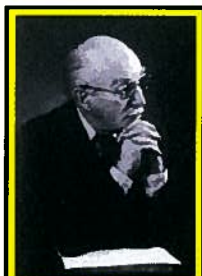
Data on data:

Presenting stakeholder alignment data on the cyberinfrastructure for earth system science

Joel Cutcher-Gershenfeld,
University of Illinois, Urbana-Champaign
Tom Altura, UCLA
Betty Barrett, MIT
Burcu Bolukbasi, UIUC
Courtney Flint, UIUC
Michael Haberman, UIUC
John King, University of Michigan
Chris Lawson, Aerospace Corporation
Barbara Lawrence, UCLA
Mark Nolan, UIUC
Barbara Mittleman, NIH
John Unsworth, Brandeis University
Illya Zaslavsky, UCSD

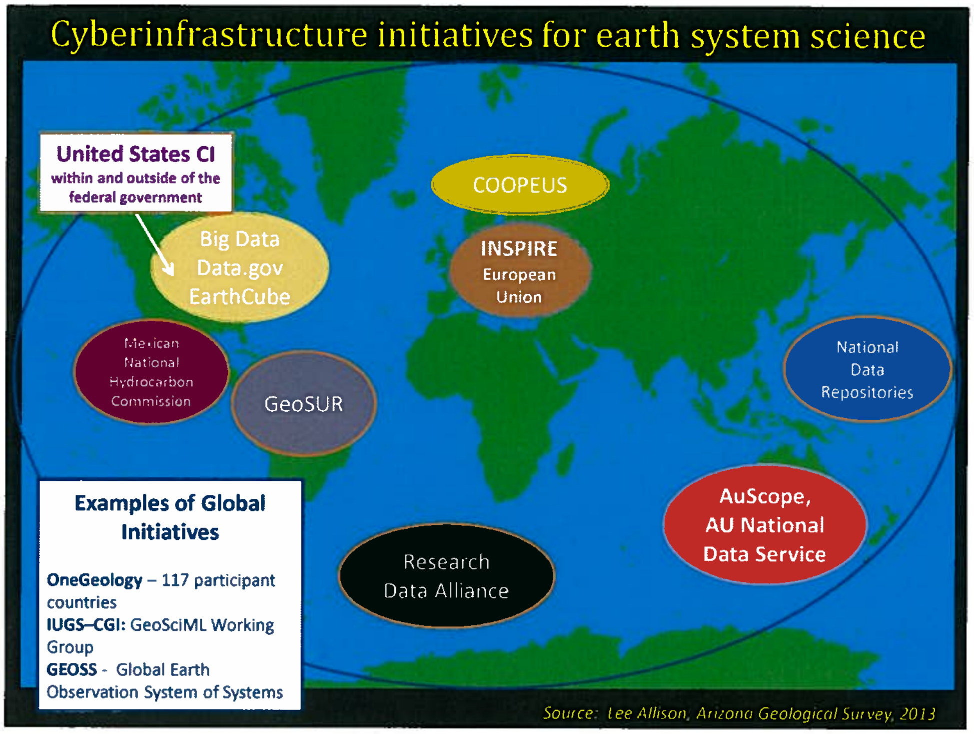
Support from the National Science Foundation is deeply appreciated (NSF-VOSS EAGER 0956472, "Stakeholder Alignment in Socio-Technical Systems," NSF OCI RAPID 1229928, "Stakeholder Alignment for EarthCube," NSF SciSPR-STS-OCI-INSPIRE 1249607, "Enabling Transformation in the Social Sciences, Geosciences, and Cyberinfrastructure")



Looking ahead . . .

". . . We are moving towards another type of society than that to which we have become accustomed. This is sometimes referred to as a new service society, the society of the second industrial revolution or the post-industrial society. There is no guarantee of our safe arrival. Not only are the interdependencies greater – they are differently structured. . . [and] demand a new mobilization of the sciences."

– Source: Eric L. Trist, from paper on "Social Aspects of Science Policy" (March, 1969) cited in *Towards a Social Ecology: Contextual Appreciation of the Future in the Present* by Fred E. Emery and Eric L. Trist (London: Plenum Press, 1973)



Institutional and systems requirements

Creating Value — *. . . expanding the "pie" and enabling systems transformation*

Mitigating Harm — *. . . anticipating and mitigating externalities and catastrophic systems failures*

Additional "ility" Challenges:

Stability . . . Agility . . . Extensibility . . . Sustainability

Why are robust institutions difficult?

- Tragedies of the commons (Hardin, 1968)
- Iron law of oligarchy (Michels, 1911)
- The logic of collective action (Olson, 1965)
- Tyrannies of majorities (Mill, 1913) and minorities (Staub and Zohn, 1980)
- Accelerating rates of technological change (Kurzweil, 1999)

Consider accelerating rates of technological change

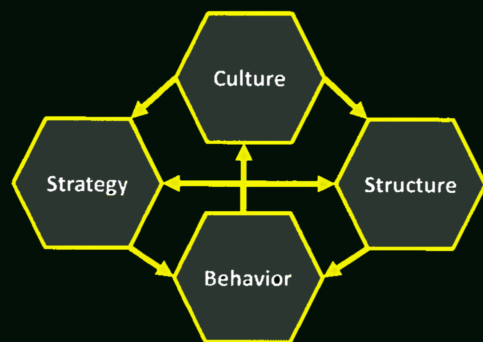
The babysitter of the future . . .

(courtesy of Steve Diggs, Scripps Institution of Oceanography)

Defining stakeholder alignment. . .

"The extent to which interdependent stakeholders orient and connect with one another to advance their separate and shared interests."

A simplified conceptual framework . . .



Selected first principles. . .

Principle 1: Institutions/systems are "socially constructed"

Principle 2: Institutions/systems are established to create value and mitigate harm

Principle 3: Institutions/systems are comprised of stakeholders and interests (separate and shared) in social and technological contexts

Principle 4: Every stakeholder has a vector of interests; every interest has a vector of stakeholders

Principle 5: Create value *and* mitigate harm by increasing stakeholder alignment to advance separate and shared interests

Corollary A: Advance separate and shared interests through visual representation, informative analytics, and constructive engagement

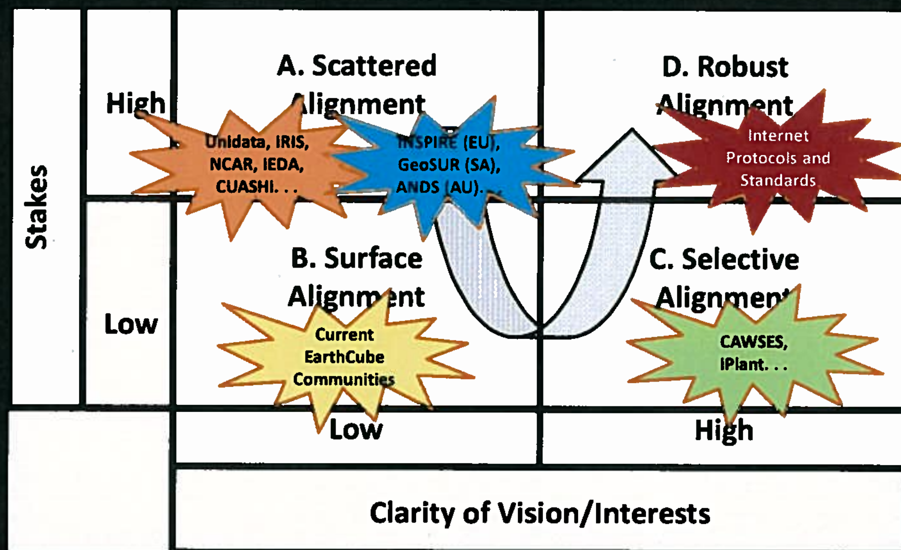
Corollary B: Leadership by influence, more than authority

Corollary C: Achieve leverage through protocols and standards

Corollary D: Enable lateral alignment through internal alignment

Corollary E: It is easier to be negative

Forms of alignment



"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*

Specify Stakeholders and Identify Interests

- Atmospheric or Space Weather scientist
- Oceanographer
- Geologist
- Geophysicist
- Hydrologist
- Critical zone scientist
- Climate scientist
- Biologist or Ecosystems scientist
- Geographers
- Computer or Cyberinfrastructure scientist
- Social scientist (Anthropologist, Economist, Psychologist, Sociologist, etc.)
- Other scientist
- Data manager
- High performance computing expert
- Software engineer
- IT user support personnel
- K-12 educator
- Designer/developer of geoscience instrumentation
- Environmental resource manager (e.g. local, state, or federal)
- Other

50+ interest questions, covering:

- Access and Utilization of Data, Observations, Visualizations, and Models – Current State and Desired State
- Increasing Uniformity and Interoperability through the EarthCube Process
- The Scope of the EarthCube Mission
- Stakeholder Relations and Governance
- Your Potential Engagement with EarthCube

Response rates and involvement with EarthCube

N=809

Data Centers (n=599)	**EC Website (n=123)**	**Domain Workshops (n=78)**
(distribution to 10,000+)	*(Pop. approx. 750 or 16%)*	

No/Low Involvement *(n=639 minus <5yrs. or 559)* — Group 1 (minus workshops)

First I have heard of EarthCube	45%
Aware of EarthCube, but no engagement	28%
Visited the EarthCube website	11%

High Involvement *(n=115 minus <5 yrs. or 110)* — Group 2 (minus workshops)

Participated in EarthCube discussions	7%
Actively involved in EarthCube communities	7%
Leadership role in EarthCube communities	2%

Early Career *(<5 yrs. exp. n=84)* — Group 3

Tectonics Workshop *(n=24)* — Group 4

EarthScope Workshop *(n=21)* — Group 5

Respondent Profile (n=809)

Geoscience	72%
Cyber/CS	6%
Both	14%
Other	8%

Under 5 years	11%
5-10 years	21%
11-20 years	29%
Over 20 years	39%

US	80%
Non US	20%

Female	27%
Male	71%

University	65%
Govt. – Federal	12%
Govt. – State/Local	2%
National Labs	7%
Industry	4%
Nonprofit/NGO	6%
Other	7%

Distribution of responses by field or discipline (n=809)

- Atmospheric or space Weather scientist — n=125 (16%)
- Oceanographer — n=99 (12%)
- Geophysicist — n=85 (11%)
- Geologist — n=152 (19%)
- Hydrologist — n=35 (4%)
- Critical Zone Scientist — n=11 (1%)
- Climate Scientist — n=60 (7%)
- Biologist or Ecosystems Scientist — n=36 (4%)
- Geographer — n=19 (2%)
- Computer or Cyber-infrastructure Scientist — n=41 (5%)
- Data Manager — n=27 (3%)

Additional 19% in other areas of expertise.

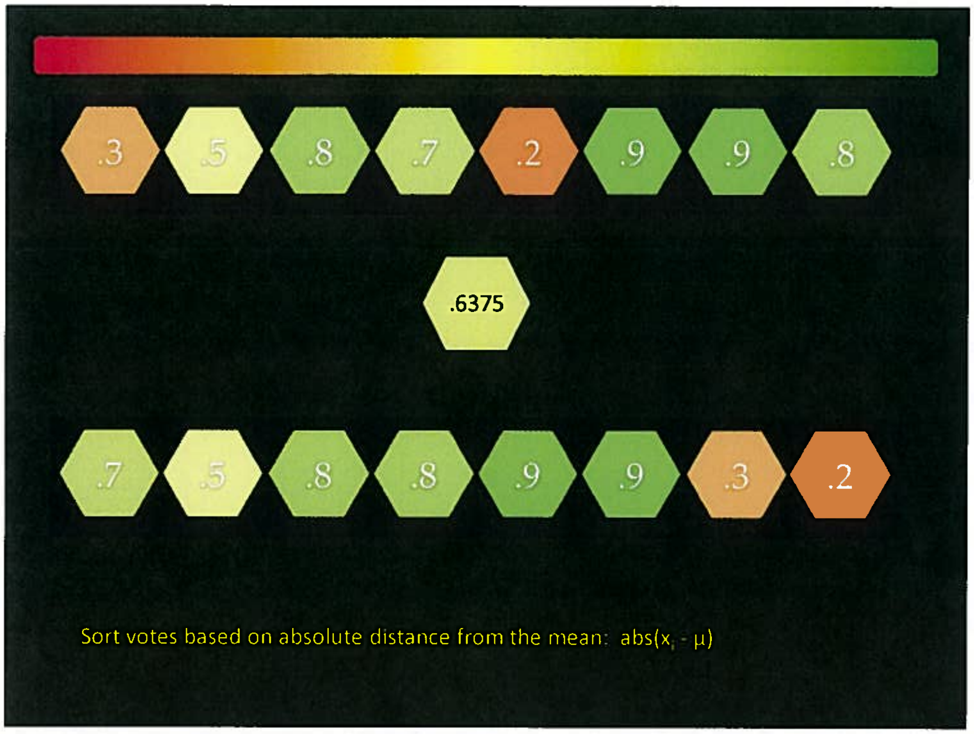
Visualizing stakeholders' interests

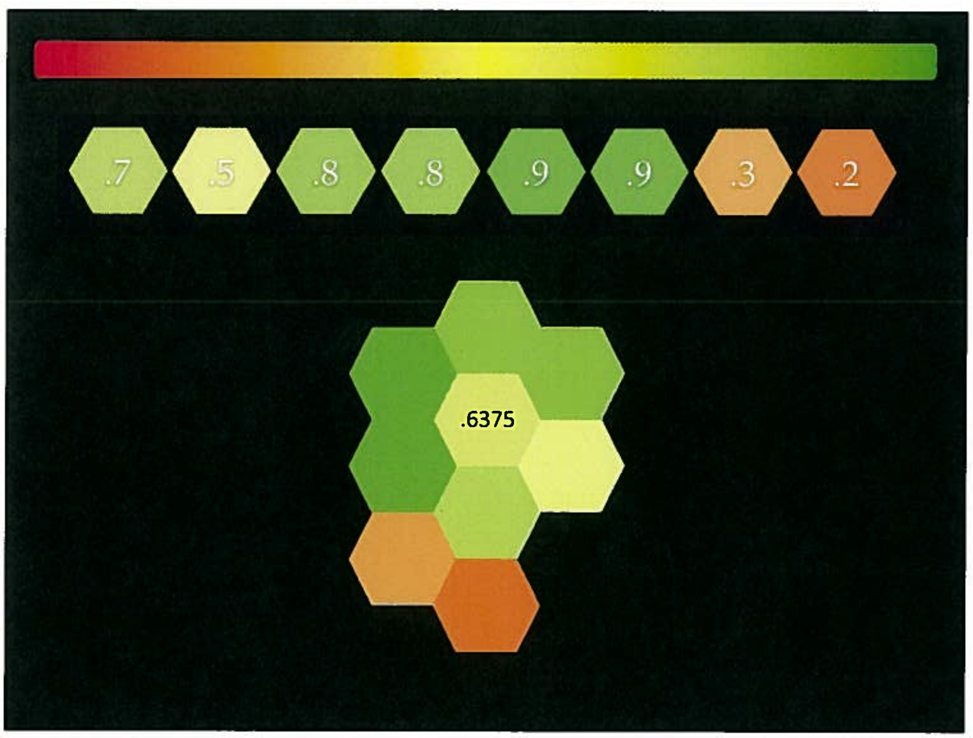
Motivation:

