EarthCube Data Science Publications

Story
Slides

Slide 1 EarthCube Data Science Publications
Slide 2 Agenda
Slide 3 Timeline
Slide 4 EarthCube 2011 Shows How Collaboration Is Changing Geoscience Research
Slide 5 EarthCube 2014 Sprint to "Stretch Goals: Open Research Data Publication and Integration"
Slide 6 EarthCube 2015 Data Science for NSF Polar Cyberinfrastructure and MIT Big Data Course Meetup
Slide 7 EarthCube May 2015 Data Science Publications
Slide 8 Key Documents: MindTouch Knowledge Base
Slide 9 Conclusions and Recommendations
Slide 10 Making the Story Behind USGS EM-EH Data Come Alive: Slides
Slide 11 What I Need From You
Slide 12 Positives of Civic Hacking
Slide 13 Data Science Publication for USGS Minerals Big Data: Story and Slides
Slide 14 Data Science for USGS Minerals Big Data: MRDS-shp in Spotfire
Slide 15 Some Conclusions and Recommendations
Slide 16 Welcome to Semantic Community.info Community Infrastructure Sandbox for 2015

Spotfire Dashboard
Research Notes
Agenda
EarthCube All Hands Meeting 2015

Mini Breakout Sessions
Building on the “Building Blocks” - supporting a Complex Research Scenario in an EarthCube multi-project synthesis demonstration
Data Science Publication for USGS Minerals Big Data

Full Breakout Sessions
Building a registry of geoscience resources: CINERGI components and how they can be used in your metadata curation system
Connecting EarthCube Projects to the Data Facilities
EarthCube Integration: Bringing Projects into a Test Environment
EarthCube Technology and Architecture: Strategic Planning and Ongoing
EarthCube Use Cases: Current Status and Forward Look
Essential Variables and Associated Challenges for EarthCube Science Drivers
Federal Agency Federal Agency Public Access to Research Results (PARR)
Getting the Word Out: EarthCube capabilities, with an example from GeoSoft
How will GeoWS help geoscientists?
Integrating Real-time Geosciences Data into the EarthCube Framework
Interoperability and Reuse of Geoscience Models
The iSamples RCN Survey & Working Group activities: How can we maximize impact for EarthCube?
Making Geoscience Resources More Discoverable: Connecting People, Data and Products Through the Semantic Web
Making it Easier to Find, Use and Share the Source Data and Outputs of Research – Engaging Geoscientists with Illustrative Use Cases
New Synergies and Lessons Learned in the Paleogeosciences through the C4P RCN

Registered Attendees

Story
Summary History and Workplan
Slides
Slide 1 EarthCube All Hands Meeting New Interest Group: EarthCube Data Science Publications
Slide 2 Summary
Slide 3 Overview
Slide 4 The EarthCube Vision
Slide 5 EarthCube Web Site
Slide 6 EarthCube Workspace
Slide 7 EarthCube Work Groups: General
Slide 8 EarthCube Work Groups: Types
Slide 9 EarthCube Work Groups: Actions
Slide 10 EarthCube Data Science Publications Interest Group
Slide 11 EarthCube Data Science Publications
Slide 12 Data Mining the EarthCube Web Site
Slide 13 Community-Developed Geoscience Cyberinfrastructure
Slide 14 Ten Simple Rules for the Care and Feeding of Scientific Data
Slide 15 My June 2014 Data Journalism Story
Slide 16 EarthCube Shows How Collaboration Is Changing Geoscience Research
Slide 17 My November 2011 Data Journalism Story
Slide 18 Data Publication in a Data Browser
Slide 19 Birds of a Feather Session Proposal
Slide 20 Some Excellent Questions and Our Answers
AGU 2014 Session Ideas - EarthCube Community Engagement Team Suggestions for AGU Sessions

[Accepted] Collaborative Community Platforms for the Earth Sciences
[Accepted] International Cooperation to Further the Earth Sciences
[Accepted] Scalable and Adaptable Architecture for Earth Science Cyberinfrastructure
[Accepted] Cyberinfrastructure for field work: data standards, computer applications, instrumentation and best practices
[Accepted] Information models and interchange formats for geoscience data
[Merged] Methods, tools and best practices of data visualization in Earth and space science
[Merged] Semantic technologies in the life cycle of geoscience data
[Accepted] Provenance enablement in Earth and space science
[Accepted] Computational intelligence in Earth and space science
[Accepted] The Brokering Approach: its services and applications for multidisciplinary science research
[Submitted] Linking Physical Objects to Cyberinfrastructure
[Submitted] An Open Source Understanding Framework for Geosciences Technologies
[Submitted] Polar Cyberinfrastructure
[Submitted] Cloud Computing for Geosciences

EarthCube Summit 2014

Agenda

Selected Slides

00-EarthCube Agenda at OGC TC, DC.ppt
  Slide 1 Title Slide
  Slide 2 Agenda

01-OGC_EarthCubeCommunityEngagement.pptx
  Slide 1 Title Slide
  Slide 2 Stakeholder Engagement (Year 1): EarthCube Assembly
  Slide 3 EarthCube Test Governance: Target Community
  Slide 4 The EarthCube Vision

02-Open_Standards_Role_in_EarthCube_IT.ppt
  Slide 1 Title Slide
  Slide 2 The EarthCube Strategy
  Slide 3 Open Data and a Culture of Cooperation
  Slide 4 Prototyping Versus Specifying
Ten Simple Rules for the Care and Feeding of Scientific Data

Article

Figures

Introduction

Figure 1. Two pages (scan) from Galilei's Sidereus Nuncius ("The Starry Messenger" or "The Herald of the Stars"), Venice, 1610.

Rule 1. Love Your Data, and Help Others Love It, Too

Rule 2. Share Your Data Online, with a Permanent Identifier

Rule 3. Conduct Science with a Particular Level of Reuse in Mind

Rule 4. Publish Workflow as Context

Rule 5. Link Your Data to Your Publications as Often as Possible

Rule 6. Publish Your Code (Even the Small Bits)

Rule 7. State How You Want to Get Credit

Rule 8. Foster and Use Data Repositories

Rule 9. Reward Colleagues Who Share Their Data Properly

Rule 10. Be a Booster for Data Science

Links to Useful Resources

A: General Data Repositories

B: Directories of Research Data Repositories

C: Workflow Management Systems

D: Source Code Repositories
Welcome

All Hands Meeting

Day 1 - Tuesday, June 24
Day 2 - Wednesday, June 25
Day 3 - Thursday, June 26

Session Descriptions

- DisConBB meets IOOS marine in-situ data
- Addressing Data Heterogeneity in EarthCube Information: A Semantic Building Blocks & Cyberinfrastructure Perspective
- Workforce Development: Panel Discussion
- Software Sustainability in Geosciences
- Aligning Representations: Practical Explorations of Metadata, Ontologies, and Semantic Web for EarthCube
- Data Assimilation Frameworks in Geosciences
- Middleware Governance
- The BCube Brokering Framework
- Inventory of EarthCube Resources: CINERGI Architecture and Workflow
- EarthCube Architecture Forum
- Social dimensions of EarthCube data and infrastructure
- Domain Registries of Information Resources: a How-To for Your Community

Workshops and Technology Demonstrations (*denotes accompanying poster)

- Crawling the Web for EarthCube*
- Ten Simple Rules for the Care and Feeding of Scientific Data
- EarthCube Cloud Commons: GeoCloud Workshop*
- Simplifying Scientific Data Management through SaaS*
- The VHub Community Cyberinfrastructure—Sharing and Growing*
- Increasing the Potential Usability, Access, and Funding of Our Geological Collections—Presentation and Open Discussion

Currently Proposed Working Groups

- EarthCube Technical Services Discussion
- EarthCube Community Workforce Development
- Domain Registries of Information Resources: a How-To for Your Community

Governance

RCNs
Building Blocks
Conceptual Designs
Bringing Geochronology into the EarthCube Framework
Articulating Cyberinfrastructure Needs of the Ocean Ecosystem Dynamics Community
Engaging the Atmospheric Cloud/ Aerosol/ Composition Community
Developing a Community Vision of Cyberinfrastructure Needs for Coral Reef Systems Science
Rock Deformation and Mineral Physics Research
Science-Driven Cyberinfrastructure Needs in Solar-Terrestrial Research
Increasing the Access to and Relevance of Marine Seismic Data

Get Involved

Introduction

1. Join the Workspace
2. Subscribe to the Newsletter
3. Get EarthCube Involved in Your Work
4. Help Build a Community Inventory of Data Resources
5. Follow EarthCube on Social Media

Funding

New Funding Opportunities

November 2013: NSF Announces EarthCube Amendment III Funding Solicitation (NSF 13-529)

2013 NSF EarthCube Awards

Governance
Research Coordination Networks (RCNs)
Building Blocks
Conceptual Designs

EarthCube Groups

Join an EarthCube Group
Create an EarthCube Group

Assembly Workshops

Test Enterprise Governance Assembly Group Workshops

News and Events

EarthCube News

EarthCube session at the OGC quarterly meeting in DC
OSTP Announces Directive on Improving Management of Scientific Collections
C4P Webinar - Neotoma and CINERGI 2/4/2014
Semantic Web/Linked Data Telecon
EC3 Field trip to Yosemite and Owens Valley
EarthCube Monthly Community Webinar

January/February Issue of D-Lib Magazine Now Available
Digital Mapping Techniques 2014 Workshop

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by Mindtouch
EarthCube Test Enterprise Governance Advisory Committee
EarthCube Monthly Community Webinar
BCube Project Webinar Set for December 17
Help Us Build a Community Catalog of Geoscience Information Resources!
ICCGI 2014 Call for Papers, Tutorials, and Panels
2013 International Data Rescue Award Ceremony to be Held at AGU
BRDI Open Data Challenge Now Accepting Entries
EarthCube Monthly Community Webinar
EarthCube: NSF funds $14.5 million in grants to improve geosciences cyberinfrastructure
End-User PI Workshop Executive Summary Now Available

Conferences
EGU 2014
2014 Ocean Sciences Meeting
American Meteorological Society Annual Meeting 2014
EarthCube at AGU

Workshops
C4P Paleobiology Workshop
End-user Communities & Professional Societies Assembly Workshop
IT/CS/FOSS Assembly Workshop
EarthCube Portfolio Assembly Group Workshop
Data Facilities Assembly Group & End-User Workshop

Upcoming Assembly Workshops: Overview

Contact
EarthCube Commons

EarthCube Shows How Collaboration Is Changing Geoscience Research

Story
The Value Proposition of Agile Analysis
ESIP Data.gov Application
Charrette Wrap Up
Post Charrette
Example Guiding Principles from Amazon's Jeff Bezos
Low Hanging Fruit "Eggs of a Feather"
EarthCube Birds of a Feather Meetings

Notes
Welcome!
Draft Charrette Agenda

Day 1, Tuesday Nov 1 From Here to EarthCube: A Call to Action;
Day 2, Wednesday Nov 2 The Challenges to Realizing EarthCube

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
Day 3, Thursday Nov 3 Mapping vision, challenges, and milestones to governance and designs
Day 4, Friday Nov 4 Moving Forward with EarthCube

Papers
EarthCube White Paper: Advancing Scientific Understanding by Communicating via Data Interactive Publications
Research Data Lifecycle Management as a Service
White Paper: National Data Infrastructure for Earth System Science
EarthCube Governance Whitepaper
What is the Technology Scope for Building a Geospatial Cyberinfrastructure to Support An Interactive Data Mining Framework for EarthCube
Use Cases to Test OGC O&M Profile
A COMMUNITY MODEL FOR WATER IN THE CONTINENTAL EARTH SYSTEM
Atmospheric Sciences and Informatics
The Brokering Approach for Earth Science
DataONE—Enabling Cyberinfrastructure for the Biological, Environmental and Earth Sciences
OGC Document 11-145 Cyberarchitecture for Geosciences
OGC Document 11-143 Governance: An OGC white paper for NSF EarthCube
Example Use Case in Climate Impacts Research
Atmospheric Sciences and Informatics EarthCube Driver Whitepaper: Use Cases
Unidata Governance: A Quarter Century of Experience 1
OGC Document 11-147 Sensor Webs: An OGC white paper for NSF EarthCube
An EarthCube Design Process: Unidata’s Perspective
Technology Solutions for Scientific Data Interoperability: Unidata’s Perspective
Provenance in Earth Science Cyberinfrastructure: A White Paper for NSF EarthCube
Semantic Web technology-driven Geosciences Network
EarthCube DeepSearch: Discovering and Integrating Data from the Geoscience Deep Web
Discovery White Paper Short From Federation of Earth Science Information Partners
Using Metadata, Data/Service Quality and Knowledge to Facilitate Better Data Discovery, Access, and Utilization for Supporting EarthCube
EarthCube White Paper A Data-Based Professional Networking System
EarthCube Design Approach
A Workflow-Based Knowledge Management Architecture for Geodynamics Data
One perspective on GeoSciences Data and Coastal Science and Sustainability: A brief to NSF’s EarthCube
A National Geoinformatics Community (NGC)
EarthCube White Paper: A Future Vision In Action at APL for Cyberinfrastructure Enabled Geoscience

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
**Data Science Publication for USGS Minerals Big Data**

The [EarthCube Data Science Publications Special Interest Group](https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications) was formed for last year's annual meeting and has been active this past year in producing a series of [Meetups](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications) featuring [Data Science Data Publications](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications) for senior federal government leaders and agencies.

This session would be lead again by Dr. Brand Niemann and Dr. Joan Aron. and feature our collaborative work with Dr. Sophia Liu, Mendenhall Postdoc Fellow at the U.S. Geological Survey, on [Data Science Publication for USGS Minerals Big Data](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications), which will be the subject of our [Federal Big Data Working Group Meetup](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications) on June 15th in the Washington, DC, area.

The agenda would be as follows:

- **Context:**
  - Dr. Brand Niemann, Director and Senior Data Scientist, Semantic Community
- **Example:**
  - Dr. Sophia Liu, Mendenhall Postdoc Fellow at the U.S. Geological Survey
- **Discussion:**
  - Dr. Joan Aron, President and Founder of Science Communication Studies

This session will update EarthCube meeting attendees on the activities of the EarthCube Data Science Publications Special Interest Group, provide an example of a Data Science Publication including a demo, and provide an opportunity to help them produce Data Science Data Publications of their work.

The presentation materials have already been posted for advance review. Questions should be sent to Dr. Brand Niemann: bniemann@cox.net

Our slides below include:

- [Side 3 Timeline](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications)
- [Slide 4 EarthCube 2011 Shows How Collaboration Is Changing Geoscience Research](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications)
- [Slide 5 EarthCube 2014 Sprint to "Stretch Goals: Open Research Data Publication and Integration](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications)
- [Slide 6 EarthCube 2015 Data Science for NSF Polar Cyberinfrastructure and MIT Big Data Course Meetup](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications)
- [Slide 7 EarthCube May 2015 Data Science Publications](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications)
- [Slide 8 Key Documents: MindTouch Knowledge Base](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications)

The last slides show the 2015 EarthCube All Hands Meeting Hackathon the author did on the 2nd day of the meeting based on the [Key Documents](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications) and other content of the first day.

Our Mini Session Conclusions and Recommendations are:

- Semantic Community uses a Semantic Wiki for Data Science Data Publications.
- The recent 2015 [NIST Big Data Framework](https://semanticommunity.info/Data_Science/EarthCube_Data_Science/Publications) documents and use cases were produced as a Data Science Data Publication to address the "Holdren Memo."
• Data Science Data Publications address most of the topics in the other All Hands Meeting sessions.
• This is a “win-win” for the EarthCube and Data Science Communities:
  ◦ Domain Scientists get increased use and visibility of their data.
  ◦ Data Scientists get access to data and subject matter expertise.
• The Federal Big Data Working Group will continue to data mine EarthCube Use Cases and Data Sets and produce Data Science Data Publications for its Meetups.
• This is Data-to-Knowledge-to-Action for Decision Making.

Our EarthCube All Hands Meeting Followup is as follows:

Volunteer to help with EarthCube 2016 to implement Roger Wakimoto’s three suggestions:
1. Scientific reproducibility (like NIH doing and NSF starting to do)
2. Citizen/Community Science (USGS is a good example)
3. External Science Advisory Committee (Data Science Community)

Our Friday Mini-Session on “Data Science Publication for USGS Minerals Big Data” included those three elements!

Help implement three suggestions that Ken Rubin liked:
• A simpler EarthCube Mission Statement: A Community of Practice for the GEO 2020 Imperatives and Frontiers
• The USGS’s new motto: “Start with Science” and produce Data Products.
• Partner with the Data Science Communities at NSF* and elsewhere.
  ◦ *NSF funds about 300 Data Science/Big Data projects.

Our Federal Big Data Working Group Meetup has over 750 members (data scientists, statisticians, etc.) and has two meetups in June (1 and 15) that feature hackathons on Homeless data and USGS Minerals Big Data, respectively.

We can do a series of EarthCube Use Case Meetup/Hackathons between now and the EarthCube 2016 All Hands Meeting.

MORE TO FOLLOW

Slides

Slide 1 EarthCube Data Science Publications

http://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications


http://www.meetup.com/Federal-Big-Data-Working-Group/events/221810524/
EarthCube Data Science Publications

Dr. Joan Aron
Dr. Sophia Liu
Dr. Brand Niemann
May 29, 2015

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
http://www.usgs.gov/centers/m<k1>rc/grou</k1>knets/201818127/

Slide 2 Agenda

Agenda

• Context:
  – Dr. Brand Niemann, Director and Senior Data Scientist, Semantic Community
• Example:
  – Dr. Sophia Liu, Mendenhall Postdoc Fellow at the U.S. Geological Survey
• Discussion:
  – Dr. Joan Aron, President and Founder of Science Communication Studies
**Side 3 Timeline**

**Timeline**

- **November 2011:**
  - EarthCube Shows How Collaboration Is Changing Geoscience Research
    (EarthCube Charrette & AOL Gov Story)
- **June 2014:**
  - EarthCube Data Science Publications Special Interest Group Breakout Session
- **February 2015:**
  - Data Science for NSF Polar Cyberinfrastructure and MIT Big Data Course Meetup
- **May 2015:**
  - EarthCube All Hands Meeting
  - Data Science Publication for USGS Minerals Data Mini Session
- **June 2015:**
  - Data Science Publication for USGS Minerals Big Data Meetup

**Slide 4 EarthCube 2011 Shows How Collaboration Is Changing Geoscience Research**


**EarthCube 2011 Shows How Collaboration Is Changing Geoscience Research**

- **Question:** When EarthCube exists and is widely useful in 2021, what does a day in the life of a scientist in your field look like? Think about your: Research, Teaching, Outreach, Workforce Development, and Interaction with the greater scientific community.
- **Answer:** I will still have some form of the tools I am using now, they will just be more integrated with one another and even more connected to many sources of information and data so I can create data stories (like I do now for AOL Government) with greater ease and frequency because the time for collection and communication is lessened and the time for analysis is maximized -- actually the analysis tool facilitates the collection and communication parts.
- **Note:** I also provided an example of the kind of agile analysis of Earth Science data that I had done earlier in the year for the Earth Science Federation Annual Conference.

http://gov.aol.com/2011/12/31/earthcube-shows-how-collaboration-is-changing-geoscience-research/

**Slide 5 EarthCube 2014 Sprint to "Stretch Goals: Open Research Data Publication and Integration"**

http://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications#Summary_History_and_Workplan
EarthCube 2014 Sprint to "Stretch Goals: Open Research Data Publication and Integration"
EarthCube May 2015 Data Science Publications

- EarthCube All Hands Meeting:
  - Mission: Community-led cyberinfrastructure that will allow for unprecedented data sharing across the geosciences.
  - Keynotes: Two best practices examples.
- Eva Zanterka, EarthCube Program Director:
  - Question: How do we get the EarthCube domain scientists to work with open data?
  - My Answer: Get data scientists to mine EarthCube GEO domain data like the NSF DataViz Hackathon for Polar Cyberinfrastructure in early November 2014 and then organize a Meetup in February 2015 to discuss the results to the Federal Big Data Working Group Meetup.
- Brand Nienhuis, Director and Senior Data Scientist, Semantic Community:
  - NSF needs a .gov for GEO that its Big Data/Data Science grant projects can use.
  - GEO needs a Data Science Data Publication Commons for its four domains: Earth, Oceans, Polar, and Atmosphere/Geospace like we have started here.

Key Documents: MindTouch Knowledge Base
Slide 9 Conclusions and Recommendations

Conclusions and Recommendations

- Semantic Community uses a Semantic Wiki for Data Science Data Publications.
- The recent 2015 NIST Big Data Framework documents and use cases were produced as a Data Science Data Publication to address the “Holdren Memo.”
- Data Science Data Publications address most of the topics in the other All Hands Meeting sessions.
- This is a “win-win” for the EarthCube and Data Science Communities:
  - Domain Scientists get increased use and visibility of their data.
  - Data Scientists get access to data and subject matter expertise.
- The Federal Big Data Working Group will continue to data mine EarthCube Use Cases and Data Sets and produce Data Science Data Publications for its Meetups.
- This is Data-to-Knowledge-to-Action for Decision Making.

Slide 10 Making the Story Behind USGS EM-EH Data Come Alive: Slides

http://semanticommunity.info/Data_Science/Data_Science_for_USGS_Minerals_Big_Data#Slides_2
Slide 11 What I Need From You

What I Need From You

- Draft a Challenge
  - National Day of Civic Hacking
  - Local Hackathon Meetups
  - Metrics for Judging
- Access to Experts
  - Domain Scientists
  - GIS Specialists
  - Web Developers
  - Typical and Potential Users

Slide 12 Positives of Civic Hacking

Positives of Civic Hacking

- Make existing USGS data more interactive from mockups to proof-of-concepts
- Integrate multiple datasets together that may have never been connected or analyzed together before
- Visualizations that reveal patterns as well as outliers, better decision-making, unexpected insights, data quality and management issues, predictors of significant trends
- Show the value of civic hackers and the need to hire data scientists, user experience designers, etc

Slide 13 Data Science Publication for USGS Minerals Big Data: Story and Slides

http://semanticommunity.info/Data_Science/Data_Science_for_USGS_Minerals_Big_Data#Story
Slide 14 Data Science for USGS Minerals Big Data: MRDS-shp in Spotfire

Web Player
Some Conclusions and Recommendations

- The USGS has evolved to Starting with Science so I started with Data Science for USGS Minerals Big Data.
- I data mined the USGS Mineral Resources Program (MRP), the USGS Commodity Statistics and Information, and the Mineral Resources On-Line Spatial Data.
- I converted PDF to MindTouch and Excel Knowledge Bases and used Spotfire for analytics and visualizations.
- The result is a Data Science Data Publication. The USGS has many more PDF publications and data sets that could be Data Science Data Publications.

Welcome to Semantic Community.info Community Infrastructure Sandbox for 2015

- The Profit and Data Enterprises:
  - Marcus Lemonis, star of The Profit, a CNBC reality show about saving small businesses through People, Process, and Products.
  - The Federal Big Data Working Group Meetup is also about helping government agencies develop:
    - People – Data Scientists/Chief Data Officers
    - Process – Data Infrastructure
    - Products – Data Publications
  - By providing MOOCs/Meetups for training and networking.
- Five MOOCs for Big Data Applications and Analytics:
  - Data Science for Mining of Massive Datasets by NiaDsp on Stanford MOOC (IN PROCESS)
  - USGS Data Science MOOC by NiaDsp (This Meetup)
  - Data Science for EDA by Data Analytics (IN PROCESS)
- Federal Big Data Working Group and Virginia Big Data Meetups:
  - June 1: Data Science for Homeless Data: EBWorks, Tableau, & Spotfire Bakeoff
  - June 15: Data Science for USGS Minerals Big Data
  - June 26: Data Science for Cyber Physical Systems: Internet of Things
  - July 10: Data Science for Affordable Care Act Data
  - Late Summer/Early Fall: BIPA6, Chief Data Scientist, Whitehouse, Linda Powell, Chief Data Officer, CTPI, etc.

Spotfire Dashboard

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Research Notes

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications

Updated: Sun, 19 May 2019 05:26:10 GMT

Powered by MindTouch™
## Agenda

Source: [https://docs.google.com/spreadsheets...#gid=422936923](https://docs.google.com/spreadsheets...#gid=422936923)

<table>
<thead>
<tr>
<th>Time</th>
<th>Room</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30 AM</td>
<td>Fitzgerald A</td>
<td>Breakfast/Registration</td>
</tr>
<tr>
<td>8:30 AM</td>
<td>Fitzgerald CDE</td>
<td>EarthCube 101</td>
</tr>
<tr>
<td>8:30 AM</td>
<td>Fitzgerald CDE</td>
<td>Opening Remarks</td>
</tr>
<tr>
<td>Keynote Speakers: Peter Wiebe (WHOI) and Cathy Constable (Scripps)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Fitzgerald CDE</td>
<td>Break</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Fitzgerald CDE</td>
<td>EarthCube Scope and Vision</td>
</tr>
<tr>
<td>EarthCube Funded Project Lightning Talks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Fitzgerald CDE</td>
<td>Lunch: Presentations and Panel by Brian Mapes (U. Miami), David Moore (U. Arizona), and Ian Foster (U. Chicago)</td>
</tr>
<tr>
<td>1:30 PM</td>
<td>Fitzgerald A</td>
<td>Making Geo. Resources</td>
</tr>
<tr>
<td>More Discoverable: Connecting People, Data and Products through the Semantic Web</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How will GeoWS help geoscientists?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EarthCube Use Cases: Current Status and Forward Look</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connecting EarthCube Projects to the Data Facilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Mayernik, Narock]</td>
<td>[Carbotte]</td>
<td>[Aronson, Hsu, Stocks, et al.]</td>
</tr>
<tr>
<td>Fitzgerald A</td>
<td>Fitzgerald B</td>
<td>Fitzgerald C</td>
</tr>
<tr>
<td>3:00 PM</td>
<td>Fitzgerald CDE</td>
<td>Break</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>3:30 PM</td>
<td>Funded Project Demonstration Showcase</td>
<td></td>
</tr>
<tr>
<td>5:00 PM</td>
<td>Reception &amp; Poster Session (Continued Showcase)</td>
<td></td>
</tr>
<tr>
<td>7:00 PM</td>
<td>End of Day</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Day 2 - Thursday, May 28</strong></td>
<td></td>
</tr>
<tr>
<td>7:30 AM</td>
<td>Breakfast/ Registration</td>
<td></td>
</tr>
<tr>
<td>8:30 AM</td>
<td>Opening Remarks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marge Cavanaugh (Deputy Assistant Director, NSF GEO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Keynote Speakers: Suzanne Pierce (U. Texas) and Heath Mills (U. Houston)</td>
<td></td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitzgerald CDE</td>
<td></td>
</tr>
<tr>
<td>10:30 AM</td>
<td>New Synergies and Lessons Learned in Paleogeosciences through the C4P RCN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The EarthCube Technology and Architecture: Strategic Planning and Ongoing Activities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Essential Variables and Associated Challenges for EarthCube Science Drivers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Making it Easier to Find, Use and Share the Source Data and Outputs of Research</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Lehnert, Noren, Uhen, et al.]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Gil, Pearlman, Yang]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Rubin, Aronson]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Khalsa, Pearlman, Duerr]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitzgerald A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitzgerald B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitzgerald C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitzgerald DE</td>
<td></td>
</tr>
<tr>
<td>12:00 PM</td>
<td>Lunch: Presentations by Emma Aronson (EC Science Committee) and Yolanda Gil &amp; Jay Pearlman (EC Technology and Architecture Committee)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitzgerald CDE</td>
<td></td>
</tr>
<tr>
<td>1:30 PM</td>
<td>EarthCube Integration: Bringing Projects into a Test Environment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Integrating Real-time Geosciences Data into the EarthCube Framework</td>
<td></td>
</tr>
<tr>
<td></td>
<td>From Invention to Innovation in EarthCube with Examples for Disseminating Software Sharing and Science Reproducibility through CINERGI Components</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Building a Registry of Geoscience Resources</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Your Metadata Curation System</td>
<td></td>
</tr>
</tbody>
</table>

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00 PM</td>
<td>Break</td>
</tr>
<tr>
<td></td>
<td>Fitzgerald CDE</td>
</tr>
<tr>
<td>3:30 PM</td>
<td>Community Input on Governance and Future Directions</td>
</tr>
<tr>
<td>5:00 PM</td>
<td>End of Day</td>
</tr>
<tr>
<td>Day 3 - Friday, May 29</td>
<td></td>
</tr>
<tr>
<td>7:30 AM</td>
<td>Breakfast/ Registration</td>
</tr>
<tr>
<td>8:30 AM</td>
<td>The iSamples RCN Survey &amp; Working Group Activities: How Can We Maximize Impact for EarthCube?</td>
</tr>
<tr>
<td></td>
<td>Interoperability and Reuse of Geoscience Models</td>
</tr>
<tr>
<td></td>
<td>[Lehnert]</td>
</tr>
<tr>
<td></td>
<td>[Gil, Malik, Peckham]</td>
</tr>
<tr>
<td></td>
<td>[Khalsa, Ahern, Zaslavsky]</td>
</tr>
<tr>
<td></td>
<td>[Niemann, Aron]</td>
</tr>
</tbody>
</table>


Federal Agency Public Access to Research Results (PARR)

[Friedlander/Zanzerkia (TBD), Bristol, de la Beaujardiere]
<table>
<thead>
<tr>
<th>Time</th>
<th>Breakout</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 AM</td>
<td>Break</td>
<td>Fitzgerald CDE</td>
</tr>
<tr>
<td>10:30 AM</td>
<td>Keynote Speakers: David Maidment (U. Texas) and Roger Wakimoto (NSF)</td>
<td>Fitzgerald CDE</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>End of Formal Meeting</td>
<td></td>
</tr>
<tr>
<td>1:00 PM</td>
<td>Engagement Team Meeting</td>
<td>Fitzgerald A</td>
</tr>
<tr>
<td></td>
<td>Liaison Team Meeting</td>
<td>Fitzgerald B</td>
</tr>
<tr>
<td></td>
<td>TAC Meeting</td>
<td>Fitzgerald C</td>
</tr>
<tr>
<td></td>
<td>Science Committee</td>
<td>Fitzgerald DE</td>
</tr>
<tr>
<td>3:00 PM</td>
<td>End of Meeting</td>
<td></td>
</tr>
</tbody>
</table>

**EarthCube All Hands Meeting 2015**

Source: [https://docs.google.com/spreadsheets.. #gid=422936923](https://docs.google.com/spreadsheets.. #gid=422936923)

Break Out Sessions: Complete List and Descriptions

**Mini Breakout Sessions**

**Building on the “Building Blocks” - supporting a Complex Research Scenario in an EarthCube multi-project synthesis demonstration**

[http://earthcube.org/forum/mini-brea...ect-synthesis](http://earthcube.org/forum/mini-brea...ect-synthesis)

Session Conveners: Siri Jodha Khalsa, Tim Ahern, Ilya Zaslavsky

Understanding and modeling earth systems is a challenge because multiple, distributed information sources with different conventions are needed for inputs to models. We will demonstrate how this problem can be addressed through integration of capabilities from three EarthCube Building Block projects: BCube, GeoWS and CINERGI. Together we show how it is possible to streamline discovery and programmatic access to multiple sources of relevant data.

BCube, through its crawler and broker components, enables discovery and data retrieval. GeoWS provides easily-consumable interfaces to discover and access large data collections. CINERGI provides a service for enhancing...
metadata descriptions to improve data selection and, thus, more effective discovery. We will demonstrate a workflow that integrates components from these projects. The final result is a dataset in a format that can be used by Earth system modelers for their research.

This demonstration illustrates one of the first building block integration activities, and will be followed by considerations regarding possible further steps forward.

**Data Science Publication for USGS Minerals Big Data**

http://earthcube.org/forum/mini-basics-als-big-data

Session Conveners: Brand Niemann and Joan Aron

The EarthCube Data Science Publications Special Interest Group was formed for last year's annual meeting (http://earthcube.org/group/earthcube...e-publications) and has been active this past year in producing a series of Meetups (http://www.meetup.com/Federal-Big-Data-Working- Group/) featuring Data Science Data Publications for senior federal government leaders and agencies (http://semanticommunity.info/).

This session will be lead again by Dr. Brand Niemann and Dr. Joan Aron. and feature our collaborative work with Dr. Sophia Liu, Mendenhall Postdoc Fellow at the U.S. Geological Survey, on Data Science Publication for USGS Minerals Big Data (http://semanticommunity.info/Data_Sc...GS_Minerals...), which will be the subject of our Federal Big Data Working Group Meetup on June 15th in the Washington, DC, area (http://www.meetup.com/Federal-Big-Da...nts/221810524/).

**Full Breakout Sessions**

**Building a registry of geoscience resources: CINERGI components and how they can be used in your metadata curation system**

http://earthcube.org/forum/full-brea...data-curation

Session Conveners: Ilya Zaslavsky, Steve Richard, Amarnath Gupta, Tanu Malik, Kerstin Lehnert, Sky Bristol

This session will introduce the Community Inventory of EarthCube Resources for Geoscience Interoperability (CINERGI) and demonstrate CINERGI metadata compilation and curation pipeline. Its purpose is to support comprehensive integrated search for geoscience information resources across domains. We invite data facility managers and other scientists interested in metadata to review the compiled inventory, discuss its components, and explore how these components can be embedded in other metadata development and curation workflows. In particular, we will discuss:

- a set of metadata enhancers that are the core of the CINERGI pipeline;
- construction of a metadata curation database;
- a cross-domain ontology to support metadata enhancement and discovery;
- metadata harvesting from data facility catalogs and other sources, which support different interfaces and metadata conventions;
- metadata validation services, and
• provenance services.

The components will be demonstrated with a geoscience data discovery use case. Perspectives several data facilities, including USGS ScienceBase, will be presented. Expected session outcomes will include refined metadata curation requirements in different domains and ways for automatically validating and enhancing metadata records maintained by data facilities.

**Connecting EarthCube Projects to the Data Facilities**

[http://earthcube.org/forum/full-brea...ata-facilities](http://earthcube.org/forum/full-brea...ata-facilities)

Session Conveners: Yolanda Gil (EC Technology and Architecture Committee co-chair), Chris MacDermaid (EC Council of Data Facilities Executive Committee)

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. This session will be an opportunity for ongoing EarthCube projects and other members of the EarthCube Technology and Architecture Committee (TAC) to engage with the CDF. There will be an introduction by the chairs of the CDF and TAC which will be followed by a panel discussion.

• Topics for the panel discussion:
  • Role of the CDF in EarthCube
  • New capabilities developed by EarthCube projects that are directly relevant to Data Facilities
  • What Data Facility resources are available to EarthCube Projects
  • What challenges are Data Facilities facing that could be addressed through EarthCube collaborations
  • Possible collaboration opportunities

**EarthCube Integration: Bringing Projects into a Test Environment**

[http://earthcube.org/forum/full-brea...st-environment](http://earthcube.org/forum/full-brea...st-environment)

Session Conveners: Emily Law (Conceptual Design Project), Marlon Pierce (Indiana University, XSEDE Science Gateway)

Session Description: TBD (see link above for more information and updates)

**EarthCube Technology and Architecture: Strategic Planning and Ongoing**

[http://earthcube.org/forum/full-brea...ing-activities](http://earthcube.org/forum/full-brea...ing-activities)

Session Conveners: Yolanda Gil, Jay Pearlman, Phil Yang

Session Description: TBD (see link above for more information and updates)
EarthCube Use Cases: Current Status and Forward Look

http://earthcube.org/forum/full-brea...s-forward-look

Session Conveners: Karen Stocks (TAC Use Case Working Group Co-Lead), Emma Aronson, (Science Committee Co-Chair), Ilya Zaslavsly (TAC member and PI on CINERGI Building Block), Leslie Hsu (C4P RCN, SEN RCN, CINERGI), and Amarnath Gupta (CINERGI, GEAR)

Session Description: TBD (see link above for more information and updates)

Essential Variables and Associated Challenges for EarthCube Science Drivers

http://earthcube.org/forum/full-brea...cience-drivers

Session Conveners: Ken Rubin (EC Science Committee member and WG Chair) and Emma Aronson (Science Committee Co-Chair)

This proposed breakout session is intended to bring together leading scientists, whose research entails the use of large and/or interdisciplinary data sets of interest to a range of Earth Science communities. The goal of the session is to identify the common, unifying essential variables for understanding Geoscience processes, and the potential challenges associated with data collection and usage within the framework of a broader EC context. The challenges might include things like data quality, sampling rate, metadata on measurement quality, location and temporal constraints, and the nature of data streams in different domains.

We anticipate 15 minute presentations by each of these speakers (Dr. Josh Fisher, Dr. Steve Goldstein, Matty Mookerjee, Kerstin Lehnert, and John Heidelberg) plus ample time for discussion. Such a discussion is aimed to entrain more domain scientists into the AHM and to help identify a focus for the emerging EC infrastructure.

The goals of this session are to host several presenters and to have a focused discussion on Science Driver topics and associated technical challenges of value to EarthCube, as identified by the Science Committee Science Strategic Planning effort and associated workshops in Berkeley, CA in March and April 2015.

Federal Agency Federal Agency Public Access to Research Results (PARR)

http://earthcube.org/forum/full-brea...h-results-parr

Session Conveners: Sky Brisol (USGS), Jeff de la Beaujardiere (NOAA), Eva Zanzerkia (NSF)

This PARR session is to present geoscience agency PARR Plans and their implementation. Speakers include Eva from NSF, Sky Bristol from USGS, Jeff DLB from NOAA. NASA, TBD.

Getting the Word Out: EarthCube capabilities, with an example from GeoSoft

http://earthcube.org/forum/full-brea...xample-geosoft

Session Conveners: Sarah Ramdeen, Stephen Slota, Mimi Tzeng, Ji-Hyun Oh, Erin Robinson, Yolanda Gil
Session Description: TBD (see link above for more information and updates)

How will GeoWS help geoscientists?

http://earthcube.org/forum/full-brea...building-block

Session Conveners: Suzanne Carbotte

Other Presenters: Chuck Meertens, Ilya Zaslavsky, Mohan Ramamurthy, Chad Trabant, Mike Gurnis, Mike Stults, Bob Arko

GeoWS is directly targeted at easing discovery, access, and usability of distributed data within the geoscience community. GeoWS partners were identified as part of a science use case articulated at the EarthScope-EarthCube End Users Workshop. GeoWS has developed standards of practice, documentation, and an approach to discovering and using data from multiple domains that can be, in theory, extended to any geo-discipline. This session will provide a brief overview of GeoWS paying particular attention to how it is driven by specific needs of the geoscientist. It will also include demonstrations by each of the 6 funded partners, and some examples of data made available from PIs and data centers that have not received EarthCube funding directly. Demonstrations will be given showing how popular client applications (GeoMapApp, IDV, and the B-Cube broker) can discover and access data from multiple GeoWS funded and unfunded

Integrating Real-time Geosciences Data into the EarthCube Framework

http://earthcube.org/forum/full-brea...cube-framework

Session Conveners: Mike Daniels, lead, V. Chandrasekar, Ken Keiser, David Arctur, Ethan Davis, Jiri Kadlec

The importance of real-time scientific data is increasing, particularly when informed decisions must be made rapidly. Such real-time data have the potential to drastically improve experiments by permitting optimal allocation of experimental resources, while simultaneously serving as a new tool for non-scientific decision makers during extreme events. The CHORDS team is addressing an urgent need for tools to bring real-time sensor web technology within reach of geoscientists lacking cyberinfrastructure skills and resources. A separate, also-important application focus taken up by the DisConBB team is high-resolution streamflow and flood forecasting at the national scale in near-real time. This research has only become practical in the last year, through the integration of advancements in models for weather forecasting, land-surface hydrology, and streamflow routing. EarthCube has enabled these two project teams to explore how their separate sciences and technologies could leverage and advance each other's work.

Due to the advancement of sensor networks and real-time access to the resulting data, many new transient phenomena in space-time have the potential to be observed which may otherwise go unnoticed. While EarthCube proposes to provide an unprecedented framework for disseminating and analyzing historical data sources, the use of real-time data raises an additional set of complex challenges: 1) “off-the-shelf” infrastructure does not exist for distributing real-time measurements made by smaller research groups, 2) sensors may fail latency issues may arise due to hardware and network constraints, 4) data and information may not be presented in a fashion that is easily analyzed and interpreted by decision-makers during a hazardous or time-critical event, 5) unexpected or new phenomenon during a field experiment
may trigger the need to rapidly shift modeling and sampling strategies and 6) standards and protocols for real-time data streams being fed to advanced visualization tools and model assimilation are limited in many cases.

