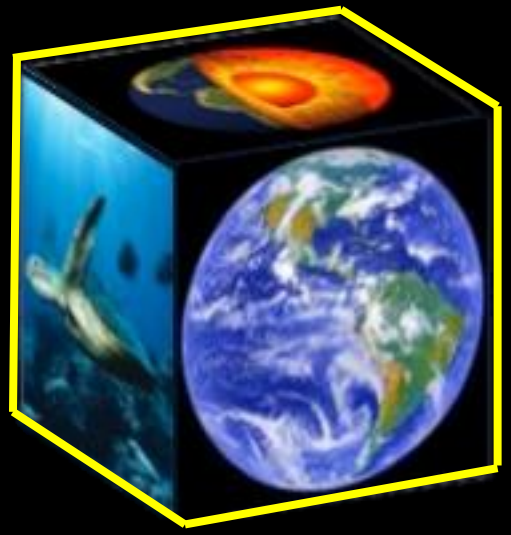
# Stakeholder Alignment for EarthCube:

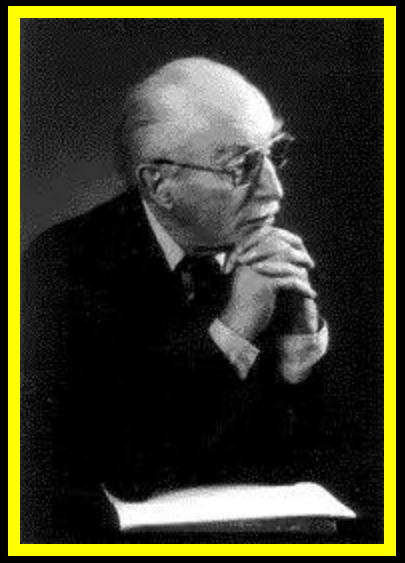
# Presentation for Community Modeling Workshop




***Joel Cutcher-Gershenfeld, UIUC***
*Betty Barrett, MIT*
*Burcu Bolukbasi, UIUC*
*Courtney Flint, UIUC*
*John King, University of Michigan*
*Chris Lawson, Aerospace Corporation*
*Barbara Lawrence, UCLA*
*Mark Nolan, UIUC*
*Barbara Mittleman, NIH*
*John Unsworth, Brandeis University*
*Ilya 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")***



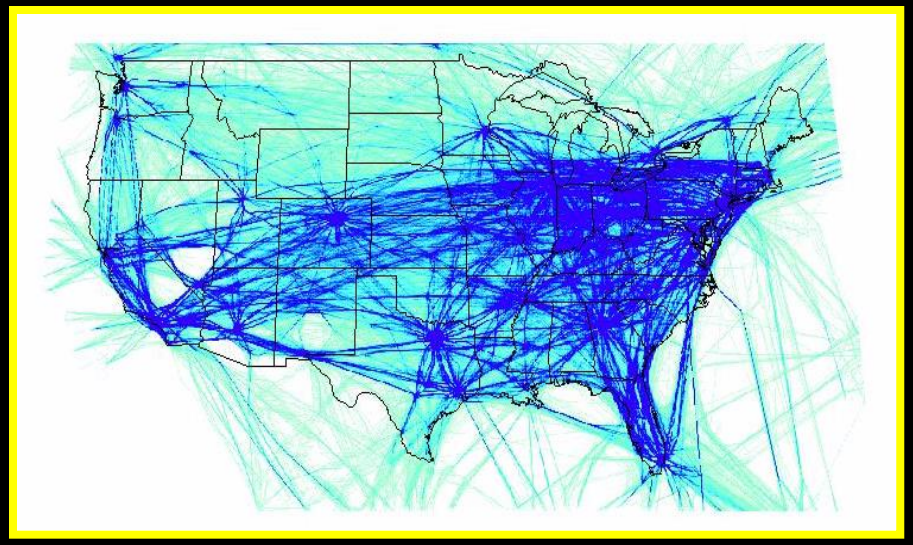
# Looking ahead . . .

". . . 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."