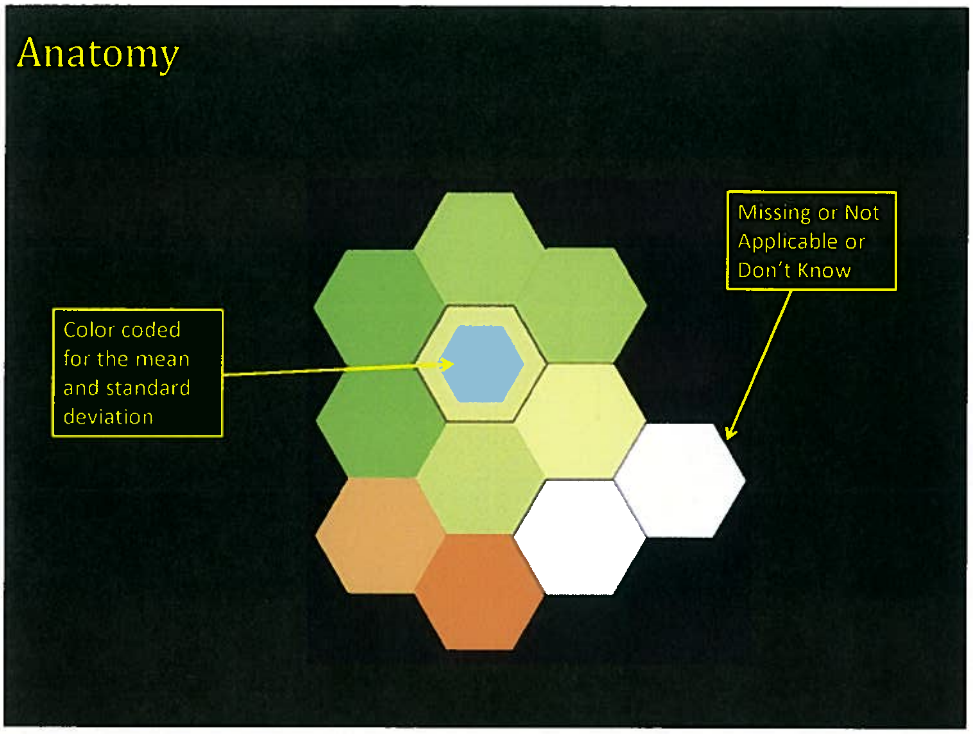
- Each stakeholder has a perspective – every vote counts
- Visualize a range of possible perspectives – communicate the distribution
  - core to outliers
  - stakeholder map

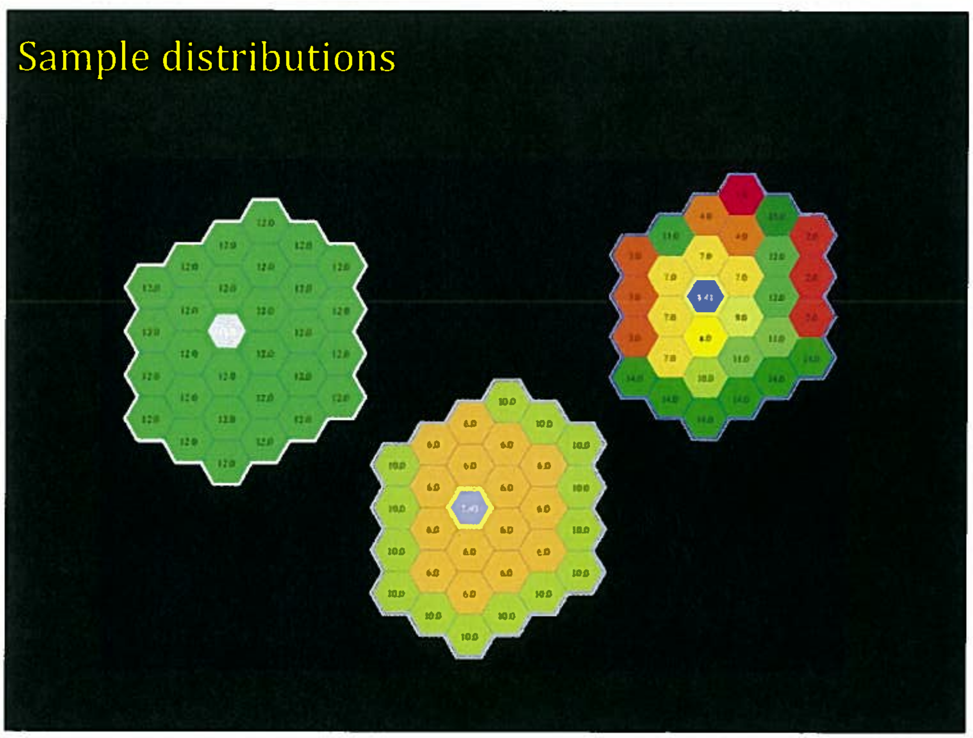
Example, growing an I-flower™

Range: [0,1], -1 (no vote, not applicable)

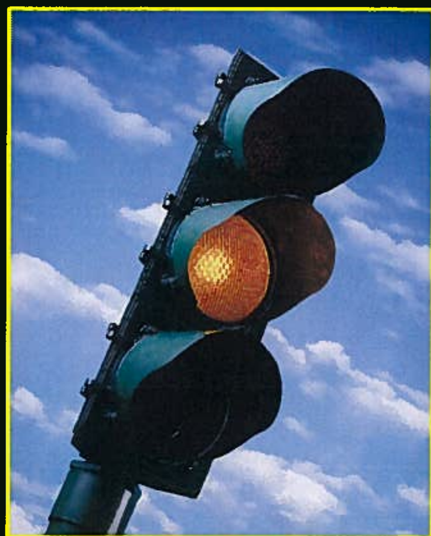

.3 .5 .8 .7 .2 .9 .9 .8
.6375
.7 .5 .8 .8 .9 .9 .3 .2
Sort votes based on absolute distance from the mean: abs(x_i - μ)


.7 .5 .8 .8 .9 .9 .3 .2
.6375


Anatomy
Missing or Not Applicable or Don't Know
Color coded for the mean and standard deviation

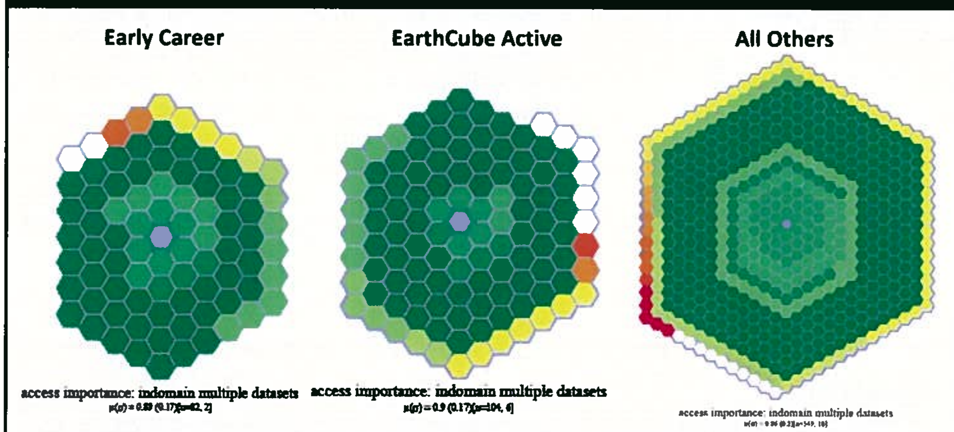

Sample distributions

Caution – construction ahead



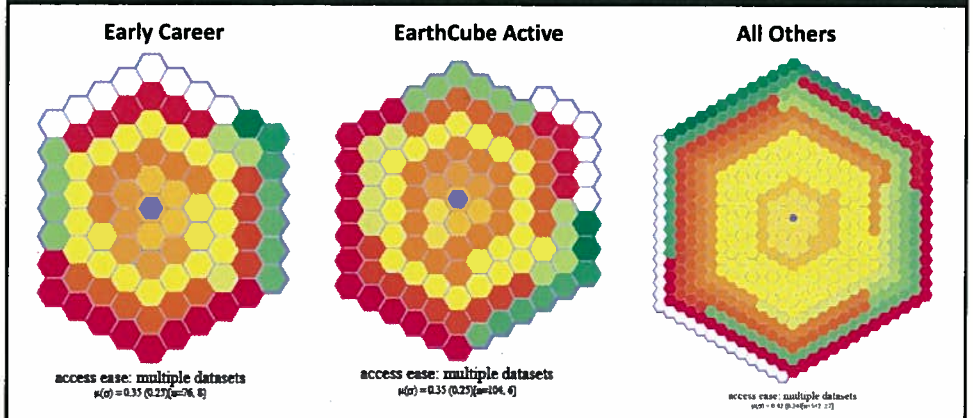
- Preliminary propositions and models
- Advances in visual representation and instrumentation
- Much work to be done testing/refining propositions, models, methods, and analytics

IMPORTANCE of integrating multiple datasets, observations, visualization tools, and/or models in your field or discipline



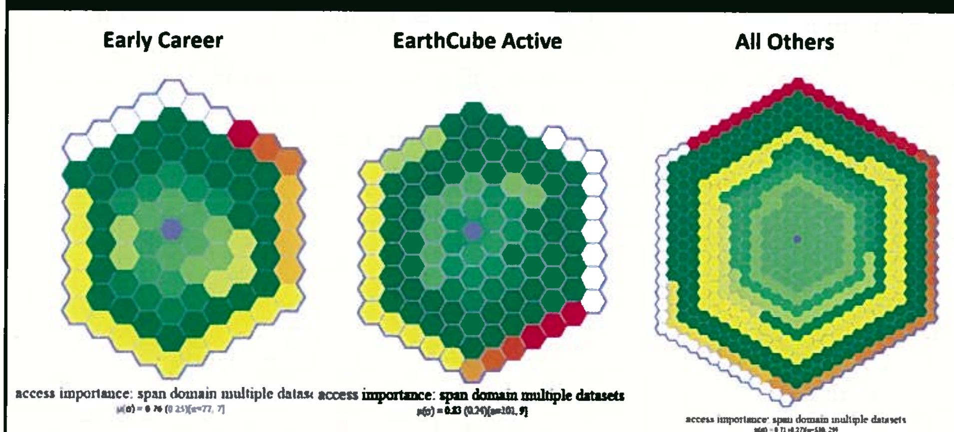
Comment: Vast majority are extremely positive on importance, with just a handful who are neutral or negative.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
How IMPORTANT is it for you to find, access, and/or integrate multiple datasets, observations, visualization tools, and/or models in your field or discipline?	.94 (.10)	.91 (.14)	.89 (.17)	.90 (.17)	.86 (.20)

EASE of integrating multiple datasets, observations, visualization tools, and/or models in your field or discipline?



Comment: Vast majority of respondents are neutral or negative, with many strongly negative; early career and EC active are most negative.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
How EASY is it for you to find, access, and/or integrate multiple datasets, observations, visualization tools, and/or models in your field or discipline?	**.42 (.25)**	**.35 (.23)**	**.35 (.25)**	**.35 (.25)**	**.42 (.24)**

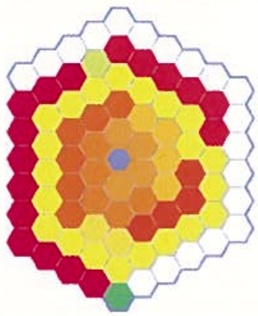
IMPORTANCE integrating multiple datasets, observations, visualization tools, and/or models spanning fields/disciplines?



Comment: EC active are most positive, though most in all groups are positive (with some neutral and negative outliers).	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
How IMPORTANT is it for you to find, access, and/or integrate multiple datasets, observations, visualization tools, and/or models that span different fields or disciplines?	**.84 (.23)**	**.74 (.27)**	**.76 (.25)**	**.83 (.24)**	**.71 (.27)**

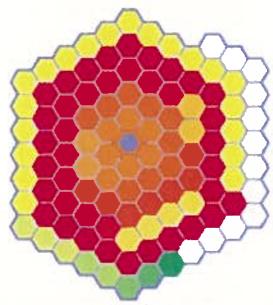
EASE of integrating multiple datasets, observations, visualization tools, and/or models that span different fields or disciplines?

