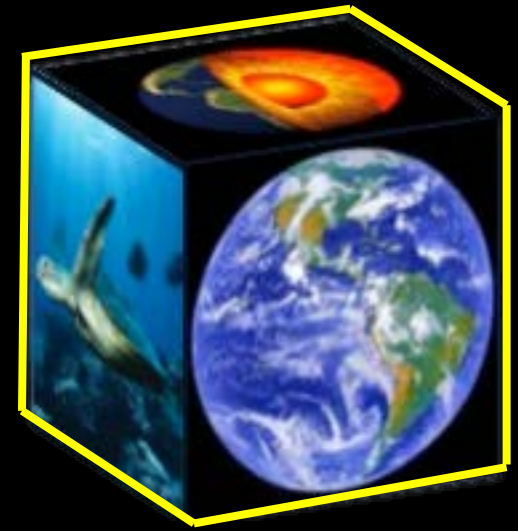


# Stakeholder Alignment for EarthCube:

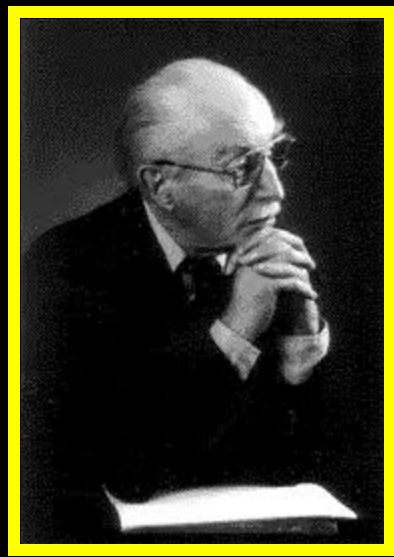
## Presentation for Atmospheric Modeling (Data Assimilation and Ensemble Prediction) Workshop



*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")*

**Joel Cutcher-Gershenfeld,  
University of Illinois, Urbana-Champaign  
November 5, 2012**





## Looking ahead . . .

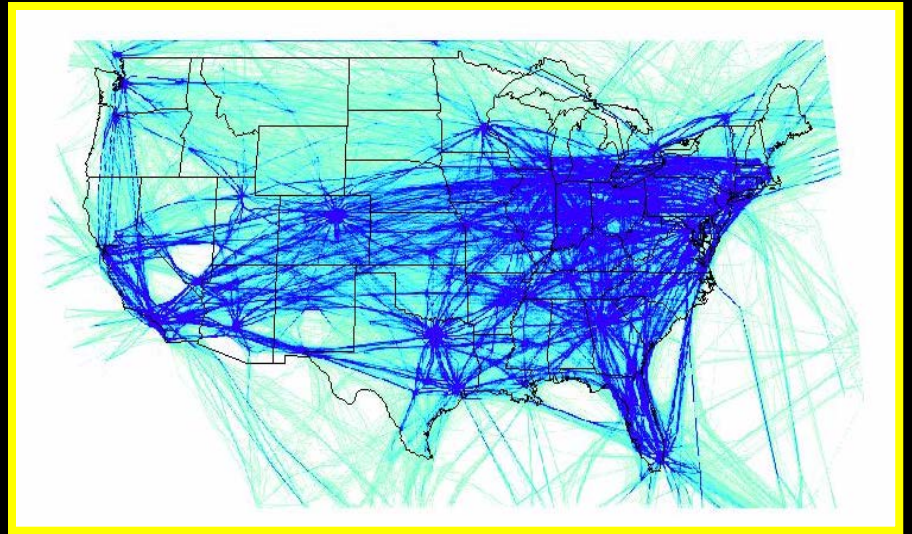
“. . . We are moving towards another type of society than that to which we have become accustomed. This is sometimes referred to as a new service society, the society of the second industrial revolution or the post-industrial society. There is no guarantee of our safe arrival. Not only are the interdependencies 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)