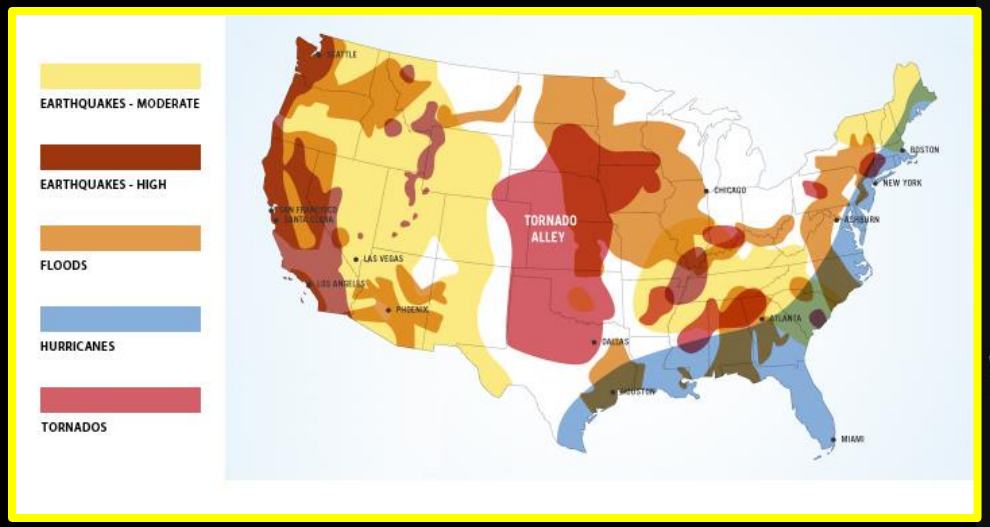
– Source: Eric L. Trist, from paper on "Social Aspects of Science Policy" (March, 1969) cited in *Towards a Social Ecology: Contextual Appreciation of the Future in the Present* by Fred E. Emery and Eric L. Trist (London: Plenum Press, 1973)

# Institutions ≠ Systems

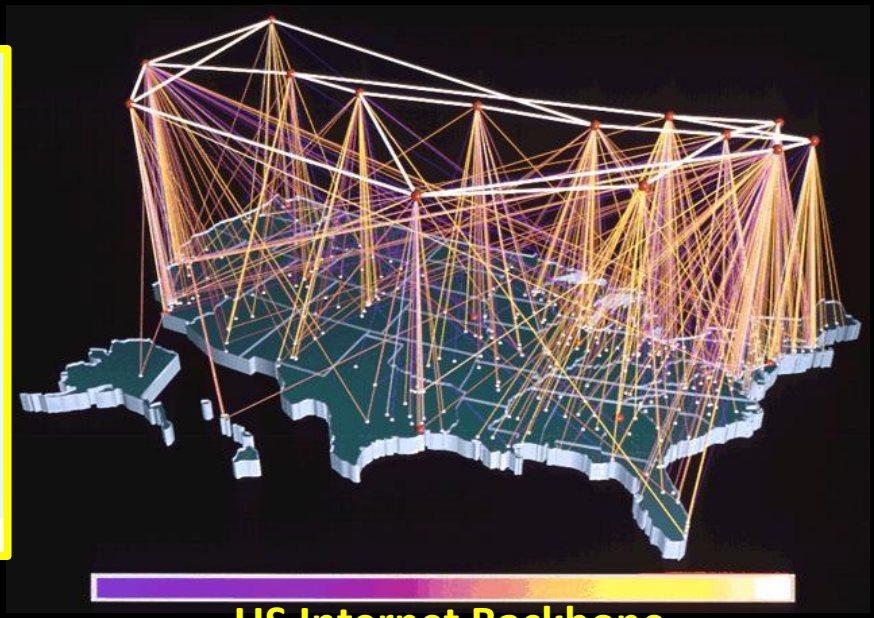
US Power Grid



US Passenger Air Transportation System




Natural Disasters



US Internet Backbone

http://www.xprt.net/~rolfsky/internetSite/internet.html

*Sources: Carolos A. Osario, ESD Doctoral Seminar, 2004, and Joel Cutcher-Gershenfeld*