Early Career



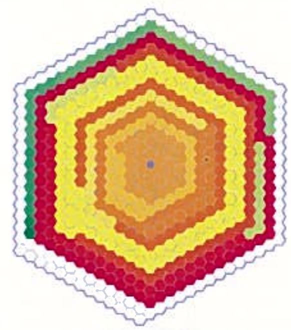
access ease: multiple datasets
μ(σ) = 0.28 (0.27)[n=65, 19]

EarthCube Active



access ease: multiple datasets
μ(σ) = 0.23 (0.22)[n=98, 12]

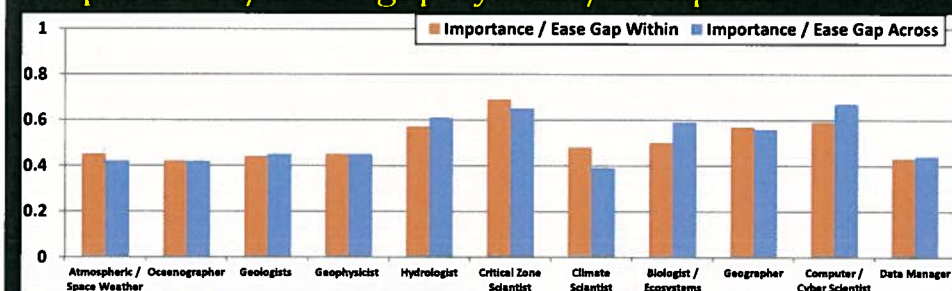
All Others



access ease: multiple datasets

Comment: EC active are most negative, though all respondents report difficulty in access (with only a handful of positive outliers).	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
How EASY is it for you to find, access, and/or integrate multiple datasets, observations, visualization tools, and/or models that span different fields or disciplines?	.32 (.24)	.27 (.25)	.28 (.20)	.23 (.22)	.32 (.23)

"Importance/Ease" gap by field/discipline

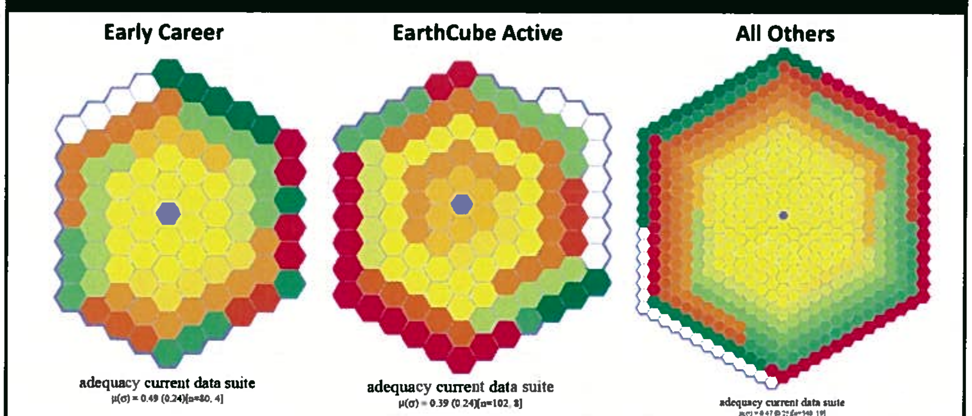


	Importance / Ease Gap Within	s.d.	Importance / Ease Gap Across	s.d.	C.V. Within	C.V. Across
Atmospheric / Space Weather Scientist (n=111)	.45	.27	.42	.34	0.60	0.81
Oceanographer (n=95)	.42	.29	.42	.31	0.69	0.74
Geologists (n=120)	.44	.31	.45	.32	0.70	0.71
Geophysicist (n=68)	.45	.30	.45	.34	0.67	0.76
Hydrologist (n=33)	.57	.25	.61	.29	0.44	0.48
Critical Zone Scientist (n=11)	.69	.30	.65	.31	0.43	0.48
Climate Scientist (n=58)	.48	.28	.39	.32	0.58	0.82
Biologist / Ecosystems Scientist (n=36)	.50	.34	.59	.31	0.68	0.53
Geographer (n=19)	.57	.26	.56	.28	0.46	0.50
Computer / Cyber Scientist (n=36)	.60	.27	.65	.33	0.45	0.51
Data Manager (n=23)	.43	.29	.44	.31	0.67	0.70

What are the implications of these selected examples of how people specify their areas of expertise?

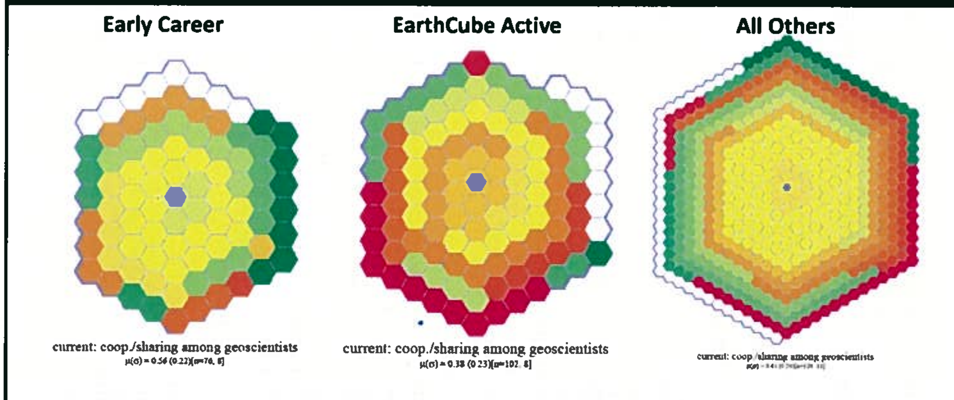
- Air Sea Interaction
- Atmospheric Radiation
- Basalt geochemistry
- Biodiversity Information Networks
- Carbonate Stratigraphy
- Chemical Oceanography
- Coastal Geomorphology
- Computational Geodynamics
- Cryosphere-Climate Interaction
- Disaster Assessment
- Ensemble data assimilation
- Geochronology
- Geoinformatics
- Geomicrobiology
- Glaciology
- Heliophysics
- Isotope Geochemistry
- "It's complicated"
- Magnetospheric Physics
- Mesoscale Meteorology
- Multibeam Bathymetric Data
- Nearshore Coastal Modeling
- Paleoceanography
- Paleomagnetism
- Permafrost Geophysics
- Planetology
- Riverine carbon and nutrient biogeochemistry
- Satellite gravity and altimetry data processing
- Tectonophysics
- Thermospheric Physics
- Watershed Management

Adequacy of current suite of tools and modeling software



Comment: Current suite seen as inadequate – motivation for EarthCube; most negative are EC active.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
Please use the scale ranging from Inadequate to Adequate to assess the present suite of publicly accessible datasets, data analysis tools, and modeling software – to what degree is it adequate for your research and education needs?	**.40 (.20)**	**.32 (.17)**	**.49 (.24)**	**.39 (.24)**	**.47 (.25)**

Cooperation and sharing of data, software among Geo



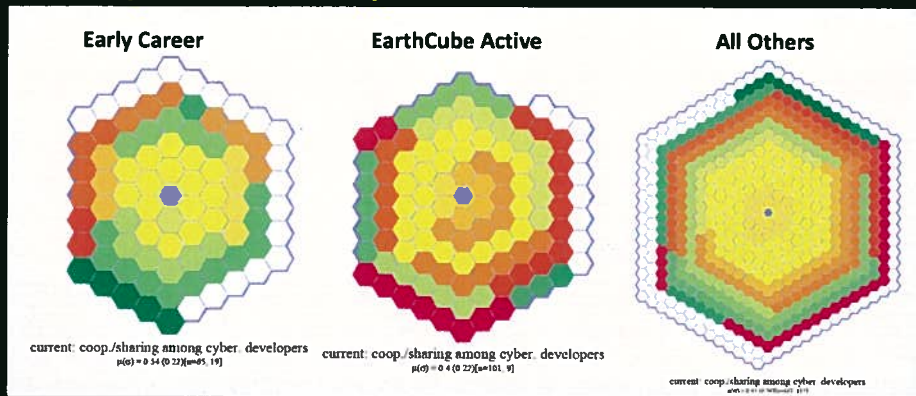
Comment: Strong negative views on cooperation and sharing in Geo community; handful of "bright spots;" most concerned views among EC active.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
There is currently a high degree of cooperation and sharing of data, models, and simulations among geoscientists	.51 (.16)	.39 (.20)	.56 (.22)	.38 (.23)	.45 (.24)

Top Ten Barriers to Sharing Data (categories):

1. No time/Needs too much QA/QC
2. No repository/No known repository
3. Inadequate standards/No standardized formats
4. Want to publish first/Don't want to be scooped
5. File size too large/Server size too small
6. Classified/proprietary/Agency or company restrictions
7. No credit/No incentive to share
8. Cost
9. Not sure what to do
10. Not sure anyone wants it

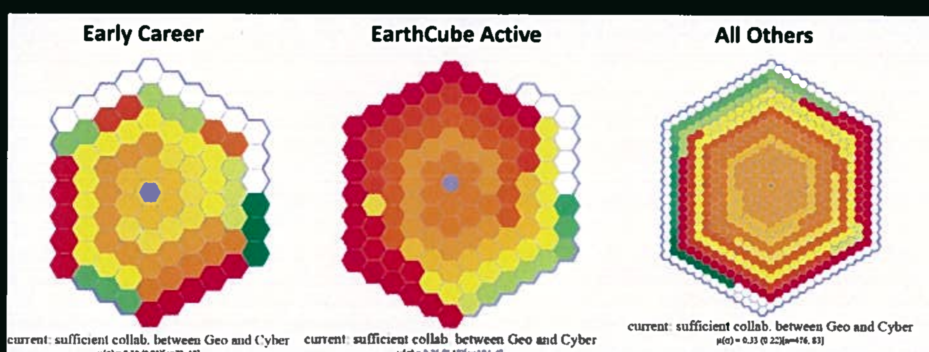
Note: Approximately 45% of respondents did not respond to the open ended question "It is difficult to share my data because. . . " and another 6% said it was easy to share their data. The balance of responses were organized into the above categories; some individuals cited more than one reason (all of which were tabulated).

Cooperation and sharing of data, software within Cyber community



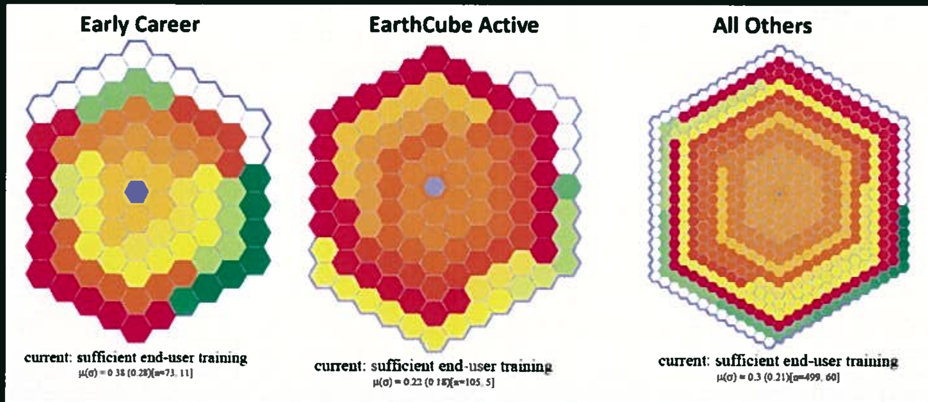
Comment: Strong negative views on cooperation and sharing in Cyber community; handful of "bright spots;" EC active most concerned.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
There is currently a high degree of cooperation and sharing of software, middleware and hardware among those developing and supporting cyberinfrastructure for the geosciences	.51 (.18)	.43 (.25)	.54 (.22)	.40 (.22)	.45 (.24)

Communication and Collaboration: Geo and Cyber

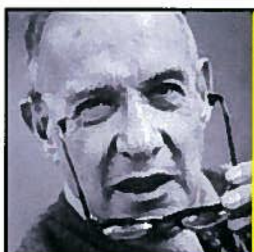


Comment: Major concerns with communication and collaboration between Geo and Cyber communities; strongest concerns among EC active.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
There is currently sufficient communication and collaboration between geoscientists and those who develop cyberinfrastructure tools and approaches to advance the geosciences	.31 (.22)	.25 (.17)	.39 (.23)	.25 (.19)	.33 (.22)

Geo and Cyber – End-user training



Comment: Major concerns end-user knowledge of Cyber by Geo, with strongest concerns among EC active.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
There is currently sufficient geoscience end-user knowledge and training so they can effectively use the present suite of cyberinfrastructure tools and train their students/colleagues in its use	.29 (.18)	.21 (.14)	.38 (.28)	.22 (.18)	.30 (.21)



"The task of leadership is to create an alignment of strengths in ways that make a system's weaknesses irrelevant."