The CHORDS and DisConBB teams are addressing these issues with easy-to-use tools that set up and run sensor webs for research with little to no technology guidance required, which then feed real-time data into downstream tools. For example, CHORDS real-time streams would feed into DisConBB mappings and tools for transforming WaterML into netCDF-CF encoded timeseries that are more easily aggregated and visualized. Applications of this capability could potentially provide important sources of observation data for comparing, validating and calibrating national streamflow and flood forecasting models that are being developed for DisConBB.

**Interoperability and Reuse of Geoscience Models**

[http://earthcube.org/forum/full-brea...science-models](http://earthcube.org/forum/full-brea...science-models)

Session Conveners: Yolanda Gil (GeoSoft), Tanu Malik (GeoDataspace), Scott Peckham (Earth Systems Bridge, GeoSoft, GeoDataspace)

This session will consider a motivating use case of geoscience model reuse and interoperability. It will demonstrate how the use case is being addressed at various technology levels in three currently funded building blocks: the Earth Systems Bridge project, the GeoSoft project, and the GeoDataspace project. The session will present and discuss work on a variety of related issues such as best practices for model sharing, publishing, and reuse, standardized interfaces to enable plug-and-play model coupling (e.g. the Basic Model Interface, BMI), standardized metadata to fully describe models (including their input and output variables and assumptions). The session will present and discuss emerging tools for packaging model and data, and acquiring, browsing, comparing and exporting the accompanying metadata. It will also explore possible synergistic points between these three building blocks that can help address larger and more realistic use cases. The objective will be to solicit feedback from the geoscience community, understand issues of geoscience model interoperability, curation, and reuse as faced by geoscientists on a daily basis.

**The iSamples RCN Survey & Working Group activities: How can we maximize impact for EarthCube?**

[http://earthcube.org/forum/full-brea...pact-earthcube](http://earthcube.org/forum/full-brea...pact-earthcube)

Session Convener: Kerstin Lehnert

This breakout session will provide an opportunity for the EarthCube community to learn about the first results of the iSamples RCN and to get engaged in the ongoing working groups. We will structure the breakout specifically to facilitate discussions between the scientists who participate in the RCN and the Building Blocks to define a next set of goals and actions to advance development and testing of data infrastructure and software tools for physical samples. The session will include:

- presentation of the results of the iSamples Stakeholder Alignment survey about sharing samples and sample data;
- overview of iSamples working groups and presentations of scientists engaged in these working groups;
- open forum for EarthCube projects to explore synergies and initiate collaborations.
Making Geoscience Resources More Discoverable: Connecting People, Data and Products Through the Semantic Web

http://earthcube.org/forum/full-brea...h-semantic-web

Session Conveners: Matt Mayernik and Tom Narock

The semantic web provides the opportunity to make the geosciences more discoverable for research, education, training and outreach. Several EarthCube projects are using semantic technologies to connect people, products and data. An important basis for this approach is incorporating, sharing, and enhancing ontologies for the semantic web. This session will discuss strategies for building effective ontologies that can be shared across the geosciences. We will also discuss the ways in which geoscientists discover and organize their resources, considering use of terms, use of information platforms (i.e. university web page, ResearchGate, LinkedIn and others) and use of identifiers (i.e., ORCID and ResearchID). We will also consider connectivity and interoperability beyond the geosciences, such as with the life sciences, chemistry and physics.

Making it Easier to Find, Use and Share the Source Data and Outputs of Research – Engaging Geoscientists with Illustrative Use Cases

http://earthcube.org/forum/full-brea...puts-research

Session Conveners: SiriJodha Khalsa, Jay Pearlman, Ruth Duerr

This session will include presentations and demonstrations by geoscientists that have applied BCube Brokering Building Block technologies. These demonstrations show how brokering and crawling middleware enables discovery and access of data sources in hydrology, weather and climate, biogeochemistry of estuaries, permafrost dynamics and physical oceanography. The demonstration include technologies that 1) find sources of geoscience data on the web through intelligent crawling; 2) support self-publication of data in a manner that satisfies recently-enacted requirements by publishers; and 3) documents data sources used in research in a manner that facilitates reproduction and verification of those sources. Interfaces with some of the existing building blocks will be discussed and illustrated.

New Synergies and Lessons Learned in the Paleogeosciences through the C4P RCN

http://earthcube.org/forum/full-brea...hrough-c4p-rcn

Session Conveners: Kerstin Lehnert, Anders Noren, Mark D. Uhen, Jack Williams, Doug

Efforts of the C4P RCN have led to new synergies between partners across the geoscience and bioscience disciplines. A desire to build greater infrastructure and integration between these disciplines has led to a need for the adoption and development integrate the needs of these diverse participants.

This session includes presentations that will highlight the development of new synergies and integrative projects, the integration of early career scientists in large-scale projects, and ongoing efforts to develop governance models to serve the diverse needs of the community. The session provides an in-depth perspective from multi-disciplinary, collaborative projects and the lessons learned while working with diverse datasets from across the geoscience and bioscience disciplines, with the goal to meet ambitious scientific research goals.
Speakers will include representatives of both established and nascent geo and bio cyberinfrastructure projects.

Registered Attendees

Source: [https://docs.google.com/spreadsheets...t#gid=52088114](https://docs.google.com/spreadsheets...t#gid=52088114)

<table>
<thead>
<tr>
<th>First Name</th>
<th>Last Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>Allison</td>
<td>Arizona Geological Survey</td>
</tr>
<tr>
<td>Arlene</td>
<td>Anderson</td>
<td>DOE Geothermal Technologies Office</td>
</tr>
<tr>
<td>David</td>
<td>Arctur</td>
<td>University of Texas, Austin</td>
</tr>
<tr>
<td>Robert</td>
<td>Arko</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Emma</td>
<td>Aronson</td>
<td>University of California Riverside</td>
</tr>
<tr>
<td>Jennifer</td>
<td>Arrigo</td>
<td>NOAA Climate Program Office</td>
</tr>
<tr>
<td>Gary</td>
<td>Berg-Cross</td>
<td>SOCoP</td>
</tr>
<tr>
<td>Asti</td>
<td>Bhatt</td>
<td>SRI International</td>
</tr>
<tr>
<td>Rachael</td>
<td>Black</td>
<td>EarthCube Office</td>
</tr>
<tr>
<td>Jeff</td>
<td>Bowman</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>James</td>
<td>Bowring</td>
<td>College of Charleston</td>
</tr>
<tr>
<td>Sky</td>
<td>Bristol</td>
<td>USGS</td>
</tr>
<tr>
<td>Scott</td>
<td>Caltagirone</td>
<td>Element 84, Inc.</td>
</tr>
<tr>
<td>Suzanne</td>
<td>Carbotte</td>
<td>Columbia University</td>
</tr>
<tr>
<td>Megan</td>
<td>Carter</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Marjorie</td>
<td>Chan</td>
<td>University of Utah</td>
</tr>
<tr>
<td>Cynthia</td>
<td>Chandler</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Chandrasekar V</td>
<td>Chandra</td>
<td>Colorado State University</td>
</tr>
<tr>
<td>Michelle</td>
<td>Cheatham</td>
<td>Wright State University</td>
</tr>
<tr>
<td>Cathy</td>
<td>Constable</td>
<td>Scripps Institution of Oceanography</td>
</tr>
<tr>
<td>Alva</td>
<td>Couch</td>
<td>Tufts University/CUAHSI</td>
</tr>
<tr>
<td>Name</td>
<td>Last Name</td>
<td>Institution</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Gary</td>
<td>Crane</td>
<td>Southeastern Universities Research Association</td>
</tr>
<tr>
<td>Chris</td>
<td>Crosby</td>
<td>UNAVCO</td>
</tr>
<tr>
<td>Joel</td>
<td>Cutcher-Gershenfeld</td>
<td>University of Illinois at Urbana–Champaign</td>
</tr>
<tr>
<td>Mike</td>
<td>Daniels</td>
<td>NCAR/EOL</td>
</tr>
<tr>
<td>Ethan</td>
<td>Davis</td>
<td>UCAR Unidata</td>
</tr>
<tr>
<td>Rowena</td>
<td>Davis</td>
<td>Arizona Geological Survey</td>
</tr>
<tr>
<td>Ibrahim</td>
<td>Demir</td>
<td>University of Iowa</td>
</tr>
<tr>
<td>Meixia</td>
<td>Deng</td>
<td>George Mason University</td>
</tr>
<tr>
<td>Liping</td>
<td>Di</td>
<td>George Mason University</td>
</tr>
<tr>
<td>Steve</td>
<td>Diggs</td>
<td>Scripps Institution of Oceanography</td>
</tr>
<tr>
<td>Scott</td>
<td>Doney</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Ruth</td>
<td>Duerr</td>
<td>NSIDC</td>
</tr>
<tr>
<td>Christopher</td>
<td>Duffy</td>
<td>Penn State University</td>
</tr>
<tr>
<td>Rocky</td>
<td>Dunlap</td>
<td>CU Boulder</td>
</tr>
<tr>
<td>Mike</td>
<td>Dye</td>
<td>NCAR</td>
</tr>
<tr>
<td>Fernando</td>
<td>Echavarria</td>
<td>US Department of State</td>
</tr>
<tr>
<td>Paul</td>
<td>Edwards</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Mostafa</td>
<td>Elag</td>
<td>University of Illinois at Urbana–Champaign</td>
</tr>
<tr>
<td>Bakinam</td>
<td>Essawy</td>
<td>University of Virginia</td>
</tr>
<tr>
<td>Douglas</td>
<td>Fils</td>
<td>Consortium for Ocean Leadership</td>
</tr>
<tr>
<td>Sean</td>
<td>Fox</td>
<td>Carleton College</td>
</tr>
<tr>
<td>Janet</td>
<td>Fredericks</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Wally</td>
<td>Fulweiler</td>
<td>Boston University</td>
</tr>
<tr>
<td>James</td>
<td>Gallagher</td>
<td>OpenDAP</td>
</tr>
<tr>
<td>Jennifer</td>
<td>Georgen</td>
<td>Old Dominion University</td>
</tr>
<tr>
<td>Name</td>
<td>Given Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Yolanda</td>
<td>Gil</td>
<td>Information Sciences Institute</td>
</tr>
<tr>
<td>Holly</td>
<td>Given</td>
<td>UCSD</td>
</tr>
<tr>
<td>Basil</td>
<td>Gomez</td>
<td>University of Hawai'i, Mānoa</td>
</tr>
<tr>
<td>Jeffrey</td>
<td>Grethe</td>
<td>University of California San Diego</td>
</tr>
<tr>
<td>Linda</td>
<td>Gundersen</td>
<td>USGS</td>
</tr>
<tr>
<td>Amarnath</td>
<td>Gupta</td>
<td>University of California San Diego</td>
</tr>
<tr>
<td>Michael</td>
<td>Gurnis</td>
<td>Caltech</td>
</tr>
<tr>
<td>Alexandra</td>
<td>Hangsterfer</td>
<td>Scripps Institution of Oceanography</td>
</tr>
<tr>
<td>Chris</td>
<td>Hanson</td>
<td>Arizona Geological Survey</td>
</tr>
<tr>
<td>John</td>
<td>Heidelberg</td>
<td>USC/Center for Dark Energy Biosphere Investigation</td>
</tr>
<tr>
<td>Emily</td>
<td>Hernandez Goldstein</td>
<td>University of Texas at Austin</td>
</tr>
<tr>
<td>Denise</td>
<td>Hills</td>
<td>Geological Survey of Alabama</td>
</tr>
<tr>
<td>Leslie</td>
<td>Hsu</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Yingjie</td>
<td>Hu</td>
<td>UC Santa Barbara</td>
</tr>
<tr>
<td>Haosheng</td>
<td>Huang</td>
<td>Louisiana State University</td>
</tr>
<tr>
<td>Elchin</td>
<td>Jafarov</td>
<td>University of Colorado</td>
</tr>
<tr>
<td>Peng</td>
<td>Ji</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Matthew</td>
<td>Jones</td>
<td>NCEAS / UC Santa Barbara</td>
</tr>
<tr>
<td>Jiri</td>
<td>Kadlec</td>
<td>Brigham Young University</td>
</tr>
<tr>
<td>Farzad</td>
<td>Kamalabadi</td>
<td>ECE Illinois</td>
</tr>
<tr>
<td>Nazifa</td>
<td>Karima</td>
<td>Wright State University</td>
</tr>
<tr>
<td>Anna</td>
<td>Katz</td>
<td>EarthCube Office</td>
</tr>
<tr>
<td>Ken</td>
<td>Keiser</td>
<td>University of Alabama in Huntsville</td>
</tr>
<tr>
<td>Anna</td>
<td>Kelbert</td>
<td>Oregon State University</td>
</tr>
<tr>
<td>Name</td>
<td>First Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Lisa</td>
<td>Kempler</td>
<td>MathWorks</td>
</tr>
<tr>
<td>SiriJodha</td>
<td>Khalsa</td>
<td>Univ. of Colorado</td>
</tr>
<tr>
<td>Huda</td>
<td>Khan</td>
<td>Cornell University</td>
</tr>
<tr>
<td>Danie</td>
<td>Kindade</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Rebecca</td>
<td>Koskela</td>
<td>DataONE</td>
</tr>
<tr>
<td>Kate</td>
<td>Kretschmann</td>
<td>EarthCube Office</td>
</tr>
<tr>
<td>Adila</td>
<td>Krisnadhi</td>
<td>Wright State University</td>
</tr>
<tr>
<td>Praveen</td>
<td>Kumar</td>
<td>University of Illinois</td>
</tr>
<tr>
<td>Kwo-Sen</td>
<td>Kuo</td>
<td>University of Maryland, College Park</td>
</tr>
<tr>
<td>Emily</td>
<td>Law</td>
<td>JPL</td>
</tr>
<tr>
<td>Erin</td>
<td>Leckey</td>
<td>STEPPE Program</td>
</tr>
<tr>
<td>Sonya</td>
<td>Legg</td>
<td>Princeton University</td>
</tr>
<tr>
<td>Kerstin</td>
<td>Lehnert</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Chris</td>
<td>Lenhardt</td>
<td>RENCI</td>
</tr>
<tr>
<td>Michael</td>
<td>Lesser</td>
<td>University of New Hampshire</td>
</tr>
<tr>
<td>Zhiyu</td>
<td>Li</td>
<td>Brigham Young University</td>
</tr>
<tr>
<td>Rui</td>
<td>Liu</td>
<td>NCSA</td>
</tr>
<tr>
<td>Miron</td>
<td>Livny</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Shane</td>
<td>Loeffler</td>
<td>University of Minnesota</td>
</tr>
<tr>
<td>Chris</td>
<td>MacDermaid</td>
<td>NOAA</td>
</tr>
<tr>
<td>David</td>
<td>Maidment</td>
<td>University of Texas at Austin</td>
</tr>
<tr>
<td>Tanu</td>
<td>Malik</td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Brian</td>
<td>Mapes</td>
<td>University of Miami</td>
</tr>
<tr>
<td>Luigi</td>
<td>Marini</td>
<td>NCSA</td>
</tr>
<tr>
<td>Chris</td>
<td>Mattman</td>
<td>University of Southern California</td>
</tr>
<tr>
<td>Name</td>
<td>Last Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Matthew</td>
<td>Mayernik</td>
<td>University Corporation for Atmospheric Research (UCAR)</td>
</tr>
<tr>
<td>Chuck</td>
<td>Meertens</td>
<td>UNAVCO</td>
</tr>
<tr>
<td>Don</td>
<td>Middleton</td>
<td>UCAR</td>
</tr>
<tr>
<td>Heath</td>
<td>Mills</td>
<td>University of Houston at Clear Lake</td>
</tr>
<tr>
<td>Ann</td>
<td>Molineux</td>
<td>University of Texas at Austin</td>
</tr>
<tr>
<td>Matty</td>
<td>Mookerjee</td>
<td>Sonoma State University</td>
</tr>
<tr>
<td>John</td>
<td>Morton</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Amy</td>
<td>Myrbo</td>
<td>LacCore/CSDCO, University of Minnesota</td>
</tr>
<tr>
<td>Tom</td>
<td>Narock</td>
<td>Marymount University</td>
</tr>
<tr>
<td>Brand</td>
<td>Niemann</td>
<td>Semantic Community</td>
</tr>
<tr>
<td>Gelu</td>
<td>Nita</td>
<td>New Jersey Institute of Technology</td>
</tr>
<tr>
<td>Anders</td>
<td>Noren</td>
<td>CSDCO/University of Minnesota</td>
</tr>
<tr>
<td>Ji-Hyun</td>
<td>Oh</td>
<td>Jet Propulsion Laboratory</td>
</tr>
<tr>
<td>Jay</td>
<td>Pearlman</td>
<td>J&amp;FE</td>
</tr>
<tr>
<td>Francoise</td>
<td>Pearlman</td>
<td>J&amp;FE</td>
</tr>
<tr>
<td>Scott</td>
<td>Peckham</td>
<td>University of Colorado at Boulder</td>
</tr>
<tr>
<td>Shanan</td>
<td>Peters</td>
<td>University of Wisconsin-Madison</td>
</tr>
<tr>
<td>Bernhard</td>
<td>Peucker-Ehrenbrink</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Suzanne</td>
<td>Pierce</td>
<td>University of Texas at Austin</td>
</tr>
<tr>
<td>Marlon</td>
<td>Pierce</td>
<td>Indiana University</td>
</tr>
<tr>
<td>Tracey</td>
<td>Pilone</td>
<td>Element 84, Inc.</td>
</tr>
<tr>
<td>Daniel</td>
<td>Pilone</td>
<td>Element 84, Inc.</td>
</tr>
<tr>
<td>Sarah</td>
<td>Purkey</td>
<td>Lamont-Doherty Earth Observatory</td>
</tr>
<tr>
<td>Mohan</td>
<td>Ramamurthy</td>
<td>Uni</td>
</tr>
<tr>
<td>Sarah</td>
<td>Ramdeen</td>
<td>University of North Carolina</td>
</tr>
<tr>
<td>Name</td>
<td>Last Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Lisa</td>
<td>Raymond</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Ying Fan</td>
<td>Reinfelder</td>
<td>Rutgers University</td>
</tr>
<tr>
<td>Erin</td>
<td>Robinson</td>
<td>Foundation for Earth Science</td>
</tr>
<tr>
<td>Ken</td>
<td>Rubin</td>
<td>SOEST/University of Hawaii</td>
</tr>
<tr>
<td>Jamie</td>
<td>Ryan</td>
<td>EarthCube Office</td>
</tr>
<tr>
<td>Jocelyn</td>
<td>Sessa</td>
<td>American Museum of Natural History</td>
</tr>
<tr>
<td>Martin</td>
<td>Seul</td>
<td>CUAHSI</td>
</tr>
<tr>
<td>Adam</td>
<td>Shepherd</td>
<td>Woods Hole Oceanographic Institution</td>
</tr>
<tr>
<td>Randy</td>
<td>Showstack</td>
<td>(Reporter) Eos/AGU</td>
</tr>
<tr>
<td>Sudhir</td>
<td>Shrestha</td>
<td>NOAA Center for Weather and Climate Prediction</td>
</tr>
<tr>
<td>Stephen</td>
<td>Slota</td>
<td>University of California Irvine</td>
</tr>
<tr>
<td>Kevin</td>
<td>Smith</td>
<td>Tufts University</td>
</tr>
<tr>
<td>Plato</td>
<td>Smith</td>
<td>University of New Mexico</td>
</tr>
<tr>
<td>Jia</td>
<td>Song</td>
<td>George Mason University</td>
</tr>
<tr>
<td>J Alex</td>
<td>Speer</td>
<td>Mineralogical Society of America</td>
</tr>
<tr>
<td>Simon</td>
<td>Spero</td>
<td>Criticollab, LLC</td>
</tr>
<tr>
<td>D. Sarah</td>
<td>Stamps</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>Karen</td>
<td>Stocks</td>
<td>Scripps Institution of Oceanography</td>
</tr>
<tr>
<td>Mike</td>
<td>Stults</td>
<td>IRIS</td>
</tr>
<tr>
<td>Chad</td>
<td>Trabant</td>
<td>IRIS Data Management Center</td>
</tr>
<tr>
<td>Mimi</td>
<td>Tzeng</td>
<td>Dauphin Island Sea Lab (AL)</td>
</tr>
<tr>
<td>Mark</td>
<td>Uhen</td>
<td>George Mason University</td>
</tr>
<tr>
<td>Anthony</td>
<td>van Eyken</td>
<td></td>
</tr>
<tr>
<td>Amanda</td>
<td>Vizedom</td>
<td>Criticollab, LLC</td>
</tr>
<tr>
<td>Christian</td>
<td>Ward-Garrison</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Last Name</td>
<td>Institution</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Peter</td>
<td>Wiebe</td>
<td>Woods Hole Oceanographic Institute</td>
</tr>
<tr>
<td>John</td>
<td>Williams</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>Xiaoqing</td>
<td>Wu</td>
<td>Iowa State University</td>
</tr>
<tr>
<td>Chaowei</td>
<td>Yang</td>
<td>George Mason University</td>
</tr>
<tr>
<td>Xuan</td>
<td>Yu</td>
<td>University of Delaware</td>
</tr>
<tr>
<td>Eugene</td>
<td>Yu</td>
<td>George Mason University - CSISS</td>
</tr>
<tr>
<td>Eva</td>
<td>Zanzerkia</td>
<td>NSF</td>
</tr>
<tr>
<td>Ilya</td>
<td>Zaslavsky</td>
<td>SDSC/UCSD</td>
</tr>
<tr>
<td>Rick</td>
<td>Ziegler</td>
<td>Environmental Protection Agency</td>
</tr>
</tbody>
</table>

**Story**

*Earth Cube Sprint to "Stretch Goals: Open Research Data Publication and Integration***
Use Case Schematic for EarthCube Governance, Architecture, and Implementation (June 25, 2014)

See Slides (Below), One Pager (Word) for EarthCube All-Hands Meeting, and Publication (PDF: 15 MB)

EarthCube Data Science Publications

The EarthCube vision is: “Over the next decade, the geosciences community commits to developing a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core, including the influences of humans and ecosystems.” GEO Vision Report of NSF Geoscience Directorate Advisory Committee, 2009.

The EarthCube Web Site features two scientific publications that I felt are very important: Community-Developed Geoscience Cyberinfrastructure and Ten Simple Rules for the Care and Feeding of Scientific Data.

The first says: EarthCube can play a major role in training the next generation of cyber-savvy geoscientists by providing intuitive, modular learning objects and self-directed lessons that can be used by teachers from K–12 through to the graduate level. In particular, an EarthCube three-dimensional virtual globe will not only be a data discovery portal for researchers but will also serve as an entry point for students to explore geoscience data. Success will be achieved when EarthCube helps everyone—including individuals outside of the geosciences—better understand how to use and interpret science data.

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch™
The second says: This article offers a short guide to the steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized. In just the past few years, hundreds of scholarly papers and reports have been written on questions of data sharing, data provenance, research reproducibility, licensing, attribution, privacy, and more—but our goal here is not to review that literature. Instead, we present a short guide intended for researchers who want to know why it is important to “care for and feed” data, with some practical advice on how to do that. This article was written collaboratively, online, in the open, using Authorea. Every iteration of the writing procedure is logged and available in the online version of this article at https://www.authorea.com/3410

The first article says what needs to be done (help geoscientists and the public better understand how to use and interpret science data) and the second article says how it should be done (steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized). So what is needed is an example of doing it. I did that when EarthCube was first launched in November 2011. See Story and ESIP Data.gov Application below.

So I just joined EarthCube and submitted a late session proposal which will be a Birds of a Feather session as follows:

- **Title:** EarthCube Data Science Publications
- **Convener:** Dr. Brand Niemann, Director and Senior Data Scientist/Data Journalist, Semantic Community
- **Co-Conveners:** Dr. Joan Aron, Global Environmental/Climate Change Scientist, Aron Consulting
- **Description:**
  - Scientific Data Publications in Data Browsers will help us "know what we know better" by mining and publishing scientific results in a fundamentally different way.
  - Scientific Data Publications in Data Browsers will give us greater return on all the investment made in scientific research and publication by enabling reuse for public education (STEM) and data science reproducibility/validation.
  - Data Publications in Data Browsers have been created for at least ten senior government officials as a way of educating and motivating them to task their staff and contractors to start doing the same.
  - A brief description of the Semantic Community services relevant to the NSF BIGDATA Project will be provided along with The Federal Big Data Initiative: Where it has been and where it is going, the Federal Big Data Working Group Meetup: Data Science Data Publications in Data Browsers, and three examples of Data Science for Scientific Data Publications in Data Browsers: NSF Grant Proposal Guide, NSF Funding for BIG DATA and Data Science, and Big Data Science for CODATA.
  - [Web Site](https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications)

We have already received some excellent questions and provided some answers:

- Who has the time these days to answer all the emails you receive when you publish a dataset online?
  - I solve that problem by providing good documentation (metadata) like I know a Data Scientist should so I have very few questions.
- How do we go beyond existing tools like Google Earth, ESRI, OPeNDAP, etc.?
  - That was a problem I as a Data Scientist had to solve.
- How do we sustain repositories of data, who provides support for long-term hosting, dataset support and update, etc.?
  - That is what EarthCube, NSF BIG DATA, etc. funding is for.

So I started to mine EarthCube content (see below) to make it a data publication for use in next weeks Birds of a Feather session. Semantic Community uses Data Science and the Cross Industry Standard Process for Data Mining.
(CRISP-DM; Shearer, 2000) to add more structure to essentially unstructured or semi-structured content, and build spreadsheets and interactive dashboards as will be seen below.

I considered several possibilities:

- Recent work on Big Data Science for CODATA in support of their Big Data for International Scientific Programmes;
- Work in progress on the Data Science for EPA EnviroAtlas;
- Paleogeoscience building on the work of Chaowei Phil Yang and Chen Xu, NSF Spatiotemporal Innovation Center George Mason University, and Carol Meyer and Erin Robinson, ESIP (See below);
- The example in the recent White House OSTP Big Data Report Fact Sheet and Big Data Report Blog Post showing a powerful tool for emergency services and communities that's driven by big data technology; and
- NSF Funded DataOne and DataNet mentioned in a recent Keynote: "NITRD Perspectives and Initiatives" by Dr. George Strawn (Director, National Coordination Office, The Networking and Information Technology Research and Development (NITRD) Program, Executive Office of the President).

AGU Digital Library: The AGU journal backfile (the Digital Library) spans more than a century of knowledge with 674,174 pages of content and is a tremendous resource to science.

MORE TO FOLLOW

NSF Strategic Plan Excerpts: Many workshops are convened each year to allow the research communities supported by GEO to articulate priorities. Of particular note in 2012 was an opportunity for considerable community planning around GEO’s EarthCube activity. This ambitious program seeks to develop the cyberinfrastructure to knit together geoscience data from a variety of disciplines, and is GEO’s contribution to NSF’s CIF21 investment, and is central to GEO’s strategy to capitalize on Big Data.

Summary History and Workplan

Summary History

Fourth Meetup: Tuesday, March 4, 6:30 p.m.: Joint NSF-NIH Biomedical Big Data Research: NSF Funding for BIG DATA and Data Science: NSF BIGDATA Grant Proposal (June 9, 2014)


All participants were asked to answer the following question and submit a two-page description of what they would do to support EarthCube which I did.

Question: When EarthCube exists and is widely useful in 2021, what does a day in the life of a scientist in your field look like? Think about your: Research, Teaching, Outreach, Workforce Development, and Interaction with the greater scientific community.

Answer: I will still have some form of the tools I am using now, they will just be more integrated with one another and even more connected to many sources of information and data so I can create data stories (like I do now for AOL Government) with greater ease and frequency because the time for collection and communication is lessened and the
time for analysis is maximized -- actually the analysis tool facilitates the collection and communication parts.

I also provided an example of the kind of agile analysis of Earth Science data that I had done earlier in the year for the Earth Science Federation Annual Conference.

EarthCube Summit 2014 (March 25, 2014, Crystal City Arlington) Virginia): Part of Open Geospatial Consortium Technical Committee Meeting (virtual meeting with 12 presenters)

Earth Cube All-Hands Meeting (June 24-26, Washington, DC): EarthCube Data Science Publications (Special Interest Group) and EarthCube Data Science Publications (Example)

Work Plan

Our 3 level pyramid architecture now has an "NSF owner": Key Note Speaker: Roger Wakimoto:

Dr. Roger M. Wakimoto is assistant director for the National Science Foundation's (NSF) Directorate for Geosciences (GEO). The GEO Directorate has an annual budget of approximately $1 billion in support of core research in the atmospheric, polar, earth and ocean sciences.

Prior to coming to NSF, Wakimoto served as director of the National Center for Atmospheric Research (NCAR), which is sponsored by NSF. Prior to becoming NCAR director, he served as associate director for NCAR's Earth Observing Laboratory. Wakimoto is a geophysicist with expertise in tornadoes, thunderstorms and other types of severe weather.

The National Science Foundation (NSF) is an independent federal agency created by Congress in 1950 "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense..." With an annual budget of $7.2 billion (FY 2014), we are the funding source for approximately 24 percent of all federally supported basic research conducted by America's colleges and universities. In many fields such as mathematics, computer science and the social sciences, NSF is the major source of federal backing.

NSF’s new Strategic Plan, Investing in Science, Engineering, and Education for the Nation’s Future: NSF Strategic Plan for FY 2014-2018, was released on March 10, 2014. It lays out two strategic goals that embody the dual nature of NSF’s mission to advance the progress of science while benefitting the Nation: Transform the Frontiers of Science and Engineering and Stimulate Innovation and Address Societal Needs through Research and Education. A third goal, Excel as a Federal Science Agency, directs NSF to hold itself accountable for achieving excellence in carrying out its mission. This goal structure enables NSF to link its investments to longer-term outcomes. To bridge the gap between these strategic goals and measurable outputs, the Strategic Plan establishes a set of strategic objectives for each strategic goal.

Dr. Wakimoto said the NSF Directorate for Geosciences (GEO) Strategic Plan: Geo-Vision 2015-2020 (in process for late summer release) will be a "living document" (might be too big for the Web) with the grand challenge being to integrate across the 4 divisions and their individual projects and results and/or how we would link these investments and what would be the "model" for doing that.
So EarthCube could/should produce Open Research Data products for individual projects and integrate across them for EarthCube success and prominence, respectively. EarthCube needs an organizational structure for the workflow to produce these results.

- **Data Management.** EarthCube will provide a shared archive in which data, tools, and services are documented and curated, enabling reuse of data sets for new analyses.
- **Resource Discovery.** Plug-in components will enable resource discovery and direct data access using scientific software in common use, such as Excel, MATLAB, Python, R, ArcGIS, or ModFlow.
- **Data Access, Integration, and Processing Tools.** User interfaces tailored for specific communities will simplify data access, visualization, and analysis using software that interoperates with EarthCube data or service providers.
- **Data Portal.** One or more portals will function as user entry points to support data exploration and access tailored for specific communities. **One that has been frequently mentioned is a three-dimensional virtual globe for data discovery and exploration, supporting the ability to spatially integrate and display geoscientific data at varying resolutions.** See Big Data Science for CODATA.

We have given an example of workflow and results: See EarthCube Special Interest Group on EarthCube Data Science Publications

This should be added to our August 4th Keynote and Panel Session as a focus to get funding going forward.

**Data Science Publication Commons** (July 17, 2014)

The product is a Data Publication Commons for EarthCube, ESIP and RDA. See Data Science for RDA.

The Data Publication Commons will maintain data publications for the websites of EarthCube, ESIP and RDA.


Answers: We will look around for the best content examples like I did for CODATA and EarthCube. ESIP did not really have any we could find yet.

The Data Publication Commons will provide tools to search through the commons (MindTouch, Excel, and Spotfire all running on servers in the Amazon Cloud)

Questions and Answers:

Can you maintain this as a volunteer? Initially, then we will need/get support from NSF, NASA, etc.

How many hours per work are involved? 1-2 days after the Fourth Plenary to build the Data Science Data Publication for it and then we will see what is next.

What about servers? See above

End of Questions Answers
The selling point to RDA is a demonstration of a technology and a resource for RDA members to gather and link information from EarthCube, ESIP and RDA.

**Slides**

**Slide 1 EarthCube All Hands Meeting New Interest Group: EarthCube Data Science Publications**

[http://semanticommunity.info](http://semanticommunity.info)

---

**Slide 2 Summary**

**Summary**

- The Semantic Community and the Federal Big Data Working Group Meetup have produced a series of Data Science Data Publications for senior government officials and programs to demonstrate Open Research Data.
- Semantic Community, founder of the FBDWG Meetup, has submitted an NSF BIGDATA Grant Proposal to continue this work.
- Semantic Community is piloting EarthCube Data Science Publications in an EarthCube Commons Interest Group and in a Birds of a Feather Session at the All-Hands Meeting.
Slide 3 Overview

Overview

- The EarthCube Vision
- EarthCube Web Site
- EarthCube Workspace
- EarthCube Work Groups:
  - General
  - Topics
  - Actions
- EarthCube Data Science Publications Interest Group
- Data Mining the EarthCube Web Site
  - Community Developed Geoscience Cyberinfrastructure
  - Ten Simple Rules for the Care and Feeding of Scientific Data
- My June 2014 Data Journalism Story
- My November 2011 Data Journalism Story
- Birds of a Feather Session Proposal
- Possibilities Under Consideration
- Some Next Steps
- Schematic Diagram for Broader Discussion: Dr. Joan Aron

Slide 4 The EarthCube Vision

The EarthCube Vision

- “Over the next decade, the geosciences community commits to developing a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core, including the influences of humans and ecosystems.”

- My Note: Keywords-Decade, Earth Core to Atmosphere-Space, and Data Science for the Geosciences Community

Slide 5 EarthCube Web Site

http://earthcube.org/
EarthCube Web Site

- Mission Statement:
  - EarthCube is a bold new initiative to create a community-driven data and knowledge management system that will allow for unprecedented data sharing across the geosciences.

- Community:
  - You can influence what EarthCube can do and will be. Join over 2,500 participants in forums, work groups, EarthCube events, and at virtual and in-person meetings across the country.

- Workspace:
  - The EarthCube Commons workspace is where EarthCube project teams and stakeholder groups collaborate. You can join working groups and see what the EarthCube community is doing in real time.


---

Slide 6 EarthCube Workspace

http://workspace.earthcube.org/

EarthCube Workspace

- Purpose:
  - Anyone with an interest in EarthCube is free to use this site to coordinate events, draft and share documents, organize meetings, and stay up to date on EarthCube activities. This workspace will be continually developed and optimized to meet the changing needs of EarthCube work groups.

- My Comment:
  - But it is not the “community-driven data and knowledge management system that will allow for unprecedented data sharing across the geosciences“ that Semantic Community proposed as a Data Management Plan in their NSF BIGDATA Grant Proposal.

  [http://workspace.earthcube.org/](http://workspace.earthcube.org/)
EarthCube Work Groups: General

- Groups are the basic structures for the conversations about EarthCube that take place on the EarthCube web platform. EarthCube groups allow virtual communities of practice within EarthCube to form around a common scientific or IT field, end-user community, and other topics of interest (e.g., data curation, interoperability, semantics, etc.). Group members can post discussion topics, comments, documents, events, and more to the group forum.

- My Comment: Place for conversations and postings, but not the place for the actual Data Science for the Geosciences Community work.

EarthCube Work Groups: Types

- Two different types of groups have been created:
  - Funded Groups: These groups represent individual National Science Foundation EarthCube awards, announced in September 2013. The NSF has awarded 15 new EarthCube projects (categorized into Governance, Building Blocks, Research Coordination Networks, and Conceptual Designs) with funds totaling $14.5 million. Scroll down for information on individual awards, projects, and personnel.
  - Interest Groups: The goal of many of these groups is to build virtual communities within EarthCube around each scientific domain, and continue conversations that started at the end-user workshops regarding what each group needs from EarthCube. Some interest groups will consist of members of EarthCube tech communities with an interest in developing key technologies for EarthCube. Visit the workspace to see what each of the special interest groups are doing right now.

- My Note: I want to create an Interest Group that becomes a Funded Group like we are doing with the Federal Big Data Working Group Meetup.
Slide 9 EarthCube Work Groups: Actions

EarthCube Work Groups: Actions

- Join an EarthCube Group:
  - Although all group discussions and other resources are publicly accessible, you must be a member of a group to post content there. You’re welcome to join as many groups as you want—to join a group, simply go to the workspace, find a group that interests you, and click a link to request membership.
- Create an EarthCube Group:
  - In an effort to keep pace with the direction that EarthCube is moving, community members will now be able to create EarthCube Groups.
  - If you would like to create a group, please consider your group's goals, and what you will do as the group leader/curator to help your group achieve these goals and keep your group active. What are your outputs? What are your outcomes? Who is your target audience? What will you do to help your group stay active? Can you enlist other group members to help you?
  - My Note: I created the EarthCube Data Science Publications Group on June 17th.

Slide 10 EarthCube Data Science Publications Interest Group

EarthCube Data Science Publications Interest Group

- Group's goals:
  - Workforce Education and Training, in Data Science Publications in Data Browse.
- What you will do as the group leader/curator to help your group achieve these goals and keep your group active?
  - Use the Meetup model we are already using successfully and obtain NSF Grant Funding and other funding.
- What are your outputs?
  - Data Science Publications in Data Browse.
- What are your outcomes?
  - Public Knowledge Base of Data Science Publications in Data Browse.
- Who is your target audience?
  - Senior Government Officials and Managers, the Public, and Especially K-12.
- What will you do to help your group stay active?
  - Use the Meetup model we are already using and obtain NSF Grant Funding and other funding.
- Can you enlist other group members to help you?
  - We will be doing that at the All Hands Meeting and in other communities of practice.

Slide 11 EarthCube Data Science Publications

http://workspace.earthcube.org/earthcube-data-science-publications
EarthCube Data Science Publications

- Data Science Publications Tasks:
  - Please see our Birds of a Feather Slides and One Pager
- Data Science Publications News and Announcements:
  - Please suggest EarthCube content for data publications and attend our Meetups
- Data Science Publications Discussion Board:
  - Please post your comments and questions.
- Data Science Publications Calendar:
  - Please attend our Federal Big Data Working Group Meetups, June 30, July 7, and August 4-6.
  

---

Slide 12 Data Mining the EarthCube Web Site

[http://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications#Story](http://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications#Story)

---

Data Mining the EarthCube Web Site

- The EarthCube Web Site features two scientific publications that I felt are very important for EarthCube Data Science Publications and my Data Journalism story:
  - Community-Developed Geoscience Cyberinfrastructure, and
  - Ten Simple Rules for the Care and Feeding of Scientific Data

[http://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications#Story](http://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications#Story)

---

Slide 13 Community-Developed Geoscience Cyberinfrastructure

**Big Data Science for CODATA** and **NSF Grant Proposal Guide**

**Community-Developed Geoscience Cyberinfrastructure**
Community-Developed Geoscience Cyberinfrastructure

- EarthCube can play a major role in training the next generation of cyber-savvy geoscientists by providing intuitive, modular learning objects and self-directed lessons that can be used by teachers from K–12 through to the graduate level.
- In particular, an EarthCube three-dimensional virtual globe will not only be a data discovery portal for researchers but will also serve as an entry point for students to explore geoscience data.
- Success will be achieved when EarthCube helps everyone—including individuals outside of the geosciences—better understand how to use and interpret science data.
- My Note: EarthCube's 3-D virtual globe for STEM is like the Semantic Community work for CODATA's Digital Earth Vision - Big Data for International Scientific Programmes and in the NSF BIGDATA Grant Proposal.
  - See: Big Data Science for CODATA and NSF Grant Proposal Guide

See Data Publication of: Community-Developed Geoscience Cyberinfrastructure

---

Slide 14 Ten Simple Rules for the Care and Feeding of Scientific Data

https://www.authorea.com/3410

Ten Simple Rules for the Care and Feeding of Scientific Data

- This article offers a short guide to the steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized.
- In just the past few years, hundreds of scholarly papers and reports have been written on questions of data sharing, data provenance, research reproducibility, licensing, attribution, privacy, and more—but our goal here is not to review that literature.
- Instead, we present a short guide intended for researchers who want to know why it is important to "care for and feed" data, with some practical advice on how to do that.
- This article was written collaboratively, online, in the open, using Authorea. Every iteration of the writing procedure is logged and available in the online version of this article at https://www.authorea.com/3410

See Data Publication of: Ten Simple Rules for the Care and Feeding of Scientific Data
Slide 15 My June 2014 Data Journalism Story

My June 2014 Data Journalism Story

- The first article says what needs to be done:
  - Help geoscientists and the public better understand how to use and interpret science data; and
- The second article says how it should be done:
  - Steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized.
- So what is needed is an example of doing it.
  - I did that when EarthCube was first launched in November 2011!
    - See Story and ESIP Data.gov Application in next slides.