# Caution – construction ahead

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

| Respondent Profile (v28, v22, v26) | Com-munity Modeling (n=42) | Petrology (n=59) | Educa-tion (n=33) | Early Career (n=37) | Earth-Cube Website (n=127) | Data Centers (n=578) |
|---|---|---|---|---|---|---|
| U.S. Institutional Affiliation | 90.5% | 98.3% | 100% | 100% | 88.1% | 77% |
| International Institutional Affiliation | 9.5% | 1.7% | 0% | 0% | 11.9% | 23% |
| | | | | | | |
| Female | 35.7% | 44.8% | 36.4% | 43.2% | 25.8% | 27.6% |
| Male | 64.3% | 55.2% | 63.6% | 56.8% | 74.2% | 72.4% |
| | | | | | | |
| Under 5 years of experience | 4.8% | 10.2% | 0.0% | 5.4% | 2.4% | 12.8% |
| 5-10 years of experience | 21.4% | 30.5% | 3% | 37.8% | 17.3% | 20.6% |
| 11-20 years of experience | 31% | 25.4% | 42.4% | 56.8% | 26.8% | 28.4% |
| Over 20 years of experience | 42.9 | 33.9% | 54.5% | 0% | 53.5% | 38.1% |

| Respondent Profile (v33) | Com-munity Modeling (n=42) | Petrology (n=59) | Educa-tion (n=33) | Early Career (n=37) | Earth-Cube Website (n=127) | Data Centers (n=578) |
|---|---|---|---|---|---|---|
| Never heard of EarthCube | 7.1% | 8.6% | 9.7% | 21.6% | 15.7% | 53.9% |
| Aware, but no direct experience | 28.6% | 25.9% | 32.3% | 32.4% | 21.3% | 29.3% |
| Visited website | 21.4% | 24.1% | 19.4% | 24.3% | 13.4% | 10.4% |
| Participated in discussions | 23.8% | 19.0% | 25.8% | 13.5% | 15.7% | 3.5% |
| Actively involved with EarthCube | 16.7% | 19.0% | 6.5% | 8.1% | 26.0% | 2.4% |
| Leadership role in EarthCube | 2.4% | 3.4% | 6.5% | 0% | 7.9% | 0.5% |

| 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.) | Com-munity Modeling (n=42) | Petrol-ogy (n=59) | Educa-tion (n=33) | Early Career (n=37) | Earth-Cube Web-site n=127 | Data Cen-ters (n=578 ) |
|---|---|---|---|---|---|---|
| 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? (v58) | .85 (.19) | .89 (.17) | .84 (.18) | .89 (.19) | .89 (.18) | .87 (.20) |
| How EASY is . . . in your field or discipline? (v59) | .44 (.24) | .44 (.24) | .40 (.22) | .33 (.30) | .41 (.25) | .42 (.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 ? (v60) | .73 (.24) | .74 (.24) | .81 (.19) | .77 (.31) | .79 (.24) | .73 (.27) |
| How EASY is it . . . spanning different fields or disciplines? (v61) | .31 (.25) | .29 (.20) | .30 (.21) | .20 (.23) | .29 (.23) | .32 (.22) |

| 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.) | Com-munity Modeling (n=42) | Petrol-ogy (n=59) | Educa-tion (n=33) | Early Career (n=37) | Earth-Cube Web-site (n=127) | Data Cen-ters (n=578) |
|---|---|---|---|---|---|---|
| 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? (v62) | .52 (.22) | .40 (.23) | .42 (.24) | .40 (.26) | .42 (.24) | .48 (.26) |
| | | | | | | |
| 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. (v98) | .61 (.25) | .78 (.17) | .66 (.22) | .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 being most positive) Mean (s.d.) | Com-munity Modeling (n=42) | Petrol-ogy (n=59) | Educa-tion (n=33) | Early Career (n=37) | Earth-Cube Web-site (n=127) | Data Cen-ters (n=578) |
|---|---|---|---|---|---|---|
| How should the EarthCube initiative balance its focus on the development of tools and approaches in support of research, on one hand (high number), and education, on the other (low number). (v94) | .63 (.16) | .65 (.19) | .50 (.20) | .61 (.20) | .64 (.20) | .65 (.20) |
| | | | | | | |
| My employer/org. will most likely value and reward my efforts in the shaping and dev. of EarthCube. (v120) | .46 (.28) | .59 (.26) | .53 (.33) | .48 (.35) | .49 (.32) | .40 (.30) |
| My contributions to the shaping and dev. of EarthCube will most likely be recognized and highly valued by colleagues in my field/domain. (v122) | .56 (.27) | .60 (.28) | .54 (.31) | .50 (.31) | .52 (.26) | .46 (.28) |
| | | | | | | |

