architecture strategic plan for interoperability and extensibility
- **Cultural Alignment**
  - 24 end user workshops and RCNs a first step toward a culture of open science

**Plans**
- About to launch the round 3 (R3) of stakeholder alignment data collection (Data from R1 n=845 and R2 n=694 now publically available)
- Developing a book on transforming science institutions (EarthCube and other such initiatives)

**Results**
- Contributed to establishment of EarthCube governance, with supporting charters
- Advancing theory of multi-stakeholder consortia in science (commentary published in *Science*)

**Illustrative Results**
From iSamples RCN stakeholder survey (n=280)
A. Have you submitted a Data Management Plan as part of funded research (current or past)? 45.0%
B. If so, did your Data Management Plan include a commitment to make descriptive data on physical samples publicly available? 30.6%
C. Also, did your Data Management Plan result in your data being discoverable through an on-line search? 19.6%
D. Finally, did your Data Management Plan result in placing physical samples in a repository? 6.9%

**Stakeholder Alignment Visualization**
- Key: Shades of Green=agree; Yellow=neutral; Red=disagree
- Each mini-cell is a stakeholder; central tendency (mean) is in the middle; spiraling out to outliers on the outer edges
Function
The Leadership Council, the elected voice of the EarthCube community, shapes the scope, vision and strategic direction for EarthCube.

Responsibilities
- Establish clear lines of authority and ensure consistency and transparency in all policies, procedures, and decision-making processes
- Articulate EarthCube’s mission and vision
- Frame the governing charter
- Outline a long-term strategy for EarthCube, including: determining what EarthCube enabled science will look like; identifying the technologies and best practices that define an EarthCube approach to the geosciences; creating the roadmap that will guide the transition from the conceptual design phase to the development of EarthCube architecture; and ensuring mechanisms exist for encouraging and sustaining productive conversations and interactions between domain scientists, technologists and the geosciences communities
- Plan for growth

Outcomes
- Policies and mechanisms for supporting integrative and collaborative efforts that add value to the work of the EarthCube community; by supporting travel, communication, collaboration, engagement, etc.
- EarthCube Charter
  [link](http://earthcube.org/document/2015/earthcube-charter)
- Strategic Science Plan that articulates the community’s vision of what EarthCube-enabled science should look like in 2020
  [link](http://earthcube.org/document/2015/earthcube-strategic-science-plan)
- EarthCube Strategic Vision document
  [link](http://earthcube.org/document/2015/earthcube-strategic-vision)
- EarthCube Strategic Plan
  (forthcoming)

Implementation
- Following elections held in the Fall of 2014, the Leadership Council was fully constituted on January 1, 2015. In addition to holding twice-monthly teleconferences, Leadership Council members held three in-person meetings to map out the tasks (including convening working groups and workshops) involved in creating the strategic documents that will sustain the organization and convey EarthCube’s mission, vision, and long-term goals to the community and the NSF.
**Function**

- The Science Committee connects the academic geoscience and technology communities in EarthCube, and ensures that end user geoscientist needs, requirements and aspirations are identified and prioritized.
- The Science Committee’s mission is to ensure community science goals are realized through the development of cyberinfrastructure that will enhance geoscientists ability to characterize and understand complex Earth systems.

**Activities**

- In response to the Leaderships Council’s charge that the Science Committee develop EarthCube’s Strategic Science Plan, Working Groups were formed to inform on: the science drivers EarthCube End-User Workshop participants identified; the alignment of funded projects objectives with those science drivers; and the intellectual aspirations that provide the broadest community motivation to pursue EarthCube.
- A March, 2015 workshop reviewed these working group’s outcomes, articulated EarthCube’s fundamental science goals, and created a final draft the Strategic Science Plan for community review.
- An April, 2015 workshop converged on a roadmap to support communication between the science and technology communities.
- The Science Committee also formed a working group to draft its Charter, and
- Funded the EarthCube Visiting Graduate Student and Early-Career Scientist Program.