Peter Drucker (2011), as quoted in interview notes Diana Whitney, Amanda Trosten-Bloom, and David Cooperrider

Top twenty-five cited sources of data

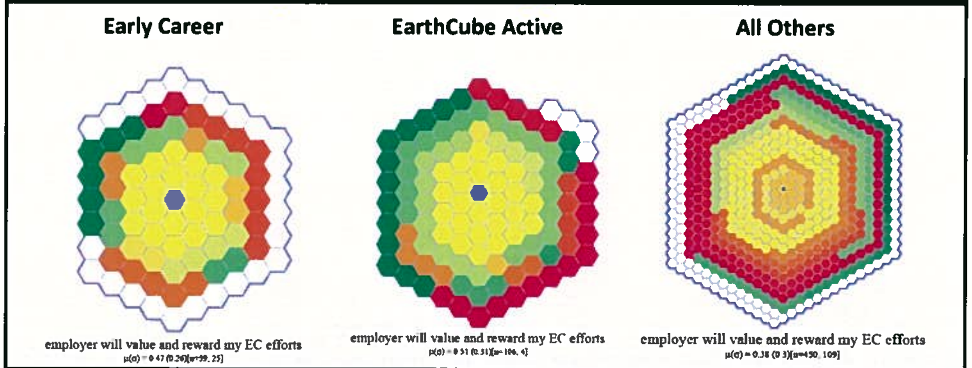
1. **NOAA (NODC, NGDC, NDBC, NCEP, etc.))** (17%)
2. **NASA (JPL, ESA, etc.)** (11%)
3. **Colleagues/Clients** (10%)
4. **The web (unspecified)** (9%)
5. **NCAR, UCAR, Unidata** (8%)
6. **Publications** (8%)
7. **USGS** (8%)
8. **IEDA (GeoRock, EarthChem, MGDS, etc.)** (8%)
9. **State and local government** (3%)
10. **International (PANGEA, etc.)** (2%)
11. **DOE** (2%)
12. **EPA** (1%)
13. **Google/Google Earth** (1%)
14. **NSIDC, NIC** (1%)
15. **USDA** (1%)
16. **IRIS, EarthScope** (1%)
17. **Neotoma, PBDB, Macrostrat** (1%)
18. **BCO-DMO, JGOFS, WOCE, CLIVAAR, Geotraces** (1%)
19. **IODP** (1%)
20. **DOD (Navy, Army, Army Corps of Engineers)** (1%)
21. **Open topography, NCALM** (1%)
22. **LTER** (1%)
23. **UNAVCO** (1%)
24. **MagIC** (under 1%)
25. **Private sector companies (IRI, ESRI)** (under 1%)

Note: All percentages rounded to the nearest whole number

What is the right balance between the social and technical?

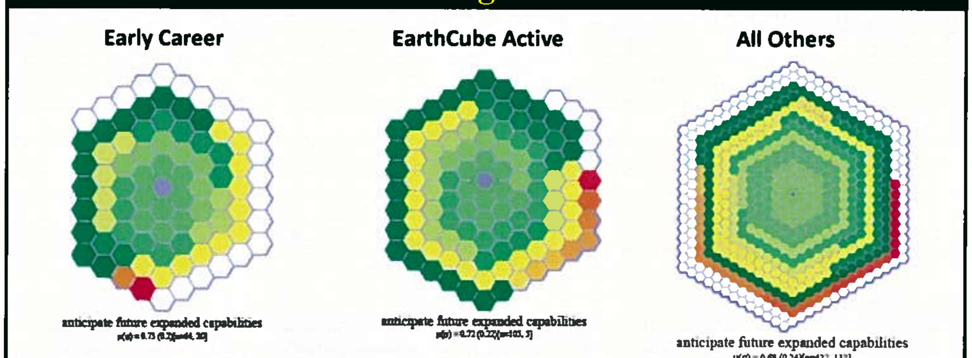
Additional Social/Institutional Elements:
- Norms for sharing data/models
- Established communications patterns
- Relations between Cyber and Geo
- Professional associations
- Promotion and tenure systems
- Roles of government agencies and private industry
- International partnerships
- Pressing societal challenges

Organizational/professional support



Comment: Positive and negative outliers need to be better understood; internal alignment may be a key barrier to full engagement with EarthCube	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
My employer/organization will most likely value and reward my efforts in the shaping and development of EarthCube	.50 (.34)	.38 (.29)	.57 (.26)	.51 (.31)	.38 (.30)
My employer/organization will most likely see my participation in the shaping and the development of EarthCube as an integral part of my job.	.45 (.33)	.35 (.29)	.40 (.27)	.50 (.30)	.31 (.29)
My contributions to the shaping and development of EarthCube will most likely be recognized and highly valued by colleagues in my field/domain	.54 (.32)	.51 (.30)	.58 (.22)	.56 (.25)	.44 (.28)

Future vision of success: Integration



Comment: A generally positive view of EarthCube's potential with some very strong positive outliers, a comparable number of neutral outliers, and a handful of very negative responses.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
In 5-7 years, I anticipate that EarthCube will result in substantially expanded capabilities to integrate multiple sources of data, datasets, observations, visualization, and models	.64 (.28)	.80 (.18)	.73 (.20)	.72 (.22)	.61 (.24)
In 5-7 years, I anticipate that EarthCube will result in substantially increased productivity for me and others whose work is similar to mine	.60 (.28)	.73 (.21)	.68 (.21)	.66 (.24)	.68 (.24)

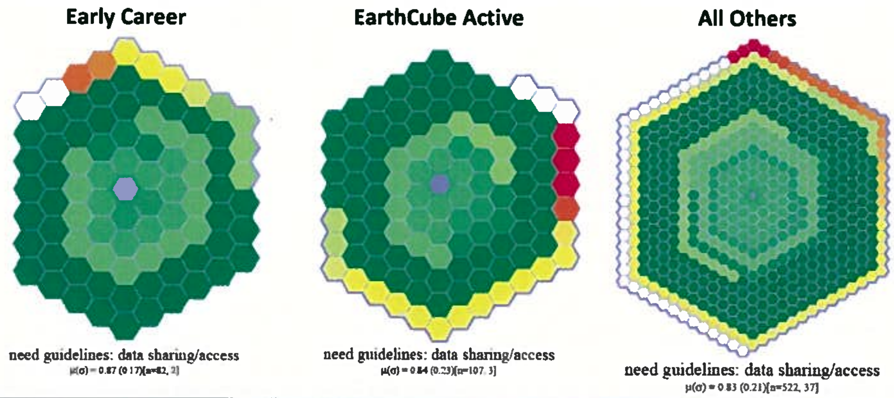
Daily/Weekly/Monthly frequency

Please indicate your work related interactions with Scientists/Educators in the following fields over the past 2-3 years:	Atmospheric or Space Weather scientist	Oceanographer	Geologist	Geophysicist	Hydrologist	Critical Zone Scientist	Climate Scientist	Biologist or Ecosystems Scientist	Geographer	Computer or Cyber-infrast. Scientist	Social Scientist	Other Scientist
Atmospheric/Space Weather Scientist	98.3	37.4	17.0	25.2	40.0	7.4	62.6	26.6	23.0	61.4	16.7	33.7
Oceanographer	49.4	97.9	49.4	36.4	28.4	16.0	59.1	76.1	10.7	37.8	9.2	36.8
Geologist	24.3	69.3	98.7	82.2	47.1	35.6	40.7	38.4	29.3	36.6	19.7	38.3
Geophysicist	14.3	35.1	78.7	95.2	28.4	25.0	36.4	25.7	13.7	47.5	10.8	42.4
Hydrologist	46.7	25.0	62.5	56.7	97.1	65.6	56.2	62.5	40.6	46.9	38.7	52.2
Critical Zone Scientist	22.2	25.0	90.0	70.0	90.0	100	60.0	70.0	50.0	50.0*	40.0	50.0
Climate Scientist	91.4	72.2	17.0	19.2	34.5	10.0	98.2	51.9	39.6	63.6	24.5	31.0
Biologist or Ecosystems Scientist	34.3	45.7	52.9	32.4	45.7	25.7	55.6	100	60.0	55.6	40.0	57.7
Geographer	58.8	47.1	50.0	35.3	52.9	35.3	58.8	77.8	100	64.7	47.1	53.3
Computer or Cyber-infrastructure Scientist	39.5	44.1	33.3	30.3	48.6	28.6	37.8	47.4	41.7	97.4	43.2	53.6
Social Scientist	0	25.0	60.0	20.0	40.0	40.0	20.0	50.0*	80.0	80.0	100	66.7
Other Scientist	22.7	40.0	65.3	49.0	41.3	23.3	51.1	60.5	41.7	40.8	27.3	72.7

The issues of how best to govern natural resources used by many individuals in common are no more settled in academia than in the world of politics. Some scholarly articles about the "tragedy of the commons" recommend that "the state" control most natural resources . . . Others recommend . . . privatization. . . What one can observe in the world, however, is that neither the state nor the market is uniformly successful in enabling individuals to sustain long-term, productive use of natural resource systems. ***Further, communities of individuals have relied on institutions resembling neither the state nor the market to govern some resource systems with reasonable degrees of success over long periods of time.***

Eleanor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, p. 1

Specifying guidelines regarding geoscience data



Comment: A "pull" for guidelines/standards (strongest by early career), with a small number of neutral or negative responses.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in discovering, accessing, sharing, and disseminating geoscience data	**.76 (.26)**	**.87 (.19)**	**.87 (.17)**	**.84 (.23)**	**.83 (.21)**
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in geoscience visualization tools	**.75 (.23)**	**.84 (.19)**	**.84 (.18)**	**.71 (.26)**	**.77 (.25)**

EarthCube Governance Considerations	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
Balancing support of research, on one hand (high number), and education, on the other (low number).	.63 (.15)	.66 (.20)	.68 (.16)	.63 (.16)	.61 (.20)	.64 (.20)	.65 (.20)
Interacting working groups – fluid, with no core group of leaders needed.	.37 (.28)	.38 (.30)	.44 (.21)	.36 (.25)	.24 (.22)	.30 (.26)	.34 (.27)
Primarily a grassroots, community-driven activity; the NSF should let it develop organically.	.56 (.24)	.63 (.22)	.65 (.19)	.59 (.23)	.49 (.24)	.57 (.24)	.58 (.26)
The NSF should play a major, active role in determining what EarthCube should be and how it should be run.	.63 (.18)	.59 (.27)	.42 (.26)	.48 (.19)	.57 (.26)	.53 (.23)	.51 (.26)
Community-elected leaders (low numbers) versus NSF-selected leaders (high numbers).	.42 (.24)	.37 (.24)	.36 (.25)	.33 (.16)	.45 (.21)	.40 (.23)	.39 (.23)
It is essential to have periodic face-to-face meetings (at least once a year) to support EarthCube.	.75 (.24)	.71 (.28)	.71 (.26)	.70 (.25)	.81 (.25)	.76 (.25)	.66 (.27)

Selected hopes and fears from Early Career workshop

HOPES

- One-stop shopping
- Improved access to data
- Better funding of data storage options
- Ease in citing data
- A "closed circle" from data production, use, review, and publication
- Aligned ontologies
- Able to "keep up" with "big data"
- Minimizing time collating data and maximizing time doing science
- "Hindcast" and predictive modeling capabilities
- International access
- Access to the general public
- Integration across fields, databases, and agencies
- A cultural shift in the field.

FEARS

- Duplication of efforts across directorates and disciplines
- Disconnect between data and science
- Data graveyard – useless collection of data
- Misuse/misinterpretation of data
- No one in our community wants to take the lead
- No incentive structure for publishing data
- Funding goes to data, not new research
- Not enough sustained funding – e.g. support data entry, curation, and storage
- Creating separation/class stratification between data generators and users
- Error propagation through datasets
- Don't lose the ability to do small projects; Don't suppress novel data collection
- Intellectual property "violations"
- No willingness to collaborate; too rigid or not rigid enough
- Vulnerability to Cyber-attacks and malicious data use

Today's most troubling and daunting problems have common features: some of them arise from human numbers and resource exploitation; they require long-term commitments from separate sectors of society and diverse disciplines to solve; simple, unidimensional solutions are unlikely; and failure to solve them can lead to disasters.

In some ways, the scales and complexities of our current and future problems are unprecedented, and it is likely that solutions will have to be iterative . . .

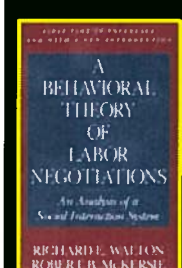
Institutions can enable the ideas and energies of individuals to have more impact and to sustain efforts in ways that individuals cannot.

From "Science to Sustain Society," by Ralph J. Cicerone, President, National Academy of Sciences, 149th Annual Meeting of the Academy (2012)

Appendix A: Internal Alignment for Lateral Alignment

"The organizations participating in labor negotiations usually lack internal consensus about the objectives they will attempt to obtain from negotiations. . . Different elements of the organization may have different ideas about the priorities assigned to various objectives. . . Disagreements can also exist around the strategies and tactics. . . Similarly, there may be a lack of consensus about what type of relationship should be developed with the other party. These are only illustrative of the internal differences that can exist over ends and/or means."

Richard Walton and Robert McKersie, *A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System*, p. 281.



Internal alignment and the Biomarkers Consortium

"[This is a] completely different way of making decisions among organizations. [I am] really dependent on getting the 'ears' of members of my own organization. "

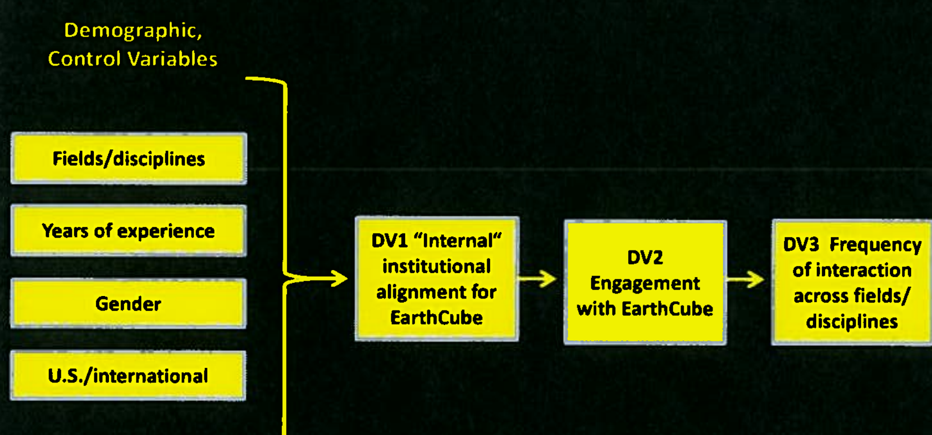
"Decision-making is facilitated because my. . .director is on the [BC] Executive Committee. This is very helpful and, as a result, there is a lot of communication. . . [It] would be a very different situation if my director was not on the EC.."

