STAKEHOLDER ALIGNMENT ENABLING A NATIONAL DATA SERVICE

Joel Cutcher-Gershenfeld
University of Illinois and WayMark Systems

NSF Research Support is deeply appreciated:

NSF-VOSS EAGER 0956472, "Stakeholder Alignment in Socio-Technical Systems,"
NSF OCI RAPID 1229928, "Stakeholder Alignment for EarthCube,"
NSF GEO-SciSIP-STS-OCI-INSPIRE 1249607, "Enabling Transformation in the Social Sciences, Geosciences, and Cyberinfrastructure,"
NSF I-CORPS 1313562 "Stakeholder Alignment for Public-Private Partnerships" leading to the establishment of WayMark Systems

Research Team

Nick Berente, University of Georgia
Burcu Bolukbasi, UIUC
Dorothy Carter, Georgia Tech University
Noshir Contractor, Northwestern University
Joel Cutcher-Gershenfeld, UIUC
Leslie DeChurch, Georgia Tech University
Courtney Flint, Utah State University
Gabriel Gershenfeld, Cleveland Indians
Michael Haberman, UIUC
John L. King, University of Michigan
Eric Knight, University of Sydney
Barbara Lawrence, UCLA
Spenser Lewis, General Dynamics
Pablo Lopez, UIUC
Ethan Masella, Brandeis University
Matt Mayernik, NCAR/UCAR
Charles Mcelroy, Case Western Reserve University
Barbara Mittleman, Nodality, Inc.
Mark Nolan, UIUC
Dechying Ruengvisesh. UIUC
Namchul Shin, Pace University
Cheryl Thompson, UIUC
Susan Winter, University of Maryland
Ilya Zaslavsky, UCSD

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)

What do institutions do?

Create Value — **. . . Expanding the "pie" *and* Enabling systems transformation**

Mitigate Harm — **. . . Anticipating and mitigating externalities *and* Preventing catastrophic systems failures**

A changing social contract with science

Office of Science and Technology Policy

The White House • Office of the Press Secretary • May 09, 2013

Executive Order -- Making Open and Machine Readable the New Default for Government Information

Section 1. General Principles. . . To promote continued **job growth, Government efficiency, and the social good** that can be gained from opening Government data to the public, the default state of new and modernized Government information resources shall be open and machine readable. Government information shall be managed as an asset throughout its life cycle to promote interoperability and openness, and, wherever possible and legally permissible, to ensure that data are released to the public in ways that **make the data easy to find, accessible, and usable**. In making this the new default state, executive departments and agencies (agencies) shall ensure that they safeguard individual privacy, confidentiality, and national security. . .

M-13-13 • **Memorandum for Heads of Executive Departments and Agencies**

. . .Specifically, this Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities. This includes using machine- readable and open formats, data standards, and common core and extensible metadata for **all new information creation and collection efforts**. . .

Accelerating rates of technological change . . .

The Babysitter of the Future

NationalDataService.org

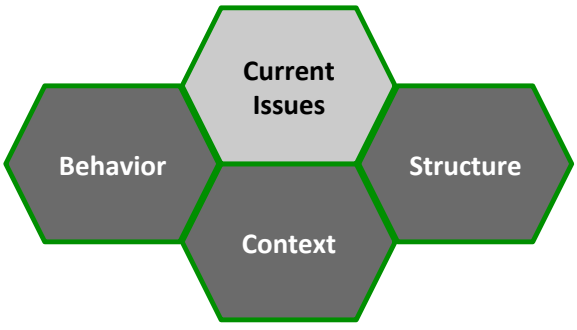
The issues of how best to govern natural resources used by many individuals in common are no more settled in academia than in the world of politics. Some scholarly articles about the "tragedy of the commons" recommend that "the state" control most natural resources . . . Others recommend . . . privatization. . . 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

Defining stakeholder alignment . . .

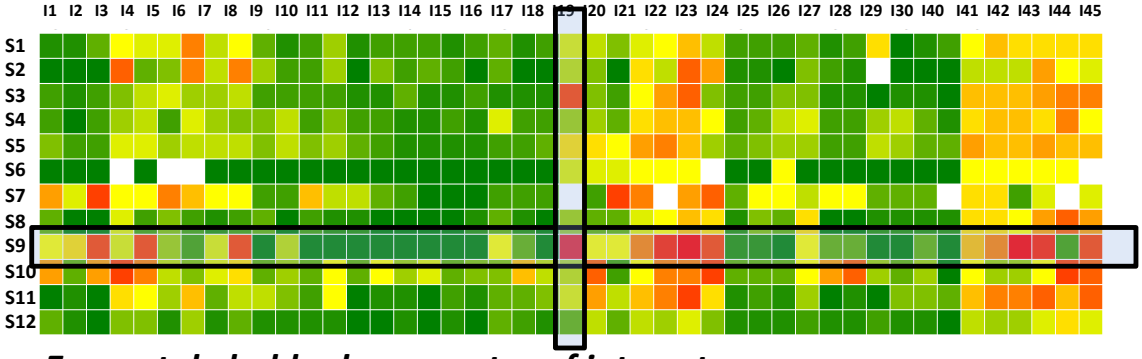
"The extent to which interdependent stakeholders orient and connect with one another to advance their separate and shared interests."

A simplified conceptual framework . . .



Conceptualize systems as a matrix of stakeholders and interests

Every Interest has a vector of stakeholders

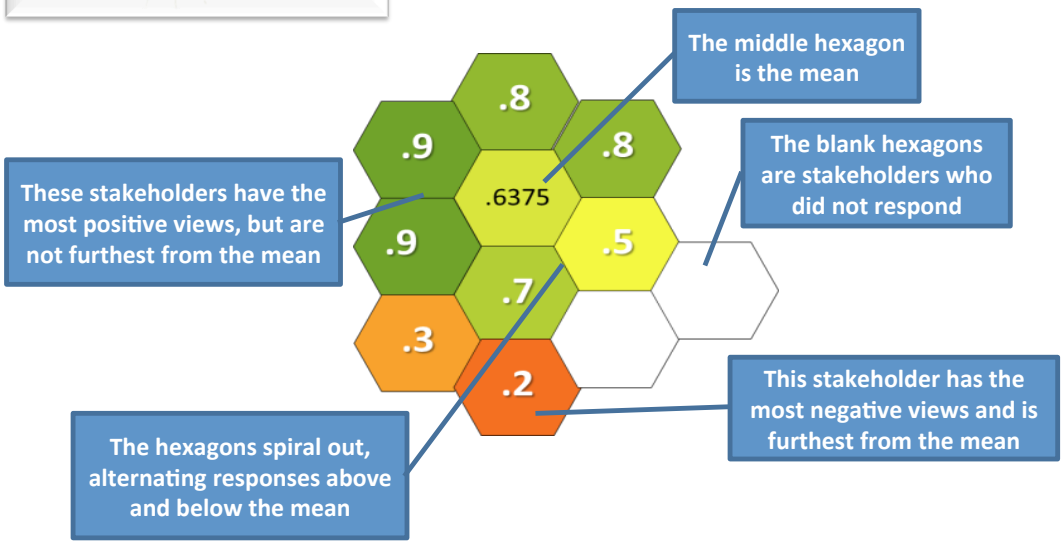


Every stakeholder has a vector of interests

Growing a z-flower™

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

Reading a z-flower™



NDS Stakeholder Alignment Map

Aim: Data-informed dialogue and action

Consider: What additional questions/data are important for the next phase of stakeholder data collection?