| 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.) | Com-munity Modeling (n=42) | Petrol-ogy (n=59) | Educa-tion (n=33) | Early Career (n=37) | Earth-Cube Web-site (n=127) | Data Cen-ters (n=578) |
|---|---|---|---|---|---|---|
| There is currently a high degree of cooperation and sharing of data, models, and simulations among geoscientists. (v69) | .56 (.20) | .48 (.24) | .50 (.22) | .40 (.23) | .40 (.25) | .47 (.24) |
| There is currently sufficient communication and collaboration between geoscientists and those who develop cyberinfrastructure tools to advance the geosciences. (v72) | .36 (.23) | .38 (.25) | .31 (.19) | .28 (.21) | .29 (.22) | .34 (.23) |
| 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. (v73) | .29 (.23) | .21 (.18) | .22 (.15) | .25 (.20) | .24 (.19) | .32 (.23) |

# 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).*

# Selected Community Modeling Responses: Difficult to share data

- It's not difficult to share my data
- It takes time to document and provide metadata, to explain to others how to use the data or product, etc. Lack of time/resources for doing his job.
- It is difficult to describe the simulation software, and inputs, and computing environment that produced the data.
- The effort required to put the data in a format understandable to others is great, and there are many other demands on my time that are more likely to advance my career.
- Model data is hard to share because it tends to consist of large data files.
- Difficulty in developing a standard format/platform
- There is no global easily accessible repository for thermochronologic data.
- Some scientists abuse the shared data, models, and tools.
- There is a limitation
- Most research is funded by industry contracts
- My field does not recognize data contributions as much as it should
- Restrictions on use

# Selected Community Modeling Responses: Difficult to Access data

- The data I need is easily accessible.
- The data (e.g. that is presented in a journal article) is not publicly available and the author is non-responsive to requests to share the data.
- Many data are either proprietary or not easily discoverable.
- Metadata are insufficient, and measurement/instrumentation limitations are not explicitly provided.
- The data are distributed across many systems, from paper to Excel to netCDF files all located at many institutions.
- Where to find it, IF I am not directly involved in the project.
- It's not organized in the way that best suits my science needs.
- There is no proper channel and better public domain, but there is lot of data unclaimed and possessed by individual scientists.
- Other disciplines use 'different language' in describing and georeferencing their data.
- Difficulty in getting permissions

# Selected Community Modeling Responses: Difficult to Access data (cont.)

- **I work at large scale, and many data sets from foreign countries are restricted. Likewise, geologists don't normally archive data, and some aren't even 'digitally minded', so it's hard to find geochron or structural or geochem data bases. Finally, we all know those who don't archive digital geophysical data in national archives, so they may retain territorial rights. That needs to stop - NSF can enforce the rules they set.**
- **I don't always know where to go; there are many historical datasets held by individual researchers who have never bothered to archive these in a way that they could be accessed (even though they say in principal that others could use these data). Many complex and sophisticated datasets require significant experience with regard to understanding data corrections and flags.**

# Community Modeling: Where do you get data?

- NEES, IRIS, COSMOS VDC, USGS, California Geological Survey
- IRIS, UNAVCO, international partners, industry, IODP, LDEO marine geophysics, Sandwell global topo-bathymetry
- NASA, USGS, NOAA, CSDMS (for models)
- NCEP, NOAA, NGDC, NHC
- CReSIS, BEDMAP, ICESAT, GRACE
- NOAA, State Geological Surveys, Macrostrat
- SOPAC, JPL
- NCAR, and personal contacts
- IRIS, UNAVCO, international partners, industry, IODP, LDEO marine geophysics, Sandwell global topo-bathymetry
- Obtain elevation data from USGS, USACE, NOAA Coastal Services Center, and State and local government agencies
- Literature, old theses and dissertations, USGS, state surveys, foreign national surveys