"I could speak for my division, regarding resources. Company reps could never speak for the commitments their company could make. Much work was being done without clear indication of the support from industry."

"What would have been helpful is for FDA internally to talk among themselves and compare perspectives. [They] didn't organize ourselves to come out with a unified FDA perspective or to present a majority and minority view. . . that could have been a useful process thing to do."

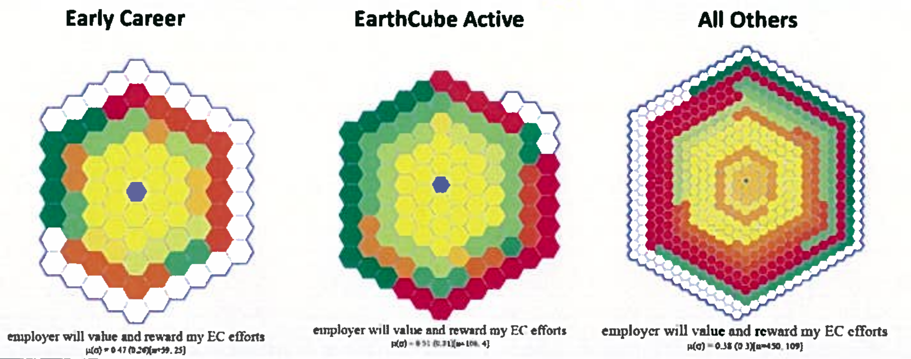
"[We] had to build up internal decision-making —it was very ad hoc at the beginning. . . [now] we all get together—there is a champion, an application process, and we decide if it is good for [the company] and finance has carved out a protected budget. [Other firms have] done the same. In the absence, the ad hoc process would be a real pain."

Preliminary causal models testing "internal alignment" proposition for EarthCube



DV2	1 = First I have heard of EarthCube	45%
	2 = Aware of EarthCube, but no engagement	28%
	3 = Visited the EarthCube website	11%
	4 = Participated in EarthCube discussions	7%
	5 = Actively involved in EarthCube communities	7%
	6 = Leadership role in EarthCube communities	2%

Return to "internal" institutional alignment – DV1



Comment: Positive and negative outliers need to be better understood; internal alignment may be a key barrier to full engagement with EarthCube	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
My employer/organization will most likely value and reward my efforts in the shaping and development of EarthCube	.50 (.34)	.38 (.29)	.57 (.26)	.51 (.31)	.38 (.30)
My employer/organization will most likely see my participation in the shaping and the development of EarthCube as an integral part of my job.	.45 (.33)	.35 (.29)	.40 (.27)	.50 (.30)	.31 (.29)
My contributions to the shaping and development of EarthCube will most likely be recognized and highly valued by colleagues in my field/domain	.54 (.32)	.51 (.30)	.58 (.22)	.56 (.25)	.44 (.28)

Average frequency of interaction across fields –DV3

(5=daily; 4=weekly; 3=monthly; 2=quarterly; 1=annual; 0=never)

	mean (s.d.)
• Geographers	2.85 (.65)
• Hydrologists	2.62 (.76)
• Critical zone scientists	2.60 (.91)
• Climate scientists	2.47 (.82)
• Biologist and ecosystems scientists	2.46 (.74)
• Computer and cyberinfrastructure scientists	2.34 (.96)
• Geologists	2.35 (.87)
• Oceanographers	2.26 (.72)
• Geophysicists	2.15 (.79)
• Atmospheric and space weather scientists	2.01 (.93)

Daily/Weekly/Monthly frequency – DV3

Please indicate your work related interactions with Scientists/Educators in the following fields over the past 2-3 years:	Atmospheric or Space Weather scientist	Oceanographer	Geologist	Geophysicist	Hydrologist	Critical Zone Scientist	Climate Scientist	Biologist or Ecosystems Scientist	Geographer	Computer or Cyber-infrast. Scientist	Social Scientist	Other Scientist
Atmospheric/Space Weather Scientist	98.3	37.4	17.0	25.2	40.0	7.4	62.6	26.6	23.0	61.4	16.7	33.7
Oceanographer	49.4	97.9	49.4	36.4	28.4	16.0	59.1	76.1	10.7	37.8	9.2	36.8
Geologist	24.3	69.3	98.7	82.2	47.1	35.6	40.7	38.4	29.3	36.6	19.7	38.3
Geophysicist	14.3	35.1	78.7	95.2	28.4	25.0	36.4	25.7	13.7	47.5	10.8	42.4
Hydrologist	46.7	25.0	62.5	56.7	97.1	65.6	56.2	62.5	40.6	46.9	38.7	52.2
Critical Zone Scientist	22.2	25.0	90.0	70.0	90.0	100	60.0	70.0	50.0	50.0*	40.0	50.0
Climate Scientist	91.4	72.2	17.0	19.2	34.5	10.0	98.2	51.9	39.6	63.6	24.5	31.0
Biologist or Ecosystems Scientist	34.3	45.7	52.9	32.4	45.7	25.7	55.6	100	60.0	55.6	40.0	57.7
Geographer	58.8	47.1	50.0	35.3	52.9	35.3	58.8	77.8	100	64.7	47.1	53.3
Computer or Cyber-infrastructure Scientist	39.5	44.1	33.3	30.3	48.6	28.6	37.8	47.4	41.7	97.4	43.2	53.6
Social Scientist	0	25.0	60.0	20.0	40.0	40.0	20.0	50.0*	80.0	80.0	100	66.7
Other Scientist	22.7	40.0	65.3	49.0	41.3	23.3	51.1	60.5	41.7	40.8	27.3	72.7

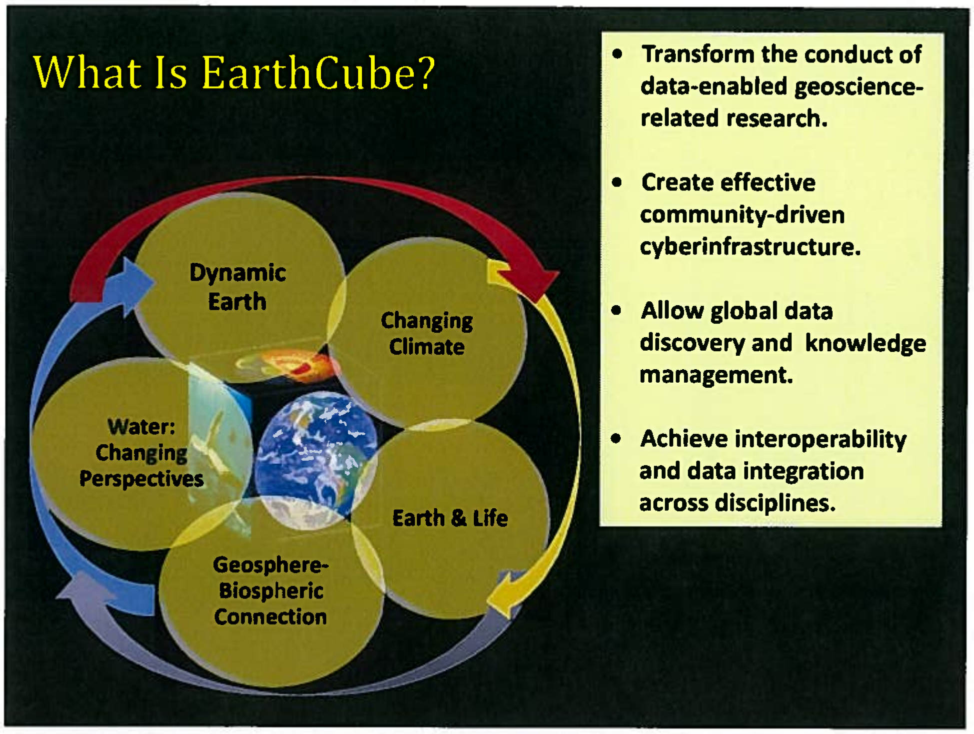
OLS Models Examining 'Internal Alignment for Lateral Alignment'

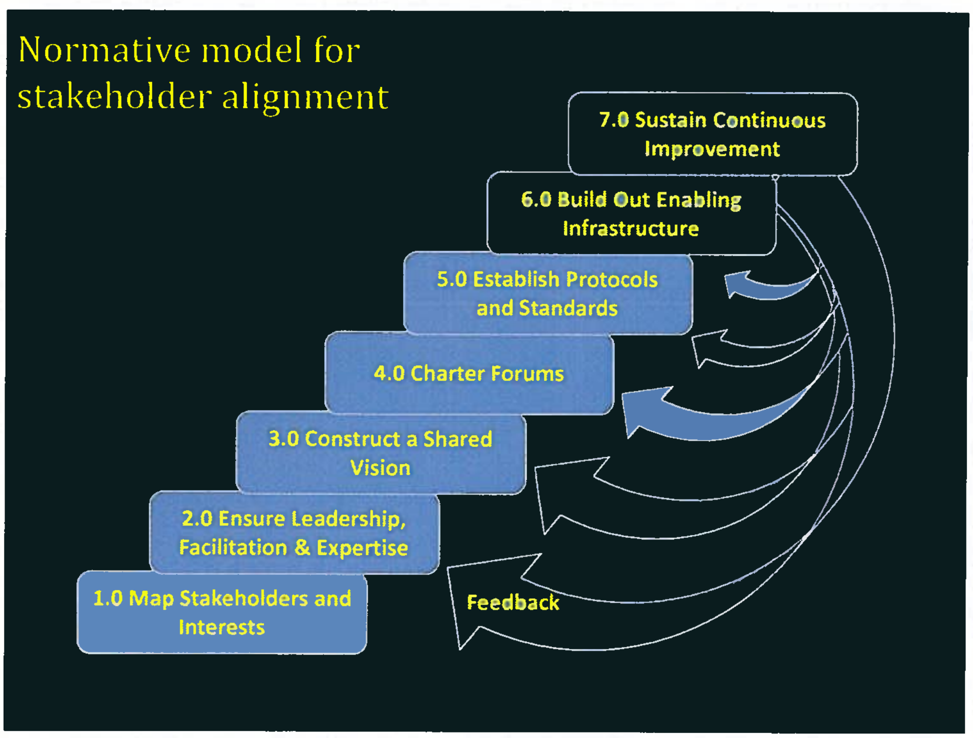
	Model 1 DV= Internal Alignment		Model 2 DV= EarthCube Engagement		Model 3 DV= Above Ave. Frequen	
Independent Variables – OLS	B	S.E.	B	S.E.	B	S.E.
Constant	.392	.050***	2.907	.252***	.278	.101**
Under 5 yrs. Experience (0,1)	.115	.039***	-.859	.191***	-.054	.070
5-10 yrs. Experience (0,1)	.093	.029**	-.558	.141***	.073	.052
11-20 yrs. Experience (0,1)	.024	.025	-.291	.122*	.107	.044*
Over 20 yrs. Experience (0,1)						
Inst. Affiliation (Intl./U.S.) (0,1)	.046	.029	-1.085	.140***	-.073	.053
Gender (Female, Male) (0,1)	.025	.024	.093	.118	.108	.043*
Atmospheric/Space Weather (0,1)	-.081	.041	-1.042	.199***	-.005	.074
Oceanographers	-.143	.043***	-1.115	.210***	-.332	.078***
Geologists (0,1)	-.111	.039**	-.620	.187***	.092	.068
Geophysicists (0,1)	-.129	.043**	-.526	.206*	.051	.075
Hydrologists (0,1)	-.137	.057*	-.515	.276	.184	.100
Critical Zone Scientists (0,1)	-.176	.096	.730	.462	.117	.168
Climate Scientists (0,1)	-.099	.048*	-1.439	.232***	.090	.087
Biologists or Ecosystems Scientists (0,1)	-.059	.057	-.859	.275**	.231	.100*
Geographers (0,1)	.022	.080	-.799	.384*	.147	.139
Social Scientists (01)	-.041	.119	-.344	.572	.020	.207
Computer/Cyber Scientists (0,1)						
Data Manager (0,1)	.038	.067	-.406	.321	.001	.116
All Others (0,1)	.015	.049	-.251	.235	-.246	.085**
"Internal" Alignment (Alpha=.85)			1.225	.199***	-.011	.074
EarthCube Engagement (1-6)					.029	.015
*p<.05 **p<.01 ***p<.001 Adj. R^2 and F	.06	F=3.11***	.23	F=11.18***	.12	F=5.29***

Implications

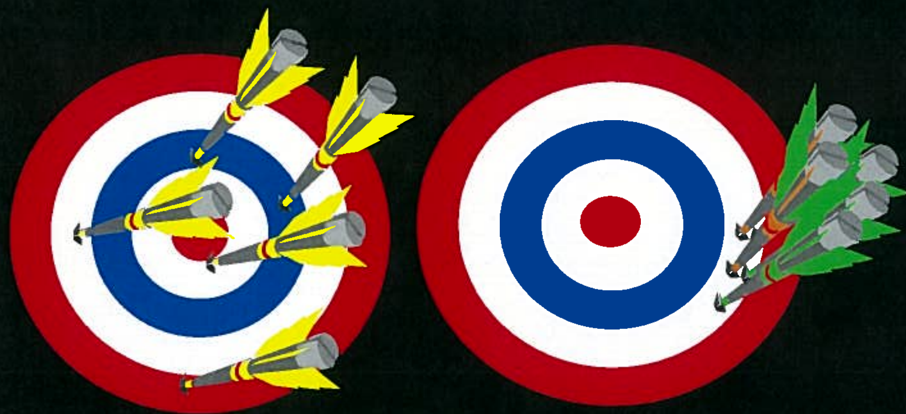
- Early career scientists perceive more institutional support for engagement than the most senior scientists, but are less engaged – risking a cyber infrastructure that doesn't meet the needs of the next generation
- Internationally based scientists are more likely to be engaged and approaching statistical significance on support for engagement and on interacting with others – making international linkages promising
- Many geoscience domains perceive less institutional support and are less institutionally engaged than computer/cyber scientists – risking a "build it and they don't come" scenario – the current array of domain workshops are crucial
- "Internal" institutional alignment in support of engagement with EarthCube is a strong, positive correlate with engagement – indicating the value of "internal alignment for lateral alignment"
- Institutional alignment for engagement with EarthCube doesn't impact frequency of interaction – a measure to track over time

Appendix B: Systems Change Materials


What Is EarthCube?
Dynamic Earth
Changing Climate
Water: Changing Perspectives
Earth & Life
Geosphere-Biospheric Connection
• 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.