Caution: *These are preliminary data, with additional perspectives needed and additional analysis to be conducted.*

Profile of NDS stakeholder alignment respondents (n=42)

Information/Computer/Cyber Scientist/Expert	41.9%
Research Scientist/Professor/Graduate Student	25.6%
Government Elected Official/Government Agency Leader/Staff	2.3%
Publisher/Professional Association Leader/Staff	11.6%
Other (please specify)	18.6%

Under 5 years	--
5-10 years	18.6%
11-20 years	32.6%
21-30 years	30.2%
Over 30 years	18.6%

Male	74.4%
Female	25.6%
US based institution	93.0%
Attending June workshop	65.1%

NDS respondent profile (cont.)

Data-producing projects (e.g. LSST, LIGO, IPlant)	39.5%
Data, model, or software sharing initiatives (national or international) (e.g. EarthCube, NEON, Virtual Observatory, DataONE, EUDAT, ANDS)	72.1%
National or multi-mission data facilities (e.g. NCAR, LLNL, NRAO, NASA IPAC, Argonne, Fermilab)	23.3%
University and/or library-based data efforts related to archiving, preservation, or discovery (e.g. University of Illinois RDS, Cornell University DataStar, etc.)	46.5%
Academic publisher, either as a volunteer (e.g. as editorial board member) or an employee (e.g. APS, Science, Elsevier, IEEE)	11.6%
Computing and/or data centers, including campus computing or data projects (e.g. NCSA, TACC, SDSC, PSC, RENCI, CUCAC)	41.9%
Tool development projects for archiving, sharing, or publishing data (e.g. Brown Dog, Globus, iRODS)	51.2%
Standards setting and/or promoting organization (e.g. IVOA, NIST, IEEE)	34.9%
University Professor	23.3%
Research Scientist in a University Facility	27.9%
Research Scientist in a Government Facility	7.0%
Research Scientist in a Commercial Facility	--
Graduate Student	--
Elected Government Official	--
Government Agency Leader	2.3%
Government Agency Staff	4.7%
Academic Publisher	7.0%
Professional Association	16.3%

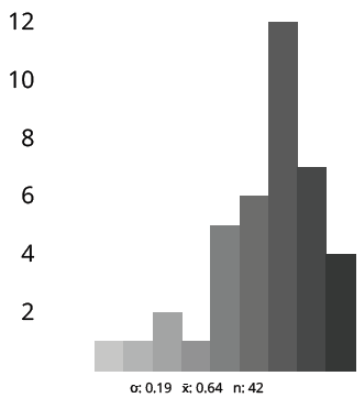
Selected respondent areas of expertise

- Web-based access to big data
- Life Sciences Data Sharing in Mental Health across many modalities (omics, clinical, imaging, neuro-signal recordings)
- Physics; information technology
- Technical data management at scale
- Astrophysics and cyberinfrastructure
- Bioinformatics
- Information description, discovery, access, and retrieval
- Information security, regulatory compliance
- Distributed and parallel computing, cyberinfrastructure, cyberGIS
- Cyberinfrastructure research, design, development and operation
- Applications of emerging technology to astronomy; software production; data management; archiving
- Information management, research networking systems, semantic web technologies, ontologies, informatics, collaboration, team science
- Data archiving and sharing, infrastructure
- Life sciences, epsec biochem, chem, but also molec bio, comp bio, struc bio, bioinformatics
- Data stewardship and exchange
- Research data curation for long-term access, discovery and reuse
- population studies
- Curation of local- and regional-scale in-situ, international, interdisciplinary scientific data

Respondent priorities for NDS

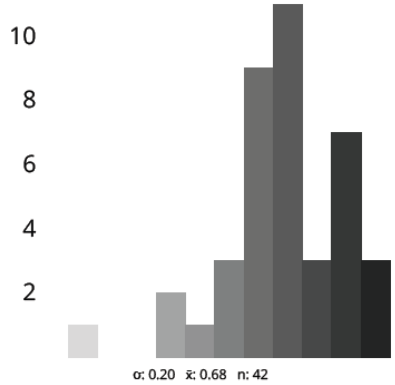
1. Access/use (290 points)
2. Discovery (258 points)
3. Storage (204 points)
4. Trust/provenance/quality (203 points)
5. Metadata correspondence (196 points)
6. Workflow integration (185 points)
7. Publishing/presenting (160 points)
8. Credit/feedback (143 points)
9. Teaching/mentoring (137 points)

EDUCATION <--•--> RESEARCH



How should NDS balance education vs research ℹ ()

PROTECTION <--•--> EASE OF ACCESS

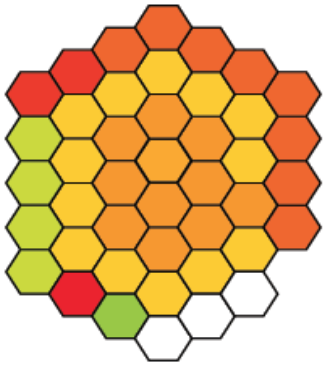


How should NDS balance protection vs ease of access ℹ ()

Indicate how NDS should balance its focus on the development of tools and approaches in support of education, on one hand, vs. research, on the other. You may see both as important, but select the point on the continuum that you believe represents the most appropriate balance. 0.64 (0.19)

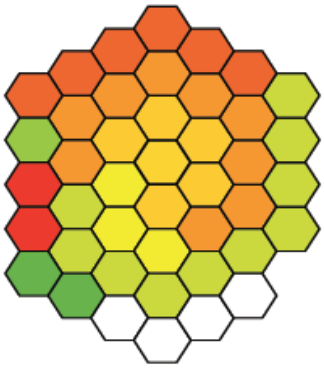
How should NDS balance between, on the one hand, strong protection against malicious attacks versus, on the other hand, ease of access? 0.68 (0.20)

NationalDataService.org



σ: 0.15 x̄: 0.34 n: 42

Currently is there cooperation and sharing of DMS among scientists ℹ ()



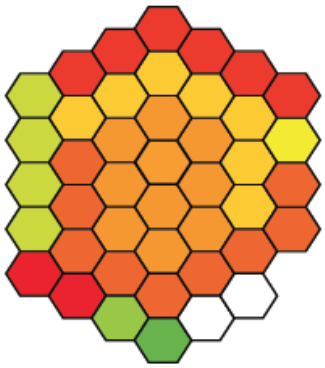
σ: 0.19 x̄: 0.42 n: 42

Currently is there cooperation and sharing of software/middleware/hardware ℹ ()

There is currently a high degree of cooperation and sharing of data, models, and simulations among end user scientists. 0.34 (0.15)