**Outcome**

- “Geoscience 2020: Cyberinfrastructure to reveal the past, comprehend the present, and envision the future” is EarthCube’s Strategic Science Plan and was approved by the Leadership Council in May 2015: http://earthcube.org/document/2015/earthcube-strategic-science-plan
- The motivation for EarthCube-enabled science is expressed in three science goals:
  - **Quantify** limits of prediction and better understand the constraints on and limits of data and model accuracy and utility
  - **Characterize** the key processes, interactions and feedbacks operating at and across different temporal and spatial scales and biological, chemical, mechanical, and physical domains
  - **Deliver** a holistic, quantitative representation of critical biological, chemical, mechanical, and physical states and fluxes, to inform fundamental science and societal decisions
- Data are the key to unlocking knowledge
- EarthCube’s short term objective is greater data availability to geoscientists; the long term objective is enhanced knowledge availability for society
- The community anticipates that EarthCube’s implementation will make geoscientists ability to discover and obtain multidisciplinary and multiscale datasets from distributed repositories an integral component of scientific workflows.
The TAC is organized into five main community-driven working groups. It has also held several workshops for community engagement. TAC working groups did a comprehensive survey of all funded projects on their approach and needs.

**Survey of Funded Projects**
With input from all the working groups, TAC issued a questionnaire to all funded projects that covered a range of technology concerns:
- Scientific domains and use cases driving the work in the project
- Core technologies and functionalities being developed
- Standards being adopted
- Requirements for deploying the technology
- Interactions with other projects
- Expected results and their benefits to scientists

**Science Use Cases and Requirements**
In collaboration with the Science Committee, the TAC Use Cases working group developed a template to collect use cases that includes:
- Science drivers and objectives
- Key people and systems involved
- Assumptions and infrastructure needed
- Measures of success
- Basic flow of the use case scenario
- Technical requirements including:
  - Data sources and their characteristics
  - Metadata and standards
  - Software and visualizations

**Technology and Architecture Strategy**
Through community involvement and input, the committee has identified several high-level principles that should drive the development of EarthCube infrastructure:
1. EarthCube infrastructure should encourage and support open science practices across geosciences
2. Knowledge-rich components should include semantic metadata to support discovery, reuse, and interoperability
3. Distributed and flexible approaches will support community contributions and sustainability
4. Fostering standards will facilitate interoperability and cross-domain integration
5. EarthCube infrastructure should be accessible to institutions with limited resources
6. Earthcube infrastructure should have value at all scales, so even when small it is useful to scientists

**Working Groups**
- Co-Chairs: Yolanda Gil (U. Southern California) and Jay Pearlman (U. Colorado)
**TAC Use Cases Working Group**
**Goals:**
- Compile use cases to engage geoscientists and cyberinfrastructure experts

**Status**
- Created use case template

**Plans**
- Interview scientists to document use cases
- Create a compendium of use case summaries
- Synthesize technical requirements for future EarthCube development

**TAC Gap Analysis Working Group**
**Goals:**
- Review existing funded project technical capabilities and their interfaces
- Identify additional capabilities needed

**Status**
- Compiled inputs from funded projects

**Plans**
- Analyze inputs from funded projects and the EarthCube community
- Report to TAC on identified gaps and make recommendations

**TAC Testbed Working Group**
**Goals:**
- Recommendations for testbed development

**Status**
- Analyzed inputs from funded projects
- Considering alternative implementations

**Plans**
- Define a high-level testbed architecture that will meet EarthCube objectives
- Formulate recommendations for testbed implementations and sustainability

**TAC Standards Working Group**
**Goals:**
- Identify standards used in the community
- Develop, and prototype a workflow for coordinating standards activities

**Status**
- Created registry of standards used in EarthCube

**Plans**
- Complete a workflow prototype and engage the community for feedback
- Written recommendations to TAC on what existing standards should be adopted.
- Written recommendations to TAC on new standards to develop

**TAC Architecture Working Group**
**Goals:**
- Address Architecture Requirements
- Recommendations for steps forward

**Status**
- High level principles defined
- Foundation for initial community inputs

**Plans**
- Workshop in June to further architecture definition
- White paper on architecture options for community and TAC review
- Architecture recommendations for EarthCube
**Functions**
Serve in a coordinating and facilitating role that includes advancing the following responsibilities and goals:

- Provide a collective voice on behalf of the member data facilities to the NSF and other foundations and associations, as appropriate.
- Engage and establish partnerships among data facilities.
- 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 support the development and utilization of shared cyberinfrastructure, professional staff development and training, and other related activities.
- Foster innovation through collaborative and interdisciplinary projects.
- Increase engagement of relevant stakeholders.

**Activities**
- Partnership with the Coalition for Publishing Data in the Earth and Space Sciences – CDF is a signatory on the Coalition Statement.
- Recruitment and categorization of CDF members into the four defined categories.
- Participation in the EarthCube Leadership Council.