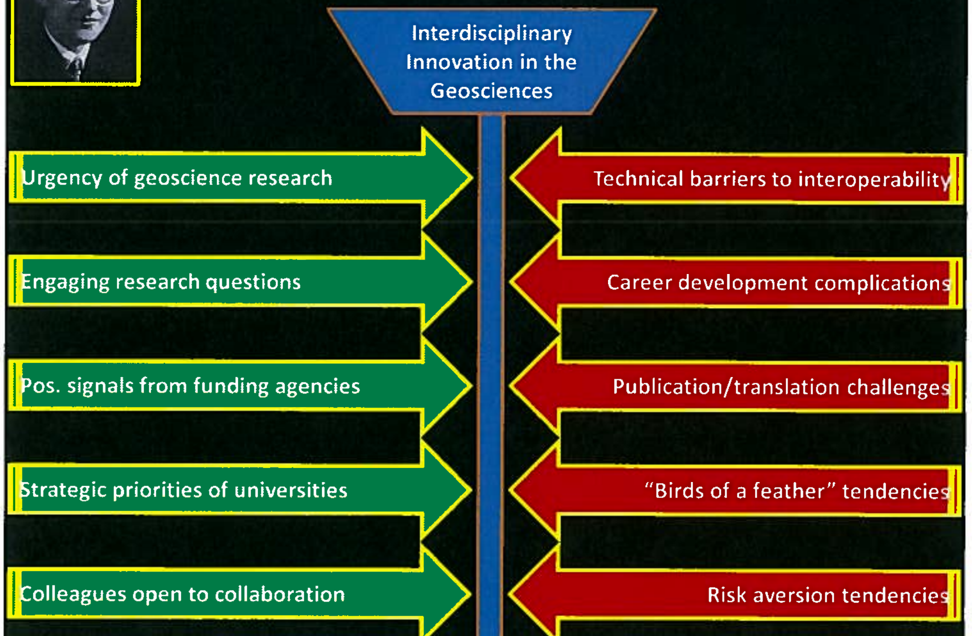

Normative model for stakeholder alignment
7.0 Sustain Continuous Improvement
6.0 Build Out Enabling Infrastructure
5.0 Establish Protocols and Standards
4.0 Charter Forums
3.0 Construct a Shared Vision
2.0 Ensure Leadership, Facilitation & Expertise
1.0 Map Stakeholders and Interests
Feedback

Stabilize before you improve

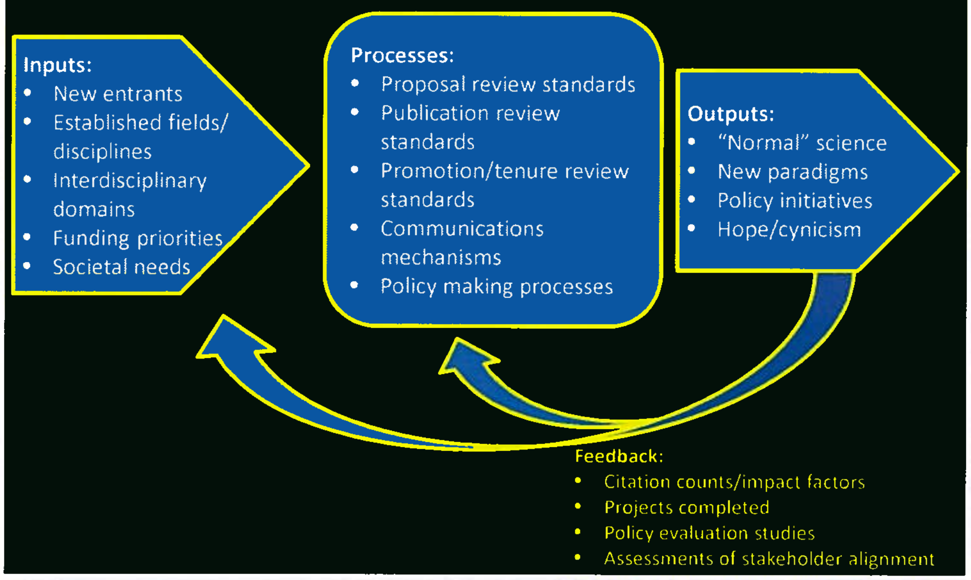


Which player did better in this round?
Who will do better in the long run?

Lewin's force field analysis



Geosciences and Cyberinfrastructure as a system



Summarizing action recommendations anchored in first principles

- Principle 1: Identify and advance public goods
  - Action: Be positive (It is easier to be negative, and it's a recipe for gridlock)
- Principle 2: Assume independent, but interdependent stakeholders
  - Action: Specify stakeholders (at appropriate levels of granularity)
- Principle 3: Value both common and competing interests
  - Action: Identify interests (current state and future potential)
- Principle 4: Data-informed decisions
  - Action: Accelerate alignment via visualization tools and methods, including visual representations of stakeholder perceptions and behaviors at the systems level (don't act on the basis of unchecked assumptions; appreciate diverse perspectives)
- Principle 5: Internal alignment for lateral alignment
  - Action: Align behaviors, structure, strategy and culture within organizations for alignment across organizations (and across layers). "n" parties need at least "n + 1" agreements – one within each party and at least one among the parties.

Tragedy of the Commons

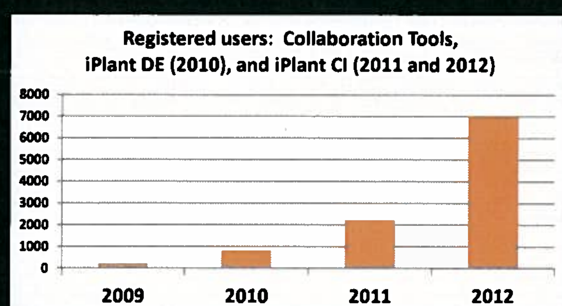
"Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit – in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. Freedom in a commons brings ruin to all."

Garrett Hardin (1968). Tragedy of the Commons. *Science*. 162, 1243-1248.

Norman W. Hudson. (1987). Soil Resources, Management and Conservation Service. Food and Agriculture Organization of the United Nations, Rome.

iPlant Collaborative™ *Empowering A New Plant Biology*

- Genotype to Phenotype (iPG2P)
- iPlant Tree of Life (iPToL)
- Seed Projects



CoGe Accelerating Comparative Genomics

PHYLOtastic

TARGeT
Maintained by: James Burnette

Galaxy

Center for Genome Research & Biocomputing

CIPRES

BrachyBio! Data Central

Appendix C: Principles of Stakeholder Alignment

Preliminary findings on **Formation** from case studies...

> Increased visibility of stakeholder interests will accelerate stakeholder dialogue and alignment – avoiding "dead ends" and pursuing opportunities

A. A shared vision of success will enable faster formation and more robust forms of stakeholder alignment

B. Lateral alignment across stakeholders will be constrained by the internal alignment within stakeholder organizations

C. Initial stakeholder alignment will depend on trust; sustained stakeholder alignment will depend on new structural arrangements (forums, roles, incentives, etc.)

Preliminary findings on **Operations** from case studies. . .

D. Sustained stakeholder alignment will require <u>leadership</u> based on influence, rather than authority

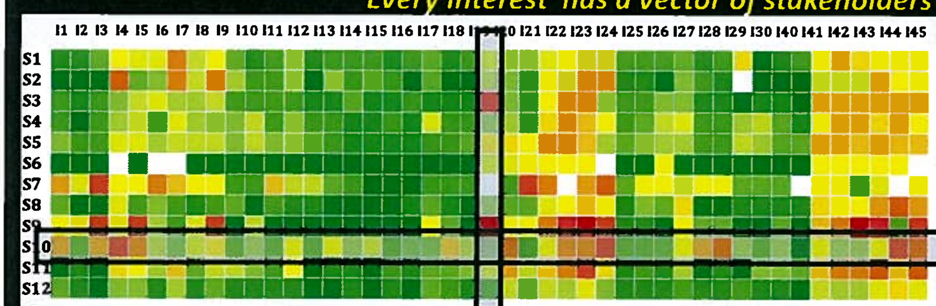
E. <u>Forums</u> that are "over specified" or "under specified" will ineffective in advancing both individual and collective interests

E. Without well-specified <u>protocols and standards</u>, top-down and bottom-up change initiatives be more variable and harder to sustain

F. Failure to deliver on both <u>individual and collective interests</u> will erode stakeholder alignment and systems success

Specify Stakeholders
Identify Interests
Conceptualize as a Matrix

Every Interest has a vector of stakeholders

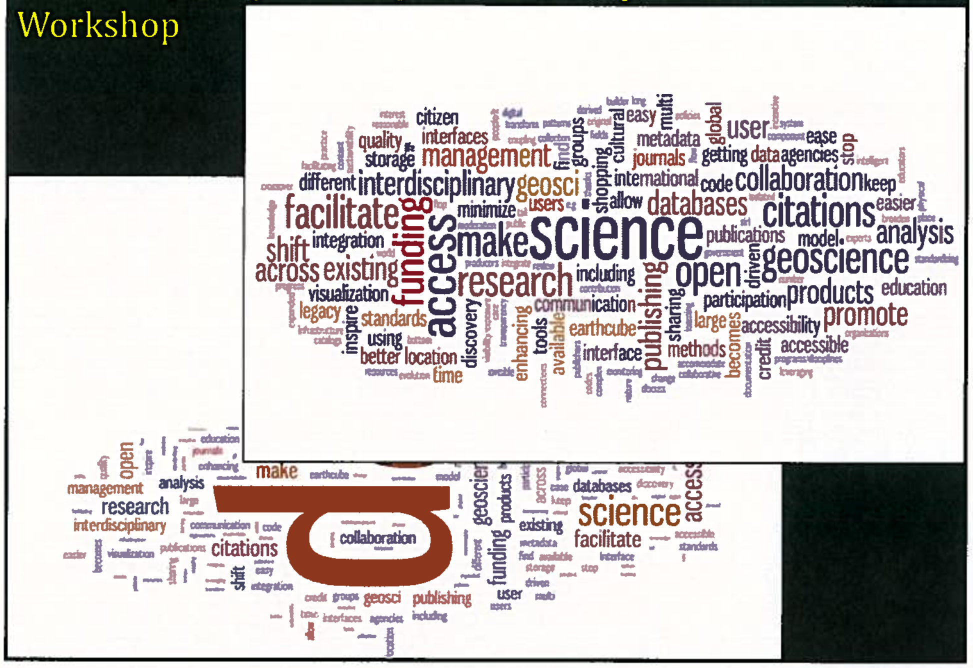


Every stakeholder has a vector of interests

Appendix D: Early Career Workshop

Word frequency for "Fears" from Early Career Workshop

Word frequency for "Hopes" from Early Career Workshop



Selected elements of success from Early Career workshop

Access/Uploading:

- *Google earth style interface*
- *Accessible data submission interface*
- *Standardized meta data on data type, data context, data provenance, etc. for field scientists (with and without internet access)*
- *Data security*
- *Public accessibility; empower non-specialists*

Utilization/Operations:

- *Community mechanisms to build tools*
- *Large data manipulation, visualization, and animation*
- *Searchable access by space, time, and context*
- *Pull up data and conduct analysis with voice commands*
- *Open source workflow management for data processing and user-contributed algorithms in order to facilitate reproducible research*
- *Cross-system comparisons; ontology crosswalks for different vocabs in different disciplines*
- *Easy integration of analytic tools (R, Matlab, etc.)*
- *NSF support for data management*

Output/Impact:

- *Mechanisms to provide credit for work done (data, models, software, etc.); ease of citations; quantify impact*
- *Promote new connections between data producers and data consumers*
- *Interactive publications from text to data*
- *Recommendations system (like Amazon) for data, literature, etc.; Flickr for data (collaborative tagging)*
- *Educational tutorials for key geoscience topics (plate tectonics, ice ages, population history, etc.)*
- *Gaming scenarios for planet management*
- *EarthCube app store; ecosystem of apps*

Appendix E: Transformative potential

Elements of a success vision for EarthCube

- That it will indeed transform the way that we do science, and that it will increase the level of respect and understanding between disciplines.
- Establishment of standards for data sharing and development of standard tools.
- Almost as easy to navigate as Google Earth
- Being able to foster greater data interoperability and positively influence stakeholders into providing transparent and unified policies for data sharing.
- That I can test out ideas quickly and learn from the data in a hierarchical manner. The data should be presented in various ways, starting with basic science questions that are answered by the data and working towards more refined questions.
- An increase in the number of cross-disciplinary, peer-reviewed research papers published using data provided via EarthCube.
- Align the needs of researchers with a K-12 outreach component so that we can better prepare the next generation of geoscientists.
- Who knows? let's saddle up and see where this horse takes us.