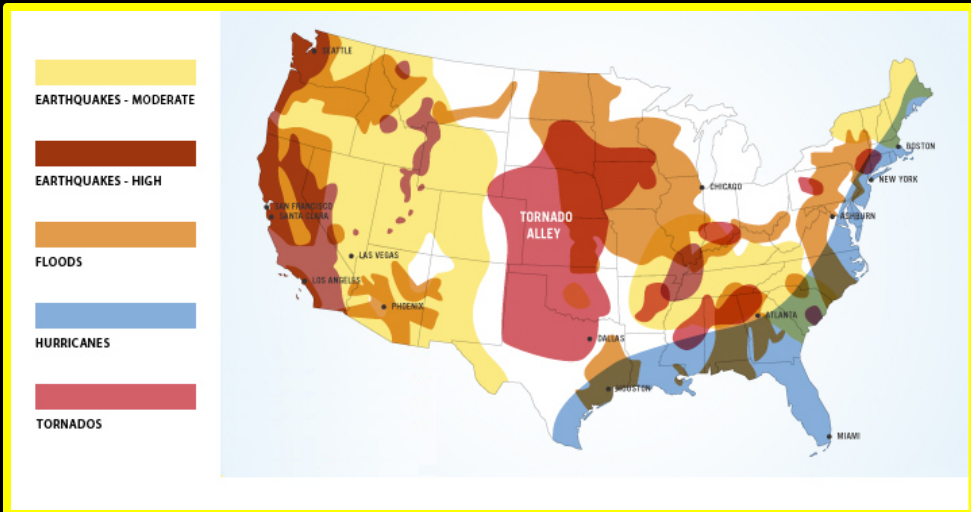
# Institutions ≠ Systems



US Power Grid



US Passenger Air Transportation System



Natural Disasters



US Internet Backbone

# Caution – construction ahead



- Preliminary comparison of survey responses
- Only descriptive stats – additional multivariate analysis needed

<b>Respondent Profile</b>	<b>Atmospheric Model Workshop (n=28)</b>	<b>Sedimentology Workshop (n=21)</b>	<b>Plate Tectonics Workshop (n=24)</b>	<b>Early Career Workshop (n=37)</b>	<b>EarthCube Website (n=126)</b>	<b>Data Centers (n=576)</b>
<b>U.S. Institutional Affiliation</b>	<b>96.4%</b>	<b>61.9%</b>	<b>100%</b>	<b>100%</b>	<b>88%</b>	<b>77.1%</b>
<b>International Institutional Affiliation</b>	<b>3.3%</b>	<b>38.1%</b>	<b>0%</b>	<b>0%</b>	<b>12%</b>	<b>22.9%</b>
<b>Female</b>	<b>21.4%</b>	<b>19%</b>	<b>0%</b>	<b>40.5%</b>	<b>26.1%</b>	<b>27.9%</b>
<b>Male</b>	<b>78.6%</b>	<b>81%</b>	<b>100%</b>	<b>59.5%</b>	<b>73.9%</b>	<b>72.1%</b>
<b>Under 5 years of experience</b>	<b>10.7%</b>	<b>23.8%</b>	<b>12.5%</b>	<b>5.4%</b>	<b>2.4%</b>	<b>12.9%</b>
<b>5-10 years of experience</b>	<b>14.3%</b>	<b>19.0%</b>	<b>12.5%</b>	<b>37.0%</b>	<b>17.5%</b>	<b>20.5%</b>
<b>11-20 years of experience</b>	<b>17.9%</b>	<b>28.6%</b>	<b>29.2%</b>	<b>56.8%</b>	<b>27.0%</b>	<b>28.5%</b>
<b>Over 20 years of experience</b>	<b>57.1%</b>	<b>28.6%</b>	<b>45.8%</b>	<b>0%</b>	<b>53.2%</b>	<b>37.9%</b>
<b>Never heard of EarthCube</b>	<b>17.9%</b>	<b>28.6%</b>	<b>12.5%</b>	<b>21.6%</b>	<b>14.3%</b>	<b>54.3%</b>
<b>Aware, but no direct experience</b>	<b>42.9%</b>	<b>47.6%</b>	<b>37.5%</b>	<b>32.4%</b>	<b>21.4%</b>	<b>29.4%</b>
<b>Visited website</b>	<b>3.6%</b>	<b>4.8%</b>	<b>12.5%</b>	<b>24.3%</b>	<b>12.7%</b>	<b>10.4%</b>
<b>Participated in discussions</b>	<b>25.0%</b>	<b>19.0%</b>	<b>16.7%</b>	<b>13.5%</b>	<b>15.1%</b>	<b>3.7%</b>
<b>Actively involved with EarthCube</b>	<b>10.7%</b>	<b>0%</b>	<b>16.7%</b>	<b>8.1%</b>	<b>28.6%</b>	<b>1.7%</b>

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one being most positive)  Mean (s.d.)	Atmos. Model Workshop (n=28)	Sedimen-tology Workshop (n=21)	Plate Tec-tonics Workshop (n=24)	Early Career Workshop (n=37)	Earth-Cube Web-site n=126	Data Cen-ters (n=576)
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?	<b>.89</b> (.17)	<b>.70</b> (.32)	<b>.91</b> (.14)	<b>.89</b> (.19)	<b>.89</b> (.18)	<b>.87</b> (.20)
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?	<b>.45</b> (.25)	<b>.40</b> (.18)	<b>.35</b> (.23)	<b>.33</b> (.30)	<b>.41</b> (.25)	<b>.42</b> (.24)
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 ?	<b>.57</b> (.32)	<b>.62</b> (.31)	<b>.74</b> (.27)	<b>.77</b> (.31)	<b>.79</b> (.24)	<b>.73</b> (.27)
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?	<b>.37</b> (.24)	<b>.28</b> (.20)	<b>.27</b> (.25)	<b>.20</b> (.24)	<b>.30</b> (.24)	<b>.32</b> (.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 – to what degree is it adequate for your research and education needs?	<b>.61</b> (.22)	<b>.33</b> (.17)	<b>.32</b> (.17)	<b>.40</b> (.26)	<b>.42</b> (.24)	<b>.49</b> (.26)