Slide 16 EarthCube Shows How Collaboration Is Changing Geoscience Research

BreakingGov.com

EarthCube Shows How Collaboration Is Changing Geoscience Research

BreakingGov.com

Slide 17 My November 2011 Data Journalism Story

http://semanticommunity.info/@api/deki/files/15653/=BrandNiemann11042011.docx
My November 2011 Data Journalism Story

- All participants were asked to answer the following question and submit a two-page description of what they would do to support EarthCube which I did. My Note: Imagine a document posted in 2011 is still available!
  - EarthCube Question: When EarthCube exists and is widely useful in 2021, what does a day in the life of a scientist in your field look like? Think about your: Research, Teaching, Outreach, Workforce Development, and Interaction with the greater scientific community.
  - My Answer: I will still have some form of the tools I am using now, they will just be more integrated with one another and even more connected to many sources of information and data so I can create data stories (like I do now for AUS Government) with greater ease and frequency because the time for collection and communication is lessened and the time for analysis is maximized—actually the analysis tool facilitates the collection and communication parts.
- I also provided an example of the kind of agile analysis of Earth Science data that I had done earlier in the year for the 2011 Earth Science Federation Annual Conference. See next slides.

Slide 18 Data Publication in a Data Browser

See Data Publication: Federation of Earth Science Information Partners in Data Browser Web Player

Data Publication in a Data Browser

Slide 19 Birds of a Feather Session Proposal

Web Site
Birds of a Feather Session Proposal

- Title: EarthCube Data Science Publications
- Convener: Dr. Brand Niemann, Director and Senior Data Scientist/Data Journalist, Semantic Community
- Co-Convener: Dr. Joan Aron, Global Environmental/Climate Change Scientist, Aron Consulting
- Description:
  - Scientific Data Publications in Data Browsers will help us "know what we know better" by mining and publishing scientific results in a fundamentally different way.
  - Scientific Data Publications in Data Browsers will give us greater return on all the investment made in scientific research and publication by enabling reuse for public education (STEM) and data science reproducibility and validation.
  - Data Publications in Data Browsers have been created for at least ten senior government officials as a way of educating and motivating them to task their staff and contractors to start doing the same.
  - A brief description of the Semantic Community services relevant to the NSF BIGDATA Project will be provided along with The Federal Big Data Initiative: Where it has been and where it is going, the Federal Big Data Working Group Meetup: Data Science Data Publications in Data Browsers, and three examples of Data Science for Scientific Data Publications in Data Browsers: NSF Grant Proposal Guide, NSF Funding for BIG DATA and Data Science, and Big Data Science for CODATA.
- Web Site

---

Slide 20 Some Excellent Questions and Our Answers

Some Excellent Questions and Our Answers

- Who has the time these days to answer all the emails you receive when you publish a dataset online?
  - I solve that problem by providing good documentation (metadata) like I know a Data Scientist should so I have very few questions.
- How do we go beyond existing tools like Google Earth, ESRI, OPeNDAP, etc.?
  - That was a problem I as a Data Scientist had to solve.
- How do we sustain repositories of data, who provides support for long-term hosting, dataset support and update, etc.?
  - That is what EarthCube, NSF BIG DATA, etc. funding is for.

---

Slide 21 Possibilities Under Consideration

Big Data Science for CODATA
Data Science for EPA EnviroAtlas
EarthCube Summit Slides
Big Data Report Fact Sheet
Big Data Report Blog Post

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
Possibilities Under Consideration

- Recent work on Big Data Science for CODATA in support of their Big Data for International Scientific Programmes;
- Work in progress on the Data Science for EPA EnviroAtlas;
- Paleogeoscience building on the work of Chao Wei Phil Yang and Chen Xu, NSF Spatiotemporal Innovation Center George Mason University, and Carol Meyer and Erin Robinson, ESIP (See EarthCube Summit Slides);
- The example in the recent White House DXP Big Data Report Fact Sheet and Big Data Report Blog Post showing a powerful tool for emergency services and communities that’s driven by big data technology; and
- NSF Funded DataOne and DataNet mentioned in a recent Keynote: “NITRD Perspectives and Initiatives” by Dr. George Straw (Director, National Coordination Office, The Networking and Information Technology Research and Development (NITRD) Program, Executive Office of the President).

Slide 22 Some Next Steps

Some Next Steps

- Semantic Community started to mine EarthCube content to make it a data publication for use in next weeks Birds of a Feather session.
- Semantic Community uses Data Science and the Cross Industry Standard Process for Data Mining (CRISP-DM; Shearer, 2000) to add more structure to essentially unstructured or semi-structured content, and build spreadsheets and interactive dashboards.

Slide 23 EarthCube Data Science Publications: Knowledge Base

http://semanticommunity.info/Data_Science/Publications
EarthCube Data Science Publications

Source: Source: http://workspace.earthcube.org/earth...e-publications

3. Summary History and Workplan (June 26, 2014):

Our 3 level pyramid architecture now has an "NSF owner": Key Note Speaker: Roger Wakimoto.

Dr. Wakimoto said the NSF Directorate for Geosciences (GEO) Strategic Plan: Geo-Vision 2015-2020 (in process for late summer release) will be a "living document" (might be too big for the Web) with the grand challenge being to integrate across the 4 divisions and their individual projects and results and/or how we would link these investments and what would be the "model" for doing that.
So EarthCube could/should produce Open Research Data products for individual projects and integrate across them for EarthCube success and prominence, respectively. EarthCube needs an organizational structure for the workflow to produce these results like what was suggested in a recent paper (PDF):

- **Data Management.** EarthCube will provide a shared archive in which data, tools, and services are documented and curated, enabling reuse of data sets for new analyses.
- **Resource Discovery.** Plug-in components will enable resource discovery and direct data access using scientific software in common use, such as Excel, MATLAB, Python, R, ArcGIS, or ModFlow.
- **Data Access, Integration, and Processing Tools.** User interfaces tailored for specific communities will simplify data access, visualization, and analysis using software that interoperates with EarthCube data or service providers.
- **Data Portal.** One or more portals will function as user entry points to support data exploration and access tailored for specific communities. One that has been frequently mentioned is a three-dimensional virtual globe for data discovery and exploration, supporting the ability to spatially integrate and display geoscientific data at varying resolutions. See Big Data Science for CODATA.

We have given an example of workflow and results: See EarthCube Special Interest Group on EarthCube Data Science Publications (this post)

2. **Recommendation for EarthCube Priority for 2014 (June 25, 2014):**

EarthCube should adopt the Data FAIRport principles so it can begin to produce EarthCube Data Science Publications,

Notes from Birds of a Feather Session on Tuesday, June 24th, 4-5 p.m.: Our BoF Session was well-received and the participants told us this was the future of EarthCube (eventually EarthCube would get to this as products) and suggested some data sources to work with to produce data publications along with the list in our Slide 21. Essentially we are doing what the participants said is leading the way (e.g. biomedical research community) and their new initiative (Data FAIRport) which was formed in the past several months with lite-weight governance, architecture, and implementation and now is focused on multi-biomedical research center hackathons to build data publications that answer important research questions. The Data FAIRport will report at the upcoming RDA Conference because it has attracted significant funding in Europe and attention in other science communities. The Federal Big Data Working Group Meetup is building a Data FAIRport,

1. **Welcome to our new Interest Group: EarthCube Data Science Publications (June 18, 2014):**

We hope to have a Birds of the Feather Session on Tuesday at 4 p.m., Wednesday at 2 p.m., and/or Thursday at 1:30 p.m.

Summary:

- The Semantic Community and the Federal Big Data Working Group Meetup have produced a series of Data Science Data Publications for senior government officials and programs to demonstrate Open Research Data.
- Semantic Community, founder of the FBDWG Meetup, has submitted an NSF BIGDATA Grant Proposal to continue this work.
- Semantic Community is piloting EarthCube Data Science Publications in an EarthCube Commons Interest Group and in a Birds of a Feather Session at the All-Hands Meeting.

A set of slides (in line) are available and in PPTX (1.2 MB)
An EarthCube Data Science Publication is available and in PDF (15 MB)

Please attend (in person or virtual) our Federal Big Data Working Group Meetups, June 30, July 7, and August 4-6, for more workforce education and training in data science, data science publications, etc. See Knowledge Base

Thank you, Brand

Dr. Brand Niemann
Director and Senior Data Scientist
Semantic Community (a non-profit devoted to providing a community service)
http://semanticommunity.info
bniemann@cox.net

Examples of Communities Served:

- CODATA: http://semanticommunity.info/Data_Science/Big_Data_Science_for_CODATA
- CKAN: http://semanticommunity.info/Build_CKAN_in_the_Cloud
- VIVO: http://semanticommunity.info/Build_VIVO_in_the_Cloud
- Chesapeake Bay Program: http://semanticommunity.info/Chesapeake_Bay_Program
- Federation of Earth Science Information Partners: http://semanticommunity.info/Federation_of_Earth_Science_Information_Par...
- First U.S.-European Conference of Metropolitan Regional Councils: http://semanticommunity.info/First_U.S._%E2%80%93_European_Conference_of...
- Growing Communities on Karst: http://semanticommunity.info/Growing_Communities_on_Karst
- Mid-Atlantic_Regional_Planning_Roundtables: http://semanticommunity.info/Mid-Atlantic_Regional_Planning_Roundtables
- Shenandoah_Valley_Science: http://semanticommunity.info/Shenandoah_Valley_Science
- Sustainable Development in the United States: A An Experimental Set of Indicators: http://semanticommunity.info/Sustainable_Development_in_the_United_State

Roger Wakimoto - Opening Keynote

Source: https://docs.google.com/document/d/1...uJxiRfn5U/edit

- GEO Vision Plan 2015-2020
  - Imperatives - GEO’s highest priorities
    - Research
    - Community Resources and Infrastructure
    - Education and Diversity
• Data and CI
  ◦ Frontiers - emerging areas of opportunity and critical for possible future science emphasis

• Data and CI
  ◦ Goals to engage the geosciences community developing a coherent and distributed framework for the open and easy discovery and access to
    • Data, software and services, computational resources
  ◦ EarthCube - community led geoscience cyberinfrastructure that serves the entire academic geosciences community
    • Important - partnership with ACI
  ◦ EX: improved visualization for tracking and cording and post analyses
    • Radar data imposed on satellite data, flight tracks of aircraft
    • PI’s dream of visualizing all this together
    • This would be a great replay for when they go back to their home institute
    • Education component - you can show this is real time or in a module to show students and educators how we can see
  ◦ EX: The Many uses of NCL - scripting language for the analysis of visualization of Geosciences data
    • One measure of success with broader impacts

• Research
  ◦ Strong support for core research (highest priority in geovision doc)
  ◦ PREEVENTS (Prediction of and Resilience of extreme Events.)
    • Prediction and Resilience to Extreme Events
    • Extreme eventss is at the heart of what GEO supports
    • PREEVENTS follows onto SEES (Sci/Engin/Edu for Sustainability) Hazards Program
  ◦ Why NSF?
    • We are the only agency that looks at this in a holistic fashion
  ◦ Consistent w/ Administration Priorities
- OSTP & OMB sci/tech priorities for FY15 budget
- Exec order on climate change (‘preparing the US for the impacts of climate change’)
- Our society is increasingly vulnerable to extreme events
  - Steady rise in impacts
    - Geological, Hydrological, Climate Change, Meteorological
  - Landslides
  - Tropical Storms and hurricanes - shows the importance of tracking these storms before they hit land

  Important Questions
  - How can adaptive sensor networks and portable systems (w/ real-time data assim.) lead to better understanding of the processes underlying extreme events and improve short-time forecasting? (EarthScope, incl. the transportable array, has provided an unprecedented view of the NA plate.
  - Have we reached the limit of radar based warnings of tornadoes or can more be achievable through a combination of numerical forecasted and improving understanding of the public response?

  Tornado Deaths
  - Has gone down over the century, decades
  - Why did we flat-line?
    - Largely it is a social science problem

  Moore Tornado Damage Tracking
  - There is a footprint of the damage path
  - As you get closer to the center you can see the damage get the strongest
  - Everything is convergent to the tornadoes path..

  Water/Food Energy
  - Current program: SEES Water, Sustainability, & Climate Initiative.
  - Follow on: Water/food/energy Nexus

- Community Resources & Infrastructure
  - MREFCs (major research equipment & facilities)
- State-of-the-art facilities (O&M, lifecycle mgmt, RCRV, Community models)
- Identify and implement efficiencies in logistics and operations
- OOI is big-ticket initiative, as well as CZO (Critical Zone Observatory), NEON, Long-Term Ecological Research (LTER)
  - Data Standards, Archives… it’s easy to say we want to integrate, but that’s the trouble (it is very difficult)
- Big challenge is sunsetting facilities - we have to in certain instances
  - Possible criteria:
    - Decadal Surveys/portfolio reviews
    - # of users including students
    - New tech available
- Support for reports that have articulated finding and recommendations regarding facilities and infrastructure
  - Decadal Survey of Ocean Sciences 2015 (a study by the ocean studies board)
- Imperative: Education and Diversity
  - Nurture the next generation of scientists
  - Broaden participation
  - Promote public and community-based science to improve public STEM literacy and advance the geosciences
    - iPhone 6 could include pressure temperature and humidity sensors
    - Vehicle infrastructure integration
  - DoT is in long-term talks w/ automakers about data collection by sensors in vehicles
    - Examples of Inadvertent community science
  - Education:
    - promote and use of community resources (facilities/instrumentation, model, research, data) for both research and educational purposes
      - No long just bringing in a thermometer
      - Now it’s bringing in aircrafts, radar sensors, etc.
• Ex: Jackson State/Univ. Neb: Classroom instruction in concert w/ hands-on experience; Student PIs deploy radar, collect & analyze data, co-author reports

• Ex: Doppler on Wheels: 2000 students participating

• GEOVision
  ◦ Living document consistent with NSF Strategic Plan
  ◦ Presents priorities. Not all inclusive (not the kitchen sink)

• First opportunity to include polar programs

• Extensive input to date (GEO advisory committee, AGU town hall, GEO program office retreat, NRC reports, National Science Board

• Continued opportunity for input this summer before we start finalizing

Questions:

• Will there be public meeting for public input?
  ◦ All the foundations have had public comment periods (publishers, societies)

• Education about how to make sense of data and models is not in our system, how can we do this more effectively? But we are receiving pushback from NSF about education within EarthCube, why is that?
  ◦ Re: gov diredict that edu sb centralized
  ◦ GEO closely involved w/ EDU directorate, but doesn’t want to duplicate effort.
  ◦ It would be nice to have an REU model, that’s the target of where we want to go
    • Target: Discipline-specific needs being protected.

• Data management plan mandate. There is a concern. Will there be a central data repository?
  ◦ DMP requirement has a bigger issue: Two aspects of public access: Publications & data. Open access for data is a major problem—too many kinds of data, too much data, too many angles & protocols.
  ◦ This is a huge problem that we are still grappling with
  ◦ The NSF has kind of put them off to be able to give them a thoughtful and real answer
  ◦ Resources for this mandate is a big issue (for the PI) and NSF is trying to address that (but can’t say more)

• DMP: Many research materials & physical samples are thrown out because of poor labeling/organization. How can they be included in the DMP mandate?
BIO put a grand challenge out to community how they include millions and millions of cases of bugs in their data archives?

Maybe NSF could do the same. Ask the community? Put out a reward?

It isn’t that we make this data available—scientists seek data. It’s our responsibility to make data available to those who normally wouldn’t be able to access. With available data, more people (the haves-nots) will be able to contribute

- How can I make my databases more available?
  - EC is one piece of NSF approach to data management & Cyberinfra.
  - There are opportunities that are aligned with EC within the CI

- Lack of funding opportunities in interagency cooperation?
  - Part of this is agencies coming together w/ an interagency plan. Several meetings w/ NOAA/NASA reps looking for synergistic opportunities
  - It is still a challenge
  - Nothing has been planned NOW, but I wouldn’t rule it out

- Do you view that the balance is shifting such that the NSF will fund more long-term observation projects?
  - The big challenge/first step for NSF is getting these organizations to link & communicate
  - NEON is an example
  - It is a data issue too
    - Long term integration issues of data over time is a hard problem

- What is your own view on the potential of EarthCube? If EC is strong enough, could it be in a linchpin in solving this data integration problem?
  - That is the stretch goal of EC

---

**AGU 2014 Session Ideas - EarthCube Community Engagement Team**

**Suggestions for AGU Sessions**

Source: [https://docs.google.com/document/d/1...5gzcdi7v8/edit](https://docs.google.com/document/d/1...5gzcdi7v8/edit)

Submission deadline: 16 April 2014, 11:59 p.m. EDT


Titles <= 300 Characters, abstracts <= 150 words

Proposed Section/Focus Group Sponsorship: ESSI

Designate Oral or Poster Session

Maximum of 4 Conveners; primary convener must have up-to-date AGU membership

At least 1 Index Term must be included: [http://onlinelibrary.wiley.com/agu/browse/taxonomy](http://onlinelibrary.wiley.com/agu/browse/taxonomy)

---

**[Accepted] Collaborative Community Platforms for the Earth Sciences**

- **Conveners/Affiliations:**
  - Suggested Conveners: RDA (Mark Parsons), ESIP (Erin Robinson), EarthCube (Kim Patten, ECTEG)
  - **Abstract:** Collaborative communities for the Earth sciences are inherently cross-domain, cross-sector, and geographically distributed. Because of this diversity, these groups have unique needs for collaboration platforms. These platforms provide opportunities to domain scientists and technology practitioners to discuss and achieve consensus (or at least identify where consensus is lacking) on some of the most pressing issues associated with today’s grand-challenge science by supporting remote and in-person collaboration. Platform is a broad term and may include off-the shelf and custom tools, diverse meeting designs and fora, and conventional and novel methods. This session seeks proposals on innovative solutions to, and examples of, collaboration platforms that improve the scalability of Earth and information sciences for moving today’s grand-challenge science forward.

- **Potential Invited Speakers:** Andres Karpistenko, Marine Explore; Bruce Caron, TNMS;

---

**[Accepted] International Cooperation to Further the Earth Sciences**

- **Suggested Conveners:** RDA (Mark Parsons), Belmont (Genevieve Peartthree/Lee Allison/Robert Gurney), CoopEUS (Bob Weller, OOI)
  - **Abstract:** Grand-challenge science doesn’t stop at national borders. Recently, a number of international initiatives have formed, such as CoopEUS, FutureEarth, and RDA, to address issues associated with international collaboration, funding of international research, collaborative science, best practices among facilities and observatories, and information management. How do we learn from past and existing initiatives like GEOSS, the IODE, the OGC and the CEOS and guide these new initiatives to truly move global science partnerships forward? How do we ensure that the activities do not duplicate, and are funded sufficiently, and
how do we manage this internationally? An activity of the Belmont Forum, a group of international funding agencies, is now starting to address the funding issues, but there remain many others. This session seeks proposals on past and current best practices in international collaboration spanning the domain sciences to information sciences to funding best practice to legal challenges associated with cross-border research.

• [Accepted] Scalable and Adaptable Architecture for Earth Science Cyberinfrastructure
  ◦ Conveners/Affiliations:
  ◦ Suggested Conveners: Ilya Z., Phil Yang, Lee Allison, Emily Law
  ◦ Abstract: Cyberinfrastructure enables and advances Earth Science research. It is the foundation of research workflows that support advanced data acquisition, processing, management, integration, analysis and visualization distributed over the internet. It also connects people, organizations, research and curricula resources in ways that advance scholarly communication and collaboration. The construction of highly scalable, adaptable scientific cyberinfrastructures within and across scientific disciplines has become critical for enabling effective Earth Science research programs such as the NSF EarthCube; particularly as the data and computational demands continue to grow. Systems must be thoughtfully architected in order to support scalability and to provide enough agility and adaptability to allow storage, processing, and analysis to occur across distributed environments. This session seeks papers that describe Earth Science cyberinfrastructure concepts; challenges; experiences, ideas and lessons learned in architecture, design and development of scalable and adaptable cyberinfrastructure.
  ◦ Emily Law/JPL: I am interested in being a co-convener of this session. I have strong interest in architecture and development of Earthcube and data intensive systems in general. My active role in Earthcube activities can facilitate and help recruit submissions.

• [Accepted] Cyberinfrastructure for field work: data standards, computer applications, instrumentation and best practices
  ◦ Conveners/Affiliations: Xiaogang Ma (RPI), Douglas J. Walker (UK), Matty Mookerjee (Sonoma U.), Boyan Brodaric (GSC)
  ◦ Abstract: Fieldwork provides primary data about the long-term history of the Earth’s lithosphere, hydrosphere, cryosphere and atmosphere. As the geoscience community continues to discuss, develop and refine its cyberinfrastructure, it is important to build the infrastructure to facilitate the inclusion of these field data. We need to promote easier data collection, sharing, discovery, accessibility and integration, i.e., we need to design a workflow that eases the transfer of field data into their appropriate databases. This may include the incorporation of mobile electronic devices and/or desktop applications that can expedite the collection,
management and processing of field data. These applications need to be designed to accommodate the continually emerging community consensus on data standards, vocabularies and metadata schemas, as well as the mechanisms of data sharing, citation and integration. We welcome submissions on conceptual model, database design, computer program, instrument, etc. associated with the fieldwork topic.

- [Accepted] Information models and interchange formats for geoscience data

(Related OGC: GeoSciML, WaterML and variations)

- Conveners/Affiliations: Xiaogang Ma, David Arctur, Ilya Z., Steve Richard

- Abstract: As geoscientists develop new types of measurements or improve understanding of existing data, formal information models and interchange formats increasingly appear, such as the North American Geological Map Data Model (NADM), the XML implementation of the OGC and ISO Observations and Measurements (O&M) conceptual model and Geography Markup Language (GML) based schemas, including GeoSciML (geological formations and borehole geometry), EarthResourceML (mineral resources exploration and exploitation), WaterML 2 (surface water time series), and GroundwaterML (aquifers and underground water flow). In order to adapt to rapidly evolving technology, a trend is to separate information models from their encoded interchange formats. Community consensus on these models and interchange formats can reduce the effort needed to reconcile datasets within a domain and to integrate data and applications across domains. Bring your experiences developing and using information models and interchange formats, and see how many common elements you can find in this session.

Possible alt version for discussion

- Electronic exchange of scientific information between interdependent science domains with varying conceptualizations of the Earth must be based on information models that are sufficiently documented that the spectrum of users can understand their meaning. Such models are the foundation of semantic interoperability. In order to actually exchange information, these models must be encoded in formats that can be generated and parsed by computer systems. It has become apparent that in order to adapt to rapidly evolving technology, it is vital to separate the abstract information model from interchange formats that implement all or part of the model. Information conforming to the same abstract model can be transformed automatically between various encoding schemes, providing utility even as technology changes.

- A growing number of information models and interchange formats are in use for geoscience data exchange. Examples include the North American Geological Map Data Model (NADM), the ISO land cover classification system (LCML), the OGC and ISO Observations and Measurements (O&M) conceptual model, and the WaterML abstract specification. Many implementations of these models are Geography Markup Language (GML)-based XML schemas, including GeoSciML (geological formations and borehole geometry), EarthResourceML (mineral resources exploration and exploitation), WaterML 2 (surface water time series observations), and GroundwaterML (descriptions of aquifers and underground water flow). Community consensus on information models and interchange formats can reduce the effort needed to reconcile datasets within a domain and to integrate data and applications across domains. Bring your experiences developing and using information models and interchange
formats for interoperability across the geosciences, and see how many common elements you can find in this session.

- **[Merged] Methods, tools and best practices of data visualization in Earth and space science**
  
  - **Conveners/Affiliations:** Xiaogang Ma (RPI), Shaowen Wang (UIUC), Wenwen Li (ASU), Linyun Fu (RPI)
  
  - **Abstract:** Data visualization plays an increasingly important role in Earth and space science. Data visualization of nowadays is not only visual displays of data, it also promotes interactions between people and the data and leads to exploration on research ideas, problem solutions and new knowledge construction, as well as the communication of scientific information with the general public. The characteristics of Earth and space science data, e.g., multi-source, multi-dimension, multi-subject and heavy volume, bring both challenges and opportunities to data visualization, and various feature methods, tools and applications have been developed, such as those in astronomy, oceanography, climatology, and geology, etc. We welcome submissions discussing recent progress on methods and tools. Experiences and feedbacks of using data visualization to tackle real-world challenges or research issues are especially encouraged to submit.

- **[Merged] Semantic technologies in the life cycle of geoscience data**
  
  - **Conveners/Affiliations:** Xiaogang Ma (RPI), Tom Narock (Marymount U.), Brian Wilson (NASA/JPL), Matthew Mayernik (NCAR)
  
  - **Abstract:** A data life cycle represents the whole process of data management and use, including concept study, data collection, processing, archiving, publication, discovery, analysis and repurposing. Concept articulation and definition can be, but is not always, a primary part in the first step of the data life cycle. For a relational database, concept study is the construction of conceptual models and logical models, and for an end-point in the Semantic Web, concept study is building ontologies and vocabularies. Though we consider concept study as an initial step of the data life cycle, the built ontologies and vocabularies are applicable to almost every step in the open-ended procedure of a life cycle. We welcome submissions that overlay semantic technologies to one or more steps in the data life cycle. Works addressing a domain-specific topic are especially encouraged to submit.

- **[Accepted] Provenance enablement in Earth and space science**
  
  - **Conveners/Affiliations:** Xiaogang Ma (RPI), Wenwen Li (ASU), Deborah McGuinness (RPI), Curt Tilmes (NASA)
Abstract: Provenance is information about entities, procedures, people and organizations involved in the production of research findings and the supporting datasets and methods. For the field of geospatial science, provenance is also known as the lineage of data. Capturing and presenting provenance have been of great importance to ensure reproducibility for many aspects of Earth and space science practice. Within the Open Geospatial Consortium (OGC), geospatial provenance has been widely discussed, such as those represented in the ISO 19115 and ISO 19115-2 lineage models. A recent effort in this field is the PROV data model and its support ontology (PROV-O) recommended by the World Wide Web Consortium (W3C), which provides an initial basis for capturing, representing and linking provenance. Best practices for implementing standards for provenance management are now emerging. This session welcomes contributions that demonstrate how provenance contributes to meeting scientific or engineering goals in Earth and space science works.

• [Accepted] Computational intelligence in Earth and space science

  Conveners/Affiliations: Xiaogang Ma (RPI), Qiuming Cheng (York U.), Shaowen Wang (UIUC), Han Wang (RPI)

  Abstract: Earth and space scientists deal with both numbers and texts in their daily work, and they use methods for both quantitative and qualitative analyses. The topic of computational intelligence aims at addressing complex real-world problems that are hard for traditional methods, and it roots in both quantitative and qualitative analyses. For the former, progress can been seen in fuzzy logic, fractal analysis, genetic algorithms, chaotic systems, probabilistic reasoning, singularity analysis, etc. For the latter, a fast rising topic is the natural language processing, which is widely used in information extraction and knowledge population with ontologies. Along with the world wide efforts in open data and the thriving social media (e.g., relevant topics in crowd-sourcing geoinformation, citizen science, etc.), computational intelligence faces a broader stage for further development and applications. We welcome submissions of methods and applications of quantitative and qualitative analyses (or both) in Earth and space science.

  Both quantitative and qualitative analyses are involved in Earth and space sciences. Computational intelligence is central to both, and has become increasingly important to addressing complex scientific problems. On the quantitative frontiers, areas such as fuzzy logic, fractal analysis, genetic algorithms, chaotic systems, probabilistic reasoning, and singularity analysis have made extensive progress. As an example of qualitative approaches, natural language processing is central to intelligent analysis of massive text data, and is widely used in information extraction and knowledge population with ontologies. Along with world wide efforts on open data and expansive social media, computational intelligence faces tremendous challenges and opportunities for further development and applications. We encourage submissions of computational intelligence theories, methods and applications from both quantitative and qualitative perspectives in Earth and space sciences.

• [Accepted] The Brokering Approach: its services and applications for multidisciplinary science research

  Conveners/Affiliations: Jay Pearlman, Siri Jodha Singh Khalsa, Stefano Nativi
• Abstract:

• [Submitted] Linking Physical Objects to Cyberinfrastructure

  Conveners/Affiliations: Kerstin Lehnert, Sarah Ramdeen, David Arctur

  Abstract: Physical objects such (specimens, cores, artifacts) are collected and studied in many Earth Science domains to generate ex-situ observations of geological and environmental features. Recent advances in physical object identifiers such as the International Geo Sample Number (IGSN), and the need to align information models for geological and biological specimens in paleobiological data systems and for environmental observations is leading to increased interdisciplinary cooperation and collaboration in the development of best practices for object identification, documentation, and citation, and for tools that facilitate the curation and digital preservation of large collections of physical specimens. This session calls for ideas, experiences, successes, and lessons learned in developing information models, databases, registries, and workflows for curation and digital preservation of physical objects that facilitate open discovery and access to the physical objects and the digital records, particularly methods and tools to generate and embed metadata collection within data capture applications.

  Sarah Ramdeen: I would like to work in some relation to this session. I am currently working on my dissertation proposal (which looks at information seeking in relation to physical objects that can not be fully replaced by a digital surrogate in CI) and this relates directly to my overall research interests/agenda.

• [Submitted] An Open Source Understanding Framework for Geosciences Technologies

  Coveners/Affiliations: Chris A. Mattmann (NASA JPL/USC), Yolanda Gil (USC-ISI), Robert Downs (Columbia University/CIESIN), Paul Ramirez (NASA JPL)

  Abstract:

  Chris A. Mattmann: I am currently funded on EarthCube Building Blocks to work with Dr. Gil on her project for Software Stewardship and Open Source geosciences software. I am a Director at the Apache Software Foundation; one of the creators of Apache Nutch/Hadoop, the progenitor of the Apache Tika framework for content detection and analysis and am active in understanding open source as an understanding framework for producers and consumers of software. I can help to recruit, and have held AGU sessions, ESIP sessions and sessions at ApacheCon in geosciences software stewardship for the last 4 years.

• [Submitted] Polar Cyberinfrastructure

  Coveners/Affiliations: Chaowei Phil Yang (GMU), Paul Morin (UMN), Wenwen Li (ASU), Lynn Yarmey (UC-Boulder)

  Abstract: We invite submissions concerning research of polar cyberinfrastructure. Cyber technologies have been positively impacting polar research in many ways. An increasing number of sophisticated sensors have
been deployed in both the Arctic and Antarctic regions. Enhanced computational power has allowed polar scientists to better capture the present state of polar regions and to unveil past trends and project future changes of process-driving quantities. Topics of interest include, but are not limited to:

1. Sensors, real-time data transmission and networking
2. Data archival and management supporting continuity of long-lived publicly accessible data sets
3. Information integration, mining, semantics, and ontologies
4. Model simulation and forecasting
5. System interoperation and software technologies

• [Submitted] Cloud Computing for Geosciences
  ◦ Coveners/Affiliations: Chaowei Phil Yang/GMU; Alex Sun (UT-Austin); Thomas Huang (NASA JPL); Wenming Ye (Microsoft)
  ◦ Abstract: Geosciences pose cyber-infrastructure and computing challenges from disruptive hazards events, complex model configuration, interoperability in data/service sharing, and many others. Cloud Computing emerged as one of the most promising solutions. This session welcomes contributions that demonstrate addressing these challenges with cloud-computing. Topics include but not limited to cloud-related research and development for

  1. data management
  2. interoperability
  3. computing support
  4. real time decision support
  5. knowledge management
  6. model configurations
  7. visual analytics, and
  8. cloud-based uncertainty quantification and sensitivity analyses.

Contributions that fuse participatory social learning into the geoscience R&D processes are also welcome.
EarthCube Summit 2014

EarthCube Summit at the Open Geospatial Consortium Technical Committee meeting in Crystal City (Arlington) Virginia, Tue March 25, 2014, 1-5pm US EDT.

Agenda

1:00pm  Welcome and introductions (David Arctur)
1:15pm  EarthCube Community Engagement (Erin Robinson)
1:40pm  EarthCube IT (George Percivall)
2:00pm  EarthCube Conceptual Designs
    2:00pm - CD-1 (Ilya Zaslavsky, Steve Richards)
    2:20pm - CD-2 (Phil Yang)
2:40pm  EarthCube Building Block – CINERGI (Ilya, Steve)
         OGC validation work in CINERGI (Luis Bermudez)
3:05pm  EarthCube BB – Integrating Discrete & Continuous Data (David Arctur)
3:20pm  EarthCube BB – OceanLink (Bob Arko)
3:40pm  ODM2 – Observation Data Model v2 (Anthony Aufdenkampe)
4:00pm  GeoViQua – User Feedback on Data Quality (Joan Maso)
4:20pm  GeoVoCamp – Geo Vocabulary Workshops (Charles Vardeman)
4:40pm  Closing (David Arctur)

Location
Lockheed Martin - Global Vision Center
2121 Crystal Drive
Arlington, VA 22202-

David K Arctur, PhD
Research Scientist, University of Texas at Austin
Jackson School of Geosciences, Cockrell School of Engineering, School of Information
EPS 1.148 - C9000, 2275 Speedway, Austin TX 78712 USA
cell +1.512.771.1434 :: office +1.512.465.2299 :: skype darctur
http://www.jsg.utexas.edu/researcher/david_arctur :: https://sites.google.com/site/dkarctur

Research/Academic Advocate, Open Geospatial Consortium (OGC)

The future is here. It's just not widely distributed yet.
- William Gibson

Selected Slides

Source: https://drive.google.com/folderview?..W8&usp=sharing
00-EarthCube Agenda at OGC TC, DC.ppt

Slide 1 Title Slide

David Arctur, George Percivall, Erin Robinson
25 March 2014
88th OGC Technical Committee
Washington, DC

Slide 2 Agenda

http://bit.ly/1fNQTXX

EarthCube Update: Activities & Interests
Tue, March 25, 1-5pm

Agenda
1:00pm - Welcome and introductions (David Arctur)
1:15pm - EarthCube Community Engagement (Erin Robinson)
1:40pm - EarthCube IT (George Percivall)
2:00pm - EarthCube Conceptual Designs
   2:00pm - CD-1 (Ilya Zaslavsky, Steve Richard)
   2:20pm - CD-2 (Phil Yang)
2:40pm - EarthCube Building Block – CINERGI (Ilya, Steve)
   OGC validation work in CINERGI (Luis Bermudez)
3:05pm - EarthCube BB – Integrating Discrete & Continuous Data (David Arctur)
3:20pm - EarthCube BB – OceanLink (Bob Arko)
3:40pm - OOM2 – Observation Data Model v2 (Anthony Aufdenkampe)
4:00pm - GeoViQua – User Feedback on Data Quality (Joan Masso)
4:20pm - GeoVoCamp – Geo Vocabulary Workshops (Charles Vardeman)
4:40pm - Closing (David Arctur)

Presentations posted on google drive folder:
http://bit.ly/1fNQTXX

01-OGC_EarthCubeCommunityEngagement.pptx
The vision…

"Over the next decade, the geosciences community commits to developing a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core, including the influences of humans and ecosystems."


02-Open Standards Role in EarthCube IT.pdf

Slide 1 Title Slide
Slide 2 The EarthCube Strategy

The EarthCube Strategy

- Engage all stakeholders: Geosciences end-users, Geosciences and CI facilities, CI and Computer Science specialists
- Build EarthCube iteratively, with community input and assessment in yearly intervals
- EarthCube built on existing resources, understanding that different geosciences communities are not uniformly served

Source of slide: Eve Zanzibbe, NSF

Copyright © 2014, Open Geospatial Consortium

Slide 3 Open Data and a Culture of Cooperation

Open Data and a Culture of Cooperation

- EarthCube is based on a network of enthusiasts willing to make the sharing of data a reality. But is just having open data enough?
  - Open data will not accelerate the process a scientist team needs to go through to understand, reformat and use the data.

- Letter to Editor response by 44 persons in EarthCube
  - "Although the question of who pays for open data is important…
  - A greater challenge lies in implementing institutional and cultural changes required before data from government-sponsored research can be openly shared"
  - Science 29 November 2013: V. 342 pp. 1041-4042

Copyright © 2014, Open Geospatial Consortium
Slide 4 Prototyping Versus Specifying

**Prototyping Versus Specifying**

- "Prototyping Versus Specifying: A Multiproject Experiment"
- 1) Prototyping yielded products with roughly equivalent performance, but with about 40 percent less code and 45 percent less effort.
- 2) The prototyped products rated some what lower on functionality and robustness, but higher on ease of use and ease of learning.
- 3) Specifying produce produced more coherent designs and software that was easier to integrate.
  - BARRY W. BOEHM, TERENCE E. GRAY, AND THOMAS SEEWALDT, IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. SE-10, NO. 3 MAY 1984
- See also Michael Schrage, e.g., **Serious Play** and "Cultures of Prototyping" in **Bringing Design to Software**

OGC

Copyright © 2012, Open Geospatial Consortium, Inc.

Slide 5 GEOSS Approach Relevant to EarthCube

**GEOSS approach relevant to EarthCube**

GEOSS connects Observations to Decisions

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
Slide 6 GEOSS Interoperability Arrangements

- Interoperability through open interfaces
  - Interoperability specifications agreed to among contributing systems
  - Access to data and information through service interfaces

- Open standards and intellectual property rights
  - GEOSS adopting standards agreed upon by consensus, preference to formal international standards
  - GEOSS will not require commercial or proprietary standards
  - Multiple software implementations compliant with the open standards should exist
  - Goal is that at least one of the implementations should be available to all implementers "royalty-free"

Slide 7 GEOSS Information System and GEOSS DataCORE

http://www.geoportal.org

- GEO Portal - new version released last week
- More than 14 million discoverable Resources
- More than 1.2 million are
- Brokered Search of >20 Community Catalogues
- Example search: 1081 datasets for Land Cover available as GEO DataCORE

03-RichardZaslavsky_EC_ConceptualDesign.pptx
Towards a Conceptual Design of a Cross-Domain Integrative Information System for the Geosciences

ILYA ZASLAVSKY, DAVID VALENTINE, AMARNATH GUPTA
San Diego Supercomputer Center/UCSD

STEPHEN RICHARD
Arizona Geological Survey

TANU MALIK
University of Chicago
Slide 3 Data, Domain Stacks, Integration

Slide 4 Geoscience Small-Scale Enterprise
We need more research to turn this into a prototype for enabling scientific research and decision support.

System to identify gaps in existing data sets and prioritize/incentivize verification of contradictory information, as well as filling gaps with new records

New educational capability that is built upon data and results drawn “live” from existing resources

Legacy and dark data incorporation – noisy signal processing.
Add the linkage for the workflow of field work to physical samples, to lab work, to education and research and to paleoscience communities to field work. Categorize the earthcube components and organize the categories close to each of the previous workflow component. Please don’t get creative to add too many arts, let’s focus on the ideas to make it clear first.

**Slide 2 Overall Aims of the Project**

The design is focused on fostering collaboration mechanisms across the geosciences disciplines, rather than creating new. Our design will be articulated in four reports:

- **Report one** provides a general context about enterprise architecture and its benefits to facilitating EarthCube enterprise. More importantly, it provides unambiguous definitions about what is EarthCube enterprise and who are the participants.
- **Report two** provides the conceptual design for EarthCube enterprise architecture. So the goal is how you should do it, not what you should do, as mentioned in Day one’s presentation. The strategy we are taking is laying out what existing in the fields, and then building connections at two levels: the data and the EarthCube participants.
- **EarthCube Systems** describes the scope and technologies of existing data/computing/experimenting infrastructures that will be incorporated into the EarthCube enterprise.
- **EarthCube Services** describes how existing resources will support the vision and requirements of EarthCube.
- **EarthCube Operational** describes the operation of EarthCube by taking advantages of existing resources.
- **EarthCube Capability** describes the capabilities that are desired by the communities, and existing capabilities. The discrepancies between the two provide evolvement direction for the EarthCube.
- **EarthCube Data and Information** describes data flow and information exchange between different layers from EarthCube Capacity to Systems.
- **Technical Standards** describes all standards that will be implemented in EarthCube and their applications.