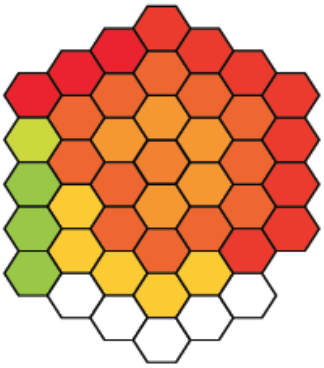
There is currently a high degree of cooperation and sharing of software, middleware and hardware among those developing and supporting cyberinfrastructure for science. 0.42 (0.19)

NationalDataService.org



σ: 0.19 x̄: 0.31 n: 42

Currently is there sufficient comm. and collab. between scientists and developers ℹ ()



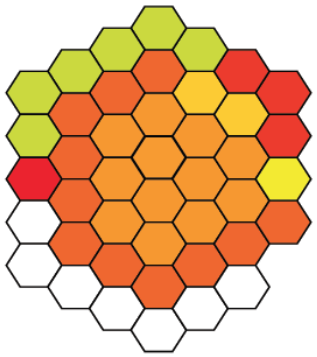
σ: 0.18 x̄: 0.25 n: 42

Currently is there sufficient science end-user knowledge and training ℹ ()

There is currently sufficient communication and collaboration between end user scientists and those who develop cyberinfrastructure tools and approaches to advance science. 0.31 (0.19)

There is currently sufficient science end-user knowledge and training so they can effectively use the present suite of cyberinfrastructure tools and train their students/colleagues in its use. 0.25 (0.18)

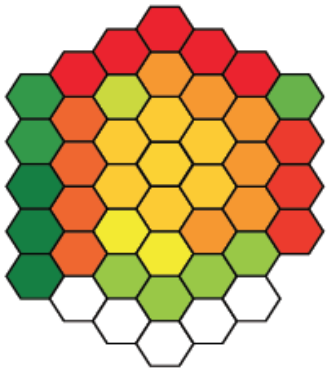
NationalDataService.org



σ: 0.16 x̄: 0.31 n: 42

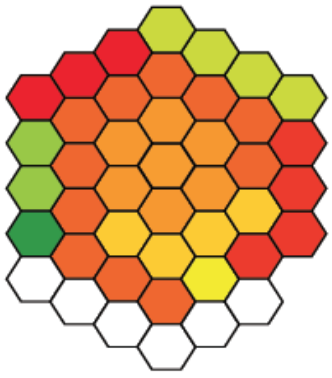
σ: 0.17 x̄: 0.32 n: 36

Are the publicly accessible DMS adequate for research & edu ℹ ()



σ: 0.31 x̄: 0.41 n: 42

Do the available resources permit easy access ℹ ()



σ: 0.21 x̄: 0.31 n: 42

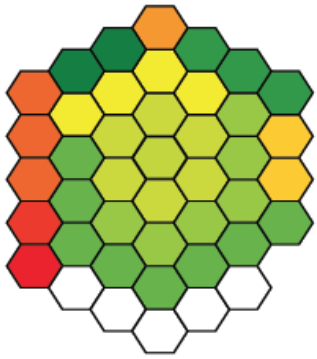
Are resources available for combining DMS from multiple sources ℹ ()

What is your assessment of the present suite of publicly accessible datasets, models, and/or software (e.g. visualization tools, middleware, etc.); to what degree is it adequate overall for research and education needs? 0.31 (0.16)

What is your assessment of the resources currently available to you for storing your research data in a manner that permits easy access and sharing. 0.41 (0.31)

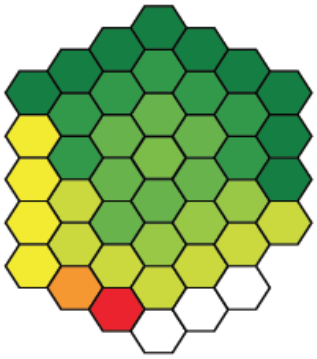
What is your assessment of the resources currently available to you for combining data, models and software tools from multiple sources. 0.31 (0.21)

NationalDataService.org



σ: 0.25 x̄: 0.61 n: 42

Should NDS specify guidelines or standards ℹ ()



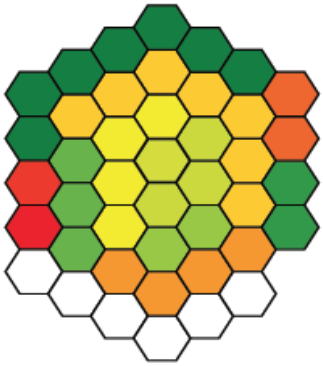
σ: 0.22 x̄: 0.76 n: 42

Should NDS use formal, internationally approved standards ℹ ()

NDS should specify guidelines or standards so there is more interoperability and uniformity in discovering, accessing, sharing, and disseminating science data. 0.61 (0.25)

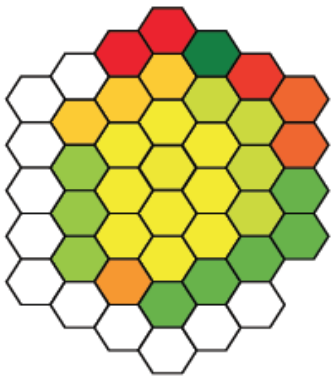
Where such standards exist, NDS should use formal, internationally approved, science-wide data access/sharing standards and protocols (e.g. ISO, OGC). 0.76 (0.22)

NationalDataService.org



σ: 0.29 x̄: 0.58 n: 42

Should NDS incorporate commercialization ℹ ()



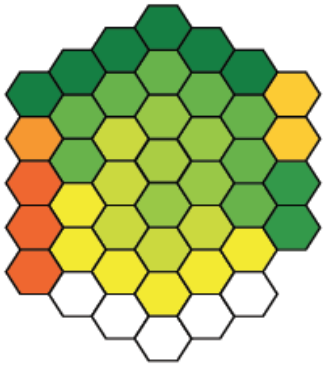
σ: 0.24 x̄: 0.52 n: 42

Should NDS generate tools that benefit commercial applications ℹ ()

NDS should incorporate commercial products or applications to reduce cost or speed development. 0.58 (0.29)

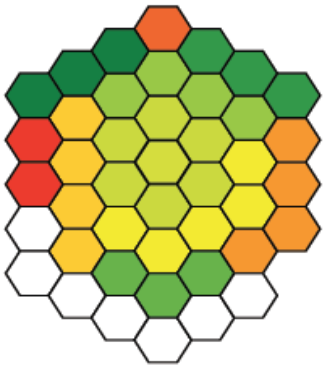
NDS should generate tools and approaches that benefit commercial products or applications. 0.52 (0.24)

NationalDataService.org



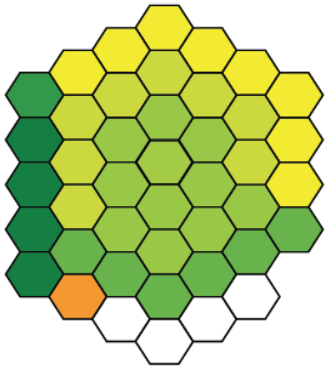
σ: 0.23 x̄: 0.67 n: 42

My employer will value my NDS based efforts ℹ()



σ: 0.25 x̄: 0.58 n: 42

My contributions to NDS will be recognized and valued by colleagues ℹ()



σ: 0.16 x̄: 0.68 n: 42

Impact of NDS on your primary organization ℹ()

My employer/organization will most likely value and reward any efforts I make in the shaping and development of NDS. 0.67 (0.23)

Any contributions I might make to the shaping and development of NDS will likely be recognized and valued by colleagues. 0.58 (0.25)

Based on what you know so far about the proposed National Data Service (NDS), please indicate what you anticipate the impact of NDS is likely to be on your primary organization. 0.68 (0.16)

What is your vision for what a successful National Data Service (NDS) would be like by the year 2025?