**Plans**
- Develop a directory of data facilities, including a catalog of resources, products, and services provided by each CDF member.
- Provide input to and help shape EarthCube architecture and its linkages to the tool, technologies, and resources at member data facilities.
- Ensure the explicit connection between the scientific process and EarthCube technical function, including facilitating testbeds and other mechanisms for evaluation of cyberinfrastructure components, and identifying and prioritizing science use cases.
- Continue to serve as an emissary between EarthCube and geoscience data and software providers and publishers, the geoscience community.
Function

- Enable broad dissemination of EarthCube information across academia, the private sector, and government
- Implement branding strategies to enable users to easily identify EarthCube results and outcomes
- Engage and support end users and stakeholders, including attracting new users to EarthCube
- Engage and support the next generation of EarthCube stakeholder leadership

Activities

- Communication and outreach
- Education
- Workshops
- Professional meeting sessions
- Professional meeting presence (e.g., booth)
- Early career incentives
- EC Ambassadors
- Best practices
- New tools, science and test use cases
- Social media
- End user feedback

Goal:

- Advance science
- Meet grand challenges
- Leverage shared cyberinfrastructure technology

“The problems of data use are half technical and half social.”
~Someone at almost every professional science meeting

Plans

- Program for building leadership
- Science/Technology collaborations
- Promote visualization formats
- Report of evaluation metrics, including discussion of successes and areas needing improvement

Results

- Result 1: Early Career Participation grants
- Result 2: Mapping the landscape
- Result 3: GSA meeting session accepted
- Result 4: Two AGU sessions proposed
- Result 5: Why EarthCube? Video
- Result 6: EarthCube Distinguished Lecture Program
### Function
The mission of the EarthCube Liaison Team is to act as a liaison to cyber-initiatives, collaborations, agencies, associations, private enterprises and other efforts and programs external to the NSF core geoscience constituency.

Consistent with this overall mission, the goal of the team is to establish partnerships, affiliations, and connections to external organizations and initiatives. Managing and maintaining formal and informal relationships.

### Activities
- Mapping the landscape of geoscience professional organizations, interoperability projects including international organizations.
- Establishing formal and informal relationships with organizations and projects that have similar goals and missions to EarthCube.

### Plans
- Develop a suite of formal and informal relationship agreements
- Continue to engage in outreach activities to establish relationships with organizations, agencies, cyber-initiatives and projects.

### Initial Landscape of Geoscience Organizations

<table>
<thead>
<tr>
<th>Function</th>
<th>Activities</th>
<th>Plans</th>
</tr>
</thead>
<tbody>
<tr>
<td>The mission of the EarthCube Liaison Team is to act as a liaison to cyber-initiatives, collaborations, agencies, associations, private enterprises and other efforts and programs external to the NSF core geoscience constituency.</td>
<td>- Mapping the landscape of geoscience professional organizations, interoperability projects including international organizations. - Establishing formal and informal relationships with organizations and projects that have similar goals and missions to EarthCube.</td>
<td>- Develop a suite of formal and informal relationship agreements - Continue to engage in outreach activities to establish relationships with organizations, agencies, cyber-initiatives and projects.</td>
</tr>
</tbody>
</table>

#### Domain: Geoscience Organizations

<table>
<thead>
<tr>
<th>Professional Organizations and Societies</th>
<th>Umbrella Organization Networks</th>
<th>Data Facilities</th>
<th>Research Infrastructure</th>
<th>Long Term Projects (5+ years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Atmosphere</td>
<td>ASPIRES/SPRIS</td>
<td>ERE</td>
<td>GIS</td>
<td>NODC</td>
</tr>
<tr>
<td>Lower Atmosphere</td>
<td>AGU-Cryosphere</td>
<td>ESCUL</td>
<td>ICOS</td>
<td>SYSCPO</td>
</tr>
<tr>
<td>Cryosphere</td>
<td>Research Station</td>
<td>CRES</td>
<td>ARA</td>
<td>IAOS</td>
</tr>
<tr>
<td>Oceans</td>
<td>ASPIRES/SPRIS</td>
<td>ERE</td>
<td>DEO</td>
<td>ASLO</td>
</tr>
<tr>
<td>Earth Surface</td>
<td>AAPG</td>
<td>STEPHIE</td>
<td>AGU</td>
<td>SVF</td>
</tr>
<tr>
<td>Deep Earth</td>
<td>AAPG</td>
<td>AAG</td>
<td>GSA</td>
<td>MSA</td>
</tr>
</tbody>
</table>