EarthCube as a top priority

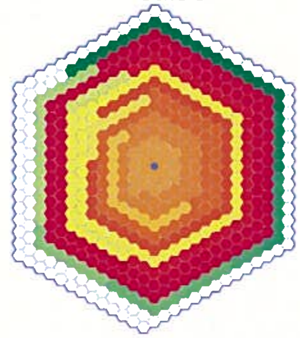
Early Career

EC will be one of my top 3 commitments
$\mu(\sigma) = 0.38\ (0.29)[n=64,\ 20]$

EarthCube Active

EC will be one of my top 3 commitments
$\mu(\sigma) = 0.58\ (0.29)[n=105,\ 5]$

All Others



EC will be one of my top 3 commitments

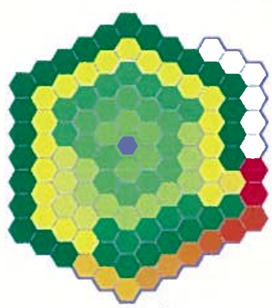
Comment: Strongly positive stakeholders among high EC involvement is to be expected; strongly negative is a surprise.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
Assisting in the development and operation of EarthCube will be one of my top three professional commitments in the next 3-5 years	.43 (.31)	.50 (.31)	.38 (.29)	.58 (.29)	.29 (.29)
I plan to devote time to help EarthCube development and operations, irrespective of whether or not I receive NSF funding or not	.44 (.24)	.46 (.24)	.48 (.25)	.51 (.29)	.38 (.29)

Transformation potential of EarthCube

Early Career

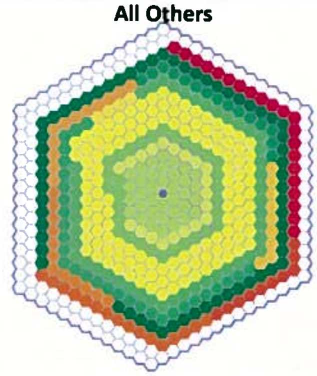
EC will transform geoscience research
$\mu(\sigma) = 0.68\ (0.2)[n=65,\ 19]$

EarthCube Active



EC will transform geoscience research
$\mu(\sigma) = 0.69\ (0.24)[n=104,\ 6]$

All Others

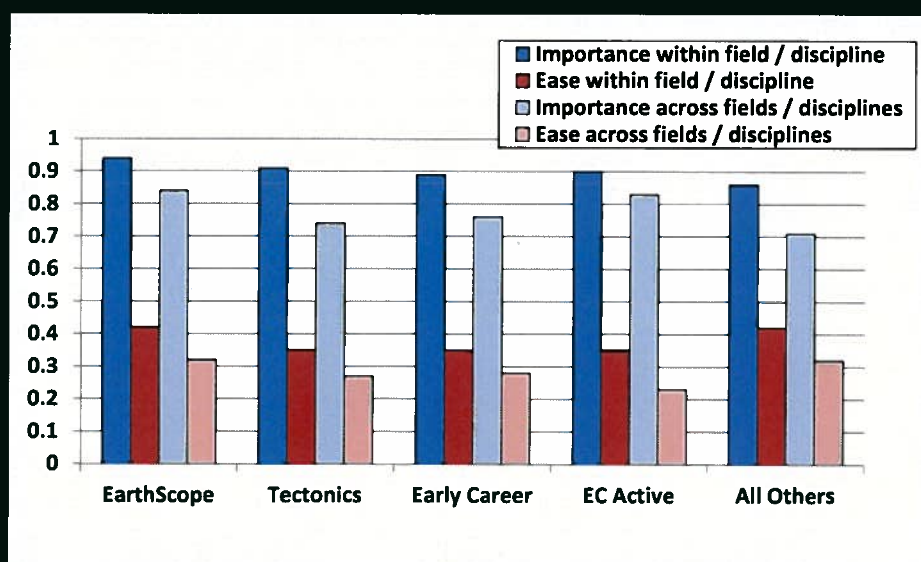


EC will transform geoscience research

Comment: Early career has few negative outliers; EarthCube active is more bimodal; overall responses are positive, but not exceedingly positive.	Earth-Scope μ (α)	Tec-tonics μ (α)	Early Career μ (α)	EC Active μ (α)	All Others μ (α)
As an integrated data and knowledge management system for the geosciences, EarthCube will transform the way geoscience research is conducted	.68 (.29)	.74 (.22)	.68 (.20)	.69 (.24)	.59 (.25)
As an integrated data and knowledge management system for the geosciences, EarthCube will transform the way geoscience education is conducted	.61 (.30)	.72 (.17)	.62 (.22)	.64 (.24)	.58 (.23)

Appendix F: "Importance and Ease" by fields and disciplines

"Importance" and "Ease" as a bar chart



Importance of access to data, observations, visualizations, tools and/or models by field/discipline (with top two highlighted in each row)

	Atmospheric/ space Weather scientist	Oceanographer	Geophysicist	Geologist	Hydrologist	Critical Zone Scientist	Climate Scientist	Biologist or Ecosystems Scientist	Geographer	Computer/Cyber-infrastructure
Atmospheric/ space weather data	.96 (.11)	.66 (.34)	.36 (.35)	.36 (.34)	.72 (.35)	.65 (.27)	.90 (.23)	.57 (.31)	.71 (.30)	.80 (.24)
Oceanographic data	.50 (.33)	.93 (.17)	.43 (.32)	.41 (.33)	.32 (.36)	.30 (.24)	.66 (.29)	.54 (.39)	.58 (.36)	.71 (.28)
Geophysical data	.60 (.32)	.46 (.35)	.94 (.14)	.64 (.29)	.76 (.29)	.59 (.35)	.56 (.35)	.51 (.32)	.66 (.28)	.65 (.30)
Geologic data	.26 (.32)	.37 (.35)	.80 (.23)	.89 (.18)	.75 (.31)	.69 (.19)	.25 (.31)	.57 (.33)	.66 (.30)	.58 (.31)
Hydrologic data	.56 (.35)	.50 (.32)	.37 (.32)	.43 (.35)	.98 (.05)	.80 (.25)	.66 (.32)	.67 (.31)	.75 (.29)	.76 (.27)
Critical zone data	.17 (.31)	.22 (.32)	.53 (.34)	.50 (.37)	.79 (.33)	.95 (.09)	.19 (.31)	.41 (.36)	.60 (.36)	.55 (.35)
Climate data	.77 (.29)	.68 (.32)	.37 (.34)	.48 (.36)	.77 (.30)	.67 (.29)	.97 (.09)	.77 (.24)	.73 (.26)	.79 (.23)
Average of means for "importance" of top three fields other than main	.64	.62	.57	.54	.77	.72	.72	.59	.68	.75

Ease of access to data, observations, visualizations, tools and/or models by field/discipline (with top two highlighted in each row)

	Atmospheric/ space Weather scientist	Oceanographer	Geophysicist	Geologist	Hydrologist	Critical Zone Scientist	Climate Scientist	Biologist or Ecosystems Scientist	Geographer	Computer/ Cyberinfrastructure
Atmospheric/ space weather data	.58 (.23)	.50 (.19)	.45 (.26)	.49 (.23)	.38 (.25)	.41 (.27)	.57 (.24)	.35 (.18)	.46 (.16)	.46 (.23)
Oceanographic data	.46 (.21)	.53 (.22)	.46 (.24)	.48 (.21)	.40 (.23)	.43 (.32)	.49 (.21)	.40 (.25)	.46 (.26)	.45 (.20)
Geophysical data	.50 (.22)	.47 (.22)	.57 (.24)	.45 (.21)	.33 (.22)	.36 (.32)	.48 (.20)	.40 (.23)	.44 (.25)	.45 (.20)
Geologic data	.41 (.22)	.40 (.18)	.45 (.24)	.53 (.21)	.36 (.23)	.38 (.21)	.45 (.22)	.41 (.22)	.47 (.22)	.40 (.19)
Hydrologic data	.44 (.25)	.42 (.21)	.38 (.24)	.47 (.24)	.45 (.21)	.45 (.23)	.43 (.21)	.38 (.26)	.52 (.26)	.44 (.24)
Critical zone data	.27 (.24)	.32 (.21)	.35 (.20)	.43 (.22)	.33 (.24)	.35 (.22)	.33 (.22)	.35 (.21)	.39 (.20)	.36 (.19)
Climate data	.52 (.23)	.50 (.19)	.47 (.24)	.48 (.23)	.43 (.24)	.34 (.23)	.58 (.21)	.46 (.25)	.48 (.20)	.42 (.25)
Average of means for "ease" of top three other than main source	.49	.49	.46	.48	.40	.41	.51	.40	.47	.45

Appendix G: Frequency of interaction and multivariate models

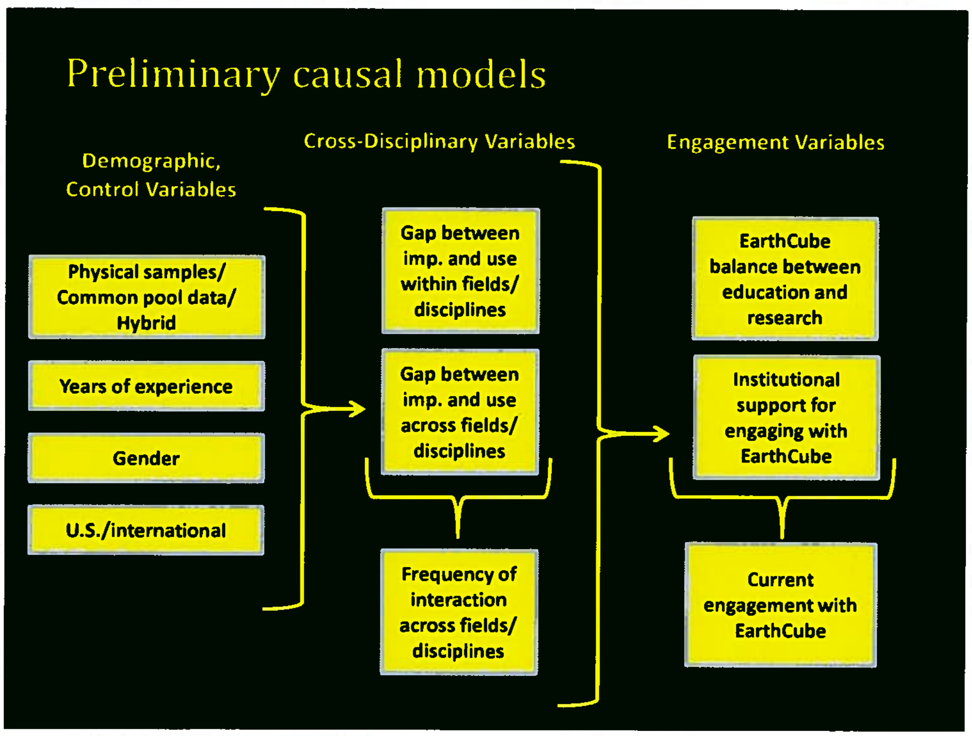
Average frequency of interaction across fields

(5=daily; 4=weekly; 3=monthly; 2=quarterly; 1=annual; 0=never)

	mean (s.d.)
• Geographers	2.85 (.65)
• Hydrologists	2.62 (.76)
• Critical zone scientists	2.60 (.91)
• Climate scientists	2.47 (.82)
• Biologist and ecosystems scientists	2.46 (.74)
• Computer and cyberinfrastructure scientists	2.34 (.96)
• Geologists	2.35 (.87)
• Oceanographers	2.26 (.72)
• Geophysicists	2.15 (.79)
• Atmospheric and space weather scientists	2.01 (.93)

Daily/Weekly/Monthly frequency

Please indicate your work related interactions with Scientists/Educators in the following fields over the past 2-3 years:	Atmospheric or Space Weather scientist	Oceanographer	Geologist	Geophysicist	Hydrologist	Critical Zone Scientist	Climate Scientist	Biologist or Ecosystems Scientist	Geographer	Computer or Cyber-infrast. Scientist	Social Scientist	Other Scientist
Atmospheric/Space Weather Scientist	98.3	37.4	17.0	25.2	40.0	7.4	62.6	26.6	23.0	61.4	16.7	33.7
Oceanographer	49.4	97.9	49.4	36.4	28.4	16.0	59.1	76.1	10.7	37.8	9.2	36.8
Geologist	24.3	69.3	98.7	82.2	47.1	35.6	40.7	38.4	29.3	36.6	19.7	38.3
Geophysicist	14.3	35.1	78.7	95.2	28.4	25.0	36.4	25.7	13.7	47.5	10.8	42.4
Hydrologist	46.7	25.0	62.5	56.7	97.1	65.6	56.2	62.5	40.6	46.9	38.7	52.2
Critical Zone Scientist	22.2	25.0	90.0	70.0	90.0	100	60.0	70.0	50.0	50.0*	40.0	50.0
Climate Scientist	91.4	72.2	17.0	19.2	34.5	10.0	98.2	51.9	39.6	63.6	24.5	31.0
Biologist or Ecosystems Scientist	34.3	45.7	52.9	32.4	45.7	25.7	55.6	100	60.0	55.6	40.0	57.7
Geographer	58.8	47.1	50.0	35.3	52.9	35.3	58.8	77.8	100	64.7	47.1	53.3
Computer or Cyber-infrastructure Scientist	39.5	44.1	33.3	30.3	48.6	28.6	37.8	47.4	41.7	97.4	43.2	53.6
Social Scientist	0	25.0	60.0	20.0	40.0	40.0	20.0	50.0*	80.0	80.0	100	66.7
Other Scientist	22.7	40.0	65.3	49.0	41.3	23.3	51.1	60.5	41.7	40.8	27.3	72.7



Preliminary hypotheses

- H1: Fields/disciplines using remote "array" data will be more cross disciplinary and more engaged with EC than those relying on "point" data
- H2: Early career scientists will be more cross disciplinary and more engaged with EC than more senior scientists
- H3: Women will be more will be more cross disciplinary, but less engaged with EC then men
- H4: International scientists will be more oriented toward will be more cross disciplinary, but less engaged with EC than U.S. scientists
- H5: Institutional support will impact views on cross disciplinary science and engagement with EC

Predicting views on the importance/ease gap (within and across fields/disciplines) and frequency of interactions with other scientists/scholars