<b>Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one being most positive)</b>  Mean (s.d.)	<b>Atmos. Model Workshop (n=28)</b>	<b>Sedimen-tology Workshop (n=21)</b>	<b>Plate Tec-tonics Work-shop (n=24)</b>	<b>Early Career Work-shop (n=37)</b>	<b>Earth-Cube Web-site (n=126)</b>	<b>Data Cen-ters (n=576)</b>
<b>In 5-7 years, I anticipate that EarthCube will result in substantially increased productivity for me and others whose work is similar to mine.</b>	<b>.64 (.26)</b>	<b>.71 (.23)</b>	<b>.73 (.21)</b>	<b>.67 (.22)</b>	<b>.65 (.22)</b>	<b>.62 (.25)</b>
<b>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.</b>	<b>.66 (.23)</b>	<b>.78 (.23)</b>	<b>.73 (.21)</b>	<b>.73 (.20)</b>	<b>.69 (.22)</b>	<b>.65 (.25)</b>
<b>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.</b>	<b>.65 (.23)</b>	<b>.77 (.20)</b>	<b>.81 (.16)</b>	<b>.68 (.22)</b>	<b>.68 (.22)</b>	<b>.67 (.23)</b>
<b>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.</b>	<b>.68 (.24)</b>	<b>.77 (.20)</b>	<b>.80 (.18)</b>	<b>.75 (.19)</b>	<b>.73 (.20)</b>	<b>.69 (.24)</b>

Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one being most positive)	Atmos. Model Workshop (n=28)	Sedimen-tology Workshop (n=21)	Plate Tec-tonics (n=24)	Early Career Work-shop (n=37)	Earth-Cube Web-site (n=126)	Data Cen-ters (n=576)
Mean (s.d.)						
My employer/org. will most likely value and reward my efforts in the shaping and dev. of EarthCube.	<b>.48</b> <b>(.25)</b>	<b>.31</b> <b>(.27)</b>	<b>.38</b> <b>(.29)</b>	<b>.45</b> <b>(.36)</b>	<b>.49</b> <b>(.32)</b>	<b>.40</b> <b>(.30)</b>
My employer/org. will most likely see my participation in the shaping and dev. of EarthCube as an integral part of my job.	<b>.52</b> <b>(.28)</b>	<b>.33</b> <b>(.34)</b>	<b>.35</b> <b>(.29)</b>	<b>.43</b> <b>(.34)</b>	<b>.43</b> <b>(.32)</b>	<b>.34</b> <b>(.29)</b>
My contributions to the shaping and dev. of EarthCube will most likely be recognized and highly valued by colleagues in my field/domain.	<b>.50</b> <b>(.24)</b>	<b>.49</b> <b>(.22)</b>	<b>.51</b> <b>(.30)</b>	<b>.48</b> <b>(.32)</b>	<b>.52</b> <b>(.26)</b>	<b>.46</b> <b>(.28)</b>
There is currently a high degree of cooperation and sharing of data, models, and simulations among geoscientists.	<b>.50</b> <b>(.24)</b>	<b>.48</b> <b>(.26)</b>	<b>.39</b> <b>(.20)</b>	<b>.40</b> <b>(.23)</b>	<b>.40</b> <b>(.25)</b>	<b>.48</b> <b>(.24)</b>
There is currently sufficient communication and collaboration between geoscientists and those who develop cyberinfrastructure tools and approaches to advance the geosciences.	<b>.34</b> <b>(.24)</b>	<b>.45</b> <b>(.28)</b>	<b>.25</b> <b>(.17)</b>	<b>.26</b> <b>(.22)</b>	<b>.29</b> <b>(.22)</b>	<b>.34</b> <b>(.23)</b>
There is currently sufficient geoscience end-user knowledge and training so they can effectively use the present suite of cyber-infrastructure tools and train their students/colleagues in its use.	<b>.35</b> <b>(.28)</b>	<b>.36</b> <b>(.25)</b>	<b>.21</b> <b>(.14)</b>	<b>.24</b> <b>(.21)</b>	<b>.24</b> <b>(.19)</b>	<b>.32</b> <b>(.23)</b>