**Report three** provides ontological description of EarthCube and the relationships between metadata of EarthCube.

**Report four** provides template and guidance for geoscience practitioners to partake the EarthCube.
Slide 3 Paleogeoscience Community Identified Resources in Categorization

DETELON (Deep Time Earth-Life Observatory Network)
http://www.paleosoc.org/DETELON_Scie...Brochure1.pdf

TRANSITIONS:
http://www.sepm.org/CM_Files/ConfSum...TIONSfinal.pdf

NROES (New Research Opportunities in the Earth Sciences)
http://www.nap.edu/openbook.php?record_id=13236

Conservation Biology Workshop Report
http://www.paleosoc.org/CP_Workshop...t_Oct_2012.pdf

IGBP PAGES Report 57
http://www.igbp.net/download/18.1b8a...t_57-PAGES.pdf

NRC 2011 Report
http://www.nap.edu/catalog/13111.html

Paleogeoscience Community Identified Resources in Categorization (Data, Computing, Services, Organization, Documents)

Documents:
- DETELON (Deep Time Earth-Life Observatory Network)
  - http://www.paleosoc.org/DETELON_Scie...Brochure1.pdf
- TRANSITIONS:
  - http://www.sepm.org/CM_Files/ConfSum...TIONSfinal.pdf
- NROES (New Research Opportunities in the Earth Sciences)
- Conservation Biology Workshop Report
- IGBP PAGES Report 57
  - http://www.igbp.net/download/18.1b8a...t_57-PAGES.pdf
- NRC 2011 Report
  - http://www.nap.edu/catalog/13111.html

Existing programs and projects of relevance to palaeoscience in general.
The proposed CI architecture provides a systematic technological framework for EarthCube by modeling the Collection, Process, and Presentation as co-equal functions in a continuous closed-loop that is driven and captured by a Collaborative Knowledge Network (CKN), a distributed collection of information that can be accessed in a coherent way.

EarthCube Enterprise Architecture Workshop at ESIP Summer Meeting

- Date: July 7, 2014
- Location: Copper Mountain, Colorado
- Content: Domain and Architecture experts to review and comment on the design
  - Pick one or more volumes to comment
  - Discuss comment at ESIP Summer Meeting
  - Provide advice on improving the EA
- Support:
  - $700/expert for up to 10 experts
- Call for participation is out and will be due April 15, 2014
- **How to apply:** Email Erin Robinson at erinrobinson@esipfed.org

05-CINERGI_OGC_TC_Mar25_2014.pptx

Slide 1 Title Slide

Community
Inventory of
EarthCube
Resources for
Geoscience
Interoperability

CINERGI data discovery is the most often cited issue in executive summaries on the EarthCube website

Ilya Zaslavsky, Steve Richard and the CINERGI team
http://workspace.earthcube.org/cinergi
Slide 2 Goals

Goals

- Large inventory of high quality information resources across disciplines, with traceable provenance, usable across EarthCube research scenarios:
  - datasets, catalogs, vocabularies, information models, services, process models, repositories, etc.
- Make it open to the community
- Organize it to enable search and integration across domains and linking between information objects
  - Plus links between resources, people/organizations, publications, models, workflows, software, activities, etc.

Slide 3 Approach

Approach

- Build on high-level resource inventory started at http://connections.earthcube.org
- Compile metadata for as many resources as we can (collect recommendations from geoscientists, harvest existing catalogs)
- Expose through simple search interface
- Use off the shelf technology: Geoportal, ISO metadata, CSW
- Make it accessible through EarthCube.org
Then add features

- Links to organizations, researchers, other systems
- Validation Services
- Deep registration of datasets/databases (at feature level)
- Data search capabilities
- Quality/interop readiness assessment
- Annotation system
Slide 2 What Tool Can We Use?

What tool can we use?

OGC Web Testing Facility

This free "self-service" testing facility can be used by any developer as often as they like to test their implementations of OGC standards. You don't need to be an OGC member, and you will always have validated access to this testing facility.

There are more than 1,000 OGC-compliant standards, but we currently can only certify implementations for the tests listed below. New tests currently under development can be seen and tried out at our [Beta Testing Facility]. We encourage you to try them out and provide feedback, as that is how standards evolve and improve.

Once your product passes the test getting it certified is easy. Visit the OGC [Compliance web page] for more details. If you need help passing the test contact the manager of the OGC forum.

We have a need for new tests and for reference implementation. If you want to contribute a reference implementation or you want to develop OGC tests, please contact the Director of the Compliance Program, [Link to contact].

<table>
<thead>
<tr>
<th>Specification</th>
<th>Version</th>
<th>Test Suite Revision</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalogue Service (CSW)</td>
<td>2.0.13</td>
<td>3.0.0-051</td>
<td>Final</td>
</tr>
<tr>
<td>Geography Markup Language (GML)</td>
<td>3.2.1</td>
<td>3.2.1-051</td>
<td>Final</td>
</tr>
<tr>
<td>Sensor Observation Service (SOS)</td>
<td>1.0.0</td>
<td>1.0.0-051</td>
<td>Final</td>
</tr>
<tr>
<td>Sensor Observation Service (SOS)</td>
<td>2.0.0</td>
<td>2.0.0-051</td>
<td>Final</td>
</tr>
</tbody>
</table>

07-DisConBB Overview, user needs.pptx

Slide 1 Title Slide

EarthCube Building Block for Integrating Discrete and Continuous Data (DisConBB)

David Maidment, University of Texas at Austin (Lead PI)
Alva Couch, Tufts University
Ethan Davis, Unidata
Dan Ames, Brigham Young University

EarthCube Summit at OGC TC, Arlington VA, 25 March 2014
David Arctur, University of Texas at Austin (Project Manager)
Slide 2 Discrete & Continuous Data: Project Outline

Discrete & Continuous Data: Project Outline

- Discrete spatial domains: GIS features (point, line and area) with observations & measurements
- Continuous spatial domains: Grids of measured or modeled variables in physical or fluid sciences
- Spatially discrete or continuous data may also vary discretely or continuously in time:
  - one-time samples vs. random points of time vs. regularly spaced intervals of time

**Common Information Model**
- Precipitation
- Evaporation
- Soil Moisture
- Streamflow
- Groundwater
- Reservoirs

**Server & user tools**

**Phase 1: Describe water & atmospheric properties over a domain of space and time**
- History
- Current conditions
- Forecasts

**Phase 2: Apply concepts and methods in other domains**
- Solid Earth
- Cryosphere
- Oceans

Slide 3 Prototype: Soil Moisture Map & Time Series

Prototype: Soil Moisture Map & Time Series

This is a common pattern across geosciences –
- **Solid Earth**: seismic activity, soil chemistry over deep time, ...
- **Oceans**: SST, acidification, ...
- **Cryosphere**: ice thickness, trapped gas content, ...

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
Overview

Goal
- Enable discovery of geoscience data and knowledge, and ultimately, integration

Strategy
- Publish content from existing network of repositories as Linked Open Data (LOD)
- Enable horizontal semantic integration
- Provide tools + services useful to working scientists

Collections
- Biological & Chemical Oceanography Data Management Office (BCO-DMO)
- Rolling Deck to Repository (R2R)
  - cruise catalog + underway enviro sensor data
- Marine Biological Laboratory / Woods Hole Oceanographic Institution (MBLWHOI) Library
  - published articles, theses, tech reports, datasets
- AGU meeting abstracts
- NSF funding award abstracts
Slide 4 Initial Results

Initial Results

"An Ontology Pattern for Oceanographic Cruises" (Krisnadhi et al.)
Technical Report and draft set of ODPs

Reuses existing patterns including
- Semantic Trajectory (Janowicz et al.)
- Information Object
- Simple Event Model
to model a Cruise and ship's track

Slide 5 Acknowledgements

Acknowledgements

“EAGER: Collaborative Research: EarthCube Building Blocks, Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery”

NSF Funding Awards:
ICER 13-54990 LDEO
ICER 13-54693 UMBC
ICER 13-54778 WSU
ICER 13-54107 WHOI

September 15, 2013 - August 31, 2014
Observations Data Model 2.0
A community information model for interoperability among feature-based earth observations

Jeff Horsburgh, USU. Project PI.
Anthony K. Aufdenkampe, Stroud Water Research Center

Kerstin Lehnert, IEDA/Columbia
Emilio Mayorga, UW-APL
Ilya Zaslavsky, SDSC
David Valentine, SDSC
David Tarboton, USU
David Lubinski, UC-Boulder

Critical Zone Science

Earth’s permeable near-surface layer from the tops of the trees to the bottom of actively cycling groundwater.

- Where rock, soil, water, air, and living organisms interact and shape the Earth’s surface.
- Critical to sustaining the earth’s sustaining services
  - Clean water
  - Productive soil
  - Balanced atmosphere
Slide 3 ODM2: Common to Most Data Types

ODM2: Common to Most Data Types

Sensor Extension  Equipment & Lab Extensions
Observations Core

Feature Model  Generic Extension

Common Semantics for Earth Observations

CUAHSI HIS  EarthChem  CZOData  IOOS

Domain Cyberinfrastructures

Slide 4 ODM2: Common to All Components

ODM2: Common to All Components

Legend

Database Encoding

Data and Metadata Transfer

Metadata Encoding

Data and Metadata Transfer

Catalog

Metadata Catalog

Metadata Harvesting

Metadata Transfer

Information Model

Data Discovery

Data Delivery

Data Server

Clients

Data Transfer
Slide 5 ODM2Core

Slide 6 NSF Scientific Software Integration

NSF Scientific Software Integration

BiG CZ SSI project (2014-2015):
The community-driven BiG CZ software system for integration and analysis of bio- and geoscience data in the critical zone

- Community Engagement in Software Design through co-design, training & testing workshops
- BiG CZ Portal web application for high-performance map-based discovery, visualization, access & publication of data on critical zone structure & function
- BiG CZ Toolbox to enable cyber-savvy CZ scientists & data managers to manage and publish the data they produce through a single scientist-focused toolkit
- BiG CZ Central software stack to bridge data systems developed for multiple critical zone domains

Slide 7 Thank You

http://github.com/UCHIC/ODM2
Future Directions for CZO Science

Report prepared by CZO community, Dec. 2010

- Develop a unifying **theoretical framework** of CZ evolution;
- Develop **coupled systems models** to explore how CZ services respond to anthropogenic, climatic, and tectonic forcings;
- Develop **four dimensional data** sets that
  - document differing CZ geologic and climatic settings,
  - inform our theoretical framework,
  - constrain our conceptual and coupled systems models,
  - test model-generated hypotheses.
Engaging the Critical Zone community to bridge long tail science with big data

Convened by A.K. Aufdenkampe, C.J. Duffy, G.E. Tucker
Univ. of Delaware: Jan. 21-23, 2013

Organizing Committee:
Kerstin Lehnert, IDEA/Columbia.
Ilya Zaslavsky, SDSC.
David Tarboton, USU.
Jeff Horsburgh, USU.
Emilio Mayorga, UW-APL

103 Participants from 16 Disciplines

- Biogeochemistry (30)
- Biology/Ecology (15)
- Biology/Molecular (3)
- Climatology/Meteorology (15)
- Data Management/Cyberinfrastructure (46)
- Engineering/Method Development (8)
- Geochemistry/Mineralogy (13)
- Geology/Chronology (14)
- Early-Career (28)
- Geomorphology (15)
- Geophysics (8)
- GIS/Remote Sensing (31)
- Hydrology (46)
- Modeling/Computational Science (36)
- Outreach/Education Research (7)
- Soil Science/Pedology (16)
- Water Chemistry (14)

10-GeoViQua – User Feedback on Data Quality.ppt
GeoViQua provides a set of scientifically developed software components and services that facilitate the creation, search and visualization of quality information on EO data integrated and validated in the GEOSS Common Infrastructure.
Slide 3 GMU Global Agricultural Monitoring and Forecasting System

George Mason University Global Agricultural Drought Monitoring and Forecasting System

Slide 4 Some Useful URL’s

http://geoviqua.org

Some Useful URL’s

- Schemas XSD and UML (ASTON)  
  - http://schemas.geoviqua.org
- User Feedback System (S&T)  
  - https://github.com/GeoVlQua/geo-userfeedback
  - http://geoviqua.stcorp.nl/home.html
- GEO label services, API and code (ASTON+52N)  
  - http://www.geolabel.net/api.html
  - http://twiki.geoviqua.org/twiki/bin/view/GEO_SIF/SIFGeoLabel
  - https://github.com/flush/geolabel-service
- Tutorial (ASTON)  
  - http://tutorial.geoviqua.org
- DAB-Q CSW, GI-CAT (CNR) description and code  
  - http://geoviqua.essi-lab.eu/dabq-demo/

11-GeoVoCamp -- Charles Vardeman
Geospatial Ontologies, Vocamps, and Computational Science

88th OGC Technical Committee
Washington, DC
Charles F. Vardeman II
25 March 2014

Slide 2 Data and Software Preservation for Open Source

http://www.daspos.org

http://www.daspos.org

Slide 3 GeoVocamps

http://vocamp.org/wiki/Main_Page
GeoVocamps

- Vocamps have been active since 2008. GeoVocamps since 2011
- I have participated in:
  - GeoVoCamp Dayton 2012
  - GeoVoCamp SOCoP DC 2012
  - GeoVoCamp Santa Barbara 2013
  - GeoVoCamp DC 2013
  - Decartes-Core GeoVoCamp Santa Barbara 2014

http://vocamp.org/wiki/Main_Page

Slide 4 Acknowledgements

<table>
<thead>
<tr>
<th>Acknowledgements</th>
</tr>
</thead>
<tbody>
<tr>
<td>UND Center for Research Computing</td>
</tr>
<tr>
<td>Dr. Jarosław Nabrzyski</td>
</tr>
<tr>
<td>Dr. Christopher Sweet</td>
</tr>
<tr>
<td>UND School of Architecture</td>
</tr>
<tr>
<td>Prof. Aimee Buccellato</td>
</tr>
<tr>
<td>UND Computer Science and Engineering</td>
</tr>
<tr>
<td>Holly Ferguson</td>
</tr>
<tr>
<td>GeoVocamps—Material Transformation</td>
</tr>
<tr>
<td>Dr. Gary Berg-Cross (SOCoP)</td>
</tr>
<tr>
<td>Prof. Pascal Hitzler (Wright State)</td>
</tr>
<tr>
<td>Prof. Krzysztof Janowicz (UCSB)</td>
</tr>
<tr>
<td>Michelle Cheatham (Wright State)</td>
</tr>
<tr>
<td>Adila Krisnadni (Wright State)</td>
</tr>
<tr>
<td>Deborah MacPherson (Cannon Design)</td>
</tr>
<tr>
<td>Lamar Henderson (CSI LEED)</td>
</tr>
<tr>
<td>Torsten Hahmann (U of Maine)</td>
</tr>
</tbody>
</table>

Funding:
- UND Center for Sustainable Energy (cSEND)
- School of Architecture
- UND College of Arts and Letters
- UND Center for Research Computing

Ten Simple Rules for the Care and Feeding of Scientific Data

Source: [http://www.ploscompbiol.org/article/...l.pcbi.1003542](http://www.ploscompbiol.org/article/...l.pcbi.1003542) (PDF)

Article
Introduction

In the early 1600s, Galileo Galilei turned a telescope toward Jupiter. In his log book each night, he drew to-scale schematic diagrams of Jupiter and some oddly moving points of light near it. Galileo labeled each drawing with the date. Eventually he used his observations to conclude that the Earth orbits the Sun, just as the four Galilean moons orbit Jupiter. History shows Galileo to be much more than an astronomical hero, though. His clear and careful record keeping and publication style not only let Galileo understand the solar system, they continue to let anyone understand how Galileo did it. Galileo's notes directly integrated his data (drawings of Jupiter and its moons), key metadata (timing of each observation, weather, and telescope properties), and text (descriptions of methods, analysis, and conclusions). Critically, when Galileo included the information from those notes in Sidereus Nuncius [1], this integration of text, data, and metadata was preserved, as shown in Figure 1. Galileo's work advanced the “Scientific Revolution,” and his approach to observation and analysis contributed significantly to the shaping of today's modern “scientific method” [2] [3].
In these pages, Galilei combines data (drawings of Jupiter and its moons), key metadata (timing of each observation, weather, and telescope properties), and text (descriptions of methods, analysis, and conclusions).

doi:10.1371/journal.pcbi.1003542.g001

Today, most research projects are considered complete when a journal article based on the analysis has been written and published. The trouble is, unlike Galileo's report in *Sidereus Nuncius*, the amount of real data and data description in modern publications is almost never sufficient to repeat or even statistically verify a study being presented. Worse, researchers wishing to build upon and extend work presented in the literature often have trouble recovering data associated with an article after it has been published. More often than scientists would like to admit, they cannot even recover the data associated with their own published works.

Complicating the modern situation, the words “data” and “analysis” have a wider variety of definitions today than at the time of Galileo. Theoretical investigations can create large “data” sets through simulations (e.g., The Millennium Simulation Project: [http://www.mpa-garching.mpg.de/galform/virgo/millennium/](http://www.mpa-garching.mpg.de/galform/virgo/millennium/)). Large-scale data collection often takes place as a community-wide effort (e.g., The Human Genome project: [http://www.genome.gov/10001772](http://www.genome.gov/10001772)), which leads to gigantic online “databases” (organized collections of data). Computers are so essential in simulations, and in the processing of experimental and observational data, that it is also often hard to draw a dividing line between “data” and “analysis” (or “code”) when discussing the care and feeding of “data.” Sometimes, a copy of the code used to create or process data is so essential to the use of those data that the code should almost be thought of as part of the “metadata” description of the data. Other times, the code used in a scientific study is more separable from the data, but even then, many preservation and sharing principles apply to code just as well as they do to data.
So how do we go about caring for and feeding data? Extra work, no doubt, is associated with nurturing your data, but care up front will save time and increase insight later. Even though a growing number of researchers, especially in large collaborations, know that conducting research with sharing and reuse in mind is essential, it still requires a paradigm shift. Most people are still motivated by piling up publications and by getting to the next one as soon as possible. But, the more we scientists find ourselves wishing we had access to extant but now unfindable data [4], the more we will realize why bad data management is bad for science. How can we improve?

This article offers a short guide to the steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized. In just the past few years, hundreds of scholarly papers and reports have been written on questions of data sharing, data provenance, research reproducibility, licensing, attribution, privacy, and more—but our goal here is not to review that literature. Instead, we present a short guide intended for researchers who want to know why it is important to “care for and feed” data, with some practical advice on how to do that. The final section at the close of this work (Links to Useful Resources) offers links to the types of services referred to throughout the text. Boldface lettering below highlights actions one can take to follow the suggested rules.

**Rule 1. Love Your Data, and Help Others Love It, Too**

Data management is a repeat-play game. If you take care to make your data easily available to others, others are more likely to do the same—eventually. While we wait for this new sharing equilibrium to be reached, you can take two important actions. First, cherish, document, and publish your data, preferably using the robust methods described in Rule 2. Get started now, as better tools and resources for data management are becoming more numerous, universities and research communities are moving toward bigger investments in data repositories (Rule 8), and more librarians and scientists are learning data management skills (Rule 10). At the very least, loving your own available data will serve you: you’ll be able to find and reuse your own data if you treat them well. Second, enable and encourage others to cherish, document, and publish their data. If you are a research scientist, chances are that not only are you an author, but also a reviewer for a specialized journal or conference venue. As a reviewer, request that the authors of papers you review provide documentation and access to their data according to the rules set out in the remainder of this article. While institutional approaches are clearly essential (Rules 8 and 10), changing minds one scientist at a time is effective as well.

**Rule 2. Share Your Data Online, with a Permanent Identifier**

Nothing really lasts forever, so “permanent” actually just means long-lasting. For example, your personal web site is unlikely to be a good option for long-term data storage (yet, in the very short run, putting your data on your site is better than doing nothing at all!). In general, although many papers include URLs to give access to datasets, most become inaccessible within a few years [5]. The best option for releasing your data with long-term guarantee is to deposit them in whatever data archive is the “go to” place for your field. A proper, trustworthy archive will: (1) assign an identifier such as a “handle” (hdl) or “digital object identifier” (doi); (2) require that you provide adequate documentation and metadata; and (3) manage the “care and feeding” of your data by employing good curation practices. If no such archive exists in your field, there are also generic (non-domain-specific) online services that can host your data and issue persistent identifiers (see Rule 8). Pointers to a few generic repositories are listed in the Links to Useful Resources (section A), and longer compilations of such services are in the Links to Useful Resources (B).

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications

Updated: Sun, 19 May 2019 05:26:10 GMT

Powered by mindtouch®
Rule 3. Conduct Science with a Particular Level of Reuse in Mind

Data from others are hard to use without context describing what the data are and how they were obtained. The W3C Provenance Group (http://www.w3.org/TR/2013/REC-prov-dm-20130430/#dfn-provenance) defines information “provenance” as the sum of all of the processes, people (institutions or agents), and documents (data included!) that were involved in generating or otherwise influencing or delivering a piece of information. Perfect documentation of provenance is rarely, if ever, attained in scientific work today. The higher the quality of provenance information, the higher the chance of enabling data reuse. In general, data reuse is most possible when: 1) data; 2) metadata (information describing the data); and 3) information about the process of generating those data, such as code, are all provided. In trying to follow the rules listed in this article, you will do best if you plan in advance for ways to provide all three kinds of information. In carrying out your work, consider what level of reuse you realistically expect and plan accordingly. Do you want your work to be fully reproducible? If so, then provenance information is a must (e.g., working pipeline analysis code, a platform to run it on, and verifiable versions of the data). Or do you just want your work to be inspectable? If so, then intermediate data products and pseudo-code may be sufficient. Or maybe your goal is that your data is usable in a wide range of applications? If so, consider adopting standard formats and metadata standards early on. At the very least, keep careful track of versions of data and code, with associated dates. Taking these steps as you plan and carry out projects will earn you the thanks of researchers, including you, looking back from the future. (Consult the Links to Useful Resources [E] for a list of tools to package all your research materials with reuse in mind.)

Rule 4. Publish Workflow as Context

Publishing a description of your processing steps offers essential context for interpreting and reusing data. As such, scientists typically include a “methods” and/or “analysis” section(s) in a scholarly article, used to describe data collection, manipulation, and analysis processes. Computer and information scientists call the combination of the collection methods and analysis processes for a project its “workflow,” and they consider the information used and captured in the workflow to be part of the “provenance” of the data. In some cases (mostly in genomics), scientists can use existing workflow software in running experiments and in recording what was done in those experiments, e.g., Gene Pattern (http://www.genepattern.org). In that best-case scenario, the workflow software, its version, and settings used can be published alongside data using the other rules laid out here. But, it is rare outside of genomics to see the end-to-end process described in a research paper run, orchestrated, and/or recorded by a single software package. In a plausible utopian future, automated workflow documentation could extend to all fields, so that an electronic provenance record could link together all the pieces that led to a result: the data citation (Rule 2), the pointer to the code (Rule 6), the workflow (this rule), and a scholarly paper (Rule 5). But what can you do now? At a minimum, provide, alongside any deposit of data, a simple sketch of data flow across software, indicating how intermediate and final data products and results are generated. If it’s feasible and you are willing to deal with a higher level of complexity, also consider using an online service to encapsulate your workflow (see Links to Useful Resources[C] for a list of services). Keep in mind that even if the data used are not “new,” in that they come from a well-documented archive, it is still important to document the archive query that produced the data you used, along with all the operations you performed on the data after they were retrieved. Keeping better track of workflow, as context, will likely benefit you and your collaborators enough to justify the loftier, more altruistic, goals espoused here.
Rule 5. Link Your Data to Your Publications as Often as Possible

Whether your “data” include tables, spreadsheets, images, graphs, databases, and/or code, you should make as much of it as possible available with any paper that presents it. If it's practical and helpful, share your data as early as possible in your research workflow: as soon as you are done with the analysis, even before you write any articles about it. Your data can even be cited before (or without) its inclusion in a paper (see Rule 7). Many journals now offer standard ways to contribute data to their archives and link it to your paper, often with a persistent identifier. Whenever possible, embed citations (links) to your data and code, each with its own persistent identifier, right into the text of your paper, just like you would reference other literature. If a journal hosting your paper doesn't offer a place for your data, and/or an identifier for it, use a repository (Rule 8) and get your own identifier (Rule 2). At a minimum, you can post, and refer to, a package of files (data, codes, documentation on parameters, metadata, license information, and/or lists of links to such) with a persistent online identifier (Rule 2). And, if your domain's journals' policies do not allow for good data–literature interlinking, try to effect change (see Rules 1 and 10).

Rule 6. Publish Your Code (Even the Small Bits)

Did you write any code to run your analysis? No matter how buggy and insignificant you may find it, publish it. Many easy-to-use source code repositories exist, which allow not only hosting of software but also facilitate collaboration and version tracking (see Links to Useful Resources [D]). Your code, even the shortest script (whether or not you are proud of its quality), can be an important component for understanding your data and how you got your results [6]. Software plays several roles in relation to data and scientific research, and norms around its publication are still evolving and differ across disciplines [7]. In some cases, software is the primary data product (e.g., new algorithms). In other cases, data are the primary research products, yet the best way to document their provenance is to publish the software that was used to generate them as “metadata.” In both cases, publishing the source code and its version history is crucial to enhance transparency and reproducibility. The use of open-source software when possible reduces barriers for subsequent users of your software-related data products [8]. The same best practices discussed above in relation to data and workflow also apply to software materials: cite the software that you use and provide unique, persistent identifiers (Rule 2) to the code you share.

Rule 7. State How You Want to Get Credit

Chances are that you want to get credit for what you share. The attribution system used for scholarly articles, accomplished via citations, often breaks in the case of data and software. When other authors reuse or cite your data or code, you may get an acknowledgment or an incoming link. If you and your colleagues have gone to the trouble to write a “data paper,” whose main purpose is to describe your data and/or code, you may also get a citation [9]. But, “data paper” writing is not always desirable, or relevant. So, how do you go about getting the full credit you deserve for your data and code? The best way is to simply describe your expectations on how you would like to be acknowledged. If you want, you can also release your data under a license and indicate explicitly in the paper or in the metadata how you want others to give you credit. But, while legal mechanisms have advantages, they can also inadvertently lead to limitations on the reuse of the data you are sharing. In any case, make information about you (e.g., your name, institution), about the data and/or code (e.g., origin, version, associated files, and metadata), and about exactly how you would like to get credit, as clear as possible. Easy-to-implement licenses, many of which offer the advantage of being machine-readable, are offered by the Creative Commons organization.
Rule 8. Foster and Use Data Repositories

Sometimes the hardest and most time-consuming step of sharing data and code is finding and deciding where to put them. Data-sharing practices vary widely across disciplines: in some fields data sharing and reuse are essential and commonplace, while in others data sharing is a "gift exchange" culture. If your community already has a standard repository, use it. If you don't know where to start looking, or you need help choosing among relevant repositories, ask an information specialist, such as a data scientist or a librarian working in your field (and consult the directories of data repositories listed in the Links to Useful Resources [B]). When choosing among repositories, try to find the one offering the best combination of ease-of-deposit, community uptake, accessibility, discoverability, value-added curation, preservation infrastructure, organizational persistence, and support for the data formats and standards you use. Remember that even if your field has no domain-based repository, your institution may have one, and your local librarian or archivist can instruct you on how to use that local resource. If neither your community nor your institution has a relevant repository, try a generic repository or consider setting up your own (see Rule 2, and Links to Useful Resources [F]).

Rule 9. Reward Colleagues Who Share Their Data Properly

Whether you do it in person at scientific meetings and conferences or by written communication when reviewing papers and grants, reward your colleagues who share data and code. Rally your colleagues and engage your community by providing feedback on the quality of the data assets in your field. Praise those following the best practices. The more the data created by your colleagues is accessible as an organized collection of some sort, the better your community's research capacity. The more data get shared, used, and cited, the more they improve. Besides personal involvement and encouragement, the best way to reward data sharing is by attribution: always cite the sources of data that you use. Follow good scientific practice and give credit to those whose data you use, following their preferred reference format and according to current best practices. Standards and practices for citing and attributing data sources are actively being developed through international partnerships.

Rule 10. Be a Booster for Data Science

As Rule 1 says, it is important not just that you love your own data, but that others love data, too. An attitude that data and code are "second-class objects," behind traditional scholarly publications, is still prevalent. But, every day, as scientists try to use the frustrating but tantalizing hodgepodge of research data available via the present ad hoc network of online systems, the value of organizing an open network of reusable data and code is becoming more and more clear, to more and more people. You, as a scientist, need to help organize your discipline and your institution to move more quickly toward a world of open, discoverable, reproducible data and research. One important step is to advocate for hiring data specialists and for the overall support of institutional programs that improve data sharing. Make sure not only advanced researchers (e.g., postdocs) experience the pleasures of doing research with freely available data and tools: explain and show the value of well-loved data to graduate and undergraduate researchers. Teach whole courses, or mini-courses, related to caring for data and software, or incorporate the ideas into existing courses. Form groups specific to your discipline to foster data and code sharing. Hold birds-of-a-feather or special sessions during large meetings demonstrating examples in which good sharing practices have led to better results and collaborations. Lead by practicing what you preach.
Links to Useful Resources

A: General Data Repositories

• Dataverse (http://thedata.org): A repository for research data that takes care of long-term preservation and good archival practices, while researchers can share, keep control of, and get recognition for their data.

• FigShare (http://figshare.com): A repository where users can make all of their research outputs available in a citable, shareable, and discoverable manner.

• Zenodo (http://zenodo.org): A repository service that enables researchers, scientists, projects, and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of existing institutional or subject-based repositories.

• Dryad (http://datadryad.org): A repository that aims to make data archiving as simple and as rewarding as possible through a suite of services not necessarily provided by publishers or institutional websites.

B: Directories of Research Data Repositories

• DataBib (http://databib.org): Databib is a tool for helping people identify and locate online repositories of research data. Users and bibliographers create and curate records that describe data repositories that users can search.

• Re3data.org (http://www.re3data.org): Re3data is a global registry of research data repositories from different academic disciplines for researchers, funding bodies, publishers, and scholarly institutions.

• Open Access Directory (http://oad.simmons.edu/oadwiki/Data_repositories): A list of repositories and databases for open data.

• Force 11 Catalog (http://www.force11.org/catalog): A dynamic inventory of web-based scholarly resources, a collection of alternative publication systems, databases, organizations and groups, software, services, standards, formats, and training tools.

C: Workflow Management Systems

• Taverna (http://www.taverna.org.uk): An open-source and domain-independent workflow management system—a suite of tools used to design and execute scientific workflows and aid in silico experimentation.

• Kepler (https://kepler-project.org): Software designed to help scientists, analysts, and computer programmers create, execute, and share models and analyses across a broad range of scientific and engineering disciplines.

• Wings (http://www.wings-workflows.org): A semantic workflow system that assists scientists with the design of computational experiments.

• VisTrails (http://www.vistrails.org): An open-source scientific workflow and provenance management system that supports data exploration and visualization.

• Knime (http://www.knime.org): A graphical workbench for the entire analysis process: data access, data transformation, initial investigation, powerful predictive analytics, visualization, and reporting.

D: Source Code Repositories

• Github (http://github.com): A web-based hosting service for software development projects that use the Git revision control system, including many open-source projects.

• Git (http://git-scm.com): A free and open-source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.
• Mercurial ([http://mercurial.selenic.com](http://mercurial.selenic.com)): A free, distributed source control management tool. It efficiently handles projects of any size and offers an easy and intuitive interface.

• BitBucket ([https://bitbucket.org](https://bitbucket.org)): A web-based hosting service for projects that use either the Mercurial or Git revision control systems.

E: Systems to Package, Access, and Execute Data and Code

• IPython Notebook ([http://ipython.org/notebook.html](http://ipython.org/notebook.html)): A web-based interactive computational environment where you can combine code execution, text, mathematics, plots, and rich media into a single document.

• ROpenSci ([http://ropensci.org](http://ropensci.org)): A suite of packages that allow access to data repositories through the R statistical programming environment.

• Authorea ([https://authorea.com](https://authorea.com)): A collaborative online word processor for scholarly papers that allows the writing of web-native, living, dynamic, "executable" articles that include text, mathematical notation, images, and data. It currently supports inclusion and rendering of d3.js plots and IPython notebooks.

• Dexy ([http://dexy.it](http://dexy.it)): A multipurpose project automation tool for working with documents via a command-line interface.

F: Software Tools to Run Your Own Document Repository

• Invenio ([http://invenio-software.org](http://invenio-software.org)): Invenio is a free software suite enabling you to run your own digital library or document repository on the web. Invenio is an ideal solution for running document repositories of moderate to large sizes (several millions of records). Invenio is codeveloped by CERN, DESY, EPFL, FNAL, and SLAC.

• Eprints ([http://www.eprints.org/software](http://www.eprints.org/software)): EPrints is one of the easiest and fastest ways to set up small to medium-sized repositories of open-access research literature, scientific data, theses, reports, and multimedia. Developed at the University of Southampton, UK.

• DSpace ([http://www.dspace.org](http://www.dspace.org)): DSpace is a turnkey institutional repository application developed by the Duraspace organization.

G: Licensing and Privacy

• Open Source Initiative ([http://opensource.org/licenses](http://opensource.org/licenses)): Open-source licenses are licenses that comply with the Open Source Definition: they allow software to be freely used, modified, and shared. These include Apache, BSD, GNU (GPL), MIT, and the Mozilla Public License.

• Privacy Tools for Sharing Research Data ([http://privacytools.seas.harvard.edu](http://privacytools.seas.harvard.edu)): A Harvard-based collaborative and multidisciplinary effort to help enable the collection, analysis, and sharing of personal data for research in social science and other fields while providing privacy for individual subjects.

Acknowledgments

This article is an outcome of an Exploratory Seminar called “What to Keep and How to Analyze It: Data Curation and Data Analysis with Multiple Phases” ([http://www.radcliffe.harvard.edu/exploratory-seminars/what-to-keep-and-how-to-analyze-it](http://www.radcliffe.harvard.edu/exploratory-seminars/what-to-keep-and-how-to-analyze-it)) organized by Xiao-Li Meng and Alyssa Goodman, held on May 9–10, 2013 at the Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Massachusetts. This article was written collaboratively, online, in the open, using Authorea. Every iteration of the writing procedure is logged and available in the online version of this article at [https://www.authorea.com/3410](https://www.authorea.com/3410).
References

   View Article
   PubMed/NCBI
   Google Scholar


   View Article
   PubMed/NCBI
   Google Scholar

   View Article
   PubMed/NCBI


11.

12.

About the Authors

Authors
Alyssa Goodman, Alberto Pepe, Alexander W. Blocker, Merce Crosas, Rosanne Di Stefano, Vinay Kashyap, Aneta Siemiginowska
Harvard University, Cambridge, Massachusetts, United States of America

Christine L. Borgman
University of California, Los Angeles, Los Angeles, California, United States of America

Kyle Cranmer, David W. Hogg
New York University, New York, New York, United States of America

Yolanda Gil
University of Southern California, Los Angeles, Los Angeles, California, United States of America

Paul Groth
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Margaret Hedstrom
University of Michigan, Ann Arbor, Michigan, United States of America

Ashish Mahabal
California Institute of Technology, Pasadena, California, United States of America

Aleksandra Slavkovic
Pennsylvania State University, State College, Pennsylvania, United States of America

Corresponding Author
Email: alberto.pepe@gmail.com

Competing Interests

The authors have declared that no competing interests exist.
Metrics

Source: http://www.ploscompbiol.org/article/...l.pcbi.1003542

Viewed

Total Article Views
12,472
Apr 24, 2014 (publication date)
through Jun 16, 2014*

<table>
<thead>
<tr>
<th></th>
<th>HTML Page Views</th>
<th>PDF Downloads</th>
<th>XML Downloads</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLOS</td>
<td>11,253</td>
<td>1,175</td>
<td>44</td>
<td>12,472</td>
</tr>
<tr>
<td>PMC</td>
<td>0</td>
<td>0</td>
<td>n.a.</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>11,253</td>
<td>1,175</td>
<td>44</td>
<td>12,472</td>
</tr>
</tbody>
</table>

10.44% of article views led to PDF downloads

MISSING GRAPHIC

*Although we update our data on a daily basis, there may be a 48-hour delay before the most recent numbers are available. PMC data is posted on a monthly basis and will be made available once received.

6 figshare
6

Cited

1 CrossRef
1

Search GoogleScholar
Search

Saved

15 CiteULike
15
69

Discussed

288 Twitter
288

115 Facebook
115

1 Comments and Notes
1
Comments

Source: http://www.ploscompbiol.org/article/...l.pcbi.1003542

None

Related Content

Source: http://www.ploscompbiol.org/article/...l.pcbi.1003542

Media Coverage of this Article

None

More by these Authors

- 13257 VIEWS
  11 CITATIONS

  How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations
  Xin Shuai, Alberto Pepe, Johan Bollen
  Research Article | 01 Nov 2012 | PLoS ONE
  doi:10.1371/journal.pone.0047523

- 2454 VIEWS
  1 CITATION

  A Measure of Total Research Impact Independent of Time and Discipline
  Alberto Pepe, Michael J. Kurtz
  Research Article | 07 Nov 2012 | PLoS ONE
  doi:10.1371/journal.pone.0046428

Related Authors

- Alyssa Goodman
- Alberto Pepe
- Alexander W. Blocker
- Aneta Siemiginowska
- Aleksandra Slavkovic
Community-Developed Geoscience Cyberinfrastructure


**Introduction**

Discoveries in the geosciences are increasingly taking place across traditional disciplinary boundaries. The EarthCube program, a community-driven project supported by the U.S. National Science Foundation, is developing an information- and tool-sharing framework to bridge between disciplines and unlock the modern geosciences' transformative potential.

The program not only is developing cyberinfrastructure for the geosciences but also is developing and testing new processes for the community that could fundamentally transform its culture. If successful, the project could outline a path for the development, deployment, and adoption of digital capabilities across the wider scientific community. The development of EarthCube and similar projects could result in changes to the practice of science comparable in speed and breadth to the spread of the Internet or other basic infrastructures.

More than 2500 Earth, atmosphere, ocean, computer, information, and social scientists, as well as educators, data managers, and others, have engaged in elucidating the EarthCube concept. This document is a synthesis of the work through August 2013 and a snapshot of an emergent end user vision for EarthCube—a vision that will continue to evolve as EarthCube grows.

**The EarthCube Vision**

The EarthCube program is a community-driven project focused on building digital infrastructure for managing, sharing, and exploring geoscience data and information with the aim of increasing scientific productivity.

EarthCube aims to foster new, transformational research by enabling scientists to gain access to formerly unconnected software, models, data, and computational resources. Science scenarios developed in conjunction with the community at end user workshops will shape EarthCube’s design to ensure that it provides concrete value for geoscientists and addresses the diverse needs of the geoscience community [see EarthCube Test Governance Project Team, 2013]. Seventeen such workshops were convened from 2012 to 2013, and the next community review of progress thus far will occur at an EarthCube All-Hands Meeting in June 2014.

The project’s goal is to design, build, and maintain an easy-to-use system based on existing resources that embraces open-source culture and methods to align technology development with scientific needs. Perhaps its greatest challenges lie in identifying key capabilities that are widely useful and in focusing efforts on implementing those capabilities without becoming so complex that it discourages use.

Ideally, the EarthCube system will constitute an integral part of everyday research and decision-making workflows, coordinating hardware, software, people, processes, data, and community. A significant part of the EarthCube system will be intangible, including specifications, policies, protocols, and communities of practice, in many ways comparable to the World Wide Web.
EarthCube Workflow

EarthCube will be organized around specific workflow activities.

**Data Management.** EarthCube will provide a shared archive in which data, tools, and services are documented and curated, enabling reuse of data sets for new analyses. The project will embrace technology that simplifies data curation for archive, publication, and reuse and that can be applied to newly acquired data as well as existing and old data sets.

