NSF EarthCube Initiative

Barbara Ransom, PhD

Program Director

National Science Foundation

A joint venture between the NSF Geosciences Directorate and the CISE Division of Advanced Cyberinfrastructure
Big Questions, Big Issues!!

- environmental change & resilience
- formation & evolution of the atmosphere & oceans
- the origin of life
- climate change
- human-earth interactions
- extreme events – causes, periodicity, & implications
- deep – surface earth Interactions & feedbacks
- resource discovery & abundance
- geohazards
- life as a geologic agent
- continental evolution & changes thru time
- future world
Community: Complicated and Multi-Faceted

- **Age of Enlightenment**
  - Computer/Cyberinfrastructure Communities

- **Modern Age**
  - Atmospheric and Climate Modeling Communities

- **Industrial Age**
  - Seismology/Earthquake and Physical Oceanography Communities

- **Bronze Age**
  - Nearly all other Geoscience groups

Present Relative State of Cyber-Sophistication and Knowledge in the Geosciences

- The 15%
- The 85%
The Problem (the 15% vs the 85%)

Two very different types of data
- sensor, bit-stream, real-time: GB/TB size (satellite, radar, seismic)
- sample-based, observations, images, multi informational, hard to describe

Two very different relationships with data
- Array-based: no ownership, don’t care about any given data point, computationally intensive processing and modeling
- Sample/observation-based: intense ownership, care deeply about each point, can interpret directly or simply

Two different levels of investment
- HPC, big iron, federal archives, modeling centers, data repositories, dedicated personnel and facilities
- spreadsheets, hero code, dark data, cultural issues, no sustainability
The 85% spend about 80% of their time looking for, collecting, and getting the necessary data together in a format they can use and about 20% of their time actually thinking/doing science.

The 15% spend an increasing amount of time having problems wrestling with unmanageably large data arrays and problems scaling from global to regional or local scales.

Neither are well integrated with each other and both types of data (array vs. point) and all of the areas of geoscience are required to solve the complex, inter-related, and pressing environmental problems we and the Earth are facing.
What Is EarthCube?

• Transform the conduct of data-enabled geoscience-related research.

• Create effective community-driven cyberinfrastructure.

• Allow global data discovery and knowledge management.

• Achieve interoperability and data integration across disciplines.
Why EarthCube?

- Nature does not recognize separate disciplines.
- EarthCube will democratize access to data.
- EarthCube will increase research time by reducing time needed to find, access, and analyze data.
- EarthCube will enable more interdisciplinary research and the pursuit of new questions.
- EarthCube will accelerate the pace of discovery.
- EarthCube will give all scientists the same chance of making major contributions regardless of institution size or institutional endowment.
Who Is EarthCube? You Are!

EarthCube CI

Solar Terrestrial
Geodesy
Aeronomy
Geophysics & Geodynamics
Cryosphere & Ice Dynamics
Geobiology & Paleontology
Igneous Petrology & Volcanology
Space Weather
Structure & Tectonics

Climate & Large Scale Dynamics
Magnetospheric Physics
Atmospheric Chemistry
Paleo-Climate

NCAR

Bio-informatics
Biological Oceanography
Physical Oceanography
Ocean Drilling & Engineering

Ecosystems
High Perf Computing
Algorithms & Data Mining
Semantics & Ontologies

Software & Modeling
Chemical Oceanography
Marine Geology

Oceanography

Biological Oceanography
Geology

EarthCube CI
An alternative approach to respond to daunting science and CI challenges

Path to the Vision
Its All about the Connections

EarthCube is an outcome AND a process

EarthCube will require NSF and broad community involvement; new ways of doing

Important Features:
- Builds off existing data/modeling systems/cyberinfrastructure investments
- Provides tools/approaches that enhance modeling results, visualization, and data discovery, access, and integration
- Leverages investments across fields
- Allows for more integrative and interdisciplinary science
Feel Our Pain!

help me!
IMPORTANCE of integrating multiple datasets, models, observations, and/or visualization tools from different fields

<table>
<thead>
<tr>
<th></th>
<th>Early Career</th>
<th>EarthCube Active</th>
<th>All Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>μ(α) = 0.76 (0.25)[n=77, 7]</td>
<td>μ(α) = 0.83 (0.24)[n=101, 9]</td>
<td>μ(α) = 0.71 (0.27)[n=530, 29]</td>
</tr>
</tbody>
</table>

Many very positive, with some neutral
Only a few negative

Data from Joel Cutcher-Gershenfeld (social science, U Illinois) stakeholder alignment study of EarthCube community

<table>
<thead>
<tr>
<th></th>
<th>Early Career</th>
<th>EC Active</th>
<th>All Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>0.76 (0.25)</td>
<td>0.83 (0.24)</td>
<td>0.71 (0.27)</td>
</tr>
</tbody>
</table>
EASE of integrating multiple datasets, models, observations, and/or visualization tools in your field

Vast majority negative, with some neutral
Only a few positive

Data from Joel Cutcher-Gershenfeld (social science, U Illinois) stakeholder alignment study of EarthCube community

<table>
<thead>
<tr>
<th></th>
<th>Early Career</th>
<th>EC Active</th>
<th>All Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ (α)</td>
<td>0.35 (0.25)</td>
<td>0.35 (0.25)</td>
<td>0.42 (0.24)</td>
</tr>
</tbody>
</table>
Many very positive, with some neutral
Only a few negative

Data from Joel Cutcher-Gershenfeld (social science, U Illinois) stakeholder alignment study of EarthCube community

<table>
<thead>
<tr>
<th></th>
<th>Early Career</th>
<th>EarthCube Active</th>
<th>All Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPORTANCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of integrating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>multiple datasets,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>models, observations,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and/or visualization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tools</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From different fields

Early Career: \( \mu (\alpha) = 0.76 \ (0.25) \ [n=77, 7] \)
EarthCube Active: \( \mu (\alpha) = 0.83 \ (0.24) \ [n=101, 9] \)
All Others: \( \mu (\alpha) = 0.71 \ (0.27) \ [n=530, 29] \)
EASE of integrating multiple datasets, models, observations, and/or visualization tools from different fields

Vast majority negative, some neutral

Data from Joel Cutcher-Gershenfeld (social science, U Illinois) stakeholder alignment study of EarthCube community

<table>
<thead>
<tr>
<th></th>
<th>Early Career</th>
<th>EarthCube Active</th>
<th>All Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ (α)</td>
<td>0.28 (0.20)</td>
<td>0.23 (0.22)</td>
<td>0.32 (0.23)</td>
</tr>
</tbody>
</table>

access ease: multiple datasets
μ(σ) = 0.28 (0.20)[n=65, 19]

access ease: multiple datasets
μ(σ) = 0.23 (0.22)[n=66, 12]

access ease: multiple datasets
μ(σ) = 0.32 (0.23)[n=466, 91]
Blue Skying the EarthCube Future

**Imagine:**
- A world without laptops and WiFi - 22 yrs ago
- A world without cell phones – 20 yrs ago
- A world without digital cameras - 11 yrs ago
- A world without public GPS - 9 yrs ago
- A world without iPhones - 6 yrs ago
- A world without iPads – 3 yrs ago

Think of how much you depend on these tools!

**Imagine:**
- What would your life/science be without them?
- What the next advance will make possible!
Now:

- Imagine a world where people can easily model their results and explore any ideas they might have.
- Imagine a world where anyone can easily plot data of interest and display it any way they want.
- Imagine a world where with easy, unlimited access to scientific data from any field.

What science could you do?

What discoveries could you make?