Interoperability and discovery across fields and disciplines:

- **"A wide variety of publicly-funded scientific data, available at different levels from student to researcher to power-user.** Interoperability of related datasets. Publishing and discovery with the same ease and acceptance as with arxiv.org today."
- "A hub for **sharing and reuse of research data across the sciences.** I think the Australian ANDS has done this well."

Established standards integrated in work flows:

- **"Integrated discovery standard.** For results, reporting of all "findings" in the general sciences, including life sciences. **Peer reviewed publication is no longer enough."**
- "Data, and then definitions of data as it relates to findings . . . **references to code, and the software, and machine images supporting the findings is also needed."**
- **"Template questions that assist researchers** in asking critical questions about the use, purpose, application, and archiving of data. What use is being made of data for policy, for example."

What is your vision for what a successful National Data Service (NDS) would be like by the year 2025?

Established standards integrated in work flows (cont.):

- **"Preserving and sharing research data is standard practice amongst scientists.** There are ample facilities to store, preserve, and share research data."
- "That the criteria we were asked to sort by order of importance earlier in this survey (storage, discovery, access, training, trust, etc.) are all met so that end users readily understand the importance of storing data for subsequent access, sharing, etc.). It may even be that performance objectives for researchers [are] created that **encourage and reward the practice of good data policies."**

Large-scale data:

- "A funded, ongoing facility for **transitioning data from large-scale scientific experiments/projects** (such as LSST) for curation and access after the experiment/project has concluded."
- "Massively-scaled, highly-integrated, most cost-effective, low-barrier, **trusted infrastructure** for support of research data life cycle interests."
- "Interoperable interfaces, **easy sharing of very large datasets, and enabling new types of discovery**. A platform for data."

What is your vision for what a successful National Data Service (NDS) would be like by the year 2025? (cont.)

Social systems:

- "By 2025 the NDS will have provided **overarching coordination for the many groups working with data** and will have created a **cultural shift in education and research towards making data matter."**
- "NDS will have commoditized basic storage and preservation services and will **support an evolving range of third-party services** that provide advanced capabilities for managing/integrating/visualizing/organizing specific data types. It will not become a monopoly portal. . ."
- **"Institutions see value and visibility for participating** are thus motivated to do so. Self-sustaining business model."
- **"The Data/CI/science communities would be organized and working collaboratively towards common goals supported by an inclusive governance structure.** Valuable existing infrastructures would be identified and integrated into a web-of-repositories model, prioritizing both specialization and broad re-use. Data and tools would be well-curated, and immediately useful to scientific, industry, education, and policy applications as needed. There would be clearly defined pathways for extending data use from initial collection through multiple re-use applications. Long-term resources would be in place. . . Research and CI would be completely partnered."

What is your vision for what a successful National Data Service (NDS) would be like by the year 2025? (cont.)

Seamless and beyond:

- "A unified, central service that any researcher knows to access in order to find a home for their data (be in via recommendation of an external repository or by deposit into an NDS repository) or knows to access in order to discover other data pertinent to their research questions. **Currently, finding a place to put data or trying to location data is a research endeavor all of its own.** The drives for Big Data and Open Data will not work and, in fact, will literally be counter productive – if they significantly slow down research projects or consume a significant fraction of researcher funding."
- **"Infrastructure as Service:** Cloud-based storage, **with no in and out costs for data transfers,** and ability to co-locate custom processing with data."
- "The scientists form in his head whatever questions they may have and **get the precise answer they are looking for through their neural interface.** However, a less ambitious vision would be a lower tech version - perhaps a spoken question is translated via AI into the right commands that query the right data sources, correlate and combine the information, and provide the answer perhaps holographically or just a plain answer such as you would get from your graduate student."

What is your vision for what a successful National Data Service (NDS) would be like by the year 2025? (cont.)

Bottom line:

- "A successful NDS would **increase data sharing and data re-use across multiple disciplines."**
- "Issues of data storage, management, access, security, reliability, etc. are **invisible to the end users. It just works."**
- "The most important development will be that **data is, by default, preserved and accessible** – much as Flickr ensures that the default for digital photos is that they are preserved and accessible. Building on this base, we will see **an explosion in interesting applications and new science based on systematic analysis and mining of large quantities of data."**

NDS Stakeholder Alignment Data

Web display of data at:
http://waymarksystems.org/presentations/nds

Again, caution: *These are preliminary data, with additional perspectives needed and additional analysis to be conducted*

Cyberinfrastructure for the Geosciences

The vision. . .

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

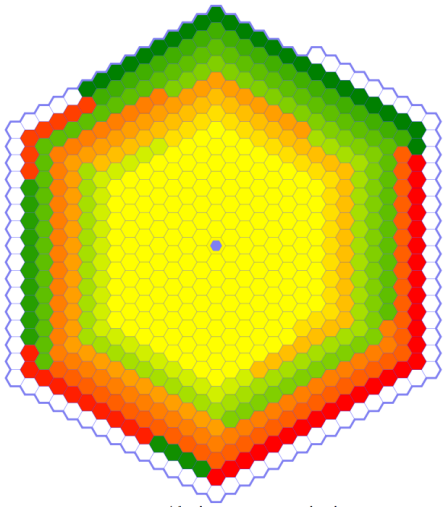
– *GEO Vision Report of NSF Geoscience Directorate Advisory Committee, 2009*

NationalDataService.org

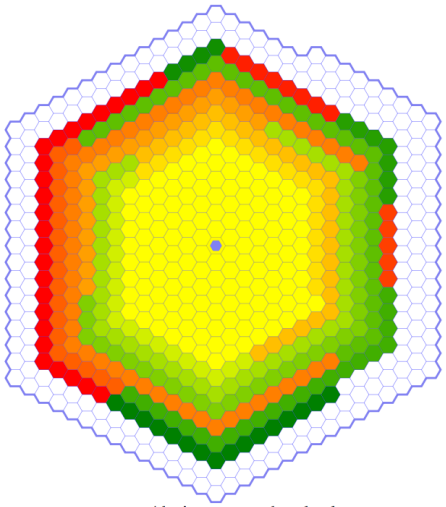
Stakeholder alignment data by End User Workshop (n=1,544)