**Resource Discovery.** Plug-in components will enable resource discovery and direct data access using scientific software in common use, such as Excel, MATLAB, Python, R, ArcGIS, or ModFlow. Such search plug-ins will be Google-like in simplicity, with text or map-based interfaces, while being efficient and accurate. Alternatively, Web services will provide machine access for searching catalogs and real-time data use.

**Data Access, Integration, and Processing Tools.** User interfaces tailored for specific communities will simplify data access, visualization, and analysis using software that interoperates with EarthCube data or service providers. Data will be linked to computational capabilities for display, exploration, real-time use, and analysis using data from different sources. Through EarthCube, it will be easy to pipe results between processes to define reproducible workflows. Real-time data sources will be used to improve event-response procedures, optimize sampling, and facilitate scientific analysis.

**Data Portal.** One or more portals will function as user entry points to support data exploration and access tailored for specific communities. These portals might be organized around various paradigms, but one that has been frequently mentioned is a three-dimensional virtual globe for data discovery and exploration, supporting the ability to spatially integrate and display geoscientific data at varying resolutions.

Technical Challenges

The dominant challenge for EarthCube is to make data management and processing easier and less time-consuming. This challenge is a large one given that the scope of EarthCube includes not only data but also models, workflows, samples, and tools.

To be used by other researchers, data must be understood sufficiently well to be trusted and processed appropriately. This presents a wide variety of technical challenges involved in building workflows for data management, including curation, documentation, access, and integration.

For example, many legacy data issues stem from the difficulty that individual researchers, operating on limited budgets, experience in trying to curate data produced by their research. As a result, data documentation is commonly insufficient to enable cross-domain use or to repurpose data obtained from repositories. In addition, using nonstandard, heterogeneous data requires significant effort. The meaning of data may be unclear because of nonstandard vocabulary usage. Inconsistent practices for data sharing make each new data acquisition a time-consuming learning experience.

EarthCube will help address these issues by encouraging the development and adoption of community standards for Web interfaces to data, metadata and data formats, and software libraries.
Challenges of Governance and Culture

The goal for EarthCube is to continue innovating while keeping scientific productivity high. To accomplish this goal, EarthCube will need to overcome challenges related to community governance, incentives, development of communities of practice, and education of current and future generations of Earth scientists.

Community representatives involved with EarthCube governance will need to coordinate programs and pilot projects, develop metrics to evaluate system components, identify capability gaps, and promote consensus on requirements. They will also need to develop a portfolio of EarthCube policies and specifications as well as set pathways for collecting, monitoring, and acting on community feedback to refine content, practices, and policies. The community will have to establish priorities for building cyberinfrastructure, making legacy data accessible, and developing standards. Recommendations for funding will need to be based on these priorities, along with usage metrics, ongoing gap analysis, and user requirements.

Social, professional, and financial incentives are necessary to motivate good data management practices, along with the sharing of data, software, or models. Data quality metrics will need to be built into the system, along with practices to minimize data misinterpretation or misuse. Access controls and respect for data ownership are necessary to ensure that data are not shared too soon or inappropriately. Adequate credit must be given for contributing data and models—its absence would be a major deterrent to participation. Cost, effort, and technical barriers, along with concerns about misuse, must not outweigh the tangible benefits of time spent on data management and publication, especially for tenure-track and project-funded researchers.

EarthCube represents a significant shift in culture and will bring new challenges to the geosciences community. To ease the transition, the program has emphasized the involvement of early-career researchers, who will be the vanguard in changing research workflows. Success will be indicated when the geoscience community identifies itself through participation in EarthCube as data providers, data consumers, system developers, maintainers, and managers.

EarthCube Beyond Research

EarthCube can play a major role in training the next generation of cyber-savvy geoscientists by providing intuitive, modular learning objects and self-directed lessons that can be used by teachers from K-12 through to the graduate level. In particular, an EarthCube three-dimensional virtual globe will not only be a data discovery portal for researchers but will also serve as an entry point for students to explore geoscience data. Success will be achieved when EarthCube helps everyone—including individuals outside of the geosciences—better understand how to use and interpret science data.

References

Welcome

Welcome to the EarthCube community workspace. Anyone with an interest in EarthCube is free to use this site to coordinate events, draft and share documents, organize meetings, and stay up to date on EarthCube activities. This workspace will be continually developed and optimized to meet the changing needs of EarthCube work groups.

All Hands Meeting

Join us as we review the draft demonstration charter for EarthCube, receive updates on multiple EarthCube projects, assess their tangible results, and share other work related to cyberinfrastructure and data management in the geosciences. This is an opportunity to help build the organization you’d like to inhabit.

The EarthCube All-Hands Meeting will bring together project institutions, partners, collaborators, and scientists from across the globe to share their progress and experience with EarthCube thus far, and discuss and plan activities for the upcoming year. This meeting will be held from June 24-26, 2014 at the Renaissance Dupont Circle hotel in Washington, D.C. To assist with travel planning, this meeting will convene at 8 am Tuesday and close at 5 pm Thursday.

This event is designed in the collaborative spirit of EarthCube and will provide multiple opportunities for networking and meaningful work, as well as the chance to share your efforts and to learn from others.

We are pleased to invite proposals for sessions for the All-Hands Meeting. Proposed sessions should relate to EarthCube, the state of cyberinfrastructure in the geosciences, and innovative geoscience contributions to data management. Sessions in multiple formats will be considered, including:
- Workshops or Hack-a-thons
- Technology Presentations
- Working Groups & Business (Project) Meetings
- Panel discussions (Plenary or Breakout)
- Presentations

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00 AM</td>
<td>EarthCube 101 Breakfast</td>
</tr>
<tr>
<td></td>
<td>Dining Room</td>
</tr>
<tr>
<td>7:30 AM</td>
<td>Registration Open</td>
</tr>
<tr>
<td></td>
<td>Lower Level Foyer</td>
</tr>
<tr>
<td>8:00 AM</td>
<td>Opening Remarks</td>
</tr>
<tr>
<td></td>
<td>Welcome and Introductions: Eva Zanzerkia (National Science Foundation) and Lee Allison (EarthCube Test Governance Project)</td>
</tr>
<tr>
<td></td>
<td>Keynote Speaker: Paul N. Edwards, University of Michigan</td>
</tr>
<tr>
<td></td>
<td>New Hampshire</td>
</tr>
<tr>
<td>9:30 AM</td>
<td>Break</td>
</tr>
<tr>
<td></td>
<td>Lower Level Foyer</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>Tackling the Tough Issues Facing EarthCube in 2014-2015</td>
</tr>
<tr>
<td></td>
<td>[Facilitators: Cutcher-Gershenfeld, Lynn]</td>
</tr>
<tr>
<td></td>
<td>New Hampshire</td>
</tr>
<tr>
<td>12:30 PM</td>
<td>Lunch</td>
</tr>
<tr>
<td></td>
<td>City Center Ballroom</td>
</tr>
<tr>
<td>2:00 PM</td>
<td>EarthCube Architecture Forum, Part 1</td>
</tr>
<tr>
<td></td>
<td>[Diggs, Richard, Percivall], [Facilitator: Jewlya Lynn]</td>
</tr>
<tr>
<td></td>
<td>New Hampshire</td>
</tr>
<tr>
<td>3:30 PM</td>
<td>Break</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>4:00:00 PM</td>
<td>Dynamic Governance: Preparing for the tough issues facing EarthCube in 2014-2015</td>
</tr>
<tr>
<td></td>
<td>Bcube Brokering Framework</td>
</tr>
<tr>
<td></td>
<td>Prototype Activities Work Group Business Meeting</td>
</tr>
<tr>
<td></td>
<td>Software Sustainability in the Geosciences</td>
</tr>
<tr>
<td></td>
<td>Birds of a Feather Session</td>
</tr>
<tr>
<td></td>
<td>[Cutcher-Gershenfeld, Lynn]</td>
</tr>
<tr>
<td></td>
<td>[Khalsa]</td>
</tr>
<tr>
<td></td>
<td>[Diggs]</td>
</tr>
<tr>
<td></td>
<td>[Gil]</td>
</tr>
<tr>
<td></td>
<td>New Hampshire Potomac Foggy Bottom Dupont Dining Room</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:30 PM</td>
<td>Break</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:00 PM</td>
<td>Poster Session &amp; Reception</td>
</tr>
<tr>
<td></td>
<td>Opening Speaker: Irene Qualters, NSF Division for Advanced Cyberinfrastructure, Director</td>
</tr>
<tr>
<td></td>
<td>City Center Ballroom</td>
</tr>
</tbody>
</table>

### Day 2 - Wednesday, June 25

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:15</td>
<td>Registration</td>
</tr>
<tr>
<td></td>
<td>Lower Level Foyer</td>
</tr>
<tr>
<td>8:00</td>
<td>Welcome, Day 2 Goals, and Selecting EarthCube's Priorities for 2014-2015</td>
</tr>
<tr>
<td></td>
<td>New Hampshire</td>
</tr>
<tr>
<td>9:30</td>
<td>Break</td>
</tr>
<tr>
<td></td>
<td>Lower Level Foyer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00</td>
<td>First Steps on EarthCube's Priorities for 2014-2015</td>
</tr>
<tr>
<td></td>
<td>EarthCube Architecture Forum, Part 2</td>
</tr>
<tr>
<td></td>
<td>[Facilitators: Cutcher-Gershenfeld, Kahn]</td>
</tr>
<tr>
<td></td>
<td>[Facilitator: Lynn]</td>
</tr>
<tr>
<td></td>
<td>New Hampshire Potomac</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>12:30</td>
<td>Lunch</td>
</tr>
<tr>
<td></td>
<td>Workshops &amp; Technology Demonstrations, Rooms TBA</td>
</tr>
<tr>
<td>2:00</td>
<td>Birds of a Feather Session</td>
</tr>
<tr>
<td></td>
<td>Architecture Forum, pt. 3</td>
</tr>
<tr>
<td></td>
<td>Social Dimensions of EarthCube data and infrastructure</td>
</tr>
<tr>
<td></td>
<td>Workforce Development Panel Discussion</td>
</tr>
<tr>
<td></td>
<td>Council of Data Facilities Business Meeting</td>
</tr>
<tr>
<td></td>
<td>New Hampshire** Potomac Foggy Bottom Dining Room Mt. Vernon</td>
</tr>
<tr>
<td>3:30</td>
<td>Break</td>
</tr>
<tr>
<td></td>
<td>Lower Level Foyer</td>
</tr>
<tr>
<td>4:00</td>
<td>Addressing Data Heterogeneity: Semantic Building Blocks &amp; CI</td>
</tr>
<tr>
<td></td>
<td>Perspective</td>
</tr>
<tr>
<td></td>
<td>Inventory of EarthCube Resources: CINERGI Architecture and Workflow</td>
</tr>
<tr>
<td></td>
<td>DisConBB meets IOOS marine in-situ data (Business Mtg)</td>
</tr>
<tr>
<td></td>
<td>Revising the Governance Model, Preparing for Day 3</td>
</tr>
<tr>
<td></td>
<td>Council of Data Facilities Business Mtg</td>
</tr>
<tr>
<td></td>
<td>New Hampshire** Potomac Foggy Bottom Dining Room Mt. Vernon</td>
</tr>
<tr>
<td>5:30</td>
<td>End of Day 2</td>
</tr>
<tr>
<td>5:45</td>
<td>Ocean Link Business Meeting</td>
</tr>
<tr>
<td></td>
<td>Software Stewardship Team Business Meeting</td>
</tr>
<tr>
<td></td>
<td>DisConBB meets IOOS marine in-situ data Business Mtg</td>
</tr>
<tr>
<td></td>
<td>Test Governance Project Advisory Committee Mtg</td>
</tr>
<tr>
<td></td>
<td>GeoWS, CINERGI, Bcube Business Mtg</td>
</tr>
<tr>
<td></td>
<td>Closed</td>
</tr>
<tr>
<td>Time</td>
<td>Location</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
</tr>
<tr>
<td>7:00 PM</td>
<td></td>
</tr>
<tr>
<td>7:30</td>
<td>Lower Level Foyer</td>
</tr>
<tr>
<td>8:30</td>
<td>New Hampshire 1-2</td>
</tr>
<tr>
<td>9:15</td>
<td>New Hampshire 1-2</td>
</tr>
<tr>
<td>10:00</td>
<td>Lower Level Foyer</td>
</tr>
<tr>
<td>10:30</td>
<td>New Hampshire 1</td>
</tr>
<tr>
<td></td>
<td>New Hampshire 2</td>
</tr>
<tr>
<td></td>
<td>Potomac</td>
</tr>
<tr>
<td></td>
<td>Dupont</td>
</tr>
<tr>
<td></td>
<td>Mt. Vernon</td>
</tr>
<tr>
<td></td>
<td>Foggy Bottom</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1:30</td>
<td>Emergent Working Group Session</td>
</tr>
<tr>
<td></td>
<td>Emergent Working Group Session</td>
</tr>
<tr>
<td></td>
<td>BCube Project Discussion (1:30-2:00)</td>
</tr>
<tr>
<td></td>
<td>Emergent Working Group Session</td>
</tr>
</tbody>
</table>
Session Descriptions

DisConBB meets IOOS marine in-situ data

Convener: Arctur, David
Co-Conveners: Emilio Mayorga (University of Washington), Ethan Davis (Unidata), Alva Couch (Tufts University)

Description: The Building Block on Integrating Discrete and Continuous Data (DisConBB) is developing a common information model and tools for converting between gridded and time series data, initially for the atmospheric and hydrologic domains. This session is for discussion to develop and agree on this common information model and tools for working with WaterML 2 and netCDF/CF, using OGC SOS, THREDDS/OPeNDAP and related web services. This supports DisConBB’s end goal to have something that can be adapted for other domains, such as oceans, solid earth, and cryosphere. Will include discussion on existing, relevant efforts that can inform DisConBB’s solution, including the Unidata Common Data Model Discrete Sampling Geometry (CDM DSG) feature types and the IOOS SOS / SWECommon profile and associated server software implementations that bridge the netCDF/CF/CDM DSG/THREDDS and OGC SOS/O&M/SWECommon/SensorML approaches.

A key issue is that while there are a plethora of options for “mapping data” between formats, the tool chains we have available for analyzing data in the new formats don’t necessarily interoperate well with the mapped data. We are interested in how these mappings would be used, in what tools, and for what outcome. One large issue is how to generate NetCDF files of WaterML data that can then be applied in compelling use cases.

Addressing Data Heterogeneity in EarthCube Information: A Semantic Building Blocks & Cyberinfrastructure Perspective

Convener: Berg-Cross, Gary (SOCoP)
Co-Conveners: Pascal Hitzler (Wright State University), Kerstin Lehnert (Columbia University), Peter Wiebe (WHOI)

Description: The EarthCube (EC) community is currently engaged in a range of activities including creation of building blocks to handle Big Data issues and meet critical needs to find, organize, interpret, and share appropriate data & resources for specific research activities across EC disciplines. Current sustainable cyberinfrastructure employs a wide assortment of traditional, but useful standards, methods, and tools. It also is starting to include aspects of the still...
maturing and promising set of tools, methods, and standards associated with Semantic Technologies. These semantic initiatives, which are being incorporated into traditional infrastructure along with supporting practices, are aimed at dealing more effectively with stubborn, central challenges that arise from the inherent complexity and heterogeneity of data that is structured and unstructured, and based on independently developed domain models. Such direct, but tractable approaches developed to address semantic interoperability, are aimed at enabling better data integration, analysis, and dissemination to support cross-disciplinary research.

This panel will provide an opportunity to review & share current progress and experience with semantic technology approaches, including linked data, the use of lightweight, opportunistic methods, bottom-up & top-down approaches, and related infrastructure. Taken together, how important are semantic building blocks and semantic technology drivers for addressing Big Data variety challenges and the handling of diverse data and infrastructure issues? What are measures of success and risk? The panel discussion is expected to cover new opportunities and issues such as: schema mapping, approaches to semantic annotation and terminology, selection of appropriate knowledge representations, development and use of ontologies and modular ontology patterns, and integration of semantic technology and its methods with traditional technology. A question for the panel: what extant work and approaches can be leveraged and how might this be done as part of plans going forward with the EarthCube community, its Research Coordination Networks, and building blocks efforts?

Workforce Development: Panel Discussion

(To be followed by a related working group session.)
Convener: Downs, Robert (CIESIN, Columbia University)
Co-Conveners: Robert S. Chen (CIESIN, Columbia University), Leslie Hsu (LDEO, Columbia University), Kerstin Lehnert (LDEO, Columbia University), Erin Robinson (Foundation for Earth Science), Ilya Zaslavsky (University of California, San Diego).

Description: The EarthCube community is faced with various challenges, opportunities, and choices for enabling Earth science. Technology is evolving rapidly and new tools and techniques are being created and enhanced to enable community members to do more individually and collaboratively. Workforce development opportunities can assist community members in their efforts to improve their capabilities and knowledge. A panel will describe workforce development opportunities that are available for the Earth science community and engage in a discussion with the audience on how the EarthCube community can improve by leveraging these opportunities.

Software Sustainability in Geosciences

Convener: Gil, Yolanda (University of Southern California)
Co-Conveners: Christopher J. Duffy (Department of Civil and Environmental Engineering, Penn State University), Chris Mattmann (Department of Computer Science, University of Southern California and NASA/JPL), Scott D. Peckham (Department of Hydrologic Sciences, University of Colorado), and Erin Robinson (Foundation for Earth Science).

Description: This panel will discuss new findings regarding the requirements of geoscientists to manage their software, best practices for software sharing and reuse, and important social considerations surrounding scientific software.
Aligning Representations: Practical Explorations of Metadata, Ontologies, and Semantic Web for EarthCube

Convener: Hitzler, Pascal (Wright State University)
Co-Conveners: Joel Cutcher-Gershenfeld (University of Illinois at Urbana Champaign), Yolanda Gil (University of Southern California), Krzysztof Janowicz (University of California, Santa Barbara), Scott Peckham (University of Colorado, Boulder), Steve Richard (Arizona Geological Survey), and Ilya Zaslavsky (University of California San Diego).

Description: Metadata, ontologies, and conceptual models for representing information play a role across many current EarthCube efforts. This shared representational theme thus has the potential to serve as a catalyst for technology integration across EarthCube, by starting to discuss possibilities for aligning these representations, or for mapping between them.

This workshop is proposed by representatives from several current EarthCube awards and the idea rose out of corresponding discussions at the EarthCube Awards Assembly in February 2014. The workshop shall map out the representational choices made by different technical approaches within EarthCube. Hands-on, workshop participants will identify practical technological solutions for obtaining alignment or mappings between representations at EarthCube scale, and will identify major technological or social obstacles which may prevent such alignment or mapping. Given enough time, the workshop participants will jointly attempt to sketch concrete technological solutions for metadata, ontologies, and conceptual models which would support the EarthCube vision.

Duration of the workshop would be a full day, but advances could also be made in a half day. The workshop would start with briefings on current representational choices made by EarthCube awards, focusing on design choices and their underlying rationales. Small breakout groups would then work on possible mappings or alignments between a subset of the presented representations. Results from these breakout groups would be reported to all participants, groups would be further combined into larger groups to incorporate further representational choices in the discussion. A plenary session would combine and record the insights made and resulting recommendations.

Data Assimilation Frameworks in Geosciences

Convener: Kelbert, Anna (Oregon State University)

Description: The ultimate goal of any modeling in geosciences is gaining a better understanding of the physical, chemical and biological processes going on in the Earth's interior and environment. This understanding is best attained through a comparison of the modeled results with measured data. A formalization of the procedure, in which measurements are driving the modeling to ensure compatibility, is known as data assimilation. Since most Earth processes are tightly interconnected, multi-physics data assimilation may sometimes be desirable to get the most of the data we have, integrating all measurable information into a merged 4D snapshot of the Earth.

I would like to use this session to get a better understanding of the following issues.

1) Which areas of geosciences would benefit from more streamlined, multi-physics data assimilation?
2) Is a general and easy to use framework for multi-physics data assimilation practical?
3) What is our community experience so far with data assimilation frameworks?
4) What are the scientific and technical challenges, and (how) can they be overcome?
Finally, I’d like to get a feel for the community consensus on the role of multidisciplinary data assimilation in Earthcube, as the program evolves.

**Middleware Governance**

Convener: Khalsa, Siri Jodha (University of Colorado)
Co-Conveners: Stefano Nativi (CNR), Jay Pearlman (J&F Enterprises)

Description: Middleware can significantly simplify creation and operation of distributed systems, as well as providing a much more efficient means of integrating legacy systems with new technology. For example, brokering middleware provides mediation and transformation services to simplify data discovery, evaluation, access and use, working as a third-party tier in a three-tier architecture (extending the Client-Server paradigm). This introduces the clear need to govern and manage a middleware tier from many perspectives such as interface, stability and sustainability.

Effective middleware governance, as part of an overall governance approach, has the potential to support longer-term development under a variety of funding models, to simplify and standardize access models and assist in establishing a basis for development and operation of an EarthCube information system. It is not, however, clear what the best practices for this governance are and how those practices shift in response to different funding and ownership models, under different architectures or as standards change. To ensure sustainable, stable development, an effective model for the governance and reuse of such middleware must be agreed upon. This will be of value not only to interoperability architects and to developers (who can plan integrated systems assuming the continued use and support of middleware) but also to system managers and end users. We propose to consider this in a session on middleware governance at the EarthCube All Hands meeting, in collaboration with the Test Governance team.

**The BCube Brokering Framework**

Convener: Khalsa, Siri Jodha (University of Colorado)
Co-Conveners: Stefano Nativi (CNR), Ruth Duerr (NSIDC), Steve Browdy (OMS Tech), Jay & Francoise Pearlman (J&F Enterprises), Oscar Schofield (Rutgers), Scott Doney (WHOI)

Description: We will demonstrate the data and services provided by the BCube Brokering Framework in support of its science scenarios. We will show how the resources of the Brokering Framework, which is hosted in the Cloud, can be accessed via disciplinary portals or desktop tools, using standard web protocols or platform independent APIs.

**Inventory of EarthCube Resources: CINERGI Architecture and Workflow**

Convener: Tanu Malik (University of Chicago)
Co-Conveners: Ilya Zaslavsky (University of California San Diego), Steve Richard (Arizona Geological Survey), David Valentine (University of California San Diego), Jeffery Grethe (University of California San Diego), and Burak Ozyurt (University of California San Diego).

Description: The CINERGI project is constructing a community inventory and knowledge base on geoscience information resources to meet the challenge of finding resources across disciplines, assessing their fitness for use in specific research scenarios, and providing tools for integrating and re-using data from multiple domains. Constructing
such a knowledge base can be challenging since data may need to be harvested from multiple sources, but more so data needs to be extracted and cleaned for easy consumption.

This technology presentation will describe in-depth the architecture, tools, and services used in creating the CINERGI repository. In particular, it will describe the how multiple information extraction pipelines are assembled using GeoPortal, a combination of customized operators/parsers, and MongoDB to curate harvested resource descriptions in a scaleable way. Given the heterogeneity of the data it is often the necessary to do post-mortem analysis of the outputs. We will describe a comprehensive provenance framework that captures extraction details from these components and answers vital questions about a resource.

This technology presentation is targeted for cyber-infrastructure specialists who manage large information extraction pipelines, and curators and users who have queries relating to provenance of resources.

EarthCube Architecture Forum

Convener: Richard, Steve (Arizona Geological Survey)
Co-Conveners: Steve Diggs (Scripps), George Percivall (Open Geospatial Consortium)

Description: Goal— promote community convergence on EarthCube architecture ideas by presenting concrete design ideas from the conceptual design projects. My observation is that many architecture discussions make no progress because people aren't talking about the same thing. The goal here is to lay some groundwork, agree on what we’re trying to do, look at some concrete proposals (from the conceptual design teams, maybe other invited ideas?) and how the current building blocks fit into that, and use that to collect input and start discussion about specifics.

Social dimensions of EarthCube data and infrastructure

Convener: Slota, Stephen (University of California, Irvine)
Co-Conveners: Siri Jodha Khalsa (University of Colorado), Geof Bowker, (University of California, Irvine)

Description: The development of new data infrastructures for collaboration and potential large-scale analysis is as much a cultural and social exercise as it is technical. Organizational, cultural, and political issues move to the forefront as various facets of wide-scale uptake, disciplinary division and data practice become the largest challenges to infrastructure adoption and use. When science or scientific outcomes are published, made open or accessible, or otherwise released into the world they bear the potential for substantial cultural, social and political effects. What consequences are there to publicly opening data and providing the basic tools for their interpretation and use? What are the social and policy outcomes of the publication of data, models and analysis? Organizational, technological and disciplinary cultures play a significant role in the use and effective re-use of scientific databases, models and tools. Social scientists have long studied the movement, storage, retrieval and re-usability of scientific data, but their role changes based on the projects and scientific communities engaged in various collaborative and interdisciplinary contexts. What are the roles of the social scientist, field scientist, and computer scientist in the design and development of new infrastructures? We propose a session to discuss these issues as they relate to the work of EarthCube and its building blocks. We will explore the social aspects of the development of technology, models and systems for supporting scientific knowledge production as well as the outcomes of that science as it moves from the lab into the world. This
session will be a panel and group discussion, followed by a description of methods and tools for assessing the social and policy impacts of 'ground-level' and 'high-level' decisions about data management, governance and re-use.

Domain Registries of Information Resources: a How-To for Your Community

(To be followed by a related working group session.)
Convener: Zaslavsky, Ilya (University of California San Diego)

Description: Creating domain catalogs of information resources, to support discovery of databases, software, models, portals, vocabularies, and other resources geoscientists used in their work, has been one of the key themes of EarthCube end-user workshops. Several workshops have assembled such initial catalogs. In addition, the EarthCube SEN and C4P Research Coordination Network projects created such catalogs for their domains. Working with the EarthCube CINERGI project, the SEN and C4P projects established a model of how such catalogs can be assembled, published, and community-curated.

This workshop, jointly presented by members of the CINERGI, C4P and SEN teams, will introduce this model, and walk you from assembling a seed catalog, to populating it with standard metadata, to publishing it on the web in a way that supports catalog visualization and community updates of resource information. The goal is to facilitate organization of working groups that would create similar catalogs in additional geoscience domains, and help them get started leveraging our software tools and experience.

In addition, the workshop will include a general introduction to issues and technologies being used to describe catalog harvesting and curation tools being developed as part of the CINERGI project.

Workshops and Technology Demonstrations (*denotes accompanying poster)

Crawling the Web for EarthCube*

Convener: Duerr, Ruth (NSIDC)
Co-Conveners: WenWen Li (Arizona State University), Siri Jodha Khalsa (University of Colorado)

Description: Structured advertisements for web services and data exist all over the web often in forms such as OpenSearch descriptions, OGC web service getCapabilities documents, WSDL documents, OAI-PMH metadata feeds and other metadata files, and semi-structured data in semantic web languages. The data and services pointed to by these documents are central to scientific analysis and decision making. Various tools now exist to allow researchers to simply fill out web forms to advertise (i.e., "publish") the existence of their data in one of these discoverable forms. Technologies such as large-scale web crawling can be used to actively search, discover and aggregate these "ads" and provide that information to catalogs and brokering services. Several EarthCube building blocks are funded to use these technologies to improve the
discoverability of distributed web resources. In this 90 min workshop these plans will be discussed, interfaces fleshed out, and feedback on the types of data and services of most important to the community solicited.

**Ten Simple Rules for the Care and Feeding of Scientific Data**

Convener: Gil, Yolanda (University of Southern California)

Description: This presentation will give practical guidelines for scientists to take care of their data. The guidelines are captured as ten simple rules, and are based on a recently published paper (My Note: See Above) that resulted from a collaborative workshop on data citation and reuse.

**EarthCube Cloud Commons: GeoCloud Workshop*\**

Convener: Law, Emily (Jet Propulsion Laboratory)  
Co-Convener: Wenming Ye (MicroSoft Research)

Description: The EarthCube Cloud Commons (ECC) working group’s objective is to investigate issues related to adoption of cloud computing by the NSF, educate and facilitate adoption of cloud computing by the EarthCube community. A pilot project is currently underway in developing a prototype cloud Virtual Machine (VM) repository with an online cloud portal that provides guidance on how to create cloud instances using ECC. A demonstration of the prototype and tutorial on how to use cloud portal will be given at the workshop. The workshop also provides an opportunity to engage the EarthCube community at large, and provides a forum for ECC discussions and inputs.

**Simplifying Scientific Data Management through SaaS*\**

Convener: Malik, Tanu (University of Chicago)  
Co-Conveners: Kyle Chard (University of Chicago)

Description: Scientific projects must deal with increasingly large and diverse data. These data encompass many files, database tables, and linked data; feature different data models and file formats; and are located in different storage systems and institutions. When data are simple and few, a researcher may track them in a notebook or spreadsheet; the data themselves may simply be stored in a common directory. For example, files related to a specific simulation run may be stored in a directory named by run number; files related to a specific field site may be stored in a directory named for that site.

The fact that current data management practices do not meet the needs of scientists is evident in that scientists spend significant amounts of their research time managing data rather than doing science, with self-reported values of 90% being common. With trends towards “big data” and the associated increases in data size, variety, and complexity, data management problems will grow yet worse. Simply keeping track of the data can quickly become an unmanageable task —greatly complicating research processes.

This technology presentation will demonstrate services for the entire data life cycle. It will first demonstrate Globus services of data movement, synchronization, and sharing. It will then focus on data publication and discovery capabilities that are being added to Globus service. These new features make it simple to describe, curate, and preserve data at

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications

Updated: Sun, 19 May 2019 05:26:10 GMT

Powered by [mindtouch](https://mindtouch.us)
desired levels of durability. They also enable rich discovery by making it possible to search, browse, and access large published data sets. Finally, we will demonstrate services that make it easy for users to track data, its transformations and view the entire provenance trail of the data.

This technology presentation will be hands-on; demonstrating and walking through with users how scientific data management tasks can be simplified drastically throughout-sourced services.

**The VHub Community Cyberinfrastructure—Sharing and Growing**

Convener: Patra, Abani (SUNY at Buffalo)

Description: A fundamental goal of cyberinfrastructure in the geosciences is to enable integration of multidisciplinary and computational thinking into research and applications. Ideally this cyberinfrastructure provides a mechanism for globally collaborative research and development of computational models of geological processes and their integration with complex geospatial, observational, and experimental data. In order to achieve these aspirations, it is critical to develop resources, within the cyberinfrastructure itself, to promote the training of the next generation of geoscientists in management of data, modeling and simulation.

The session will start with a broad introduction of the successful VHub (Vhub.org) collaboratory that has a significant fraction of the volcanology community using it for various services including but not limited to model and data sharing, online simulation, project based collaboration and community organizational efforts. We will follow with more participatory exercises on using the VHub platform for shared projects including modeling software development.

In this session we will focus on the use of cyberinfrastructure platforms, like VHub, to enhance the academic experience of students, convey concepts in code verification and validation, and to train new researchers in modeling and simulation. We are interested in highlighting innovative methods to ease students and other learners through the process of development of new research skills: from running code and manipulating outputs, considering model verification and validation, through to modifying code and contributing new code to the research community. Development of this perspective is essential for Earthcube, not only for students but to increase the flexibility and long-term sustainability of cyberinfrastructure and software used in the geosciences, as in other disciplines.

**Increasing the Potential Usability, Access, and Funding of Our Geological Collections— Presentation and Open Discussion**

Convener: Timm, Sarah (ECLIPSE)

Description: Geological materials are used every day, in research, as educational materials, as ornamental pieces, and when broken down to their finest components are incorporated into objects we use on a daily basis. For years the accumulation of geological materials in institutional and individual collections has continued without sufficient thought to long-term preservation, access, and usability. Now these collections are at a pivotal point: with geological localities all over the world being shut down, reclaimed, and exhausted, storage limits being exceeded, and limited funding for collections management, physical geological materials are becoming endangered. If action is not taken soon to preserve these collections, then there is great risk of losing them forever.
Due to the technological revolution we have had, the potential for preserving these materials has increased dramatically. Paper labels can only communicate so much information and only to someone nearby the sample. Digital records for specimens can be shared across continents and inform scientists years down to road of analyses preformed today. As part of my master’s thesis at Virginia Tech I developed a database program for geological collections called EGEMS which stand for electronic geological management system. I have spent the majority of the last two years focused on improving access, usability, and funding for geological collections through the use of EGEMS. I have worked with Virginia Tech and the Virginia Museum of Natural History to incorporate existing data and add much more, including data for research materials which could not have easily been done using their previous system. It has been an incredibly rewarding experience but it has had a much smaller impact than it could have given that most academic institutions have geological materials of some kind. I am now working on a web interface called GeoCat to share this data with institutions worldwide.

Currently Proposed Working Groups

EarthCube Technical Services Discussion

Convener: Crane, Gary (SURA)
Co-Conveners: Kevin Tyle (University at Albany), Aaron Piña (Colorado State University), Jeremy Cothan (SECOORA), Jim Tung (MathWorks), and Gus Alaka (Colorado State University)

Description: This session will explore the need for the addition of a technical services function to the Earth Cube portfolio. Attendees will be provided with an overview of a possible suite of Earth Cube technical support services and then engaged in an open dialogue to explore the value and demand for a range of Earth Cube technical support capabilities that could be included as a core function of EarthCube.

This could range from best practices documents, technical hangouts, forums or chat rooms that might facilitate peer support, research data management, and assistance with access to and use of high performance computing systems, all the way up to a staffed Earth Cube technical support help desk.

EarthCube Community Workforce Development

(Follows a related presentation.)
Convener: Downs, Robert (CIESIN, Columbia University)
Co-Conveners: Robert S. Chen (CIESIN, Columbia University), Leslie Hsu (LDEO, Columbia University), Kerstin Lehnert (LDEO, Columbia University), Erin Robinson (Foundation for Earth Science), Ilya Zaslavsky (University of California, San Diego).

Description: The EarthCube community is faced with various challenges, opportunities, and choices for enabling Earth science. Technology is evolving rapidly and new tools and techniques are being created and enhanced to enable community members to do more individually and collaboratively. Workforce development opportunities can assist community members in their efforts to improve their capabilities and knowledge. A panel will describe workforce development opportunities that are available for the Earth science community and engage in a discussion with the audience on how the EarthCube community can improve by leveraging these opportunities.
Domain Registries of Information Resources: a How-To for Your Community

(Follows a related presentation.)
Convener: Zaslavsky, Ilya (University of California San Diego)

Description: Creating domain catalogs of information resources, to support discovery of databases, software, models, portals, vocabularies, and other resources geoscientists used in their work, has been one of the key themes of EarthCube end-user workshops. Several workshops have assembled such initial catalogs. In addition, the EarthCube SEN and C4P Research Coordination Network projects created such catalogs for their domains. Working with the EarthCube CINERGI project, the SEN and C4P projects established a model of how such catalogs can be assembled, published, and community-curated.

This workshop, jointly presented by members of the CINERGI, C4P and SEN teams, will introduce this model, and walk you from assembling a seed catalog, to populating it with standard metadata, to publishing it on the web in a way that supports catalog visualization and community updates of resource information. The goal is to facilitate organization of working groups that would create similar catalogs in additional geoscience domains, and help them get started leveraging our software tools and experience.

In addition, the workshop will include a general introduction to issues and technologies being used to describe catalog harvesting and curation tools being developed as part of the CINERGI project.

Governance

Source: http://workspace.earthcube.org/test-governance

The Draft Governance Charter is now available for review, please see the Charter Review page.

EarthCube Test Enterprise Governance: An Agile Approach

The EarthCube Test Enterprise Governance award will implement an agile approach to engage multiple, broad, and diverse geo and cyber science communities to determine an appropriate governing system for EarthCube. In Year 1, three governance scenarios will evaluate and choose a governance framework to be tested in Year 2. This will be coordinated by a Secretariat to carry out project management duties, an Assembly of seven EarthCube stakeholder groups to evaluate potential governance models, and crowdsourcing mechanisms to review and vet governance models with the broad EarthCube community. In Year 2, a prototype demonstration governance framework will facilitate community convergence on a reference architecture, procedures for standards, and coordination among emerging technical elements.

This workspace will be used for cross-project collaboration among members of the Test Governance Secretariat, Advisory Committee, and Assembly teams.

Upcoming Events

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch™
RCNs

Source: [http://workspace.earthcube.org/rcns](http://workspace.earthcube.org/rcns)

Research Coordination Networks (RCNs)

As of September 2013, NSF awarded 15 new EarthCube projects (categorized into Governance, Building Blocks, Research Coordination Networks, and Conceptual Designs) with funds totaling $14.5 million. The three current RCN awards are:

- **EarthCube RCN: C4P: Collaboration and Cyberinfrastructure for Paleogeosciences.** Kerstin Lehnert, Columbia University, Lamont Doherty Earth Observatory.
- **Building a Sediment Experimentalist Network (SEN).** Wonsuck Kim, University of Texas at Austin.
- **EC3: Earth-centered Communication for Cyberinfrastructure: Challenges of Field Data Collection, Management, and Integration.** Matty Mookerjee, Sonoma State University.

Follow along with the C4P Webinars at the C4P [YouTube Channel](http://workspace.earthcube.org) or view the EarthCube [YouTube Channel](http://workspace.earthcube.org).

Building Blocks

Source: [http://workspace.earthcube.org/building-blocks](http://workspace.earthcube.org/building-blocks)

As of September 2013, NSF awarded 15 new EarthCube projects (categorized into Governance, Building Blocks, Research Coordination Networks, and Conceptual Designs) with funds totaling $14.5 million. The nine current Building Blocks awards are:

- **Deploying Web Services Across Multiple Geoscience Domains.** Tim Ahern, Incorporated Research Institutions for Seismology.
- **Specifying and Implementing ODSIP, a Data-service Invocation Protocol.** David Fulkner, Open Source Project for Network Data Access Protocols.
- **Software Stewardship for the Geosciences.** Yolanda Gil, University of Southern California:
- **A Broker Framework for Next Generation Geoscience (BCube).** Siri-Jodha Khalsa, University of Colorado.
- **Integrating Discrete and Continuous Data.** David Maidment, University of Texas at Austin.
• **Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery (OceanLink)**. Thomas Narock, Marymount University.

• **A Cognitive Computer Infrastructure for Geoscience**. Christopher Re, Stanford University.

• **Earth System Bridge: Spanning Scientific Communities with Interoperable Modeling Frameworks**. Scott Peckham, University of Colorado.

• **Community Inventory of EarthCube Resources for Geoscience Interoperability (CINERGI)**. Ilya Zaslavsky, University of California, San Diego.

Upcoming Events

Task List

1

Recent Activity

A Few

---

### Conceptual Designs

Source: [http://workspace.earthcube.org/conceptual-designs](http://workspace.earthcube.org/conceptual-designs)

As of September 2013, NSF awarded 15 new EarthCube projects (categorized into Governance, Building Blocks, Research Coordination Networks, and Conceptual Designs) with funds totaling $14.5 million. The current Conceptual Design awards are:

• **Developing a Data-Oriented Human-Centric Enterprise Architecture for EarthCube**. Chaowei Yang, George Mason University.

• **Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences**. Ilya Zaslavsky, University of California, San Diego.

Task List

No Tasks

---

### Interest Groups

Source: [http://workspace.earthcube.org/interest-groups](http://workspace.earthcube.org/interest-groups)

Special Interest Groups are virtual teams working together on EarthCube-related projects. The goal of many of these groups is to build communities within EarthCube around each scientific domain, and continue conversations that started at the end-user workshops regarding what each group needs from EarthCube. Other interest groups consist of members of the tech community with an interest in developing key technologies for EarthCube.

My Note: See Example Below of Many, Many, etc.
Recent Activity

**DRIHM Summer School 2014 Announced**

Posted by Rachael Black on June 10, 2014


---

**Hubs**

Source: [http://workspace.earthcube.org/hubs](http://workspace.earthcube.org/hubs)

No Tasks

---

**Earth Cube Web Site**

Source: [http://earthcube.org/](http://earthcube.org/)

---

**Welcome to EarthCube**

EarthCube is a bold new initiative to create a community-driven data and knowledge management system that will allow for unprecedented data sharing across the geosciences. See [About](http://earthcube.org/page/about) Below

---

**Join Our Community**

You can influence what EarthCube can do and will be. Join over 2,500 participants in forums, work groups, EarthCube events, and at virtual and in-person meetings across the country. See [Get Involved](http://earthcube.org/page/get-involved) Below

---

**EarthCube Commons**

The EarthCube Commons workspace is where EarthCube project teams and stakeholder groups collaborate. You can join working groups and see what the EarthCube community is doing in real time. See [Earth Cube Commons](http://earthcube.org/page/earthcube-commons) Below

---

**About**

Source: [http://earthcube.org/page/about](http://earthcube.org/page/about)
What is EarthCube?