* p < .05 ** p < .01 *** p < .001	Model 1 Imp./Ease Within		Model 2 Imp./Ease Across		Model 3 Above Ave. Frequency	
Independent Variables – OLS	B	S.E.	B	S.E.	B	S.E.
Constant	.429	.040***	.470	.047***	.148	.072*
Array Data (Atm, Geophys, Climate) (0,1)	-.034	.025	-.107	.030***	.121	.042**
Point Data (Geol, Hydr, Crit Zone) (0,1)	-.001	.027	-.028	.032	.193	.045***
Hybrid Data (Ocean, Geog) (0,1)						
Under 5 yrs. Experience (0,1)	.133	.037***	.079	.045	-.068	.064
5-10 yrs. Experience (0,1)	.075	.029**	.066	.034	.071	.048
11-20 yrs. Experience (0,1)	.091	.026***	.073	.032*	.132	.044**
Over 20 yrs. Experience (0,1)						
Inst. Affiliation (Intl./U.S.) (0,1)	.047	.028	.040	.033	-.086	.045
Gender (Female, Male) (0,1)	-.068	.024**	-.047	.029	.136	.040***
Imp./Ease Within Fields/Disciplines					-.061	.082
Imp./Ease Across Fields/Disciplines					.159	.072*
Model Adj. R^2 and F	.04	F=5.11***	.03	F=3.48***	.07	F=6.03***

Preliminary Findings:

- ***The level of "pain" as indicated by the "importance/ease" gap is lowest for the most experienced***
- ***The "hybrid data" scientists have a greater "importance/ease" gap than the "array data" scientists***
- ***Both the "array data" and the "point data" scientists interact with others more than "hybrid data"***
- ***Compared with the most experienced, 11-20 years is above average in interactions with others***
- ***Women are more likely to experience the "importance/ease" gap, but less likely to interact with others***
- ***Those with higher "importance/ease" gaps are more likely to interact with others***

Predicting engagement with EarthCube and views on its transformational potential

* p < .05 ** p < .01 *** p < .001	Model 1 Engage EarthCube		Model 2 EC Trans. Research		Model 3 EC Trans. Education	
Independent Variables – OLS	B	S.E.	B	S.E.	B	S.E.
Constant	-1.882	.289***	.189	.051***	.459	.051***
Array Data (Atm, Geophys, Climate) (0,1)	-.322	.134*	-.001	.024	.003	.024
Point Data (Geol, Hydr, Crit Zone) (0,1)	-.010	.145	.028	.025	.012	.026
Hybrid Data (Ocean, Geog) (0,1)						
Under 5 yrs. Experience (0,1)	-.792	.214***	.070	.039	-.005	.038
5-10 yrs. Experience (0,1)	-.632	.153***	.063	.027*	.043	.027
11-20 yrs. Experience (0,1)	-.337	.138*	.017	.024	.009	.024
Over 20 yrs. Experience (0,1)						
Inst. Affiliation (Intl./U.S.) (0,1)	-1.045	.152***	.039	.026	.053	.026*
Gender (Female, Male) (0,1)	.128	.128	-.036	.023	.009	.023
Imp./Ease Within Fields/Disciplines	-.267	.259	.050	.046	.067	.045
Imp./Ease Across Fields/Disciplines	1.102	.227***	.166	.040***	.098	.040*
Above Average Freq. of Interactions (0,1)	.247	.122*	-.013	.021	-.024	.021
Institutional Support Scale (alpha=.85)	1.212	.219***	.357	.039***	.270	.039***
Balance between Research & Education	-.063	.287	.229	.050***	-.176	.050***
Model Adj. R^2 and F	.23	F=13.39***	.31	F=17.62***	.18	F=9.18***

Preliminary findings:

- ***"Array data" scientists are less likely to be engaged with EarthCube than "hybrid data" scientists***
- ***Over 20 years experience is more likely to be engaged with EarthCube than all other experience levels***
- ***Scientists experiencing higher "importance/ease" gaps across fields, engaged in more interactions with others, and receiving more institutional support are more likely to engage with EarthCube***
- ***Institutional support impacts views on the potential for EarthCube***
- ***Views on the balance between education and research via EarthCube impact views on overall potential***

Preliminary implications from OLS models

- While very senior scientists are more likely to be engaged with EarthCube, they are also less likely to experience the "importance/ease" gap
- Institutional support for engagement with EarthCube is essential
- Women are more likely to be positive toward EarthCube, but less frequently interacting with others
- Internationally-based scholars are less likely to be engaged with EarthCube, but more likely to value the educational mission
- There is a clear distinction between those valuing the research and educational missions of EarthCube

Appendix H: Difficult to get and share data

"It is difficult to access the data I need . . ." *(sample items)*

1. There are few centralized repositories of earth science data
2. There are too many sources and formats
3. Documentation is often HORRIBLE
4. People do not share their data willingly
5. Initial data discovery can be difficult, if it is housed in a non-standard archive
6. Numerous large files make downloading, storage, retrieval, etc. very time consuming
7. I am a user of fortran-77. I have difficulty using netcdf files
8. The data is available for free if you have the correct top level domain name in your computer name. . . our IT department does not want to name the desktop machines
9. NSF has an unwritten policy that one can hold on to data for a very long time if still working on it
10. The weather data is important, but they are extremely huge in the amount for storage and maintain. . . It is so difficult to find the regular measured data a week after they were measured

"it is difficult to share my data because . . ." *(sample items)*

1. No suitable platform exists
2. Lack of guidelines
3. Lack of incentive from current institution
4. Data ownership issues
5. Authorship
6. Existing standards often apply poorly to specialized research data sets
7. The time it would take to thoroughly document the data and develop a clean web interface for it
8. Lack of resources to manage the data
9. Institutional fear of liability
10. The files are too big to upload
11. No one else is using the data at the resolution I am currently using
12. Simulation of fundamental physical problems does not fit well into a specific geographical location
13. To publicly share mass spectrometry data would mean having an interplay system able to standardize many procedures of data acquisition, correction, reduction, sharing
14. I work for the Department of Defense

Selected pre "use cases" on the use of data from more than one discipline

- atmospheric chemistry, oceanography, meteorology, sea ice
- geophysical, meteorological, oceanographical
- oceanography, climate data, seismology, geodesy
- water quality, water quantity, land use, meteorologic data, air quality
- plant phenology, energy uses
- geophysical, geochemical, geochronological, geological, climate
- space physics, atmospheric Science, Solid Earth Science
- GPS, seismology, geology
- surficial geology, sea bottom type and composition, benthos, landforms, bathymetry, satellite altimetry, ocean color, ocean temperature, fishing data, biological observations, human uses
- ecology, hydrology, micrometeorology

Appendix I: Domain Workshops

Respondent Profile	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
Domestic Institutional Affiliation	96.4%	61.9%	95.2%	100%	100%	88%	77.1%
International Institutional Affiliation	3.3%	38.1%	4.8%	0%	0%	12%	22.9%
Female	21.4%	19%	28.6%	0%	40.5%	26.1%	27.9%
Male	78.6%	81%	71.4%	100%	59.5%	73.9%	72.1%
Under 5 years of experience	10.7%	23.8%	4.8%	12.5%	5.4%	2.4%	12.9%
5-10 years of experience	14.3%	19.0%	19.0%	12.5%	37.0%	17.5%	20.5%
11-20 years of experience	17.9%	28.6%	14.3%	29.2%	56.8%	27.0%	28.5%
Over 20 years of experience	57.1%	28.6%	61.9%	45.8%	0%	53.2%	37.9%
Never heard of EarthCube	17.9%	28.6%	0%	12.5%	21.6%	14.3%	54.3%
Aware, but no direct experience	42.9%	47.6%	4.8%	37.5%	32.4%	21.4%	29.4%
Visited website	3.6%	4.8%	43.8%	12.5%	24.3%	12.7%	10.4%
Participated in discussions	25.0%	19.0%	31.3%	16.7%	13.5%	15.1%	3.7%
Actively involved with EarthCube	10.7%	0%	6.3%	16.7%	8.1%	28.6%	1.7%
Leadership role in EarthCube	0%	0%	12.5%	4.2%	0%	7.1%	0.5%

Responses on for you to find, access, and/or integrate multiple datasets, observations, visualization tools, and/or modes (all responses normalized on a scale of zero to one, with one as most positive)	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
How IMPORTANT is it . . . in your field or discipline? (Q23)	.89 (.17)	.70 (.32)	.94 (.10)	.91 (.14)	.89 (.19)	.89 (.18)	.87 (.20)
How EASY is it . . . in your field or discipline? (Q24)	.45 (.25)	.40 (.18)	.42 (.25)	.35 (.23)	.33 (.30)	.41 (.25)	.42 (.24)
How IMPORTANT is it . . . spanning different fields or disciplines ? (Q25)	.57 (.32)	.62 (.31)	.84 (.23)	.74 (.27)	.77 (.31)	.79 (.24)	.73 (.27)
How EASY is it . . . Spanning different fields or disciplines? (Q26)	.37 (.24)	.28 (.20)	.32 (.24)	.27 (.25)	.20 (.24)	.30 (.24)	.32 (.22)
Please use the scale ranging from "Inadequate" to "Adequate" to assess the present suite of publicly accessible datasets, data analysis tools, and modeling software (Q27)	.61 (.22)	.33 (.17)	.40 (.20)	.32 (.17)	.40 (.26)	.42 (.24)	.49 (.26)

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one as most positive)	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
In 5-7 years, I anticipate that EarthCube will result in substantially increased productivity for me and others whose work is similar to mine. (Q31)	.64 (.26)	.71 (.23)	.60 (.28)	.73 (.21)	.67 (.22)	.65 (.22)	.62 (.25)
In 5-7 years, I anticipate that EarthCube will result in substantially expanded research opportunities for me and others whose work is similar to mine. (Q32)	.66 (.23)	.78 (.23)	.63 (.30)	.73 (.21)	.73 (.20)	.69 (.22)	.65 (.25)
In 5-7 years, I anticipate that EarthCube will result in substantially expanded educational tools for me and others whose work is similar to mine. (Q33)	.65 (.23)	.77 (.20)	.66 (.24)	.81 (.16)	.68 (.22)	.68 (.22)	.67 (.23)
In 5-7 years, I anticipate that EarthCube will result in substantially expanded capabilities to integrate multiple sources of data, datasets, observations, visualization, and models. (Q34)	.68 (.24)	.77 (.20)	.64 (.28)	.80 (.18)	.75 (.19)	.73 (.20)	.69 (.24)

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one as most positive)	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in discovering, accessing, sharing, and disseminating geoscience data. (Q35)	.85 (.18)	.79 (.19)	.76 (.26)	.87 (.19)	.88 (.23)	.84 (.23)	.84 (.21)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in geoscience data analysis tools, methods, and/or models. (Q36)	.79 (.26)	.67 (.26)	.75 (.24)	.82 (.21)	.84 (.19)	.76 (.27)	.79 (.25)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in geoscience visualization tools. (Q37)	.79 (.24)	.64 (.22)	.75 (.23)	.84 (.19)	.81 (.20)	.75 (.26)	.78 (.25)

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one as most positive)	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
There are presently substantial unresolved issues around the access and use of geoscience data housed in federal government repositories. (Q43)	**.66 (.29)**	**.73 (.26)**	**.69 (.22)**	**.68 (.24)**	**.67 (.24)**	**.77 (.24)**	**.67 (.24)**
There are presently substantial unresolved issues around the access and use of data held by investigators funded by NSF and other federal agencies. (Q44)	**.76 (.26)**	**.85 (.20)**	**.68 (.29)**	**.74 (.22)**	**.66 (.29)**	**.68 (.26)**	**.61 (.25)**
There are presently substantial unresolved issues around the attribution/authorship of data in the use of data housed or retrieved by data aggregating systems like EarthCube. (Q45)	**.76 (.30)**	**.73 (.23)**	**.61 (.30)**	**.72 (.17)**	**.73 (.23)**	**.63 (.25)**	**.59 (.23)**

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one as most positive)	Atmos-pheric Modeling Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
My employer/organization will most likely value and reward my efforts in the shaping and development of EarthCube. (Q61)	.48 (.25)	.31 (.27)	.50 (.34)	.38 (.29)	.45 (.36)	.49 (.32)	.40 (.30)
My employer/org. will most likely see my participation in the shaping and dev. of EarthCube as an integral part of my job. (Q62)	.53 (.28)	.33 (.34)	.45 (.33)	.35 (.29)	.43 (.34)	.43 (.32)	.34 (.29)
My contributions to the shaping and development of EarthCube will most likely be recognized and highly valued by colleagues in my field/domain. (Q63)	.50 (.24)	.49 (.22)	.54 (.32)	.51 (.30)	.48 (.32)	.52 (.26)	.46 (.28)

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one as most positive)	Atmos-pheric Modelin g Work-shop (n=28)	Sedimen-tology Work-shop (n=21)	Earth-Scope Work-shop (n=21)	Plate Tec-tonics Work-shop (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Centers (n=576)
There is currently a high degree of cooperation and sharing of data, models, and simulations among geoscientists.(Q52)	**.50 (.24)**	**.48 (.26)**	**.51 (.16)**	**.39 (.20)**	**.40 (.23)**	**.40 (.25)**	**.48 (.24)**
There is currently a high degree of cooperation and sharing of software, middleware, and hardware among those developing and supporting cyberinfrastructure for the geosciences. (Q53)	**.47 (.19)**	**.45 (.28)**	**.51 (.18)**	**.43 (.25)**	**.38 (.25)**	**.43 (.25)**	**.46 (.24)**
There is currently sufficient communication and collaboration between geoscientists and those who develop cyberinfrastructure tools and approaches to advance the geosciences. (Q55)	**.34 (.24)**	**.36 (.25)**	**.31 (.22)**	**.25 (.17)**	**.26 (.22)**	**.29 (.22)**	**.34 (.23)**
There is currently sufficient geoscience end-user knowledge and training so they can effectively use the present	**.35 (.28)**	**.31 (.27)**	**.29 (.18)**	**.21 (.14)**	**.24 (.21)**	**.24 (.19)**	**.32 (.23)**