EarthCube Website	(n=164)	
Data Centers	(n=578)	
Early Career	(n=37)	Oct. 17-18, 2012
Structure and Tectonics	(n=24)	Nov. 19-20, 2012
EarthScope	(n=22)	Nov. 29-30, 2012
Experimental Stratigraphy	(n=21)	Dec. 11-12, 2012
Atmospheric Modeling	(n=29)	Dec. 19, 2012
OGC	(n=14)	Jan. 13, 2013
Critical Zone	(n=39)	Jan. 21-23, 2013
Hydrology / Envisioning a Digital Crust	(n=23)	Jan. 29-31, 2013
Paleogeoscience	(n=40)	Feb. 3-5, 2013
Education & Workforce Training	(n=33)	Mar. 3-5, 2013
Petrology & Geochemistry	(n=59)	Mar. 6-7, 2013
Sedimentary Geology	(n=50)	Mar. 25-27, 2013
Community Geodynamic Modeling	(n=45)	Apr. 22-24, 2013
Integrating Inland Waters, Geochemistry, Biogeochem …	(n=46)	Apr. 24-26, 2013
Deep Sea Floor Processes and Dynamics	(n=29)	June 5-6, 2013
Real-Time Data	(n=25)	June 17-18, 2013
Ocean 'Omics	(n=42)	Aug. 21-23, 2013
Coral Reef Systems	(n=44)	Sept. 18-19/Oct. 23-24, 2013
Geochronology	(n=66)	Oct. 1-3, 2013
Ocean Ecosystem Dynamics	(n=36)	Oct. 7-8, 2013
Clouds and Aerosols	(n=39)	Oct. 21-22, 2013
Rock Deformation and Mineral Physics	(n=35)	Nov. 12-14, 2013

Cooperation/sharing among geoscientists
Cooperation/sharing among cyber-developers



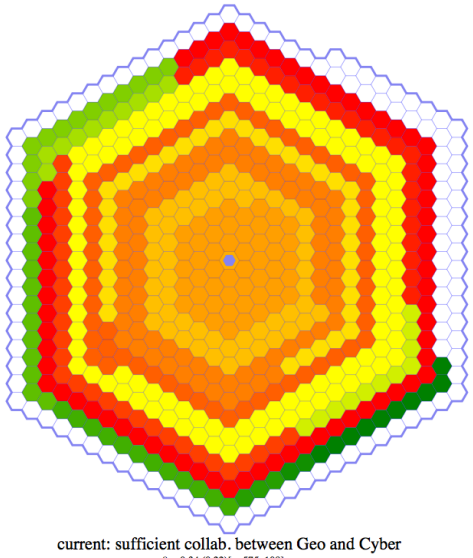
current: coop./sharing among geoscientists
() = 0.49 (0.23)[n=611, 72]



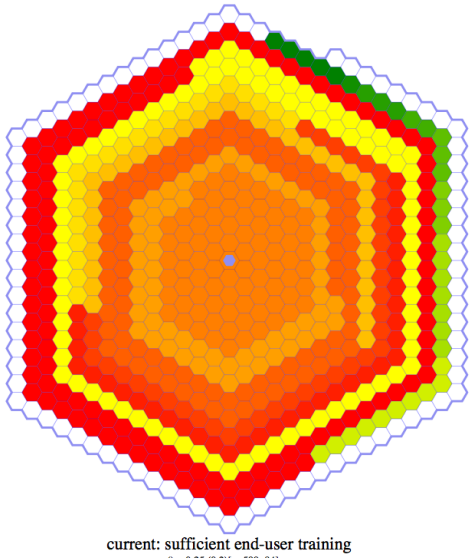
current: coop./sharing among cyber. developers
() = 0.49 (0.23)[n=489, 194]

There is currently a high degree of sharing of data, models, and software among geoscientists. (v69)
There is currently a high degree of sharing of software, middleware and hardware among those developing and supporting cyberinfrastructure for the geosciences. (v70)

Collaboration between geo and cyber
Sufficient end user training



current: sufficient collab. between Geo and Cyber
() = 0.34 (0.22)[n=575, 108]



current: sufficient end-user training
() = 0.25 (0.2)[n=599, 84]

There is currently sufficient communication and collaboration between geoscientists and those who develop cyberinfrastructure tools and approaches to advance the geosciences. (v72)
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. **(v73)**

NationalDataService.org

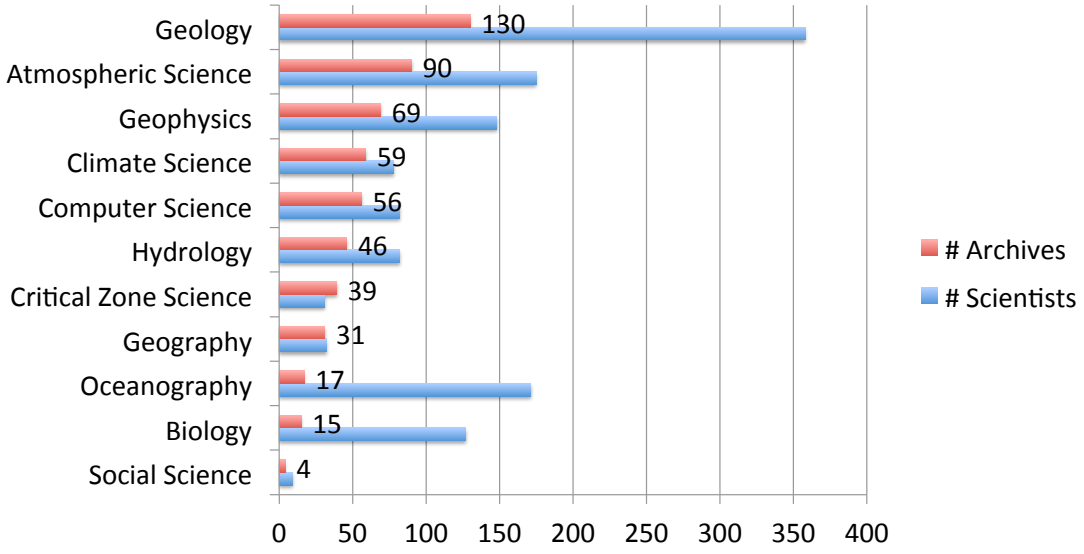
Top ten barriers to sharing data (categories):

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).