# 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).*

<b>Responses on Data Access, Use, and EarthCube (all responses normalized on a scale of zero to one, with one being most positive)</b>  Mean (s.d.)	<b>Atmos. Model Workshop (n=28)</b>	<b>Sedimen-tology Workshop (n=21)</b>	<b>Plate Tec-tonics (n=24)</b>	<b>Early Career Work-shop (n=37)</b>	<b>Earth-Cube Web-site (n=126)</b>	<b>Data Cen-ters (n=576)</b>
There are presently substantial unresolved issues around the access and use of geoscience data housed in federal government repositories.	<b>.66</b> (.29)	<b>.73</b> (.26)	<b>.68</b> (.24)	<b>.67</b> (.24)	<b>.77</b> (.24)	<b>.67</b> (.24)
There are presently substantial unresolved issues around the access and use of data held by invest. funded by NSF and other federal agencies.	<b>.76</b> (.26)	<b>.85</b> (.20)	<b>.74</b> (.22)	<b>.66</b> (.29)	<b>.68</b> (.26)	<b>.61</b> (.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.	<b>.76</b> (.30)	<b>.73</b> (.23)	<b>.72</b> (.17)	<b>.73</b> (.23)	<b>.63</b> (.25)	<b>.59</b> (.23)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in discovering, accessing, sharing, and disseminating geoscience data.	<b>.85</b> (.18)	<b>.79</b> (.19)	<b>.87</b> (.19)	<b>.88</b> (.23)	<b>.84</b> (.23)	<b>.84</b> (.21)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in geoscience data analysis tools, methods, models.	<b>.79</b> (.26)	<b>.67</b> (.26)	<b>.82</b> (.21)	<b>.84</b> (.19)	<b>.76</b> (.27)	<b>.79</b> (.25)
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in geoscience visualization tools.	<b>.79</b> (.24)	<b>.64</b> (.22)	<b>.84</b> (.19)	<b>.81</b> (.20)	<b>.75</b> (.26)	<b>.78</b> (.25)

# Selected comments why it is difficult to access needed data from Atmospheric Modeling Workshop:

- Different formats, capacity and networking issues, discoverability, cumbersome interfaces
- It can be difficult to locate less commonly used datasets.
- There is no central repository that has everything I need. I am often unaware of all possible places to find what I need.
- . . . [T]he challenge is the human tendency (to borrow an overused phrase): "if I have a hammer, everything looks like a nail." People use the data they are comfortable with and have experience with generally even if a better option may be available.
- It is difficult to find, it is scattered around the internet, it is in non-standardized formats, it is not well documented or quality controlled.
- 1. Reliability is an issue. 2. Bandwidth/speed is limited. 3. Difficult to understand metadata, filename meaning, organization of data. 4. Some technologies for transfer are difficult to understand or I am using the wrong tools to access the data.
- Restrictions on government data sites.
- I often need historical runs of real-time analysis systems, which aren't as readily available as more recent time periods
- Formatting!

# Selected elements of a success vision for EarthCube from Atmospheric Modeling Workshop:

- A common format for data that is easily read and used within many software packages. For example that temperature is called the same thing in all different datasets.
- The creation of a single repository from which all different datasets can be found, plus software such as data converters that make it very easy to convert all types into whatever type is needed, especially for use in visualization systems.
- Recognize successful efforts of this type already underway and support those rather than attempting to reinvent the wheel. Improved online tools for training people how to use data and tools. The presumption that by enabling/creating interdisciplinary catalogs of data and tools that people are going to creatively use them is a fallacy. People don't conduct research by looking at catalogs. It requires communication via conferences, seminars, research papers where data and tools are used. Then, the light bulbs go off. . .
- A unified distribution center for data, meeting compatible standards with typically used models, data assimilation systems, and visualization tools.
- Collaborative network for advancing ensemble modeling

# Selected elements of a success vision for EarthCube from Atmospheric Modeling Workshop (cont.):

- 1. Infuse more intelligence and ease-of-use in finding data that is needed. 2. Encourage contributions from a broader range of scientists, not just a few powerful, well-connected groups. 3. Evaluate participation based on the merits of an idea, not on political connections. 4. Keep disinterested, arrogant, clueless people off the review panels. Do not shoot down a proposal/idea because only 1 reviewer is too lazy/confused to fairly evaluate it.
- Right now I am a bit cynical. I think the plan is too broad, too grand, and too likely to fall victim to budget cuts before its goals can ever be effectively implemented.
- If we can bring together the data providers, data users and those with unique ways to interrogate/mine/visualize the data, and provide an information clearinghouse for how to get at data, that would be a major step in itself. It is critically important this be a long-term vision, not something just for demonstrations and show.
- This can be highly successful if we find common interests for data sharing between public and private sector.
- Centralized data access and ease of various data usage
- Cannot be a top-down approach



**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)*