Co-Chairs: Lindsay Powers (The HDF Group) and Rick Ziegler (Environmental Protection Agency).
### Ibrahim Demir
**Dept. of Civil and Environmental Engineering, University of Iowa**

Ibrahim first got involved with the EarthCube to develop the Roadmap for Workflows in Geosciences in 2012. Since then, he has been active in the Technology and Architecture Committee, and currently co-chairs the Standards Working Group.

“I think information technology is going to change the geoscience world, and EarthCube will be at the front and center, and we need to make sure it addresses our needs.”

“My research interests are environmental information systems, scientific visualization of geo-spatial data, user interface design, and information communication. I dream of a future when the data from all environmental observations across the US are readily available and understandable to everyone.”

### Anna Kelbert
**College of Earth, Ocean, and Atmospheric Sciences, Oregon State U.**

Anna is co-PI of the Earth Systems Bridge project. She is a member of the Science Committee and the Technology and Architecture Committee, where she co-chairs the Funded Projects Gap Analysis Working Group. She recently published an EOS Transactions article titled “Science and Cyberinfrastructure: The Chicken and Egg Problem.”

“Study of deep Earth processes is by necessity dominated by numerical modeling, validated by interdisciplinary surface measurements. As I gathered from detailed community discussions, we need to streamline scientific reproducibility (including benchmarking); interdisciplinary communication; integration of data and models; data and software management procedures. As a geophysicist, I have had ample first-hand exposure to these issues myself. This equips me with a clear vision for EarthCube.”

### Branko Kerkez
**Civil and Environmental Engineering University of Michigan**

Branko’s involvement with EarthCube began while he was still a graduate student at Cal Berkeley. He is a co-PI on the real-time data building block project, where he works with a number of geoscientists and cyberinfrastructure experts on building the next generation of real-time data and sensor tools for EarthCube.

“EarthCube will allow me to tackle fundamental knowledge gaps in the hydrologic sciences by enabling a suite of adaptive experiments that will be informed by real-time hydrologic and meteorological data.”

“EarthCube will also change how we make non-scientific decisions by feeding into applications such as disaster prediction and sustainability.”
Raleigh Martin  
Dept. of Atmospheric and  
Oceanic Sciences, University of  
California Los Angeles  
Raleigh is a co-PI of the Sediment Experimentalist Network (SEN) Research Coordination Network (RCN) and a member of the EarthCube Science Committee. He is interested in developing data management, archiving, and analysis resources for earth-surface scientists collecting large but idiosyncratic datasets.  

“As a graduate student, I conducted a variety of flume experiments on sediment transport, and now as a postdoc I am performing field campaigns on the dynamics of wind-blown sand. In both cases I have developed novel methods and unique dark-tail datasets not amenable to existing data management approaches. Through the SEN RCN and EarthCube, I am learning ways to better manage my data and develop tools for other earth-surface scientists facing similar challenges.”

Matty Mookerjee  
Dept. of Geology,  
Sonoma State University  
Matty is the PI of the EC3 Research Coordination Network to facilitate conversations between the field-based geoscience community and the cyberscience community. EC3 organized a field trip in 2014 and has an upcoming trip in 2015 that takes computer scientists, data scientists, library and social scientists in the field to discuss field data collection, management, and integration.

“My research is in structural geology, and it encompasses field work, lab work, numerical modeling, and software development. I am very interested in developing a database for our community along with data standards, best practices, and workflows. In EC3, we are communicating with computer scientists about our data requirements and logistical considerations associated with data collection in the various field–based subdisciplines. We are also learning from them about the cutting edge technologies that we could use in the field.”

Ji-Hyun Oh  
Science Data Modeling and Computing Group,  
NASA/JPL  
Ji-Hyun is a post-doctoral researcher in the GeoSoft project. She is interested in learning about open source software best practices and applying them to climate research. She spends a lot of time developing software to handle very large amounts of data in various formats.