EarthCube is community-led cyberinfrastructure that will allow for unprecedented data sharing across the geosciences. Its aim is to develop a framework over the next decade to assist researchers in understanding and predicting the Earth system from the sun to the center of the Earth. EarthCube has the potential to:

- Create effective, community-driven cyberinfrastructure
- Allow global data discovery and knowledge management
- Achieve interoperability and data integration within and across disciplines

Who is EarthCube?

Formed in 2011, EarthCube is a collaborative partnership between NSF’s Directorate of Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI). It’s also a virtual community with over 2,500 participants, including atmosphere, ocean, computer, information, and social scientists, as well as educators, data managers, and other diverse contributors.

This community has done a great deal of collaborative work since EarthCube was launched in mid-2011. As of September 2013, a new round of NSF awards was made to develop key technologies, promote community building, explore integrative systems, and prototype a governance structure.

In addition to the funded project teams, several Special Interest Groups are in place as virtual teams working together on EarthCube-related projects. Visit the EarthCube workspace to see what these teams are doing right now.
What are the Goals of EarthCube?

EarthCube aims to:

- Transform research and data management practices within the geosciences community over the next decade
- Provide unprecedented new capabilities, including access to data and visualization tools, to researchers and educators
- Vastly improve the productivity of the geosciences community
- Accelerate research on the Earth system
- Provide a knowledge management framework for the geosciences

History and Progress


EarthCube was launched in mid-2011 as a collaborative partnership between NSF’s Directorate for Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI), with the goal of creating a more sustainable future through improvements in our understanding of Earth as a complex and changing planet. EarthCube is a cornerstone of NSF’s Cyberinfrastructure for the 21st Century (CIF21) initiative, whose chief objective is to develop a nationwide, sustainable, and community-based cyberinfrastructure for researchers and educators.

Initial Planning

The GEO Directorate and Office of Cyberinfrastructure of the National Science Foundation (NSF) released a [Dear Colleague Letter: The "Earth Cube"—Towards a National Data Infrastructure for Earth System Science](http://earthcube.org/page/history-and-progress) in June 2011 announcing their partnership in starting EarthCube and their initial goals for EarthCube. Several webinars followed, and an additional document: [EarthCube Guidance to the Community](http://earthcube.org/page/history-and-progress) gave more-detailed guidance on EarthCube.

EarthCube held its first charrette (community event) November 1–4, 2011, with the goal to gather as many ideas as possible from a broad representation of potential EarthCube participants (Earth scientists, IT experts, information scientists, and other interested parties) to jump-start the planning and development processes for EarthCube. Goals of this charrette were to:
1. Identify commonalities and differences in EarthCube vision and challenges,
2. Craft explicit statements of challenges and develop a list of possible milestones
3. Map the vision, challenges, and milestones to governance and potential EarthCube designs
4. Develop a rough understanding of what the next steps must address.

In advance of this event, NSF solicited a series of White Papers for EarthCube. These were written by interested individuals and organizations across the domain and computer sciences. These white papers provided a forum for people to propose what EarthCube should look like in terms of science requirements, technology solutions, designs, and governance.

Following the November 2011 charrette, NSF solicited Expressions of Interest, which allowed the EarthCube community to contribute their ideas and visions on what EarthCube should consider or contain, which formed the foundation for the creation of the EarthCube groups. NSF formed and funded several EarthCube Community Groups and Concept Award Teams by merging the authors of Expressions of Interest under common themes. The goal of the Community Groups (Data, Governance, Semantics & Ontologies, and Workflow) was to gather EarthCube user requirements through broad engagement of the geosciences and IT communities, while the Concept Teams (Brokering, Earth System Model, Layered Architecture, Cross-Domain Interoperability, Web Services, Open Hydrospheric Modeling Framework (OHMF), and Dark Geo Data) evaluated and prototyped innovative key technologies.

Between March and August 2012, each of the Community Groups and Concept Teams was tasked with writing Roadmaps regarding 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 identify initial EarthCube stakeholders and cyberinfrastructure components. They served to collectively 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 in order to forge a common vision and create a cohesive set of milestones to move EarthCube forward. Activities and discussions were focused on identifying common themes, challenges, and synergies, in order to merge these roadmaps into one common roadmap for EarthCube.
Meetings of the EarthCube Community Groups and Concept Award Teams Principal Investigators were held in June and October 2012. The first meeting was held in order to further the discussions and roadmap integration efforts which began at the June 2012 charrette; several ideas regarding use cases, reference architecture, governance and timelines were crafted by meeting participants. The follow up meeting continued to integrate the roadmaps and thereby develop a more cohesive vision on how to move EarthCube forward. While significant steps have been made towards achieving this goal, a comprehensive technical roadmap for EarthCube is still in progress.

**The EarthCube Community**

One of the principal conclusions from the June 2012 charrette was that a community engagement and outreach effort must be undertaken to grow the EarthCube community. The community engagement plan undertaken by EarthCube participants to expand the EarthCube community and gather user requirements for EarthCube includes several components: the Stakeholder Alignment Survey, EarthCube End-User Workshops, workshops and presentations at EarthCube professional conferences, publications in trade and scholarly journals, and several online mechanisms, including social media ([Facebook](http://facebook.com), [Twitter](http://twitter.com), [YouTube](http://youtube.com)), wikis, webinars, and a large-scale virtual kickoff event to mark the new round of 2013 NSF EarthCube awards. While many of these components are ongoing, some have already provided important and useful insights for the growth and development of EarthCube and its community.

**Stakeholder Alignment Survey**

Complementing the efforts of the Community and Concept Groups is an on-going EarthCube Stakeholder Alignment Survey. The success of EarthCube will depend on the ability of diverse stakeholders to orient and connect with one another in new ways that advance both individual and shared objectives. This process of stakeholder alignment is the focus of a current study of "Stakeholder Alignment for EarthCube" (OCI 12-29928), which includes the connections within and across disciplines, fields, and areas of expertise (Geo and Cyber), as well as the new institutional arrangements that emerge to sustain EarthCube as a transformational initiative. A visual and interactive representation of the data collected in this survey is available [here](https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications). (This version requires the Microsoft Silverlight plug-in. A plug-in free version is also available.) Additional feedback provided in the months and years to come from this and upcoming surveys will continue to shape EarthCube by allowing the wants and needs of stakeholders to be included in the development of EarthCube.
End-User Workshops

NSF funded a series of two dozen EarthCube domain end-user workshops beginning in June 2012. These workshops targeted a broad spectrum of Earth, atmosphere, ocean, and related scientists, including senior and early career scientists. The purpose of these workshops was to allow geoscience communities to articulate and document their cyberinfrastructure needs and what they would like to do in the future, both within their disciplines and in terms of accessing data and information from other fields. An additional goal of these workshops was to gather requirements on EarthCube science-drivers, data utilities, user-interfaces, modeling software, tools, and other needs so that EarthCube can be designed to help geoscientists more easily do the science they want and need to do. Community input and engagement programs will help direct and shape the function and form of what EarthCube will be. A complete listing of all of these workshops, including an executive summary of each is now available.

Special Interest Groups

Several Special Interest Groups have been created since the June 2012 charrette in order to address issues not covered by the existing Community and Concept Groups. The Special Interest Groups include, but are not limited to: Education and Workforce Development, Software Engineering Community of Practice, Website and Collaboration, Physical Samples as part of Cyberinfrastructure, and Data Integration for Biology and Earth Science.

The EarthCube community is still in the process of identifying gaps and overlaps among these groups, and new groups will likely be created as additional needs are identified. If you see need for an additional Special Interest Group, and would like to create one, please contact us and we can assist you in setting up a virtual collaboration space.

Moving Forward

In September 2013, the NSF 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 the paleogeosciences, field data collection, and experimental sedimentary processes, as well as to help these communities move forward to prioritize specific software development projects to engage with EarthCube. A single project to develop and test a prototype governance framework for EarthCube was also funded. In November 2013, the NSF released a call for additional

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by Q1mindtouch
EarthCube RCNs and Building Blocks, with proposals due in March 2014. (For more information on this solicitation, please see: the NSF EarthCube Website.) Descriptions of current EarthCube projects are available, this list will be updated as more funds are awarded.

A great deal of collaborative work has been done to further the advancement of EarthCube, but the process is far from over. In the coming years the EarthCube community will continue to build and expand upon the work already done.

**End-User Workshops**

Source: [http://earthcube.org/page/end-user-workshops](http://earthcube.org/page/end-user-workshops)

NSF funded a series of two dozen EarthCube domain end-user workshops targeting a broad spectrum of Earth, atmosphere, ocean, and related scientists, including senior and early career scientists. The purpose of these workshops was to allow geoscience communities to articulate and document their cyberinfrastructure needs and what they would like to do in the future in terms of accessing data and information within and outside their disciplines. The workshops also focused on gathering requirements on EarthCube science-drivers, data utilities, user-interfaces, modeling software, tools, and other needs so that EarthCube can be designed to help geoscientists more easily do the science they want and need to do. Community input from these workshops and other community engagement programs will help direct and shape the function and form of what EarthCube will be. Additionally, these domain workshops served to introduce EarthCube to end-users, and were designed to listen to the needs of the geosciences end-user groups to better understand how data-enabled science can help them achieve their scientific goals.

Executive summaries for each of these workshops are available below.

**Geochemistry**

**June 28, 2012, Montreal, Canada**
Organizer: Kerstin Lehnert, Lamont-Doherty Earth Observatory at Columbia University

**Executive Summary**

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

**Executive Summary**

**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

**Executive Summary**
Structural Geology and Tectonics

October, 20-21 2012, Chicago, IL
Organizer: J. Douglas Walker
Executive Summary

EarthScope

October 29-30, 2012, Arizona State University, Tempe, AZ
Organizers: Ramon Arrowsmith, Arizona State University and the EarthScope Cyberinfrastructure Subcommittee (ECISC)
Executive Summary

Experimental Stratigraphy

December 11-12, 2012, University of Texas at Austin
Organizers: Wonsuck Kim, Leslie Hsu, Raleigh Martin, and Brandon McElroy; University of Texas at Austin
Executive Summary

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

December 17-18, 2012 UCAR, Boulder, CO
Organizer: Mohan Ramamurthy, UCAR
Executive Summary

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

January 21-23, 2013, University of Delaware
Organizers: Anthony Aufdenkampe, Christopher Duffy, and Gregory Tucker; Stroud Water Research Center
Executive Summary

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, Norman Jones, and Ying Fan Reinfelder; Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI)
Executive Summary

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
Executive Summary

Education

March 4-5, 2013, Scripps Institution of Oceanography, La Jolla, CA
Organizers: Kim Kastens (Education Development Center), Ruth Krumhansl (Education Development Center), Cherry Peach (Scripps)

Executive Summary

Petrology and Geochemistry

March 6-7, 2013, National Museum of Natural History, Washington, DC
Organizers: Chuck Conner (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)

Executive Summary

Community Modeling

April 22-23, 2013, Boulder, CO
Organizers: Louise Kellogg (University of California, Davis), Jennifer Arrigo (CUAHSI), Lorraine Hwang (Computational Infrastructure Dynamics), Scott Peckham (Community Surface Dynamics Modeling System), David Tarboton (Utah State University)

Executive Summary

Integrating Inland Waters, Geochemistry Biogeochemistry and Fluvial Sedimentology Communities

April 24-26, Boulder, CO
Organizers: Emilio Mayorga (University of Washington) and Albert Kettner (University of Colorado)

Executive Summary

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

Executive Summary

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
Executive Summary

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)

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
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 and Shanan Peters (University of Wisconsin-Madison)

Articulating Cyberinfrastructure Needs of the Ocean Ecosystem Dynamics Community
October 21-22, 2013, Woods Hole Oceanographic Institution, Woods Hole, MA
Organizers: Danie Kinkade and Peter Wiebe (WHOI)

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)

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

Source: http://earthcube.org/page/get-involved

Introduction

1. Join the Workspace

EarthCube needs your input and expertise. We invite you to take a hands-on role in shaping what EarthCube can do and will be. Please take a moment to visit our workspace to join working groups, participate in forums, and attend virtual and in-person events. EarthCube is a large-scale effort that depends on active collaboration from members of the entire geosciences community, who are actively engaged in research that involves collecting and using data, simulating processes, and visualizing complex interacting systems. To do this, there needs to be close interaction between end users and the software creators who will make sure this cyberinfrastructure is responsive to diverse user needs and can innovate within the changing landscape of computer science and computing hardware.

2. Subscribe to the Newsletter

We'll provide regular updates on EarthCube events, funding opportunities and more. Newsletter coming soon.

3. Get EarthCube Involved in Your Work

The workspace’s goal is to serve as information hub for the entire EarthCube community, and that means you can use it as a venue for showcasing your team’s progress and inviting others to get involved with what you’re doing. EarthCube.org disseminates information via a newsletter, online news feed, social media pages, and other publications, and our engagement team will be happy to help promote your team’s announcements and events.
4. Help Build a Community Inventory of Data Resources

The EarthCube CINERGI (Community Inventory of Earthcube Resources for Geoscience Interoperability) project is assembling a large metadata registry that will integrate information from data repositories in multiple domains, datasets submitted by researchers directly, as well as information resources such as vocabularies, ontologies, web services, software, catalogs, formal information models and schemas for different types of data.

Have you created such datasets or other resources that you’d like to share with the community? Do you know of resources that should be included in such an inventory? If so, please enter them using the form found at metadata.earthcube.org. If the data you are submitting is not yet in a managed data repository, we can also help you connect with repository managers. Don't forget to check back often to explore the inventory as it grows.

5. Follow EarthCube on Social Media

Facebook
Twitter
YouTube
SlideShare

Stay up to date by liking EarthCube on Facebook or following us on Twitter @EarthCube.

You can also check out the EarthCubeNSF YouTube channel to meet the current funded project teams, see highlights from earlier projects, and take a look at the future of EarthCube.

Funding

Source: http://earthcube.org/page/funding

New Funding Opportunities

A key component of earthcube.org is to provide updates on funding opportunities and provide a forum in which potential collaborators can find each other. Check this page for information on upcoming NSF funding solicitations related to EarthCube.

November 2013: NSF Announces EarthCube Amendment III Funding Solicitation (NSF 13-529)

View details at NSF.gov
Due dates:

Full Proposal Deadline: March 12, 2014
EarthCube Building Blocks
Full Proposal Deadline: March 12, 2014
EarthCube Research Coordination Networks

Synopsis:

EarthCube is a community-driven activity sponsored through a partnership between the NSF Directorate of Geosciences and the Directorate for Computer & Information Science & Engineering Division of Advanced Cyberinfrastructure to transform research in the academic geosciences community. EarthCube aims to create a well-connected and facile environment to share data and knowledge in an open, transparent, and inclusive manner, thus accelerating our ability to understand and predict the Earth system.

Achieving EarthCube will require a long-term dialog between the NSF and the interested scientific communities to develop cyberinfrastructure that is thoughtfully and systematically built to meet the current and future requirements of geoscientists. New avenues will be supported to gather community requirements and priorities for the elements of EarthCube, and to capture the best technologies to meet these current and future needs. The EarthCube portfolio will consist of interconnected projects and activities that engage the geosciences, cyberinfrastructure, computer science, and associated communities. The portfolio of activities and funding opportunities will evolve over time depending on the status of the EarthCube effort and the scientific and cultural needs of the geosciences community.

This umbrella solicitation for EarthCube allows funding opportunities to be flexible and responsive to emerging needs and collaborative processes. The EarthCube vision and goals do not change over time, and this section of the solicitation will remain constant. Funding opportunities to develop elements of the EarthCube environment will be described in Amendments to this solicitation. Amendments will appear in the Program Description Section of the solicitation and will include details on the parameters, scope, conditions, and requirements of the proposal call. Researchers who receive alerts related to solicitation releases will receive notification when the EarthCube solicitation is updated with an Amendment.

2013 NSF EarthCube Awards

The next phase of EarthCube is here! To foster a dialogue among geo-, bio-, and cyberscientists to create an EarthCube framework, the National Science Foundation (NSF) has made a new round of awards as of September 2013. NSF has awarded these 15 new EarthCube projects (categorized into Governance, Building Blocks, Research Coordination Networks, and Conceptual Designs) with funds totaling $14.5 million.

Governance

Lee Allison, University of Arizona: EarthCube Test Enterprise Governance: An Agile Approach

The EarthCube Test Enterprise Governance award will implement an agile approach to engage multiple, broad, and diverse geo and cyber science communities to determine an appropriate governing system for EarthCube. In Year 1,
three governance scenarios will evaluate and choose a governance framework to be tested in Year 2. This will be coordinated by a Secretariat to carry out project management duties, an Assembly of seven EarthCube stakeholder groups to evaluate potential governance models, and crowdsourcing mechanisms to review and vet governance models with the broad EarthCube community. In Year 2, a prototype demonstration governance framework will facilitate community convergence on a reference architecture, procedures for standards, and coordination among emerging technical elements.

**Research Coordination Networks (RCNs)**

Kerstin Lehnert, Columbia University, Lamont Doherty Earth Observatory: [EarthCube RCN: C4P: Collaboration and Cyberinfrastructure for Paleogeosciences](https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications)

This project establishes and operates the EarthCube Research Coordination Network (RCN) Collaboration and Cyberinfrastructure for Paleogeosciences to advance the role of cyberinfrastructure in unraveling 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, 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 better analyze data within the paleogeosciences and other domains and disciplines. [more]

Wonsuck Kim, University of Texas at Austin: [RCN: Building a Sediment Experimentalist Network (SEN)](https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications)

This RCN will form a Sediment Experimentalist Network (SEN) to help integrate the efforts of sediment experimentalists and build a knowledge base for guidance on best practices for data collection and management. The project will also facilitate cross-institutional collaborative experiments and communicate with and educate the research community about data and metadata standards 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. Major outcomes will be the creation of a knowledge base, coordination of experimental collaboratories, and integration of educational efforts and data standards development with tools for propagating new technology and methods. [more]

Matty Mookerjee, Sonoma State University: [EC3: Earth-centered Communication for Cyberinfrastructure: Challenges of Field Data Collection, Management, and Integration](https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications)

This RCN will facilitate digitization of geological field data in an effort to develop communication between the cyberinfrastructure community and those involved in field-based, solid earth geoscience. Researchers will take steps to document what exists currently for field data collection; assemble a community for discussing and exploring field data collection issues, specifically targeting young investigators; motivate distinct communities to work together on common issues associated with digitization; and evaluate what is missing in the creation of open and accessible data. To facilitate this, the RCN will conduct a series of workshops and townhalls at national meetings (GSA, AGU, AAPG) to foster community awareness, catalog resources, and investigate data collection and sharing scenarios. [more]
Building Blocks

Tim Ahern, Integrated Research Institutions for Seismology: Deploying Web Services Across Multiple Geoscience Domains

This Building Blocks project intends to extend the promotion of simple web services to simplify the task of discovering, accessing and using data from multiple sources. Investigators will promote the use of this system to manage data from the long tail of science and make it discoverable, removing it from the domain of "dark data". The project will extend its approach of exposing data sets through web services to those managed by non-NSF data centers both within the United States as well as international data sets by providing resources to stand up web services to expose the data holdings of other centers. This building block is an effort to engage EarthCube cyberinfrastructure in developing, establishing and adopting international standards to allow geoscientists to focus on science and increase productivity. [more]

David Fulkner, Open Source Project for Network Data Access Protocols: Specifying and Implementing ODSIP, a Data-service Invocation Protocol

This EarthCube building blocks project intends to build ODSIP (Open Data Services Invocation Protocol) in order to provide an array of open specification in client/server libraries. This project also seeks to provide a system in which EarthCube can be built effectively around clients and servers that employ common and conceptually rich protocols for data acquisition. [more]

Yolanda Gil, University of Southern California: Software Stewardship for the Geosciences

This building blocks project is to research and develop a system whereby geoscience and environmental software are generated effectively by geoscientists themselves, so that the software can be captured, curated, managed, and made available to all parties upon request. This project will begin this process by building partnerships between computer scientists, software developers, and scientists across all geoscience domains with the goal or creating a software ecosystem and a culture of software stewardship that will empower geoscientists and others to make their software accessible and manage it as a valuable scientific asset. [more]

Siri-Jodha Khalsa, University of Colorado: A Broker Framework for Next Generation Geoscience (BCube)

To address complex Earth system issues such as climate change and water resources, geoscientists must work across disciplinary boundaries, which requires them to access data outside of their fields. This award brings together an internationally recognized team of geo- and social-scientists, cyberinfrastructure experts and educators to explore how expert systems can mediate interactions and improve access between scientific fields. The initial focus is on hydrology, oceans, polar and weather, with the intent to make the technology applicable and available to all the geosciences. The team’s social scientists and educators will research how technology can improve knowledge exchange between scientific communities. [more]

David Maidment, University of Texas at Austin: Integrating Discrete and Continuous Data

This project builds upon previous work focused on integrated discovery of common information themes including precipitation in discrete data from the CUAHSI hydrologic information system and continuous data from the Unidata THREDDS data server. Investigators will advance that work by exploring the creation of new technologies for publishing
and discovery of information through the Global Earth Observation System of Systems (GEOSS) Common Infrastructure, the definition of a Common Information Model for discrete and continuous data, development of shared software tools for using this Common Information Model, and extension of the concepts to similar information in the Polar, Ocean and Solid Earth Sciences. [more]

Thomas Narock, University of Maryland, Baltimore County: Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery

A wide spectrum of maturing methods and tools, collectively characterized as Semantic Web Technologies, is enabling machines to complete tasks automatically that previously required human direction. For the Geosciences, Semantic Web Technologies will vastly improve the integration, analysis and dissemination of research data and results. This EAGER project will conduct exploratory research applying state-of-the-art Semantic Web Technologies to support data representation, discovery, analysis, sharing and integration of datasets from the global oceans, and related resources including meeting abstracts and library holdings. A key contribution will be semantically enabled cyberinfrastructure components capable of automated data integration across distributed repositories. [more]

Christopher Re, Stanford University: A Cognitive Computer Infrastructure for Geoscience

Today, access to information is often less of a problem than our ability to discover, process and use it. Geoscience currently lacks a sustainable cyberinfrastructure that can efficiently and with high precision and accuracy find, extract, and organize data that are critical to advancing many areas of science and leveraging current and past investments in data acquisition. We are developing a geoscience-oriented trained computing system, powered by world-class high throughput computing infrastructure, that will serve as a cross-disciplinary tool for finding, extracting, and organizing “dark data” from the text, tables, and figures of hundreds of thousands of documents. Early results indicate that our system can perform tasks of data identification and extraction reliably and at a fraction of the time and cost of humans. We hope to produce a dependable EarthCube building block that can serve many scientific communities. [more]

Scott Peckham, University of Colorado: Earth System Bridge: Spanning Scientific Communities with Interoperable Modeling Frameworks

This EarthCube Building Blocks project will draw from significant disciplinary and interdisciplinary expertise in the development, implementation and support of geoscientific modeling architectures and in the adoption of community standards in model development and data management. This team will integrate existing model architectures, model coupling standards, and data standards into a set of open-source Earth System Bridge building blocks that will transform the process of Earth system model coupling, and bridge the present technological gap. [more]

Ilya Zaslavsky, University of California, San Diego: Community Inventory of EarthCube Resources for Geoscience Interoperability (CINERGI)

This Building Blocks project focuses on constructing a community inventory and knowledge base on geoscience information resources to meet the challenge of finding resources across disciplines, assessing their fitness for use in specific research scenarios, and providing tools for integrating and re-using data from multiple domains. The project team envisions a comprehensive system linking geoscience resources, users, publications, usage information, and cyberinfrastructure components. This system would serve geoscientists across all domains to efficiently use existing and emerging resources for productive and transformative research. [more]
Conceptual Designs

Chaowei Yang, George Mason University: Developing a Data-Oriented Human-Centric Enterprise Architecture for EarthCube

This EarthCube conceptual design project is to develop a data-oriented and human-centric EarthCube enterprise architecture for achieving the goal of EarthCube as a community-driven activity to transform the conduct of geoscience research and education. The proposed EarthCube enterprise architecture will have geoscientists and domain experts at its center and facilitate them to communicate and collaborate through data sharing, and ultimately bring geosciences forward in a holistic fashion. This project seeks to design a conceptual architecture that can bring geoscientists, computing scientists, and social scientists together to collaborate on networks of data, technology, applications, business models, and stakeholders. [more]

Ilya Zaslavsky, University of California, San Diego: Enterprise Architecture for Transformative Research and Collaboration Across the Geosciences

This EarthCube conceptual design project is to generate an innovative EarthCube enterprise architecture that fosters inter-community research collaboration and data exchange across the geosciences. The main feature of this EarthCube enterprise architecture is that it enhances existing patterns of data and information exchange through the ability to evolve by factoring in the impact of maturing movements of data and collaboration. This project seeks to integrate traditional cyberinfrastructure components with other CI components 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. [more]

EarthCube Groups

Source: http://earthcube.org/page/earthcube-groups

Groups are the basic structures for the conversations about EarthCube that take place on the EarthCube web platform. EarthCube groups allow virtual communities of practice within EarthCube to form around a common scientific or IT field, end-user community, and other topics of interest (e.g., data curation, interoperability, semantics, etc.). Group members can post discussion topics, comments, documents, events, and more to the group forum.

Two different types of groups have been created:

**Funded Groups:** These groups represent individual National Science Foundation EarthCube awards, announced in September 2013. The NSF has awarded 15 new EarthCube projects (categorized into Governance, Building Blocks, Research Coordination Networks, and Conceptual Designs) with funds totaling $14.5 million. Scroll down for information on individual awards, projects, and personnel.

**Interest Groups:** The goal of many of these groups is to build virtual communities within EarthCube around each scientific domain, and continue conversations that started at the end-user workshops regarding what each group needs from EarthCube. Some interest groups will consist of members of EarthCube tech communities with an interest in developing key technologies for EarthCube. Visit the workspace to see what each of the special interest groups are doing right now.
Join an EarthCube Group

Although all group discussions and other resources are publicly accessible, you must be a member of a group to post content there. You’re welcome to join as many groups as you want—to join a group, simply go to the workspace, find a group that interests you, and click a link to request membership.

Create an EarthCube Group

In an effort to keep pace with the direction that EarthCube is moving, community members will now be able to create EarthCube Groups.

If you would like to create a group, please consider your group’s goals, and what you will do as the group leader/curator to help your group achieve these goals and keep your group active. What are you outputs? What are your outcomes? Who is your target audience? What will you do to help your group stay active? Can you enlist other group members to help you?

Those interested in creating a new group on the workspace can contact us for more information.

Assembly Workshops

Source: http://earthcube.org/page/assembly-groups
The EarthCube Test Enterprise Governance award is seeking community members to serve on one of six Assembly groups. These groups will serve as a primary venue for evaluating and testing governance models leading to a prototype demonstration governance framework for EarthCube. The six Assembly groups have been designed to provide the broad EarthCube community with a voice in the test governance process. The groups are arranged into the following categories:

**EarthCube Portfolio**
This group is made up of EarthCube related Research Coordination Networks, Conceptual Design awards, and Building Blocks award teams funded by the National Science Foundation (NSF) in September, 2013.

**EarthCube End-User Communities & Workshop Participants**
Beginning in June, 2012, the NSF funded a series of workshops with domain specific topics. Each of these workshops was organized by a Principal Investigator (PI) and steering committee and included between 50 and 100 attendees. This group will be comprised of representatives of that community, whether it is the PI of the organized workshop, workshop participants, or other representatives of that end-user community. This group may also include representatives of end-user communities not included in the initial workshop series.

**Professional Societies**
This group is comprised of scientific and technical professional societies whose members constitute one of the primary
foci of EarthCube. Most scientists belong to at least one professional organization; thus, this group targets professional organization administration, executives, and section chairs (e.g. AGU’s ESSI chair or GSA’s Geoinformatics Chair).

**Information Technology and Computer Sciences**
While broad in definition, this Assembly Group targets members of the technical communities that develop core computer and information science elements underlying EarthCube functionality, but are not the prime end-users of EarthCube. This group is designed to facilitate the communication between the primary end-users – scientific and academic research communities – and the developers that assist in the creation of data and documentation workflows.

**Data Facilities and Users**
This Assembly Group is comprised of the data facilities responsible for housing, curating, or making data available for the research science and academic community. These may be federally Funded Research and Development Centers (FFRDCs), NSF funded initiatives (such as IEDA, CUAHSI-HIS, or BCO DMO), or other data facilities (publically or privately funded).

**Industry & Free and Open Source Software (FOSS): Instrumentation, Software, and Technology Developers**
This Assembly Group is comprised of the developers and manufacturers of data collection instrumentation, data management software, and technology. This facilitates the creation of appropriate workflows and metadata for more easily integrating datasets for the scientific and research communities. The group targets the front line manufacturers responsible for the data collection mechanisms used in cutting-edge research.

To facilitate the work of these groups, the Test Governance award team is coordinating a series of in-person workshops between January and March, 2014:

- **Data Facilities**, January 15-17, 2014 in Washington, DC
- **EarthCube Funded Projects**, February 12-14, 2014 in Boulder, CO
- **Information Technology and Free and Open Source Software and Instrumentation**, March 5-7, 2014 in Boulder, CO
- **End-User Communities & Professional Societies**, March 18-20, 2014 in Washington, DC

Registration for the End-User Communities & Professional Societies Assembly Workshop is now open. If you would like to attend, please register using [this form](http://earthcube.org/page/news-and-events).

---

### News and Events

**Source:** [http://earthcube.org/page/news-and-events](http://earthcube.org/page/news-and-events)

**EarthCube News**

**EarthCube session at the OGC quarterly meeting in DC**
Tuesday, March 25, 2014
To all interested in finding out what's going on with the US National Science Foundation's (NSF) EarthCube initiative:

Come to OGC's quarterly technical meeting in the DC area on Tuesday, March 25, 1-5pm. A webcast of the session will be available:

**GOTOMEETING WEBCON**
Meeting URL: [https://www4.gotomeeting.com/join/460445031](https://www4.gotomeeting.com/join/460445031)
Audio: Use your headset/microphone and speakers (VoIP), or call in using your telephone.
United States: +1 (213) 493-0014

**OSTP Announces Directive on Improving Management of Scientific Collections**
Thursday, March 20, 2014

Today, the [US White House Office of Science and Technology Policy](https://www.whitehouse.gov) issued a directive to federal agencies that own or support scientific collections calling for improved management and access to these collections.

**Call for Papers: FOSS4G Conference 2014**
Friday, March 7, 2014

This year's FOSS4G Conference will take place in Portland, Oregon from September 8-13, and has recently issued a call for workshops, presentations, and academic papers. Put on by OSGeo, FOSS4G is the international gathering of open source and geospatial tribes. From its beginnings the FOSS4G conference has been the gathering of the spatial tribes and has a reputation of being a melting pot for great ideas in the spatial industry and a catalyst for many successful geospatial products, standards and protocols.

**C4P Webinar - Neotoma and CINERGI 2/4/2014**
Wednesday, February 5, 2014

The steering committee of the Collaboration and Cyberinfrastructure for Paleogeosciences (C4P) project would like to invite you our webinar series. The first webinar is Tuesday, February 4th, 4-5pm Eastern Time.

This webinar series will address Advances And Emergent Needs in Paleogeoscience Cyberinfrastructure - Community Inventory of EarthCube Resources for Geosciences Interoperability (CINERGI) and the Neotoma Database.

The topics for the first C4P Webinar are the Neotoma Database (Eric Grimm) and the CINERGI EarthCube project (Ilya Zaslavsky).

**Semantic Web/Linked Data Telecon**
Friday, January 31, 2014

The EarthCube program will host a **telecon Friday, 1/31, at 2:30 EST** provide a forum for any interested researchers, whether they are developers of technology or geosciences community members, to encourage dialog and collaborative solutions regarding semantic web and linked data concepts that have emerged as a potential mechanism to connect geosciences researchers, data and information that were not previously linked.
EC3 Field trip to Yosemite and Owens Valley
Thursday, January 30, 2014

The EC3 – Earth-Centered Communication for Cyberinfrastructure project is organizing two field excursions to facilitate a dialogue between field-based geologists, and computer and cognitive scientists concerning the unique problems faced by the field-based geological community with respect to data format, standards, management, representation, and integration. This project aims to break down the artificial barriers between subfields within the Earth Sciences, allowing us to ask new types of questions, and providing the means to contend with previously unanswerable ones. Field Trip Application Deadline: March 10th, 2014.

EarthCube Monthly Community Webinar
Thursday, January 23, 2014

This month's EarthCube Community Webinar will feature Mike Potter and Karen Borchert of Phase2, who will lead a demonstration and introduction into Open Atrium (the Drupal-based portal package our site runs on), followed by a brief Q & A session about the EarthCube Workspace. Check out the website at workspace.EarthCube.org. We look forward to 'seeing' you on the 23rd!

January/February Issue of D-Lib Magazine Now Available
Friday, January 17, 2014

The January/February 2014 issue of D-Lib Magazine (http://www.dlib.org/) is now available. This is a special issue on the topic of the Research Data Alliance (RDA) with guest editors Fran Berman, Rensselaer Polytechnic Institute; Ross Wilkinson, Australian National Data Service; and John Wood, The Association of Commonwealth Universities. The issue contains five articles and two conference reports. The 'In Brief' column presents five short pieces and excerpts from recent press releases. In addition you will find news of upcoming conferences and other items of interest in the 'Clips and Pointers' column. This month, D-Lib features DataVis.ca, courtesy of Michael Friendly, Professor of Psychology, Chair of the graduate program in Quantitative Methods at York University.

Digital Mapping Techniques 2014 Workshop
Thursday, January 9, 2014

Delaware State Geologist David R. Wunsch has announced that the Delaware Geological Survey will be hosting the 2014 Digital Mapping Techniques meeting on June 1-4, 2014, at the University of Delaware's John M. Clayton Hall Conference Center. Details and registration information can be found at http://www.dgs.udel.edu/digital-mapping-techniques-2014-workshop.

EarthCube Test Enterprise Governance Advisory Committee
Friday, December 20, 2013

The EarthCube Test Enterprise Governance Advisory Committee (AC) has now been formed. This group collectively represents the core stakeholder communities of EarthCube, and will help guide the process of managing and setting up
a demonstration governance framework for EarthCube over the next two years, as part of NSF-funded project EarthCube Test Enterprise Governance: An Agile Approach.

EarthCube Monthly Community Webinar
Thursday, December 19, 2013

This month’s EarthCube Community Webinar will feature NSF EarthCube Program Officer Eva Zanzerkia, who will give a brief overview of, and field questions about, NSF EarthCube Program Solicitation 13-529. This solicitation calls for proposals for Research Coordination Networks (estimated number of awards: 6-8) and EarthCube Building Blocks (estimated number of awards: 4-7). Proposals are due March 12, 2014. Please bring your questions. We look forward to seeing you on the 19th!

BCube Project Webinar Set for December 17
Friday, December 13, 2013

A webinar is organized to summarize the BCube kickoff meeting and introduce the BCube project to the Earthcube community at large. This webinar is scheduled for December 17 from noon to 1:00PM Eastern time (9:00 to 10:00AM Pacific time). This is a one-hour-long virtual meeting that will include an audience QA session. See the agenda and WEBEX information here.

Help Us Build a Community Catalog of Geoscience Information Resources!
Friday, December 13, 2013

The EarthCube CINERGI (Community Inventory of Earthcube Resources for Geoscience Interoperability) project is assembling a large metadata registry that will integrate information from data repositories in multiple domains, datasets submitted by researchers directly, as well as information resources such as vocabularies, ontologies, web services, software, catalogs, formal information models and schemas for different types of data.

ICCGI 2014 Call for Papers, Tutorials, and Panels
Sunday, December 1, 2013

The International Conference on Computing in the Global Information Technology (ICCGI) has announced an opportunity to submit and publish original scientific results to ICCGI 2014. Authors of selected papers will be invited to submit extended article versions to one of the journals of the International Academy, Research, and Industry Association (IARIA).

Visit the IARIA website for details on proposed topics and submission requirements. The deadline for submission is January 28, 2014.

2013 International Data Rescue Award Ceremony to be Held at AGU
Wednesday, November 27, 2013
The 2013 International Data Rescue Award in the Geosciences was created by IEDA and the Elsevier Research Data Services to recognize efforts that improve preservation and access of research data, particularly of dark data, and share the varied ways that these data are being processed, stored, and used. Contest participants will showcase their data rescue projects during the reception, to take place on Monday, December 9, 2013, 7:00-8:30 PM, at the Intercontinental Hotel on 888 Howard Street, Twin Peaks room. [More]

**BRDI Open Data Challenge Now Accepting Entries**
Friday, November 22, 2013

The NAS Board on Research Data and Information is holding an open challenge to increase awareness of current issues and opportunities in research data and information. Letters of intent are requested by Dec. 1; final entries by May 15. [more]

**EarthCube Monthly Community Webinar**
Monday, November 18, 2013

November 22, 2013 from 11:30am to 12:30pm (EST)

Starting this November, the EarthCube Monthly Community Webinar will provide an update on EarthCube awards and other EarthCube progress. This webinar is open to all who are interested in finding out what's going on in EarthCube. Please join us!

**EarthCube: NSF funds $14.5 million in grants to improve geosciences cyberinfrastructure**
Thursday, October 24, 2013

Read on for NSF’s official press release on the new EarthCube awards. [more]

**End-User PI Workshop Executive Summary Now Available**
Tuesday, October 15, 2013

The workshop report is now available! Check it out for a synthesis of needs, challenges, science and action items from previous end-user workshops, as well as potential paths forward and potential elements of a vision of success for EarthCube. [more]

**Conferences**

Source: [http://earthcube.org/page/conferences](http://earthcube.org/page/conferences)

**EGU 2014**

April 27, 2014 to May 2, 2014
The European Geosciences Union 2014 General Assembly will convene this April 27 in Vienna, Austria. EarthCube will be represented in a number of poster and oral sessions. Click here to access the full EGU session programme.

EGU2014-15681
Lee Allison and Eva Zanzerkia, “Seeking consensus for cyberinfrastructure governance in the USA.” Wednesday, April 30, 2014. 08:45-09 ORAL

EGU2014-16027

EGU2014-11611
Kerstin Lehnert, Lee Allison, David Arctur, Jens Klump, and Christopher Lenhardt, “Geosamples.org: Shared Cyberinfrastructure for Geoscience Samples,” European Geophysical Union General Assembly, Vienna, Austria, April 29, 2014 POSTER

EGU2014-4792

2014 Ocean Sciences Meeting
February 24, 2014 to February 28, 2014
EarthCube will be at the 2014 Ocean Sciences Meeting in Honolulu from February 23-28. In addition to the exhibition booth (E42), an EarthCube workshop will take place early in the week, entitled “Data at Your Fingertips: NSF’s EarthCube and the Ocean Sciences.” This workshop brings in featured speaker Deborah Bronk from NSF’s Division of Ocean Sciences, who will discuss the potential role of marine science domains in building EarthCube.

Town Hall: Data at Your Fingertips: NSF’s EarthCube and the Ocean Sciences
Date: Monday, February 24, 2014
Time: 6:30:00 PM HST
Location: 313 A

EarthCube is a direct and proactive federal funding agency response to the data deluge that is on the increase due to technological advances from the sensor to the user in all of the geosciences in general, and marine sciences in particular. The challenge facing geosciences has become how to discover, integrate, and analyze existing discipline-specific data, large or small, in order to advance necessary interdisciplinary research. Current EarthCube activities are working towards integrate these rapidly growing data streams and technologies into an open, adaptable, sustainable framework for data-intensive, multi-disciplinary geoscience research compatible with high-performance computing, large and small data sets, scalable software, and modeling.

American Meteorological Society Annual Meeting 2014
February 2, 2014 to February 6, 2014
The 94th American Meteorological Society Annual Meeting will be held 2–6 February 2014 at the Georgia World Congress Center in Atlanta, Georgia. The EarthCube outreach team will be at Booth #907 answering questions about EarthCube, holding a contest to win an iPod Nano, and handing out those coveted stress cubes. Come find us to talk about data challenges and solutions with open data, interoperability, and cross-domain data sharing!