# Community Modeling: Where do you get data? (cont.)

- **Oil & gas industry**
- **Mostly others data. Texas department of transportation, Tiger data, us dept of agriculture, NCTCOG, cities, etc**
- **Yes. The published literature and sites such as the HOT and BATS site and BCO-DMO.**
- **Yes. From USGS National Water information System. From USGS National Hydrography Dataset. From NASA North American Land Data Assimilation System.**
- **Marine Geoscience Data System, peer reviewed literature, and personal communication with collaborators.**
- **PBO website for geodetic data**
- **Many sources, planetary data system node, national geophysical data center, colleages websites, petdb, rockdb, neic, noaa, ...**
- **Satellite remote sensing data (altimetry, SST, Coastal HF radar: various web sites); geomagnetic data (observatories--NGDC, magnetotelluric data--IRIS DMC)**

| 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.) | Com-munity Modeling (n=41) | Petrol-ogy (n=51) | Educa-tion (n=29) | Early Career (n=37) | Earth-Cube Web-site (n=126 | Data Cen-ters (n=576 |
|---|---|---|---|---|---|---|
| There are presently substantial unresolved issues around the access and use of geoscience data housed in federal government repositories. (v66) | .59 (.28) | .71 (.30) | .65 (.26) | .71 (.25) | .69 (.27) | .65 (.25) |
| There are presently substantial unresolved issues around the access and use of data held by invest. funded by NSF and other federal agencies. (v67) | .72 (.23) | .76 (.22) | .72 (.25) | .74 (.22) | .78 (.24) | .69 (.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. (v68) | .68 (.27) | .74 (.26) | .83 (.16) | .67 (.24) | .77 (.23) | .67 (.24) |

| 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.) | Com-unity Modeling (n=41) | Petrol-ogy (n=51) | Educa-tion (n=29) | Early Career (n=37) | Earth-Cube Web-site (n=126 | Data Cen-ters (n=576 |
|---|---|---|---|---|---|---|
| As an integrated data and knowledge management system for the geosciences, EarthCube will transform the way geoscience RESEARCH is conducted. (v108) | .58 (.23) | .66 (.25) | .62 (.19) | .66 (.29) | .68 (.26) | .61 (.25) |
| As an integrated data and knowledge management system for the geosciences, EarthCube will transform the way geoscience EDUCATION is conducted. (v109) | .55 (.23) | .64 (.22) | .68 (.20) | .73 (.23) | .63 (.25) | .59 (.23) |
| | | | | | | |
| The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in discovering, accessing, sharing, and disseminating geoscience data. (v99) | .71 (.24) | .77 (.18) | .78 (.18) | .88 (.23) | .84 (.24) | .84 (.21) |

# 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*

# Selected Elements of Success for Community Modeling Workshop Participants

- I don't know, but I wish you well!
- I am still trying to understand what EarthCube is and is trying to become, so I cannot really answer this question.
- An environment where one can browse data for ideas in a similar way we browse online media such as Youtube.
- To be a first-stop location for obtaining well documented data.
- Strong leadership by NSF that ensures equal participation, attribution, minimal abuse, and utility for all.
- More and better data its visualization
- A full integrated data-model portal that will interface with existing data bases.
- Common platform and good initiative..
- Data that are not in software-specific formats and support for commercial software.
- The ability to integrate 'long tail' data at multiple scales of space and time with 'big data'. The ability to utilize 'long tail' data of others in modeling and statistical/comparative studies.

# Selected Elements of Success for Community Modeling Workshop Participants (cont.)

- **If EarthCube made an index that pointed to different data, that would be one step toward success.**
- **All fields of geosciences running on a Google-Earth like digital Earth, and accessed by scientists for research and by anyone else for information.**
- **A seamless, logical map-based infrastructure (like a mega-google earth) with data available at a few clicks.**
- **Being at the hub of linking the data sets, high standards of staff because data quality is important**
- **A new way to do science, where publishing in Science and Nature is not the ONLY measure of success**
- **Developing a network with clear-cut goals so that scientists from different disciplines can access the facility to share the data as well as clarify their research questions with their peers**
- **High involvements of scientists, easy to use products**

# Selected Elements of Success for Community Modeling Workshop Participants (cont.)

- **Networking, Collaboration, and Standardization**
- **Providing such a permanent data and computer model repository, with user support (workshops at AGU, GSA, etc)**
- **Most data are freely available (or at least timely, e.g., within one year of collection) and easily discoverable (i.e., geospatial and well described/good metadata), with standardized ways of properly attributing the data collector, data processor, and data distributor.**
- **If EarthCube will adopt current industry standards for data exchange and archiving, and not try to come up with its own, then I see some potential for success. If EarthCube funding can be spread around to a much more broad group of investigators and institutions and not concentrate funding in already well-funded institutions, and NSF is willing to take more risks, then I see potential for an effect on the community.**

**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)***