“I am not an expert programmer, but I have learned to write code to process data in NetCDF format and other common formats in my field. I also spend a lot of time writing codes to visualize my research outputs. I wish I could easily find and reuse software that would make these tasks easier. I also wish I could learn to make my software more scalable, and run it easily on efficient infrastructure deployed in the cloud. EarthCube could help make climate research more productive and cross-disciplinary.”
Allen Pope
National Snow and Ice Data Center, U. Colorado Boulder

Allen is a member of the Early Career Advisory Committee of the EarthCube GeoSoft project. He is participating in the Geoscience Paper of the Future activity, where he is learning best practices for software and data sharing and applying them in his ongoing work. He plans to disseminate those best practices through his involvement in the Association of Polar Early Career Scientists.

“Investing time to learn about how to incorporate new technologies in my work is always hard, but it always pays off. I think it is important to make our research reproducible, and to share the results of our work not just with our colleagues but with researchers in other disciplines. EarthCube wants to make it easy for all of us to build on other research. My work on glaciology, polar remote sensing, and supraglacial lakes would benefit from the integration of additional data and models from other disciplines.”

Plato Smith
Council on Library and Information Resources and Digital Library Federation, U. New Mexico

Plato is involved in the EarthCube Technology and Architecture Committee, where he co-chairs the Funded Projects Gap Analysis Working Group. He recently led the Committee’s work on a survey of all EarthCube funded projects regarding their requirements and capabilities. He also participates in the EarthCube Engagement Team.

“EarthCube is leading the way in transforming how science communities interact with infrastructure. Scientists are articulating their requirements, learning about new technologies, and expressing the benefits that the vision of EarthCube could have for scientific research.”

“As a data management specialist, I work with the research personnel of various labs in writing metadata records for their publicly-funded data. It takes a lot of effort to keep all the datasets well documented and accessible, and my hope is that EarthCube will give us improved technology to make this easier. Once annotated, all this data should be accessible to anyone who might need it. EarthCube wants to make data easy to find, easy to access, easy to query, easy to integrate, and easy to reuse. I want to help EarthCube make that vision happen.”

Mimi Tzeng
Dauphin Island Sea Lab

Mimi is an active member of the EarthCube Science Committee, where she led the Funded Projects Analysis Working Group and now the Web Services task to attract scientists to participate in the committee. She is co-author of the Committee’s report “EarthCube Strategic Science Plan: Geoscience 2020”.

“EarthCube is leading the way in transforming how science communities interact with infrastructure. Scientists are articulating their requirements, learning about new technologies, and expressing the benefits that the vision of EarthCube could have for scientific research.”

“My research is on data management and curation, and through my involvement in EarthCube I can study how scientists collaborate, collect, integrate, manage, analyze, visualize, share, and use data across disciplines.”
Join EarthCube!

Become a member
Subscribe to the newsletter
Contribute to the working groups

Be the future

http://www.earthcube.org
Back cover: Geoscience Papers of the Future

To learn best practices of software and data sharing, thirteen members of the Early Career Advisory Committee of the EarthCube GeoSoft project are publishing a Geoscience Paper of the Future (GPF). This implies publishing all the software and data used to produce the results of the paper, as well as detailed workflows and provenance of how they were generated. They were trained by GeoSoft project members on best practices for software and data sharing, open source software, and provenance. The papers will appear in a special issue of the AGU Earth and Space Science journal. The special issue will be widely advertised by AGU calling for contributions from the community. Training sessions will be given by GeoSoft experts and by already trained GPF participants at the ESIP Summer Meeting, AGU Fall Meeting, and other community events.

The figure includes pictures of the participants, workflow and provenance diagrams, and some of the visualizations of their results. From top to bottom, left to right: Cedric David, NASA/JPL (hydrologic modeling); Ibrahim Demir, U. Iowa, (hydrology sensor networks); Robinson W. Fulweiler, Boston U. (biogeochemistry in marine ecology); Jonathan Goodall, U. South Carolina (hydrology visualizations); Leif Karlstrom, U. of Oregon (volcanic vent clustering); Kyo Lee, NASA/JPL (regional climate model evaluation); Heath Mills, U. Houston (geochemistry and marine microbiology); Ji-Hyun Oh, NASA/JPL (tropical meteorology); Suzanne Pierce, U. Texas Austin (hydrogeology for decision support); Allen Pope, U. Colorado Boulder (glaciology); Mimi Tzeng, Dauphine Sea Lab (ocean fisheries); Sandra Villamizar, U. California Merced (river ecohydrology); Xuan Yu, U. Delaware (hydrologic modeling).