**EarthCube at AGU**

December 9, 2013 to December 13, 2013

**EarthCube is coming to San Francisco!**

This year's [AGU Fall Meeting](#) will play host to several EarthCube-related presentations, posters, and town halls. We hope you'll be able to get involved by attending one of the scheduled sessions listed below. (For detailed presentation listings and abstracts, [click here](#).)

We'll also have a final events list and a few interesting projects going on at the EarthCube community engagement booth--don't forget to stop by!

**All Week**

**The EarthCube Data Race**

Join the EarthCube Data Team in locating as much data as possible at AGU. As you walk the poster floor and attend sessions be on the lookout for the data behind the science. Send links for any data you find to EarthCube for your chance to win devices that will enhance your digital life. Including an iPod and a Kindle Fire. [Click here](#) for more information and official rules.

**EarthCube in the Exhibit Hall**

Visit us at booth 106.

**Monday, December 9**

IN11D [Presentations]
Title: **Data Curation, Credibility, Preservation Implementation, and Data Rescue to Enable Multi-Source Science I & II**
Conveners: B. Branch, H. Ramapiyan, S. Kempler, K. Lehnert
Time: Time: 8:00 am – 12:20 am
Location: 2020 [Moscone West]

IN12A [Posters]
Title: **Data Curation, Credibility, Preservation Implementation, and Data Rescue to Enable Multi-Source Science III**
Conveners: B. Branch, H. Ramapiyan, S. Kempler, K. Lehnert
Time: 1:40 pm – 6:00 pm
IN13B [Posters]
Title: Persistent Identifiers and Citation in Earth Science Information Infrastructure
Conveners: J. de la Beaujardiere, A. P. Privette, K. Lehnert, J. Klump
Time: 1:40 pm – 6:00 pm
Location: Hall A-C [Moscone South]

V12A [Presentations]
Title: The Geoinformatics Revolution: Data-Driven Science, Databases, and Data Systems for Thermodynamics and Geochemical and Geophysical Modeling I
Conveners: M. Ghiorso, K. Lehnert, J. Bass, E. Cottrell
Time: 10:20 am – 12:20 am
Location: Room 306 [Moscone South]

V13A [Posters]
Title: The Geoinformatics Revolution: Data-Driven Science, Databases, and Data Systems for Thermodynamics and Geochemical and Geophysical Modeling II
Conveners: M. Ghiorso, K. Lehnert, J. Bass, E. Cottrell
Time: 1:40 pm – 6:00 pm
Location: Hall A-C [Moscone South]

TH12E [Town Hall]
Title: Research Data Alliance
Time: 12:30 pm – 1:30 pm
Location: Room 2006 [Moscone West]

TH12G [Town Hall]
Title: Physical Samples in the Digital Age
Conveners: K. Lehnert
Time: 12:30 pm – 1:30 pm
Location: Room 2008 [Moscone West]

TH15F [Town Hall]
Title: Connecting Data Stakeholders for a Long-term Vision of Data Stewardship
Conveners: D. Hills, S. Ramdeean, A. Wilson, C. Meyer, E. Robinson
Time: 6:15 pm – 7:15 pm
Location: Room 306 [Moscone South]

[Reception/Showcase]
Title: 2013 International Data Rescue Award in the Geosciences
Time: 7:00 pm – 8:30 pm
Location: Twin Peaks Room [Intercontinental Hotel]
Tuesday, December 10

IN23B [Posters]
Title: Collaborative Frameworks and Experiences in Earth and Space Science
Conveners: M. Allison, I. Zaslavsky, E. Law, E. Robinson
Time: 1:40 pm – 6:00 pm
Location: Hall A-C [Moscone South]

IN23D [Posters]
Title: International Cross-Project Collaboration and Interoperability of Data Management Systems
Conveners: B. Ritschel, T. Iyemori, A. Yatagai
Time: 1:40 pm – 6:00 pm
Location: Hall A-C [Moscone South]

TH22F [Town Hall]
Title: Publishing Research Data: Evolving Policies, Procedures, Best Practices
Conveners: F. Murphy, K. Lehnert, J. Tedds, S. Callaghan
Time: 12:20 pm – 1:30 pm
Location: 2006 [Moscone West]

Wednesday, December 12

IN31B [Posters]
Title: Future of Earth Science Informatics in Access, Visualization, and Analysis of Large-Scale Data I
Conveners: I. Demir, W. Krajewski, H. El-Askary, M. Kafatos
Time: 8:00 am – 12:20 pm
Location: Hall A-C [Moscone South]

[Town Hall]
Belmont Forum Town Hall Meeting
Time: 9:00 am – 11:00 pm
Location: City Room [Westin San Francisco Market Street]

IN33C [Presentations]
Title: Future of Earth Science Informatics in Access, Visualization, and Analysis of Large-Scale Data II
Conveners: I. Demir, W. Krajewski, H. El-Askary, M. Kafatos
Time: 1:40 pm – 3:40 pm
Location: 2020 [Moscone West]

IEDA Data Publication Workshop
Time: 12:15 – 1:15 pm
Location: City Room [Westin San Francisco Market Street]
TH32D [Town Hall]
Title: Building a Sediment Experimentalist Network
Conveners: L. Hsu, W. Kim, R. Martin, B. McElroy
Time: 12:30 pm – 1:30 pm
Location: 2005 [Moscone West]

COOPEUS Partners and Stakeholders General Meeting
Time: Open Session, 2:00 pm – 4:00 pm; Restricted Session, 4:00 pm – 6:00 pm
Location: San Francisco Marriott Marquis, 780 mission St., Room: Foothills G

[Town Hall]
EarthCube Town Hall Meeting
Time: 3:00 pm – 5:00 pm
Location: City Room [Westin San Francisco Market Street]

Thursday, December 12

IN41A [Posters]
Title: Emerging Concepts for Cyberinfrastructure in the Geosciences I
Conveners: M. Allison, D. Arctur, K. Lehnert, E. Robinson
Time: 8:00 am – 12:20 pm
Location: Hall A-C [Moscone South]

IN41B [Posters]
Title: Enabling Science Through Reuse of Data and Free Open Source Software I
Conveners: J. Klump, C. Mattman, R. Downs, P. Leowe
Time: 8:00 am – 12:20 pm
Location: Hall A-C [Moscone South]

IN41C [Posters]
Title: Environmental Sensor Networks and Polar Cyberinfrastructure I
Conveners: K. Martinez, C. Yang, J. Hart, W. Li
Time: 8:00 am – 12:20 pm
Location: Hall A-C [Moscone South]

IN42A [Presentation]
Title: Leptoukh Lecture
Conveners: M. Piasicki, K. Lehnert, T. Narock
Time: 10:20 am – 11:20 am
Location: 103 [Moscone South]

IN43B [Presentations]
Title: Emerging Concepts for Cyberinfrastructure in the Geosciences II
Conveners: M. Allison, D. Arctur, K. Lehnert, E. Robinson
Time: 1:40 pm – 3:40 pm
Location: 103 [Moscone South]

IN43C [Presentations]
Title: Enabling Science Through Reuse of Data and Free Open Source Software I
Conveners: J. Klump, C. Mattman, R. Downs, P. Leowe
Time: 1:40 pm – 3:40 pm
Location: 2020 [Moscone West]

IN44A [Presentations]
Title: Environmental Sensor Networks and Polar Cyberinfrastructure II
Conveners: K. Martinez, C. Yang, J. Hart, W. Li
Time: 4:00 pm – 6:00 pm
Location: 2020 [Moscone West]

TH42C
Title: Software and Research Town Hall
Time: 12:30 pm – 1:30 pm
Location: 2004 [Moscone West]

Friday, December 14

IN51A [Posters]
Title: Approaches, Architectures and Standards for End to End Brokering I
Conveners: L. Wyborn, R. Woodcock, S. Nativi, and M. Santoro
Time: 8:00 am – 12:20 am
Location: Hall A-C [Moscone South]

IN51C [Presentations]
Title: Adopting Cloud Computing for Environmental Decision Support Systems I
Conveners: A. Sun, C. Yang
Time: 8:00 am – 10:00 am
Location: 2010 [Moscone West]

IN51D [Presentations]
Title: Data Stewardship in Theory and Practice I
Conveners: C. Chandler, D. McGuinness, L. Wyborn, D. Wright
Time: 8:00 am – 10:00 am
Location: 2020 [Moscone West]

IN52B [Presentations]
Title: Semantically Enabling Annotation, Access, and Integration of Scientific Data I
Conveners: B. Wilson, X. Ma, T. Narock, P. Fox
Time: 10:20 – 12:20
Location: 2020 [Moscone West]

IN53A [Posters]
Title: Adopting Cloud Computing for Environmental Decision Support Systems II
Conveners: A. Sun, C. Yang
Time: 1:40 pm – 6:00 pm
Location: Hall A-C [Moscone South]

IN53C [Posters]
Title: Data Stewardship in Theory and Practice II
Conveners: C. Chandler, D. McGuinness, L. Wyborn, D. Wright
Time: 1:40 pm – 6:00 pm
Location: Hall A-C [Moscone South]

IN53E [Presentations]
Title: Approaches, Architectures and Standards for End to End Brokering II
Conveners: J. Pearlman, L. Wyborn, R. Woodcock, S. Nativi
Time: 1:40 pm – 3:40 pm
Location: 2020 [Moscone West]

Workshops

Source: [http://earthcube.org/page/workshops](http://earthcube.org/page/workshops)

C4P Paleobiology Workshop

May 21, 2014 to May 23, 2014

The first C4P RCN Workshop on Paleobiology will facilitate the global integration of paleobiological data from the full array of cyberinfrastructure resources, with an emphasis on sample-based information. We are inviting an array of experts from across the geoscience and cyberscientific communities. The workshop will take place May 21-23 at the Consortium for Ocean Leadership.

C4P's large scale goal is to build and grow a new community of practice, bringing together scientists, CI experts and computer scientists to articulate and address the CI challenges in the paleobiosciences and allied domains. Towards this goal, this workshop is part of a series of webinars and workshops to both gather ideas and to share them with the broader paleogeoscience community.

For the latest information, please see the [C4P Paleobiology Workshop website](http://earthcube.org/page/workshops) or contact Mark Uhen or Jack Williams.
End-user Communities & Professional Societies Assembly Workshop

March 18, 2014 to March 20, 2014

Summary:
The End-user Communities & Professional Societies workshop is part of a series of Stakeholder Assembly Workshops convened by the EarthCube Test Enterprise Governance Team. The goal of this workshop is to facilitate communication, collaboration, and coordination among the scientific and technical professional societies whose membership includes people engaged with geoscience communities. The intended outcome of this workshop is to provide a venue in which participants may determine how they might interact with each other to achieve individual goals, as well as determine how these goals might align with the long-term goals of EarthCube.

This workshop brings together two distinct EarthCube Stakeholder Assemblies:

- **Professional Societies**: This group is comprised of scientific and technical professional societies whose members constitute one of the primary foci of EarthCube. Most scientists belong to at least one professional organization; thus, this group targets professional organization administration, executives, and section chairs (e.g. AGU’s ESSI chair or GSA’s Geoinformatics Chair).

- **End-user Communities and Workshop Participants**: Beginning in June, 2012, the NSF funded a series of workshops with domain specific topics. Each of these workshops was organized by a Principal Investigator (PI) and steering committee and included between 50 and 100 attendees. This group will be comprised of representatives of that community, whether it is the PI of the organized workshop, workshop participants, or other representatives of that end-user community. This group may also include representatives of end-user communities not included in the initial workshop series.

Location:
The workshop will be held March 18-20 at the American Geophysical Union headquarters in Washington, D.C.

Registration:
Registration for this workshop is now open to the public.
If you would like to attend, please register using this registration form.

Lodging:
The Courtyard by Marriott Washington, DC/ Dupont Circle is offering a special rate of $224 per night for registered workshop participants. Please book your room using this link to ensure you receive this rate.

If you have questions regarding this workshop, please contact us.

IT/CS/FOSS Assembly Workshop

March 5, 2014 to March 7, 2014

Summary:
The IT/CS/FOSS workshop is part of a series of Stakeholder Assembly Workshops convened by the EarthCube Test Enterprise Governance Team. The mission of these events is to engage EarthCube stakeholders in the co-creation of EarthCube, and to provide resources and facilities for each stakeholder community to institute its own governance/
collaboration system for EarthCube that meets the long-term goals of the project while recognizing the diverse governance needs of each unique community.

This workshop brings together two distinct EarthCube Stakeholder Assemblies: Information Technology and Computer Sciences (CS); and Industry & Free and Open Source Software: Instrumentation, Software, and Technology Developers (FOSS).

• The IT/CS Assembly targets members of the technical communities that develop core computer and information science elements underlying EarthCube functionality, but are not the prime end-users of EarthCube. This group is designed to facilitate the communication between the primary end-users – scientific and academic research communities – and the developers that assist in the creation of data and documentation workflows.

• The FOSS Assembly is comprised of the developers and manufacturers of data collection instrumentation, data management software, and technology. The group targets the front-line manufacturers responsible for the data collection mechanisms used in cutting-edge research.

Location:
The workshop will be held March 5-7, 2014 at the Millennium Harvest House in Boulder, Colorado.

Registration:
Registration for this workshop is now open to the public.
If you would like to attend, please register using this registration form no later than February 24.

Lodging:
The Millennium Harvest House is offering a special rate of $111 per night for registered workshop participants. Please book your room using this link to ensure you receive this rate. (Reference group code: 1403EARTHI)

If you have questions regarding this workshop, please contact us.

EarthCube Portfolio Assembly Group Workshop

February 12, 2014 to February 14, 2014

**Presentation slides, summaries, reports, and group updates will be posted on the Portfolio Workspace page as they become available.

Summary:
The EarthCube Portfolio is made up of EarthCube related Research Coordination Networks, Conceptual Design awards, and Building Blocks award teams funded by the National Science Foundation (NSF) in September, 2013.
This collaboration-building workshop will be held February 12-14, 2014 at the NEON Headquarters in Boulder, CO, under the auspices of the EarthCube Test Enterprise Governance project. The event will focus on the development of a jointly authored charter representing the EarthCube Building Blocks, Research Coordination Networks, and Conceptual Design awards, at the urging of NSF. The charter will be presented to the broader community at the EarthCube All Hands Meeting next June, and used during the demonstration phase of the EarthCube Test Enterprise Governance phase in 2015. The workshop will also identify collaboration opportunities between the existing awards and help inform community engagement and outreach for EarthCube.

This is a working session and the charter will develop organically with portfolio input with guidance from social science facilitators. Currently, the agenda is set for two full days of discussion on governance issues, decision making implications, and charter development, with the final half day dedicated to collaboration techniques and opportunities. We request attendance from each of the funded projects, from either the PI, Governance Liaison, or their designee. Attendees are welcome to provide feedback on the agenda and suggest key discussion topics for the event.

**Location:**
NEON Inc., headquarters, 1685 38th St., Suite 100, Boulder, CO, in the classroom.

**Steering Committee:**
Jennifer Arrigo (CUAHSI), Jim Bowring (College of Charleston), Cecelia DeLuca (NOAA), Steve Diggs (Scripps), Siri Jodha Khalsa (NSIDC), Kerstin Lehnert (LDEO), George Percivall (OGC), Steve Richard (AZGS), Erin Robinson (ESIP), Ilya Zaslavsky (SDSC).
Workshop and Hotel Registration:
Workshop registration is available through this link. A special lodging rate at the Millennium Harvest House, 1345 28th Street, has been arranged for federal per diem of $111 per night for all attendees (reservations can be made by calling the hotel directly at 303-443-3850; a web link will be provided to attendees shortly). Accommodations include shuttle service between the hotel and NEON HQ. A light breakfast and boxed lunch will be served for all attendees. Attendees are responsible for their lodging and travel costs as part of the EarthCube coordination requirements within your awards.

Virtual Participation:
Click here for Webex details.

Data Facilities Assembly Group & End-User Workshop
January 15, 2014 to January 17, 2014

**Presentation slides, summaries, reports, and group updates will be posted on the Data Facilities Workspace page as they become available.

Summary:
This workshop will provide a forum to discuss the unique requirements and challenges associated with developing the communication, collaboration, interoperability, and governance structures that will be required to build EarthCube in conjunction with existing and emerging NSF/GEO facilities. In addition, the workshop will inform attendees of new developments within the National Science Foundation’s EarthCube initiative, including ways to align with EarthCube components.

Steering Committee: Jennifer Arrigo (CUAHSI), Tim Ahern (IRIS), Sky Bristol (USGS), Danie Kinkade (WHOI), Cindy Chandler (WHOI), Mohan Ramamurthy (UCAR), Don Middleton (UCAR), Kerstin Lehnert (LDEO), Steve Diggs (Scripps), Lee Allison (AZGS)

Agenda, in brief:
January 15, 8:30am to 5:00pm: Common Challenges
January 16, 8:30am to 5:00pm: Collaborative Solutions
January 17, 8:30am to 12:00pm: Governance

Location:
Hilton Arlington
950 North Stafford St. Arlington, VA 22203

Virtual Participation/Call-in Information:
Go to https://earthcube.webex.com/earthcube/j.php?ED=259563852&UID=1734374932&PW=NYWQ4Zjg5Mjh&RT=MixMQ%3D%3D
Call-in toll-free number (US/Canada): 1-855-244-8681
Call-in toll number (US/Canada): 1-650-479-3207

Attendee access code: 222 033 01

Workshop and Hotel Registration:
Please use this form to register for the workshop: http://bit.ly/1fn12wu. We have secured a block of rooms at the Hilton Arlington at a discounted rate for this event. To ensure you receive this rate, please make your hotel reservation no later than December 26, 2013, at this link: http://bit.ly/19yd6dH

Upcoming Assembly Workshops: Overview

The EarthCube Test Enterprise Governance Project is hosting a number of Stakeholder Assembly Workshops in the coming months. The six Assembly groups have been designed to provide the broad EarthCube community with a voice in the test governance process. If you are interested in participating in an Assembly Group and attending one of these workshops, please let the Test Governance Project Team know by filling out an expression of interest form.

EarthCube Portfolio
This group is made up of EarthCube related Research Coordination Networks, Conceptual Design awards, and Building Blocks award teams funded by the National Science Foundation (NSF) in September, 2013. This workshop will take place February 14-17, 2014 in Boulder, CO.

IT and Computer Sciences
This group is designed to facilitate the communication between the primary end-users – scientific and academic
research communities – and the developers that assist in the creation of data and documentation workflows. This workshop will occur in tandem with the FOSS workshop on **March 5-7 in Boulder, CO**.

**Industry & Free and Open Source Software (FOSS): Instrumentation, Software, and Technology Developers**

This Assembly Group is comprised of the developers and manufacturers of data collection instrumentation, data management software, and technology. The group targets the front line manufacturers responsible for the data collection mechanisms used in cutting-edge research. This event will occur in tandem with the IT/CS workshop on **March 5-7 in Boulder, CO**.

**EarthCube End-User Communities & Workshop Participants**

This group will comprise representatives from EarthCube's End-User workshops including workshop PIs, participants, or other representatives of that end-user community. This group may also include representatives of end-user communities not included in the initial workshop series. This event will take place in tandem with the Professional Societies Workshop on **March 18-20, 2014 in Washington, DC**.

**Professional Societies**

This group is comprised of scientific and technical professional societies whose members constitute one of the primary foci of EarthCube. This group targets professional organization administration, executives, and section chairs. The workshop will take place in tandem with the End-User Communities Workshop on **March 18-20, 2014 in Washington, DC**.

Contact

Source: [http://earthcube.org/contact](http://earthcube.org/contact)

If you have any questions, comments, or feedback regarding EarthCube, please use the form below to contact a project coordinator.

EarthCube Commons

Source: [http://workspace.earthcube.org/](http://workspace.earthcube.org/)

My Note: See Above

**EarthCube Shows How Collaboration Is Changing Geoscience Research**


*By Brand Niemann*

Published: November 11, 2011
Last week I attended the EarthCube Charrette with about 140 geoscientists from the National Science Foundation (NSF), US Geological Survey, academic institutions, and industry. Estimates are there are about 100,000 geoscientists in the world.

The goal of EarthCube is to transform the conduct of research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the geosciences.

The charrette is aimed at architecting the design, not just another workshop. Attendees were asked to use language understandable to the public and avoid jargon.

A website has been set up to foster community collaboration, and will provide updated information, resource documents, and discussion forums so that community groups, consortia, researchers, and educators can share ideas, introduce concepts, and find and develop collaborative efforts. All the papers were provided ahead of time and I uploaded them to my social knowledgebase page for this activity. NSF seeks transformative concepts and approaches to create integrated data management infrastructures across the Geosciences.

In a new partnership, the Geosciences Directorate (GEO) and the Office of Cyberinfrastructure (OCI) recognize the multifaceted challenges of modern, data-intensive science and education. The organizations envision an environment where low adoption thresholds and new capabilities act together to greatly increase the productivity and capability of researchers and educators working at the frontiers of Earth system science.

They also plan to adopt NSF's version of “social networking” for this discipline! All participants were encouraged to follow the guiding principles from Amazon's Jeff Bezos:

1. **All teams will henceforth expose their data and functionality through service interfaces.**

2. **Teams must communicate with each other through these interfaces.**

3. **There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever. The only communication allowed is via service interface calls over the network.**

4. **It doesn't matter what technology they use. HTTP, Corba, Pubsub, custom protocols -- doesn't matter. Bezos doesn't care.**

5. **All service interfaces, without exception, must be designed from the ground up to be extensible. That is to say, the team must plan and design to be able to expose the interface to developers in the outside world. No exceptions.**

6. **Anyone who does not do this will be fired.**

7. **Thank you; have a nice day!**

All participants were asked to answer the following question and submit a two-page description of what they would do to support EarthCube which I did.

**Question:** When EarthCube exists and is widely useful in 2021, what does a day in the life of a scientist in your field look like? Think about your: Research, Teaching, Outreach, Workforce Development, and Interaction with the greater

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
scientific community.

Answer: I will still have some form of the tools I am using now, they will just be more integrated with one another and even more connected to many sources of information and data so I can create data stories (like I do now for AOL Government) with greater ease and frequency because the time for collection and communication is lessened and the time for analysis is maximized -- actually the analysis tool facilitates the collection and communication parts.

I also provided an example of the kind of agile analysis of Earth Science data that I had done earlier in the year for the Earth Science Federation Annual Conference.

---

**Story**

**AOL Government**

![AOL Government](image)

Last week I attended the EarthCube Charrette with about 140 geoscientists from the National Science Foundation (NSF), US Geological Survey, academic institutions, and industry. Estimates are there are about 100,000 geoscientists in the world. Charrette is architecting the design, not just another workshop and attendees were asked to use language understandable to the public and avoid jargon.

The goal of EarthCube is to transform the conduct of research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences. A website has been set up to foster community collaboration, and will provide updated information, resource documents, and discussion forums so that community groups, consortia, researchers, and educators can share ideas, introduce concepts, and find and develop collaborative efforts. All the papers were provided ahead of time and I uploaded them to my social knowledgebase page for this activity.

NSF seeks transformative concepts and approaches to create integrated data management infrastructures across the Geosciences. In a new partnership, the Geosciences Directorate (GEO) and the Office of Cyberinfrastructure (OCI) recognize the multifaceted challenges of modern, data-intensive science and education and envision an environment where low adoption thresholds and new capabilities act together to greatly increase the productivity and capability of researchers and educators working at the frontiers of Earth system science. This is aimed at all three goals (see below) and is NSF’s version of "social networking" for this discipline!

All participants were encouraged to follow the Guiding Principles from Amazon’s Jeff Bezos:
1. All teams will henceforth expose their data and functionality through service interfaces.

2. Teams must communicate with each other through these interfaces.

3. There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever. The only communication allowed is via service interface calls over the network.

4. It doesn't matter what technology they use. HTTP, Corba, Pubsub, custom protocols -- doesn't matter. Bezos doesn't care.

5. All service interfaces, without exception, must be designed from the ground up to be extensible. That is to say, the team must plan and design to be able to expose the interface to developers in the outside world. No exceptions.

6. Anyone who does not do this will be fired.

7. Thank you; have a nice day!

All participants were asked to answer the following question and submit a two-page description of what they would do to support EarthCube which I did.

Question: When EarthCube exists and is widely useful in 2021, what does a day in the life of a scientist in your field look like? Think about your: Research, Teaching, Outreach, Workforce Development, and Interaction with the greater scientific community.

Answer: I will still have some form of the tools I am using now, they will just be more integrated with one another and even more connected to many sources of information and data so I can create data stories (like I do now for AOL Government) with greater ease and frequency because the time for collection and communication is lessened and the time for analysis is maximized - actually the analysis tool facilitates the collection and communication parts.

I also provided an example of the kind of agile analysis of Earth Science data that I had done earlier in the year for the Earth Science Federation Annual Conference.

The Value Proposition of Agile Analysis

Becoming more innovative with Andy Burnett, Knowinnovations, introduce yourself and tell one "stupid human trick"

My Businesss Model (in response to USGS does not procure semantics - provide analysis/analytics)
Next: Share your group’s story with others and get feedback to report to the entire group (this is the Knowinnovation facilitation process for the Charette).

Question: When EarthCube exists and is widely useful in 2021, what does a day in the life of a scientist in your field look like? Think about your: Research, Teaching, Outreach, Workforce Development, and Interaction with the greater scientific community.

Answer: I will still have some form of the tools I am using now, they will just be more integrated with one another and even more connected to many sources of information and data so I can create data stories (like I do now for AOL Government) with greater ease and frequency because the time for collection and communication is lessened and the time for analysis is maximized - actually the analysis tool facilitates the collection and communication parts.

Internet Community Building, Steve Wolff (Cisco now, former NSF, etc.)

People: Senator Al Gore, Congressman Rick Boucher, NSF Director Erich Boch, and Assistant Director Goordon Bel, Bill Wolf, Chuck Browster, Industrial collaborators (Liberman, Bowman, Wei, Armstrong), Investigation (Gordon Cook, NSF Inspector General)

NSF is a remarkable agency - it reports to the President!

Rules: Cooperative agreements, Outsourcing NSF Functions (Regional network of NSFNet, Proposal management review and award testbeds, GEN, etc.

Standards: So many to choose from, ...


A few rules of engagement: Don't stop - keep going!

My NGA - NCOIC - Google experience - where is the win-win? Google is driven by the bottom line (and does evil things).
ESIP Data.gov Application

Source: Federation of Earth Science Information Partners

See Temperature Anomalies Tab Below

Note: For full-screen display use the Web Player. Also see An Interface to a Digital Library of the Atlas of Science Submission and Slides.

Charrette Wrap Up

Productive past 3 days. Accomplished goals. Defined initial scope of EarthCube - far exceeded NSF expectations. See content on walls and web site (I picked up handouts) which will be put on the web site for all to see. Next activities for grants and collaboration. Will accept two papers afterwards: what you want to do and how you want to do it.

Two-pager Instructions:

• Page One - one or more of these topics
  ◦ 1. Strategic Organization Framework for EarthCube
  ◦ 2. Development of a New Capability
  ◦ 3. Progress on a Critical Milestone for current capabilities

• Page Two
  ◦ Scientific motivation, resultant advance and team

Post Charrette

NSF 11-085

Opportunity to Submit an EAGER

November to April

Awards are listed on the NSF Awards Website

Other funding opportunities

Additional information from the Charrette will appear on the EarthCube web pages

Next major step is to reassess and extend the capabilities as a followup event: May 2012
Example Guiding Principles from Amazon's Jeff Bezos

1. All teams will henceforth expose their data and functionality through service interfaces.

2. Teams must communicate with each other through these interfaces.

3. There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever. The only communication allowed is via service interface calls over the network.

4. It doesn't matter what technology they use. HTTP, Corba, Pubsub, custom protocols -- doesn't matter. Bezos doesn't care.

5. All service interfaces, without exception, must be designed from the ground up to be extensible. That is to say, the team must plan and design to be able to expose the interface to developers in the outside world. No exceptions.

6. Anyone who does not do this will be fired.

7. Thank you; have a nice day! Not his, but rather Rip Rowan.

Low Hanging Fruit "Eggs of a Feather"

- Concept of an EarthCube Collaboration MOU
  - Define and commit to guiding principles
  - Possible short term action items
    - Common Data Model (or the idea of some commonly understood data framework that has semantic mappings. Does this facilitate or present a barrier?)
    - Data Services (providing data that is easily understood, e.g. has semantic mappings such as ascii with a metadata header)
    - Catalog of data and services
    - Demonstrate interoperable systems
  - Inter-Institutional Collaboration Approaches
    - E.g. Engage existing NSF-funded facilities and groups
    - Develop a Formal Communication process; workshops get started
  - Define a Technological Framework for collaboration
<table>
<thead>
<tr>
<th>Title</th>
<th>Day</th>
<th>Room</th>
<th>Host First Name</th>
<th>Host Last Name</th>
<th>Title</th>
<th>Link to Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>EarthCube Manifesto - building a cultural phenomenon out of this endeavor</td>
<td>Wednesday</td>
<td>Belmont</td>
<td>Sky</td>
<td>Bristol</td>
<td>EarthCube Manifesto - building a cultural phenomenon out of this endeavor</td>
<td><a href="http://earthcube.ning.com/group/bofs_on-out-of-this">http://earthcube.ning.com/group/bofs_on-out-of-this</a></td>
</tr>
<tr>
<td>Data publication: rethinking what it means to access and share data</td>
<td>Wednesday</td>
<td>Piedmont 1</td>
<td>Shanan</td>
<td>Peters</td>
<td>Data publication: rethinking what it means to access and share data</td>
<td><a href="http://earthcube.ning.com/group/bofs_cess-and-share">http://earthcube.ning.com/group/bofs_cess-and-share</a></td>
</tr>
<tr>
<td>EarthCube as an engine for augmented reality systems</td>
<td>Wednesday</td>
<td>Piedmont 2</td>
<td>Michael</td>
<td>Tuite</td>
<td>EarthCube as an engine for augmented reality systems</td>
<td><a href="http://earthcube.ning.com/group/bofs_ality-systems">http://earthcube.ning.com/group/bofs_ality-systems</a></td>
</tr>
<tr>
<td>Education and training in Geoinformatics</td>
<td>Wednesday</td>
<td>Chrysialis</td>
<td>Chaitan</td>
<td>Baru</td>
<td>Education and training in Geoinformatics</td>
<td><a href="http://earthcube.ning.com/group/bofs_geoinformatics">http://earthcube.ning.com/group/bofs_geoinformatics</a></td>
</tr>
<tr>
<td>Seriously, can we really do semantics for all of GEO?</td>
<td>Wednesday</td>
<td>Sully</td>
<td>Peter</td>
<td>Fox</td>
<td>Seriously, can we really do semantics for all of GEO?</td>
<td><a href="http://earthcube.ning.com/group/bofs_for-all-of-geo">http://earthcube.ning.com/group/bofs_for-all-of-geo</a></td>
</tr>
<tr>
<td>Professional networking (and social media and social networking)</td>
<td>Thursday</td>
<td>Belmont</td>
<td>Karen</td>
<td>Remick</td>
<td>Professional networking (and social media and social networking)</td>
<td><a href="http://earthcube.ning.com/group/bofs_nal-networking">http://earthcube.ning.com/group/bofs_nal-networking</a></td>
</tr>
<tr>
<td>Geo-spatial data discovery tools and ideas BOF</td>
<td>Thursday</td>
<td>Piedmont 1</td>
<td>Andy</td>
<td>Maffei</td>
<td>Geo-spatial data discovery tools and ideas BOF</td>
<td><a href="http://earthcube.ning.com/group/bofs_overy-tool-bof">http://earthcube.ning.com/group/bofs_overy-tool-bof</a></td>
</tr>
</tbody>
</table>
Enable user access to the scientific content of data

<table>
<thead>
<tr>
<th>Thursday</th>
<th>Piedmont 2</th>
<th>Michael</th>
<th>Ruohoniemi</th>
</tr>
</thead>
</table>

Computer scientists sharing experiences working with specific science domains

<table>
<thead>
<tr>
<th>Thursday</th>
<th>Chrysialis</th>
<th>James</th>
<th>Bowing</th>
</tr>
</thead>
</table>

Complex networks trumps traditional governance - power to the people

<table>
<thead>
<tr>
<th>Thursday</th>
<th>Sully</th>
<th>Peter</th>
<th>Fox</th>
</tr>
</thead>
</table>

Note: There are 15 at: http://earthcube.ning.com/group/bofs/forum?groupUrl=bofs&%2F=id=6435147%3AGroup%3A7137&page=1

Notes

NSF Overview: Alan Blatecky and Tim Killeen and Setting the Stage: Cliff Jacobs, Eva Zanzerkia, Jennifer Schopf

There are lots of NSF folks here. This is a Charette (not a workshop) which is to architect/design the EarthCube to be an integrated cyberinfrastructure

There are 111 white papers on design and use cases - some even provide full designs.

In 6-9 months we will have a "sandpit" - an ideas activity to populate the roadmap.

EarthCube is a community-led and community-architected activity supported by multiple agencies - NSF, NASA, NOAA, and USGS.

We will go through several iterations to get to a sustainable model with buy-in ownership by narrowing the options.

The NSF Geosciences Program is $100 M/year program that wants to create more opportunities for multi-disciplinary work through GEO and OCI Partnership: Tim Killeen, GEO, and Alan Blatecky, OCI

We want to get to an Internet for Understanding Complex Systems.

Knowinnovation.com is facilitating this for us.
Open to questions from the audience.

My question is which of the three goals is this aimed at:

Answer: Aimed at all three goals:

1. Define new research program and solicitations
2. Better focus existing research programs
3. Better integrate research programs across the multiple agencies mentioned.

NSF’s version of "social networking" for this discipline!

Breakouts: Earth, Atmosphere, and Oceans

In your teams present the capability clusters which your team requires from EarthCube

Keep a composite list of capabilities your scientific field requires from EarthCube

Discuss these (in subgroups) and develop a list of barriers we must overcome in order to achieve these capabilities

Capture a composite list of challenges

List Capabilities and Barriers

Present a list of Capabilities we want EarthCube to have

A list of Barriers we face and must overcome

Welcome!

- Posted by EarthCube Project Team on September 12, 2011 at 11:00am
- View Blog

Welcome to EarthCube! The goal of EarthCube is to transform the conduct of research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management across the Geosciences, and is described in a white paper on EarthCube.

The design and implementation of EarthCube will require leadership from and continual engagement with the community. This website offers the opportunity to express to NSF your present and future needs for research and your interest in attending the EarthCube Design Charrette, which is planned for Nov. 1-4 in the Washington, DC area. We are gathering initial use scenarios and requirements through a geosciences user questionnaire https://www.surveymonkey.com/s/GeoSciRequirements. The EarthCube Design Charrette will be a community event to share ideas, enhance or create partnerships, and for focused discussions on EarthCube. Read more about NSF expectations for the EarthCube Charrette and guidance to the community to prepare for the
event in the White Paper available Here or register your interest in attending the Charrette at https://www.surveymonkey.com/s/EarthCubeInterest.

In addition, we have formed initial discussion groups on the EarthCube website to encourage and facilitate a dialog among community members in the following areas:

- **Science Requirements**, which will include not only use scenario requirements that EarthCube should strive to satisfy but also discussions of approaches to reach a wide set of users and meet these needs;
- **Technology and CI Solutions**, which deals with the components of EarthCube; and
- **EarthCube Designs**, which will develop a partial or complete set of design concepts that will characterize EarthCube.

The intent of these groups is to stimulate a robust dialog in all three areas. All three groups allow community members to upload white papers and comment on ideas openly. Just register for each group you want to follow and post away. More groups will be added as the community identifies needs for additional areas of concentrated discussion.

This is an exciting time for Office and Cyberinfrastructure (OCI) and the Geosciences Directorate (GEO), and EarthCube represents a major opportunity to strengthen the vital partnership between OCI and GEO. EarthCube will contribute to the goals of Cyberinfrastructure for the 21st century, which is a major NSF research thrust, and will create an unprecedented opportunity for the geosciences to come together to enable a transformation in research practices. We hope to see many of you online and at the Charrette in November.

Sincerely, Tim Killeen, Assistant Director, GEO, and Alan Blatecky, Director, OCI

---

**Draft Charrette Agenda**

**Day 1, Tuesday Nov 1 From Here to EarthCube: A Call to Action**;

End of Day 1 Goal: Commonalities and differences in vision and challenges

7:00 Registration available

8:00-8:30 Gathering your thoughts

8:30 Welcome

8:35-8:55 NSF Overview: Alan Blatecky and Tim Killeen

8:55-9:15 Setting the Stage: Cliff Jacobs, Eva Zanzerkia, Jennifer Schopf

9:15-9:45 Instructions for first breakouts (Andy)

9:45-12:00 First breakout - Vision to capability to obstacle (Break as needed)
12:00-1:30  Lunch on your own
1:30-2:30  Internet Community Building, Steve Wolff
2:30-3:45  Breakout - Capabilities and obstacles
4:00-5:30  Report back from breakouts, next steps
5:30-7:30  Reception

Day 2, Wednesday Nov 2 The Challenges to Realizing EarthCube
End of Day 2 Goal: Explicit statements of challenges and possible milestones

7:30  Coffee available
8:00-8:30  How do yesterday’s challenges map in NSF’s mind. Commonalities with whitepapers
8:30-10:00  Breakouts – Either discussion of challenges, or holistic challenge-timeline mapping
10:00-10:30  Break
10:30-12:00  Report back from breakouts
12:00-1:30  Lunch on your own
1:30-4:00  Breakouts: What solutions exist for each of the challenges/NSF design
4:00-5:00  Breakout report back
5:00-5:30  Concluding Day 2
6-8pm  Evening white paper talks, Birds of Feathers

Day 3, Thursday Nov 3 Mapping vision, challenges, and milestones to governance and designs
End of Day 3 goal: Understanding the EarthCube basics

7:30  Coffee available
8-12  Plenaries and breakouts map vision, challenges, and milestones to solutions and possible EarthCube Design Approaches
12:00-1:30  Lunch on your own
Day 4, Friday Nov 4 Moving Forward with EarthCube

End of Day 4 goal: Rough understanding of what next steps must address

8:30-12:00 Clarification on next steps and paths forward

Papers

Total: 111 (see attached below)

Note: Work in process to put extract from each into the wiki to understand where the participants are thinking and proposing.

EarthCube White Paper: Advancing Scientific Understanding by Communicating via Data Interactive Publications

Last modified: 17 October 2011
Ben Domenico
Unidata Program Center University Corporation for Atmospheric Research (UCAR) Sponsored by NSF Atmospheric and Geospace Sciences (AGS)
(Note: This EarthCube whitepaper is an abridged and generalized version of an earlier whitepaper prepared for the Unidata Policy Committee written before the author was aware of the Earthcube initiative. The online version of the paper illustrates the data interactivity aspect of the publication, whereas the Microsoft Word and PDF versions have some live links but the server processing that generates the display figures is not run "on the fly." Printed copies of course are just printed copies. The white paper is also referenced by the OpenGeospatial Consortium as OGC Document 11-146. The reader is encouraged to work with the online version of the whitepaper at https://sites.google.com/site/datain...e-publications)

Research Data Lifecycle Management as a Service

Ian Foster and the Globus Online team (http://www.globusonline.org)
Computation Institute
University of Chicago and Argonne National Laboratory

Introduction

Big increases in data generated within research laboratories and demands for more careful data management lead to increased pressure on investigators. Researchers need not storage, but full-service data lifecycle management processes, encompassing data collection, storage, sharing, metadata, search, archiving, provenance, assignment of
DOIs, security, etc. Establishing such processes would demand substantial time and resources that most researchers do not have, and cannot easily acquire.

We believe that the solution to this problem is not simply to define “best practices”—nor to provide researchers with software. Once defined, best practices must still be implemented. Software still must be installed, operated, and maintained. Those implementation, installation, and operations steps are precisely where many investigators run into problems. Instead, we should aim to outsource the entire lifecycle management process to a third party Research Data Lifecycle Management service.