NationalDataService.org

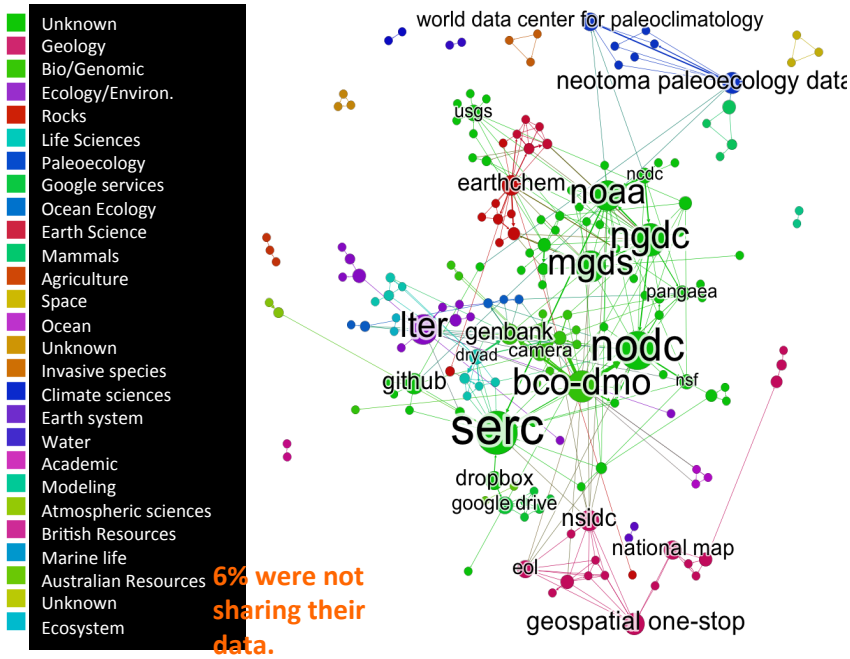
Unique repositories/locations for accessing data



Source: Preliminary analysis by Cheryl Thompson, University of Illinois

NationalDataService.org

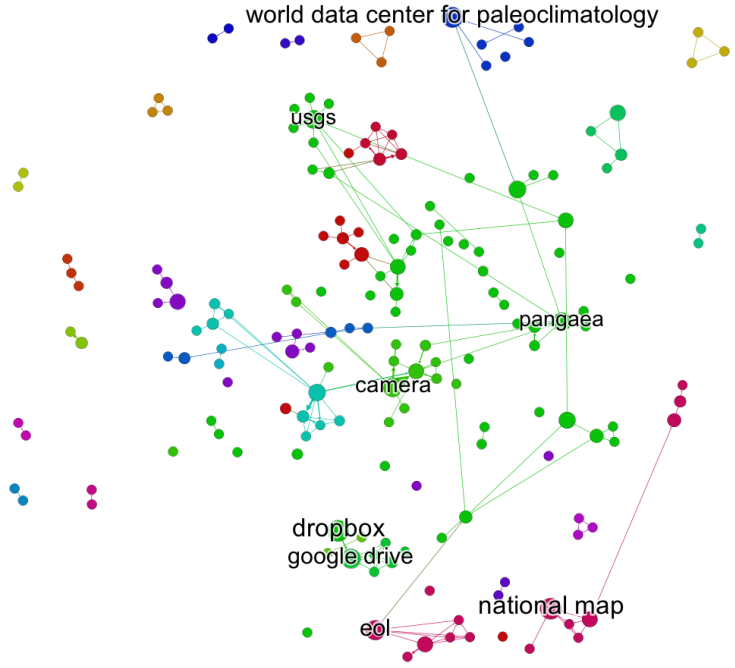
How are scientists sharing data?



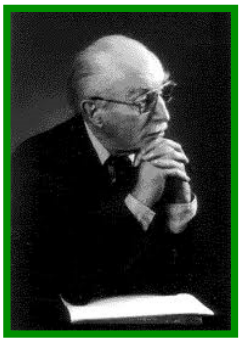
Source: Preliminary analysis by Cheryl Thompson, University of Illinois

NationalDataService.org

Removing highly connected players



Source: Preliminary analysis by Cheryl Thompson, University of Illinois



"... 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)

Predictable tensions in governance

Charters for forums

National Council of Data Facilities

I. Preamble
II. Vision
III. Mission and goals
IV. Definition
V. Membership
VI. Roles and responsibilities
VII. Operations
VIII. Coordination with EarthCube
IX. Signatures (charter members)

III. Mission and goals

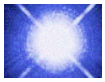
The mission of the Council of Data Facilities is to serve in a coordinating and facilitating role that includes advancing the following goals:

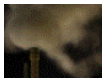
- Providing a collective voice on behalf of the member data facilities to the NSF and other foundations and associations, as appropriate.
- Identifying, endorsing, and promoting standards and best or exemplary practices in the organization and operation of a data facility.
- Identifying and supporting the development and utilization of shared infrastructure services, including computing services, professional staff development and training services, and related activities.
- Fostering innovation through collaborative projects.
- Collaborating with standard-setting bodies with respect to standards for data sharing and interoperability, metadata, and related matters.

In advancing this mission, the Council of Data Facilities is committed to working with relevant agencies, professional associations, initiatives, and other complementary efforts.

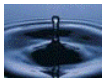
Most important challenges of the 21st Century, as identified by NAE


Make solar energy economical


Provide energy from fusion


Develop carbon sequestration methods


Manage the nitrogen cycle


Provide access to clean water


Restore and improve urban infrastructure


Advance health informatics


Engineer better medicines


Reverse-engineer the brain


Prevent nuclear terror


Secure cyberspace


Enhance virtual reality


Advance personalized learning


Engineer the tools of scientific discovery

NationalDataService.org

Source: http://www.engineeringchallenges.org/

APPENDIX

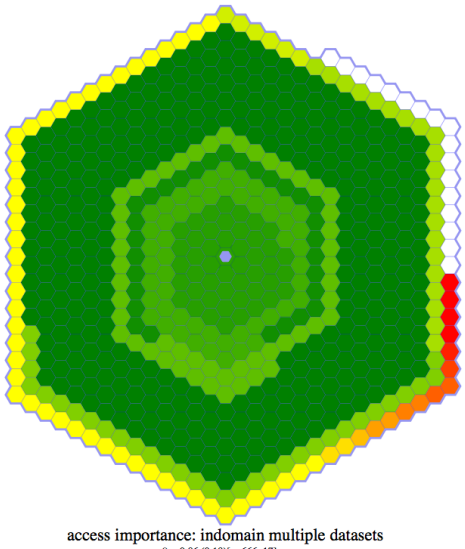
Cyberinfrastructure for the Geosciences

The vision. . .

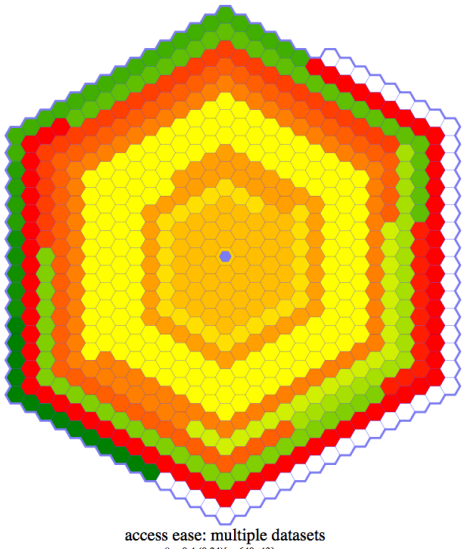
"Over the next decade, the geosciences community commits to developing a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core, including the influences of humans and ecosystems."