White Paper: National Data Infrastructure for Earth System Science

Reagan W. Moore
Arcot Rajasekar
Mike Conway
University of North Carolina at Chapel Hill
Wayne Schroeder
Mike Wan
University of California, San Diego

A national data infrastructure is emerging through the federation of data management infrastructure used by National Science Foundation research projects, the National Oceanic and Atmospheric Agency archive, and NASA data grids. The approach is based on the use of collection virtualization technology to enable interoperability between existing data management systems. The technology that enables federation is the iRODS integrated Rule-Oriented Data System [1,2,3]. The capabilities provided by the iRODS data grid are being used today to implement national data infrastructure in the United States, Europe, the Far East, and Australia. The central concept is the realization that collaboration requires the formation of a shared collection. Researchers rely upon a common name space for identifying files, common semantic terms to understand the context associated with each file, shared procedures for manipulating and analyzing the data, and a shared consensus on the policies for managing the data. The shared context enables a designated community to understand and use the shared collection.

EarthCube Governance Whitepaper

Realizing expectable returns on EarthCube investments in community building and democratic governance.

Bruce Caron, PhD
bruce@nmri.org
New Media Research Institute, Santa Barbara, CA
September 2011

What follows is a white paper pressed mainly from some years of reflections on the Cybersocialstructure blog (http://cybersocialstructure.org).

1. The "goods" of community-based governance
One of the "rst conversations I have with people who have been tasked to build or manage a virtual organization centers on the cost/bene"t issues of democratic governance. Given the usual shortage of funding and time, they have real concerns about the effort required to build a community-based governance system. These concerns are usually layered on top of the more general concern that the community (or rather, certain activists within the community) may use the governance system to push the organization's goals toward their own interests.
What is the Technology Scope for Building a Geospatial Cyberinfrastructure to Support EarthCube? A white paper for EarthCube revised from a review paper Chaowei Yang (GMU), Robert Raskin (NASA JPL), Michael Goodchild (UCSB), Mark Gahegan (Auckland)

1. Introduction
A Cyberinfrastructure (CI) is a combination of data resources, network protocols, computing platforms, and computational services that brings people, information, and computational tools together to perform science or other data-rich applications in this information-driven world (NSF 2007 and NSF 2009). Most science domains adopt intrinsic geospatial principles (such as spatial constraints in phenomena evolution) for large amounts of geospatial data processing (such as geospatial analysis, feature relationship calculations, geospatial modeling, geovisualization, and geospatial decision support). Geospatial CI (GCI) refers to CI that utilizes geospatial principles and geospatial information to transform how research, development, and education are conducted within and across science domains (such as the environmental and Earth sciences). GCI is based on recent advancements in geospatial science, information technology, computer networks, sensor networks, Web computing, CI, and e-research/escience. GCI provides significant improvements to how the sciences that need geospatial information will advance. The evolution of GCI will produce platforms for geospatial science domains and communities to better conduct research and development and to better collect data, access data, analyze data, model and simulate phenomena, visualize data and information, and produce knowledge. To achieve these transformative objectives, collaborative research and federated developments are needed for the building a sustainable CI to support the new Earth sciences, geospatial applications, and education in the 21st century. This white paper is an excerpt of a full review paper about the research, development, education, and other efforts that have contributed to building GCI in terms of its history, objectives, architecture, supporting technologies, functions, application communities, and future research directions.

An Interactive Data Mining Framework for EarthCube
Joseph B.H. Baker and J. Michael Ruohoniemi
Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg Virginia
Naren Ramakrishnan
Department of Computer Science, Virginia Tech, Blacksburg, Virginia

Vision: An EarthCube Data Mining Framework for Geosciences Research and Life-Long Learning

The goal of EarthCube is to transform the conduct of geosciences research by supporting the development of community-guided cyberinfrastructure to integrate data and information for knowledge management. Compiling the multitude of geosciences datasets into one single EarthCube* will undoubtedly increase scientific productivity by making it easier for researchers to access multiple datasets quickly. However, a major challenge in geosciences is the reliable extraction of recurrent features from the massive archive of multidimensional datasets, many of which exhibit a large degree of spatiotemporal sparseness. This challenge exists today, and will only become much more pronounced once EarthCube is fully operational. It is our thesis that the majority of geosciences datasets are currently under-utilized - not simply because of issues with data access and distribution - but rather because the geosciences community does not yet have sufficient computational capabilities in automated event detection and feature classification to extract the fullest quantitative information from the datasets. What is really required to allow EarthCube to become more than just a comprehensive data clearing house is the development of sophisticated new tools for interactive visualization and mining of multiple datasets for physical content. Such an effort will require an intellectual partnership between the geosciences and computer science research communities. Several existing databases can function as test-beds for development of new EarthCube data mining algorithms, allowing the effort to proceed in parallel with the roll-out of EarthCube cyberinfrastructure. Once fully incorporated into EarthCube the new algorithms will automate extraction of important recurrent features across multiple datasets and thereby improve scientific productivity. Furthermore, the new data mining framework will also function as a flexible machine learning environment for students of all ages, allowing them to become "citizen scientists" in much the same way that the internet has enabled the rise of a generation of

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch
“citizen journalists” (i.e. bloggers). In summary, if EarthCube is to be fully successful in producing transformative change in the conduct of geosciences research and education, then the design needs to explicitly include plans to develop a dedicated interactive framework for cutting-edge data mining.

Geosciences: An Emerging Front in Data Analytics

Use Cases to Test OGC O&M Profile

Hooper

Background
As we develop experience in having scientists both use CUAHSI HIS and publish data in CUAHSI HIS, we are finding some difficulties arising in interpreting the data when a simple location/time/parameter name isn’t adequate to describe the data. Scientists need to know more about the context of the data collection—this could include the intent of the scientists who collected the data (what does this measurement represent to the scientist?) and the environmental setting of the measurement (what is the “site type”?).

This issue has not been very fully explored with the HIS project because we have considered the stream gage as the archetype of our measurement location. In this regard, the most important underlying independent variables describing the gage data are i) location, ii) sampling time. A gage on a river provides a great deal of context by its location—one can readily understand the position of the gage in a river network by looking at a map, streamflow data is understood to reflect hydrologic conditions at that point of the river and to convey information about the hydrologic status of the watershed it drains. The terrestrial setting is defined by properties of that watershed, such as land cover.

However, as we extend to different measurement types, the structure and efficient transmission of information about the environmental context and interpretation of the data differ for each type of data. The attached “flower” diagram from the Critical Zone Exploration Network (www.czen.org) is a useful depiction of the range of data types that exist to describe the critical zone.

The intent of this discussion piece is to pose some use cases to see how the Open Geospatial Consortium’s Observations and Measurement (O&M) profile can be adapted to capture this broader contextual information about data.

A COMMUNITY MODEL FOR WATER IN THE CONTINENTAL EARTH SYSTEM

Larry Murdoch (Clemson U.), Jay Famiglietti (U.C. Irvine), Larry Band (U. North Carolina), Venkat Lakshmi (U. South Carolina), Rick Hooper (CUASHI), Ying Fan Reinfelder (Rutgers U.), Witold Krajewski (U. Iowa), Chris Duffy (Penn State)

Oct 2011

SUMMARY

The ability to simulate water in the continental Earth system is scattered among myriad disciplines and this lack of focus is impeding scientific progress. We suggest that EarthCube should develop a Community Water Model designed to bring the scientific community together to link data and computations that will provide the best available simulations of the hydrologic cycle and coupled water-related processes. It will be an open-source code managed by the scientific community and its capabilities will grow through contributions from the community. The Community Water Model can serve as the connection point for other disciplines that require prediction of water movement, including
general atmospheric and ocean circulation, landform evolution, biogeochemical, vegetation, ecological, geodetic and economic models.

Atmospheric Sciences and Informatics

EarthCube Driver Whitepaper: Technical Infrastructure
Beth Plale1, Rich Clark2, Craig Mattocks3, Keith Brewster4, Rebecca Barthelmie5, Kelvin Droegemeier6, Dennis Gannon7, Sara Graves8, Scott Jensen1, William Mahoney9, Sara Pryor5, Rahul Ramachandran8, Robert McDonald9, Mohan Ramamurthy10, Bob Wilmington11, Ming Xue4, Sepi Yalda2
1 School of Informatics and Computing and Data To Insight Center, Indiana University
2 Dept. of Earth Science, Millersville University
3 Rosenstiel School of Marine & Atmospheric Science, University of Miami
4 Center for Analysis and Prediction of Storms (CAPS), University of Oklahoma
5 Dept. of Geography, Indiana University
6 School of Meteorology, University of Oklahoma
7 Microsoft Research
8 Information Technology and Systems Center (ITSC), University of Alabama
9 National Center for Atmospheric Research (NCAR)
10 Indiana University Libraries and Data To Insight Center, Indiana University
11 Dept of Atmospheric Science, University of Illinois and NCSA
10 October 2011

1. Introduction
In response to the EarthCube call for transformative concepts and approaches to create the next generation of integrated data management infrastructures across the Geosciences, the above team of atmospheric, climate, emergency management, transportation, and informatics researchers has defined principles and technologies for technical infrastructure that are described in this document. We have defined a set of science scenarios as well, and these appear in an accompanying whitepaper by the same set of authors. The authors draw in part from the NSF large ITR Linked Environment for Atmospheric Discovery (LEAD) (2003-2009)(Droegemeier et al. 2005). LEAD was one of the pioneers of the Science Gateway, a portal serving a community of researchers and educators that was one of the first to bring high performance computing resources into the hands of users in an on-demand way. The science gateway concept became so successful that in 2010, it supported 30% of all TeraGrid users1. While the LEAD cyberinfrastructure was very successful in demonstrating key advances in technology, it is not being proposed as a solution here. This whitepaper does not advocate specific technologies per se. It represents new thinking, new people, new research questions drawn from our vast experience in data driven science, largely in the atmospheric sciences, but relevant to other geosciences as well.

The Brokering Approach for Earth Science

Cyberinfrastructure
Stefano Nativi1, SiriJodha Khalsa2, Ben Domenico3, Max Craglia4, Jay Pearlman5, Paolo Mazzetti1, Russ Rew3
1 National Research Council of Italy -Institute of Atmospheric Pollution Research (CNR-IIA)
2 National Snow and Ice Data Center (NSIDC), University of Colorado
3 Unidata, University Corporation for Atmospheric Research (UCAR)
4 EC-Joint Research Center (JRC) -SDI Unit
5 IEEE

Abstract
There are significant benefits in relieving individual information systems from the complexities of implementing interoperability in the environment where interdisciplinary cooperation is key to scientific understanding. These include increased information access and use, sustainability of crossdomain discovery and facilitating future technology insertion in a consistent manner. The Brokering approach to interoperability uses a central mechanism to converge disparate vocabularies, promote translatable standards and enable uniformity of search and access in divergent operating environments. The broker approach and implementation is maturing rapidly with embedded capabilities to include new technologies. Already, the Brokering approach can address web 2.0 interfaces, support for model webs and advanced information flows. As a proposed element of an Earth Cube Data Infrastructure, we argue that the Broking approach supports the current and future needs of the science and applications communities.

DataONE—Enabling Cyberinfrastructure for the Biological, Environmental and Earth Sciences

William K. Michener1,2, Rebecca Koskela1,2, Matthew B. Jones2,3, Robert B. Cook2,4, Mike Frame2,5, Bruce Wilson2,4, and David A. Vieglais2,6

1 University Libraries, MSC04 2815, University of New Mexico, Albuquerque, NM 87131-0001
2 DataONE, University of New Mexico, Albuquerque, NM 87131-0001
3 NCEAS, University of California Santa Barbara, Santa Barbara, CA 93101
4 Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6301
5 U.S. Geological Survey, Core Science Systems, Biological Informatics Program, Building 1916T2, 230 Warehouse Road: Mailing Address: P. O. Box 6015, Oak Ridge, TN 37831
6 Biodiversity Research Center, University of Kansas, Lawrence KS 66045-7593

Keywords

Data centers, Federated data systems, User-centered design, Analysis, Data integration, Data life cycle

Introduction

The scope and nature of biological, environmental and Earth science research are evolving in response to environmental challenges such as global climate change, invasive species, and emergent diseases. In particular, scientific studies are increasingly focusing on long-term, broad-scale, and complex questions that require massive amounts of diverse data collected by remote sensing platforms and embedded environmental sensor networks; collaborative, interdisciplinary science teams; and new approaches for managing, preserving, analyzing, and sharing data. Here, we describe the design of DataONE (Data Observation Network for Earth)—a cyberinfrastructure platform developed to support rapid data discovery and access across diverse data centers distributed worldwide and designed to provide scientists with an integrated set of familiar tools that support all elements of the data life cycle (e.g., from planning and acquisition through data integration, analysis, and visualization). Ongoing evolution of the DataONE architecture is based on participatory, user-centered design processes including: (1) identification and prioritization of
stakeholder communities; (2) developing an understanding of their perceptions, attitudes, and user requirements; (3) usability analysis and assessment; and (4) engaging science teams in grand challenge science exemplars.

OGC Document 11-145 Cyberarchitecture for Geosciences

Amy Apon, Clemson Univ. Lorenzo Bigagli, CNR-IIA
Jack Cothren, CAST Phillip Dibner, Ecosystems
Gerd Heber, The HDF Group Craig Lee, Aerospace
Fred Limp, CAST Yong Liu, NCSA
Jay Pearlman, IEEE Sven Schade, JRC
Ingo Simonis, IGSi Don Sullivan, NASA
Dawn Wright, Esri/Oregon St. Yan Xu, Microsoft
Editor: George Percivall, OGC

Introduction
The National Science Foundation (NSF) is developing “EarthCube” - Towards a National Data Infrastructure for Earth System Science1. In a new partnership between GEO and OCI, NSF seeks transformative concepts and approaches to create a sustained, integrated data management infrastructure spanning the Geosciences. Meeting the challenges in geoscience research requires innovation and paradigm shifts in cyberinfrastructure. Information technology must advance to meet the emerging approaches to science. A cyber-architecture identifies repeatable patterns, reusable components, and open standards that provide starting point for innovative developments.

This white paper was written by Open Geospatial Consortium (OGC) members and associates to contribute to development of the NSF EarthCube. This document does not represent an official position of the OGC. However, the discussions in this document could very well lead to NSF developments and subsequent OGC documents. Recipients of this document are invited to reply to the authors’ with notification of any relevant patent rights of which they are aware and to provide supporting documentation.

OGC Document 11-143 Governance: An OGC white paper for NSF EarthCube

Editors: David Arctur, OGC
Robert Downs and Robert Chen, CIESIN
Stefano Nativi, CNR-IIA
David Schell, OGC

Scope
This white paper describes a proven framework of tools, policies, and procedures for governance of long-term, interdisciplinary and multi-sector collaboration for high-impact scientific research for societal benefit. In fields of scientific research the systems, frameworks and issues to be governed can vary considerably, ranging from choices of abstract information models and taxonomies, to operational procedures for data management and curation, and even to enabling adaptive evolution of the framework itself due to changing conditions and requirements. To accomplish and coordinate research, prototyping, and deployment, governance processes include patterns for describing complex systems, coordinating across institutional boundaries, forming and managing working groups, enabling rapid mobilization of research findings into broader community applications, developing compliance guidelines, advancing the design and lifecycle of relevant standards and best practices, and many other functions. An effective approach must consider cultural, social, legal, and other institutional issues, including trust, uncertainty, authentication and authorization for discovery and access. The approaches used for governance of the research process can profoundly influence all facets of the NSF EarthCube project. This white paper presents an approach based on the governance process used in the Open Geospatial Consortium which has achieved broad usage and trust across many disciplines and business sectors (government, commercial, academic, and other non-profit).
A critical next step for OGC is to be more closely engaged and further influenced by the academic community, so that further development of standards-based architecture and communications is increasingly driven by the research community for purposes of societal benefit. OGC not only serves the research process but is ready to be embedded within the research community to become part of the core technology and institutional infrastructure, and a resource for advancing research objectives.

This white paper was written by Open Geospatial Consortium (OGC) members and associates to contribute to development of the NSF EarthCube. This document does not represent an official position of the OGC. However, the discussions in this document could very well lead to NSF developments and subsequent OGC documents. Recipients of this document are invited to reply to the authors notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Example Use Case in Climate Impacts Research

National Center for Atmospheric Research

The central element of the system we’re envisioning is methodology upload. Our community already has pretty solid cyberinfrastructure in place for the sharing of geoscience data. For example, in NARCCAP[1], we’re publishing CF-compliant NetCDF data through ESG[2], and that seems like a generally effective solution. It doesn't achieve full interworkability, but it feels like the core elements all exist and are evolving in the right direction. What we propose here is the development of a compatible system to handle the sharing of analysis and visualization, plus an affiliated layer of opinion and interpretation.

This is an illustrative story about how such a software system would work and how a relative novice would use it.

Atmospheric Sciences and Informatics EarthCube Driver Whitepaper: Use Cases

Beth Plale1, Rich Clark2, Craig Mattocks3, Keith Brewster4, Rebecca Barthelmie5, Kelvin Droegemeier6, Dennis Gannon7, Sara Graves8, Scott Jensen1, William Mahoney9, Sara Pryor5, Rahul Ramachandran8, Robert McDonald10, Mohan Ramamurthy11, Bob Wilhelmson12, Ming Xue4, Sepi Yalda2
1 School of Informatics and Computing and Data To Insight Center, Indiana University
2 Dept. of Earth Science, Millersville University
3 Rosenstiel School of Marine & Atmospheric Science, University of Miami
4 Center for Analysis and Prediction of Storms (CAPS), University of Oklahoma
5 Dept. of Geography, Indiana University
6 School of Meteorology, University of Oklahoma
7 Microsoft Research
8 Information Technology and Systems Center (ITSC), University of Alabama
9 National Center for Atmospheric Research (NCAR)
10 Indiana University Libraries and Data To Insight Center, Indiana University
11 Unidata
12 Dept of Atmospheric Science, University of Illinois and NCSA

1. Introduction
In response to the EarthCube call for transformative concepts and approaches to create the next generation of integrated data management infrastructures across the Geosciences, the above team of atmospheric, climate, emergency management, transportation, and informatics researchers has assembled a set of science scenarios that are described in this document. We have also described a technical infrastructure, and this appears in an accompanying whitepaper by the same set of authors. The authors draw in part from the NSF large ITR Linked Environment for Atmospheric Discovery (LEAD) (2003-2009)(Droegemeier et al. 2005). LEAD was one of the pioneers of the Science
Gateway, a portal serving a community of researchers and educators that was one of the first to bring high performance computing resources into the hands of users in an on-demand way. The science gateway concept became so successful that in 2010, it supported 30% of all TeraGrid users. While the LEAD cyberinfrastructure was very successful in demonstrating key advances in technology, it is not being proposed as a solution here. This whitepaper does not advocate specific technologies per se. It represents new thinking, new people, new research questions drawn from our vast experience in data driven science, largely in the atmospheric sciences, but relevant to other geosciences as well.

Unidata Governance: A Quarter Century of Experience

October 2011
1 This white paper draws mainly upon the experiences of Unidata, a community-driven data facility for the geosciences. Some of the material in this paper is based on previously published information.
Unidata Program Center
UCAR Office of Programs P.O. Box 3000 Boulder, CO 80307-3000
Mohan Ramamurthy, Director

1 Introduction
The overarching goals of EarthCube are to build a unified, adaptive, and scalable cyberinfrastructure framework for enabling transformative advances in geosciences research and education and realizing the vision articulated in the Geo Vision report. In the process, EarthCube aims to create a knowledge management system and infrastructure that integrates all Earth system and human dimensions data in an open, transparent, and inclusive manner.

The project of building a user-driven framework that addresses the computational, data, software, knowledge management, and user services needs of the diverse geoscience community is enormous. But the biggest challenges may be sociological rather than technical. Broad adoption and use of the resulting system will require users to adopt new methods in place of their current workflows. EarthCube’s governance model must recognize and address the need to bridge cultural differences between disparate disciplines and developer as well as user communities, bringing all stakeholders together to create system-wide solutions to community problems.

Effective governance for EarthCube will:
a) actively engage its diverse users
b) provide strong leadership and oversight of the project to forge close cooperation, coordination, and collaboration among distributed development activities and the principal EC groups
c) facilitate alignment of program plans and priorities with the needs of the community
d) help the successful execution of the mission, meeting stakeholder obligations

This whitepaper describes Unidata’s governance structure, process, and key elements, and illustrates how that governance has successfully stewarded the program and provided careful oversight for it as the community broadened. The paper concludes with a suggestion on how Unidata’s experiences with community-driven governance, with appropriate changes, may also benefit the EarthCube project. At the same time, we recognize that there exist other models and implementations of governance within the scientific cyberinfrastructure arena, notably, the governance structures for Open Geospatial Consortium, TeraGrid, Open Grid Forum, and the Federation of Earth Science Partners.

OGC Document 11-147 Sensor Webs: An OGC white paper for NSF EarthCube

Editors: Liping Di, GMU
Scott Bainbridge, GBROOS
Luis Bermudez, OGC
Tony Fountain, UCSD
Janet Fredericks, WHOI
Gerd Heber, The HDF Group
Yong Liu, NCSA
Abstract
This white paper covers the topic of Sensor Webs for providing information to National Science Foundation (NSF) EarthCube program. The paper reviewed the current status of Sensor Web technology, discussed the requirements for geospatial domain, and proposed a solution to advance existing NSF funded capabilities for building an integrated Earth observation Sensor Web System to support Earth system science research and education.

An EarthCube Design Process: Unidata’s Perspective
October 2011
Ethan Davis, Douglas Dirks, Linda Miller, and Tom Yoksas
Unidata Program Center
UCAR Office of Programs P.O. Box 3000 Boulder, CO 80307-3000
Mohan Ramamurthy, Director

1 Unidata’s Experience
For over a quarter of a century, UCAR’s Unidata Program Center has served a diverse, worldwide geoscience community of researchers and educators. Unidata works to advance data services and tools that help realize the community’s vision of Geoscience at the Speed of Thought.
In an era of increasing data complexity and volume and need for multidisciplinary integration, the Unidata community has developed and profited from a rich set of data services and tools. The Unidata Program Center leads this community effort by:
- Exploring new technologies, technological standards, and tools that affect the geosciences community
- Supporting tools developed by others and developing tools to meet unmet needs
- Advocating for free and open access to geosciences data on behalf of the community by serving as liaison to data providers
- Providing solutions in new and creative ways
- Working for open standards, interoperability, and open-source approaches

This whitepaper presents Unidata’s perspective on the EarthCube project design process, based on our experience providing community-centered cyberinfrastructure.

Technology Solutions for Scientific Data Interoperability: Unidata’s Perspective
October 2011
Russ Rew
Unidata Program Center
UCAR Office of Programs P.O. Box 3000 Boulder, CO 80307-3000
Mohan Ramamurthy, Director

1 Introduction
The Unidata Program has been navigating the rapids of technological change for over twenty-five years, while building and delivering cyberinfrastructure solutions for a growing community of researchers and educators. From its perspective as a small organization providing software, services, and support to a growing geosciences community, Unidata is in a
position to offer recommendations for applying technology solutions to overcome barriers that hamper interdisciplinary research.

To better understand the nature of such barriers, consider a scenario involving multiple islands, each with their own ocean model. The models are tailored for the needs of island inhabitants, so they differ in various respects, such as which physical quantities are output, how time is represented, and what spatial coordinate systems are used.

On each island, an evolving collection of software applications depend on local model outputs, satisfying specific needs of the island’s inhabitants and decision makers. These applications also make use of archives of past observations and model outputs to detect trends and derive information that will be useful for estimating fish populations, seasonal wave heights, potential for beach erosion, and other such useful knowledge.

The plot thickens when a “cross-island” researcher needs to perform an analysis that requires accessing output from multiple models to create a single visualization.

Approaches to solving such interoperability problems include:

---

**Provenance in Earth Science Cyberinfrastructure: A White Paper for NSF EarthCube**

Liping Di and Peng Yue  
Center for Spatial Information Science and Systems (CSISS)  
George Mason University  
4400 University Drive, MS 6E1, Fairfax, VA 22030  
Phone: 703-993-6114; Fax: 703-993-6127  
Email: ldi@gmu.edu, http://csiss.gmu.edu

1. **Introduction**

Provenance, also called lineage, records the derivation history of a data product (Figure 1). The history could include the algorithms used, the process steps taken, the computing environment run, data sources input to the processes, the organization/person responsible for the product, etc. Provenance provides important information to data users for them to determine the usability and reliability of the product. In the science domain, the data provenance is especially important since scientists need to use such information to determine the scientific validity of a data product and to decide if such a product can be used as the basis for further scientific analysis. It can be further used to address a series of cyberinfrastructure-related issues, including transparency in data sharing and processing, proper credits to data and algorithm contributors, and reproducibility and trust-ability of scientific results.

---

**Semantic Web technology-driven Geosciences Network**

Hassan A. Babaie – Department of Geosciences, Georgia State University, Atlanta, GA 30303 Email: h巴巴ie@gsu.edu

**Summary**

In this white paper, it is argued that the design and architecture of the geosciences network, as envisioned in the EarthCube framework, can be based on the Semantic Web (SW) technology. The SW technology offers the best solutions to enable the required semantic-based integration of data by providing formal (i.e., machine readable) and explicit meaning for geoscience data, thus enhancing their understanding by geoscientists. The Web Ontology Language (OWL) and its underlying RDF and RDFS languages, with the help of the Unique Resource Identifier (URI) and the HTTP protocol, allow geoscientists in different fields to build and deploy ontologies. The main purpose of building ontologies is to represent and manage domain knowledge, and to discover new knowledge, with the help of the semantic rules embedded in these languages, applying the reasoning power of the SW query languages such as SPARQL. The Semantic Web technology enables us to reach the cardinal goal of global integration of geosciences data, advancing knowledge, building effective search and query tools, and achieving interoperability and intercommunication among geoscientists and their resources at all levels in each community. We argue that the self-
similarity that characterizes the research of interacting geosciences communities, and that of many geological processes, requires a fractal structuring of resources, such as software, database, ontology, and tools, over a wide range of scales, from individuals to progressively larger communities, on the proposed network.

---

**EarthCube DeepSearch: Discovering and Integrating Data from the Geoscience Deep Web**

Chaitan Baru, David Nadeau

San Diego Supercomputer Center, UC San Diego

I. Introduction

The NSF Workshop on Envisioning a National Geoinformatics System for the United States, held in Denver, CO on March 14, 2007, described “…a future in which someone can sit at a terminal and have easy access to vast stores of [geoscience] data of almost any kind, with the easy ability to visualize, analyze and model those data.” This is more than just a vision. Indeed, such a capability is essential if future scientists are to conquer and make effective use of the vast amounts of data that we are producing, and will continue to produce, in the geosciences and elsewhere. There is a critical need for a “search engine” capability that makes discovery of data easy, and enables users to access the data in a convenient form for immediate use in downstream processing with other tools and software applications. The current situation where scientists just “know” where to find the data that matches their needs, and “know” how to navigate the specific—and sometimes arcane—interfaces at each data provider site, is untenable. There will be too many sources of data and information in future. Future researchers and future graduate students cannot, and ought not to, spend much of their research hours learning all of these details of data access. With the amount and variety of data that will be available, this would leave relatively little time for actually working with the data to do novel research. Just as the Google and Bing services enable easy access to information on the Web, a discovery service such as the proposed EarthCube DeepSearch is needed to enable easy access to the large and increasing amounts of geoscience data that are being generated and made available online. The DeepSearch service should be as easy and efficient to use as current Web search tools. It should provide access to the metadata associated with datasets as well as to the actual data itself, wherever possible.

---

**Discovery White Paper Short From Federation of Earth Science Information Partners**

Original Longer Version

Contents

1 Title: A Lightweight Approach to Earth Science Data Discovery
1.1 Authors: The Earth Science Information Partners Discovery Cluster (ESIP-DC)
2 The Challenge and Vision
3 Requirements
4 The ESIP Federation and Discovery Cluster
5 CI Architecture Design, Development, and Integration
5.1 Federated Search Framework
5.2 Data and Service Casting Frameworks
5.3 Beyond Discovery - Integrating Frameworks Together
5.4 Technical Governance
6 Looking to the future
7 Discovery and Earth Science Collaboratory Cluster Participants

Title: A Lightweight Approach to Earth Science Data Discovery
Authors: The Earth Science Information Partners Discovery Cluster (ESIP-DC)
The Challenge and Vision
To truly understand the Earth system requires that we integrate increasingly diverse and complex data from all disciplines. But first one must find the data and see if they are useful. These seemingly simple first steps to data integration and workability remain significant challenges especially when seeking data across myriad disciplines, each with their own cultures of data collection, structure, and description. Traditional cataloging methods to enable discovery are reliant on cumbersome centralized registries, do not scale across disciplines, and do not allow adequate description of data to diverse audiences. The Federation of Earth Science Information Partners (ESIP) have, instead, developed a basic, grassroots, federated approach to discovery that does not rely on centralized registries. Further, we recognize that data are often useless without associated services, so we also enable discovery of data services and associate those services with relevant data.

Using Metadata, Data/Service Quality and Knowledge to Facilitate Better Data Discovery, Access, and Utilization for Supporting EarthCube
Chaowei Yang (GMU CISC), Myra Bambacus (NASA), Karl Benedict (UNM), Doug Nebert (FGDC), Doug Mochuney and Sue Hazlett (USGS), Paul Houser (GMU CISC), Robert Raskin (JPL), Yan Xu and Daniel Fay (Microsoft), Abdelmounaam Rezgui, Quanying Huang and Chen Xu (GMU CISC)

1. Introduction
We have been collecting large amounts data about our home planet from the upper atmosphere (such as ozone change) to well beneath the surface (such as the ocean floor and deep wells). The data record of the Earth in its spatiotemporal context is of great value in understanding the past of our Earth system and to predict the future by analyzing change (Donner 2009). A variety of services and procedures have been built to disseminate, analyze, and utilize data through online tools. Relevant cyberinfrastructures (CI) are deployed to support access to services by relevant disciplines, agencies, and countries. All the data, service, and CI resources built to share and use the Earth data are critical to solve 21st century challenges of geosciences.

Globalization requires sharing and coordinating physical and social resources across the planet (NRC 2003). The globalization also bring many grand challenges, for example, 1) climate is changing and sea level rise is occurring with more polar/glacier ice melt (NRC 2009b; Yang, Nebert, and Fraser, 2011c); 2) severe weather and disasters, such as hurricanes and tsunami, is more frequent (NRC 2011); 3) global social activities speed up the transmission and outbreak of contagious diseases (NRC, 2009a, such as SARS and H1N1); 4) the desertification of land surface, deforestation, and fresh water shortages are becoming more severe. To better address these global challenges, the data about our planet should be shared across domains and across jurisdiction boundaries to enable global geoscience research and emergency response.

The NSF (2009b) CI vision and NSF (2009a) geoscience vision highlight the need for Earth system monitoring; EarthCube (NSF 2011) is proposed at the right time for integrating the resources required to address the grand challenges. Metadata and semantics for describing the resources are central to this integration process to enable users to discover and evaluate the information resources that exist for Earth science studies and applications. To enable across domain and jurisdiction boundary sharing, the standardization (by FGDC, OGC, ISO, ANSI and others) of metadata is critical for identifying interoperable resources and accessing and using these data on the fly. The GEOSS (Global Earth Observation System of Systems) Clearinghouse provides an exemplar global system to harvest and search the distributed metadata collections using standardized interfaces and formats/contents. Data and service quality also plays a significant role to help end users choose the resources that best fit their research, application, or educational needs. The ontology and semantics linked (directly or indirectly) to the resources can help improve the discovery process with better accuracy (Zhang et al., 2010). To address millions of concurrent users when EarthCube becomes operational, support for a spatiotemporal metadata structure, index, ranking, and searching will be crucial. This white paper synthesizes a number of collaborative research and development projects across a variety of agencies.

https://semanticommunity.info/Data_Science/EarthCube_Data_Science_Publications
Updated: Sun, 19 May 2019 05:26:10 GMT
Powered by mindtouch 193
organizations, and companies to demonstrate the utilization of metadata, semantics, and quality to better discover, access, and utilize geoscience resources.

EarthCube White Paper A Data-Based Professional Networking System

James Long (jlong@iarc.uaf.edu) and Karen Remick (kremick@iarc.uaf.edu)
International Arctic Research Center – University of Alaska Fairbanks

Introduction
Today, scientific cyber-infrastructure is confronted with three primary challenges. The first is that, due to technological advances, scientific studies are generating data at an unprecedented rate. These data sets must be properly organized, stored, and distributed so that they can be verified and built upon. The second is that interdisciplinary studies are becoming more and more important, thus the organization of data and metadata needs to be flexible enough to address the many forms of data organization used by different disciplines. Finally, because there are more researchers and data sets than is knowable by any one person, a professional networking organization needs to be designed to allow researchers to know who is doing what, and how work from other people affects their research plans. This white paper discusses an approach that addresses these challenges.

There are two primary ways of determining what data is accepted into a data center, subject and location. Which type is used depends on the purpose and location of the data facility. If a data center is subject oriented, such as one found at a medical facility, its data organization is tailored be useful to the relevant discipline. However its use is then restricted to researchers in that subject and it makes cross disciplinary studies difficult. If a data center is location based, and only accepts data from a specific institution or geopolitical location, such as at a university, it needs flexibility of organization to allow for cross disciplinary studies, but restricts its members to seeing only locally produced results. The federation of data sites can overcome the boundaries for both organization schemes, allowing institutions to develop naturally in a way that is beneficial to them, yet allows the free exchange of information.

EarthCube Design Approach

Hannes E. Leetaru, Illinois State Geological Survey, University of Illinois
Michael Welge, National Center For Supercomputing Applications, University of Illinois
Bernie Ács, National Center For Supercomputing Applications, University of Illinois
Shaowen Wang, Department of Geography, University of Illinois
Yu-Feng Forrest Lin, Illinois State Water Survey, University of Illinois
Kenton McHenry, National Center For Supercomputing Applications, University of Illinois

VISION FOR EarthCube
A description of the envisioned scope of EarthCube and how will it transform geosciences research, including the functionality EarthCube can provide to the whole Geosciences community.

The vision of EarthCube presented here is an open access archive of all publicly available geosciences data in the United States and eventually the world, capable of supporting the technical needs to store and offer easy access to petabyte-size datasets in the next decade, offering long-term format translation, rich provenance and metadata, advanced cross-disciplinary portals that can be tailored for use from the citizen scientist to graduate students to expert users, offers community features to facilitate and promote collaboration, and has strong governance and sustainability models to ensure long-term success. A hybrid architecture of a centralized data system with multiple distributed nodes for legacy data and specialized portals combines ease-of-integration of distributed data with the centralized needs of governance. There will be a common template for domain portals that will enable transparent access for the new user and still enable experienced users to use advanced features unique to their disciplines. EarthCube development will use an iterative approach with rapid development of prototypes used to test technical feasibility and suitability to science needs.
Important, but easily forgotten, is the need for easy access and analysis capabilities for the “Citizen Scientist” and K-12 students to leverage the same resources as domain scientists, but through interfaces and curriculums tailored for their use. These non-scientists have recently made significant contributions to science that would otherwise not happened. EarthCube will offer unified search for across attributes, geographic region, specific methodologies, provenance, or earth characteristics. Ultimately, we envision all geoscience projects funded by the US government and nongovernmental organizations including USDOE, USEPA, corporations, and USGS covering non-classified data would integrate their data into EarthCube. Finally, to encourage initial adoption and to maximize the coverage of the system, connectors will support the integration of numerous online data environments, as well as integration with NSF XSEDE to allow for seamless application of HPC and service oriented approaches to synthesizing data and associated analytics.

A Workflow-Based Knowledge Management Architecture for Geodynamics Data

Jian Qin John D'Ignazio
School of Information Studies
Syracuse University
jqin, jadignaz@syr.edu
Suzanne L Baldwin
Department of Earth Sciences
College of Arts and Sciences
Syracuse University
sbaldwin@syr.edu

Abstract
A prototype project made up of Syracuse University earth sciences and information science researchers is exploring the cognitively demanding set of workflows required to make science advances in thermochronology and plate tectonics. The participants aim to understand the data practices and requirements that support publication in this research domain and use them to suggest knowledge management tool and standard development on two levels: 1) raw data file management and 2) knowledge management related to steps in sample and data processing. Enhancements in data management and curation functions are meant to standardize or automate processes in data verifiability, interworkability, analyzability, and interoperability, leaving researchers free to conduct higher levels of analysis and synthesis. This prototype would therefore offer critical lessons for EarthCube to increase the productivity and capability of geoscientists.

One perspective on GeoSciences Data and Coastal Science and Sustainability: A brief to NSF’s EarthCube

Brian Blanton1, Harvey Seim2, Stan Ahalt1
1Renaissance Computing Institute, University of North Carolina at Chapel Hill
2Department of Marine Sciences, University of North Carolina at Chapel Hill

A substantial percentage of the world population lives within 10-100 kilometers of the ocean and in lowlying, flat coastal regions (Small and Nicholls, 2003; United Nations Ocean Atlas, 2010). This population segment, the land, and the supporting infrastructure represent a complex and diverse system that is particularly vulnerable to coastal hazards including storm surge inundation, precipitation-based flooding, waves, and coastal erosion. Further, this coastal zone (CZ) is also the buffer and filter between terrestrial and oceanic waters, through which most international commerce is transported. It also supports a vibrant economy that mixes together, among other ingredients, tourism, military assets, critical fisheries, and environmentally sensitive areas. Furthermore, in many cases, CZ population density continues to increase. In the coming decades and centuries, the CZ and its population and resources will become increasingly more at risk as the consequences of a changing climate (e.g., elevated sea levels, increased precipitation, potentially increased storm intensity and frequencies (Najjar, et al, 2000; Nicholls, 2004)) are felt throughout coastal systems. Ensuring the long-term sustainability and increasing the resiliency of the CZ and its economies, cultures, and ecosystems is thus one of the greatest challenges faced by the United States and the world in general.
A National Geoinformatics Community (NGC)

J. Douglas Walker (University of Kansas)
Linda Gundersen (USGS)
Lee Allison (Arizona Geological Survey)
Hassan Babaie (Georgia State University)
Doug Fils (Ocean Leadership)
Steve Richard (Arizona Geological Survey)
Ramon Arrowsmith (Arizona State University)
Cinzia Cervato (Iowa State University)
Steve Whitmeyer (James Madison University)

Summary
The motivation for further developing and involving the National Geoinformatics Community (NGC) concept in EarthCube is:

I can't integrate what I can't find
I can't use something I don't understand
I don't want to use something I don't trust
I can't use something that isn't there any more

The scientific, societal, and educational problems the geoscientific community is being asked to address are increasingly complex and thus require the application of multiple datasets and advanced computational capabilities for integrated data analysis. However, we cannot integrate data we cannot find, understand/evaluate, or trust. It is also critical that data, tools, and services are persistent and do not disappear at the end of projects or with time. We cannot jointly analyze multiple datasets without the required tools. The National Geoinformatics Community effort is aimed at addressing these critical issues and developing and offering community-based solutions.

Framing Statement
For the past 10 years, Geoinformatics has evolved into a critical and growing activity in the Earth Sciences. It has potential transformative impacts in all areas of research, outreach, and education. Aspects of Geoinformatics include not only data discovery, access, and delivery, but also data, model, and system interoperability, data management, and archiving of past, current, and future research and education studies. This is emphasized by the new and emerging data reporting requirements being enforced by the US National Science Foundation and other funding agencies. It also appears to be the motivation behind the EarthCube effort.

EarthCube White Paper: A Future Vision In Action at APL for Cyberinfrastructure Enabled Geoscience

Lars Dyrud1
Jonathan Fentzke1
Robin Barnes1
Brian Anderson1
Larry Paxton1
Elsayed Talaat1
Daniel Morrison1
Shin Oshanti1
Jesper Gjerloev2
1 Space Dept. . Johns Hopkins University Applied Physics Laboratory
2 Depart. of Physics and Technology - University of Bergen

In order to create a cyberinfrastructure for the geosciences that enables transformative research and new understanding of the environment.
based tools in a must. At APL we have been actively involved in meeting the future challenges of the geosciences by developing
data management architectures for ingesting, processing, and serving higher level data products to the community for the various NSF and internal programs we actively manage including: AMPERE, SuperDARN, SuperMAG, GAIA, and GEOScan. A goal of existing programs is to create tools that reduce cost and save user/provider time through common interfaces and flexible applications, which have been instrumental in producing synergistic discovery. This white paper provides an overview of our existing capabilities and ties them to a proposed future vision that will help enable the broad community initiative for enhanced cyberinfrastructure in the geosciences: EarthCube.