– *GEO Vision Report of NSF Geoscience Directorate Advisory Committee, 2009*

Accessing data, models, and software within fields/disciplines: Importance and ease



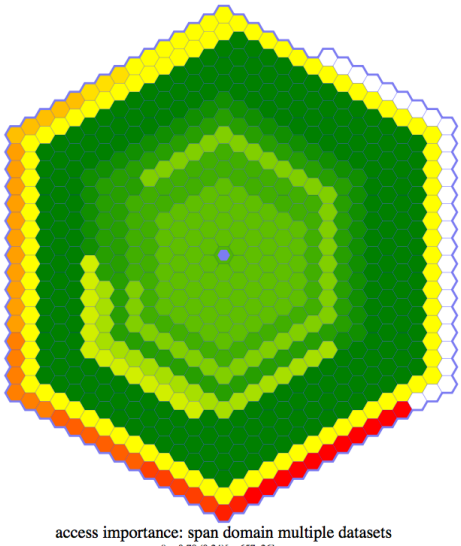
access importance: indomain multiple datasets
() = 0.86 (0.18)[n=666, 17]



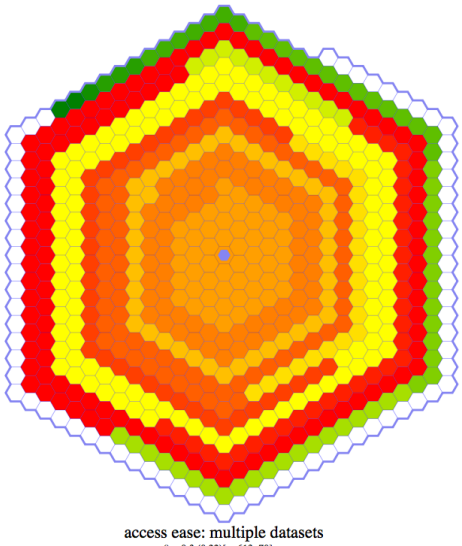
access ease: multiple datasets
() = 0.4 (0.24)[n=640, 43]

How IMPORTANT is it for you to find, access, and/or integrate multiple datasets, models, and/or software (e.g. visualization tools, middleware, etc.) in your field or discipline? (v58)
How EASY is it for you to find, access, and/or integrate multiple datasets, models, and/or software (e.g. visualization tools, middleware, etc.) in your field or discipline? (v59)

Accessing data, models, and software across fields/disciplines: Importance and ease



access importance: span domain multiple datasets
() = 0.78 (0.24)[n=657, 26]

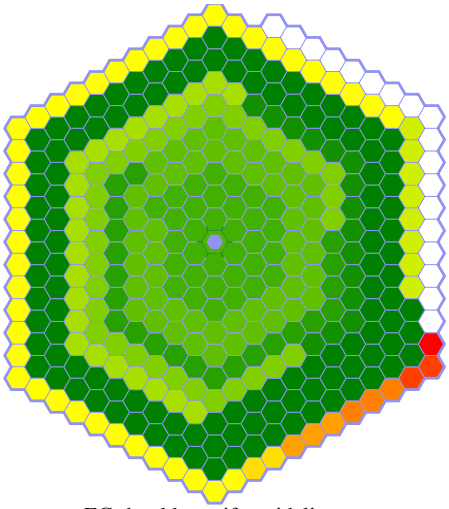


access ease: multiple datasets
() = 0.3 (0.22)[n=613, 70]

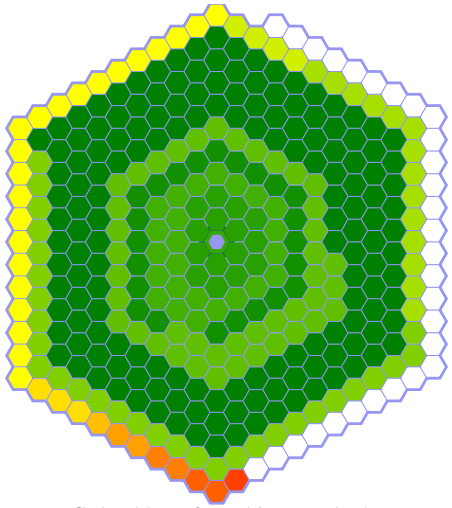
How IMPORTANT is it for you to find, access, and/or integrate multiple datasets, models, and/or software (e.g. visualization tools, middleware, etc.) that span different fields or disciplines? (v60)
How EASY is it for you to find, access, and/or integrate multiple datasets, models, and/or software (e.g. visualization tools, middleware, etc.) that span different fields or disciplines? (v61)

NationalDataService.org

Support for EarthCube specifying guidelines
Support for guidelines using international standards



EC should specify guidelines
μ(σ) = 0.79 (0.19)[n=353, 18]

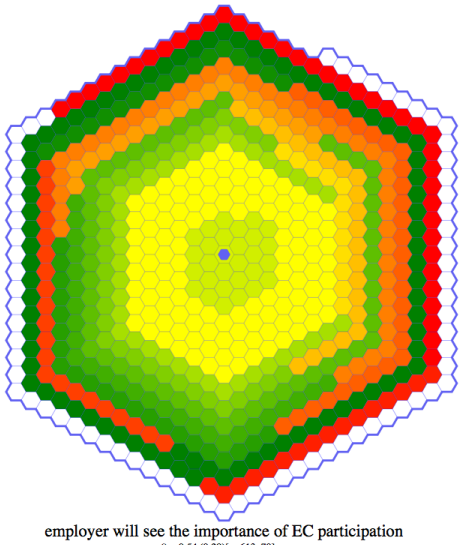


EC should use formal int. standards
μ(σ) = 0.84 (0.18)[n=342, 29]

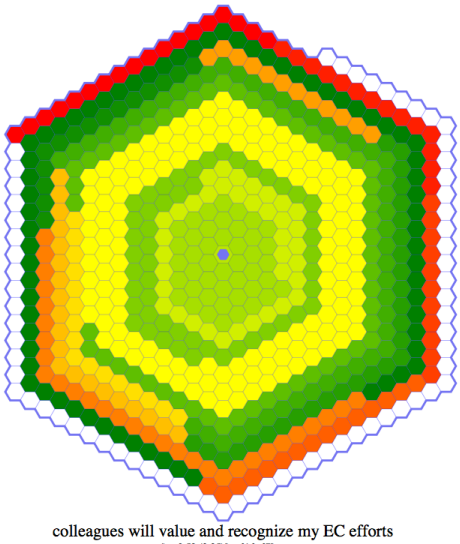
The EarthCube initiative should specify guidelines so there is more interoperability and uniformity in discovering, accessing, sharing, and disseminating geoscience data. (v99)
Where such standards exist, EarthCube should use formal, internationally approved, geoscience-wide data access/sharing standards and protocols (e.g. ISO, OGC). (v100)

NationalDataService.org

EarthCube
Support for sharing from employer and colleagues



employer will see the importance of EC participation
() = 0.54 (0.29)[n=613, 70]

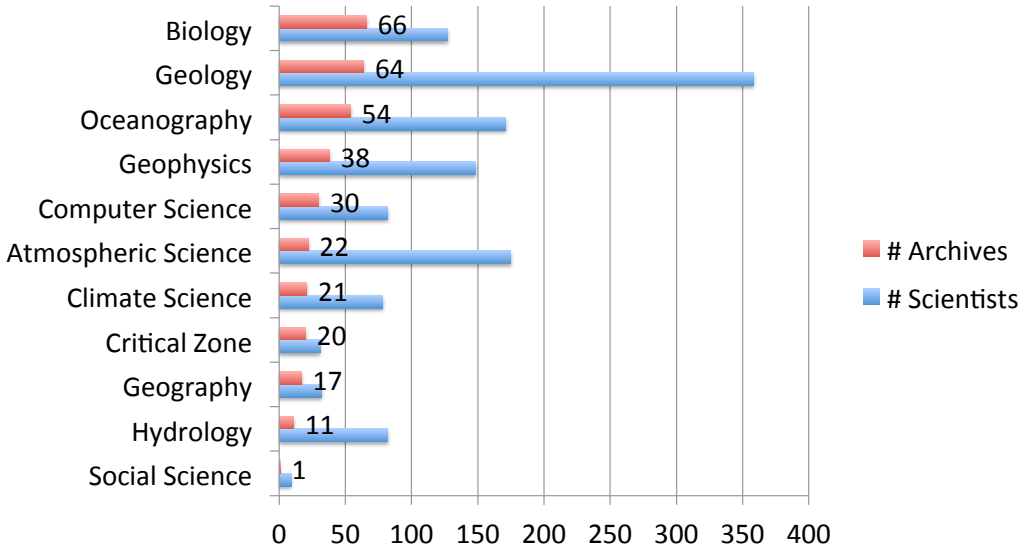


colleagues will value and recognize my EC efforts
() = 0.59 (0.25)[n=616, 67]

My employer/organization will most likely value and reward any efforts I make in the shaping and development of EarthCube (v120). Any contributions I might make to the shaping and development of EarthCube will likely be recognized and valued by colleagues in my field/discipline (v122).

NationalDataService.org

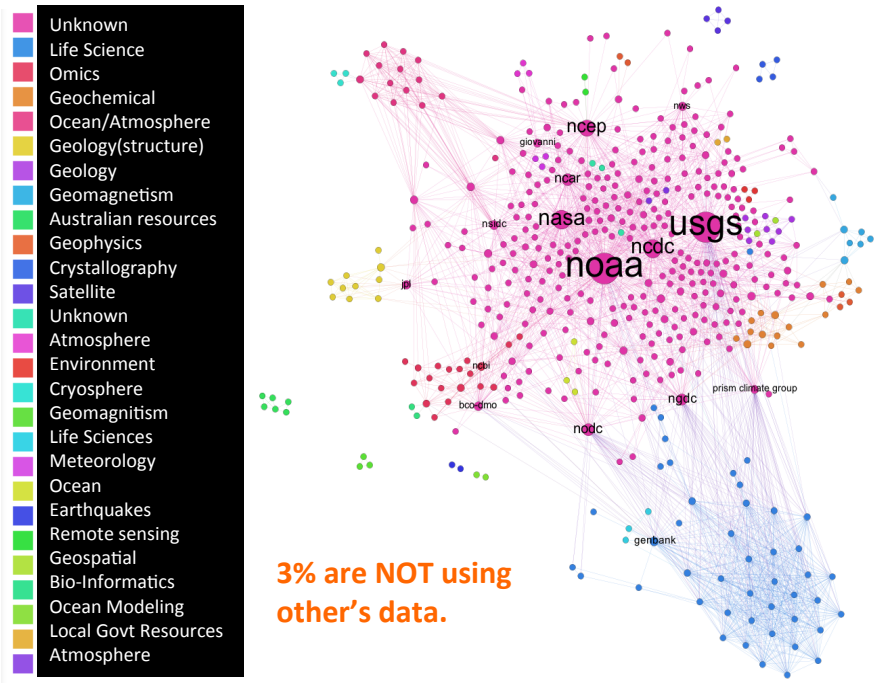
Unique repositories/locations for sharing data



Source: Preliminary analysis by Cheryl Thompson, University of Illinois

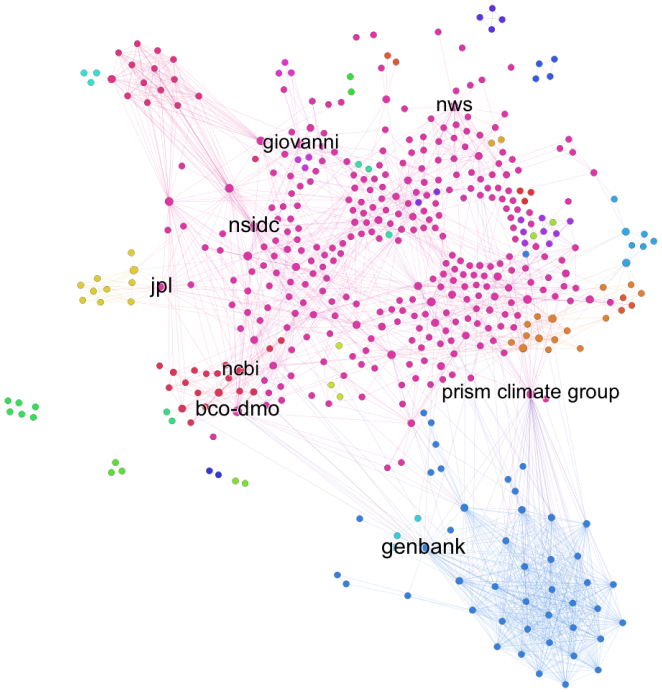
NationalDataService.org

Where Scientists Get Data?



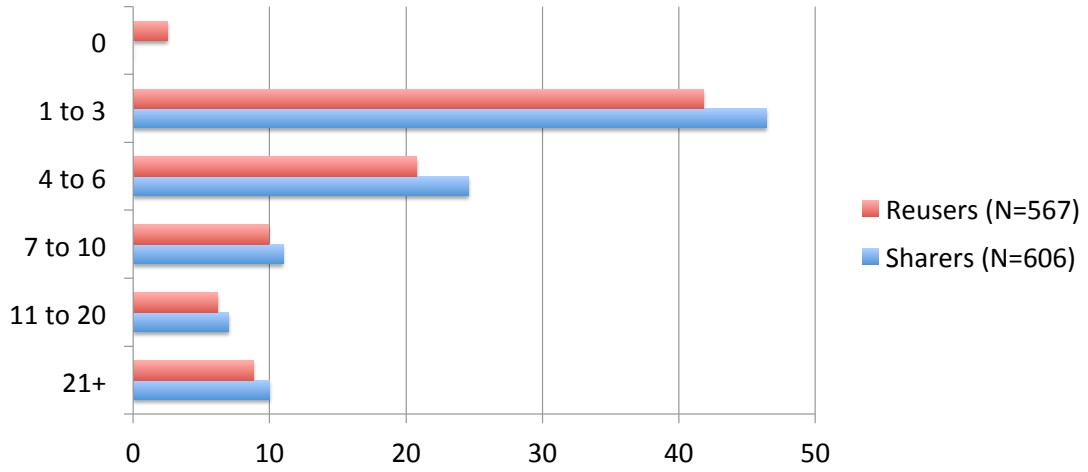
Source: Preliminary analysis by Cheryl Thompson, University of Illinois

Removing highly connected players



Data products made publicly available

During the past five years, approximately how many data sets, models, or software have you made publically available to others?

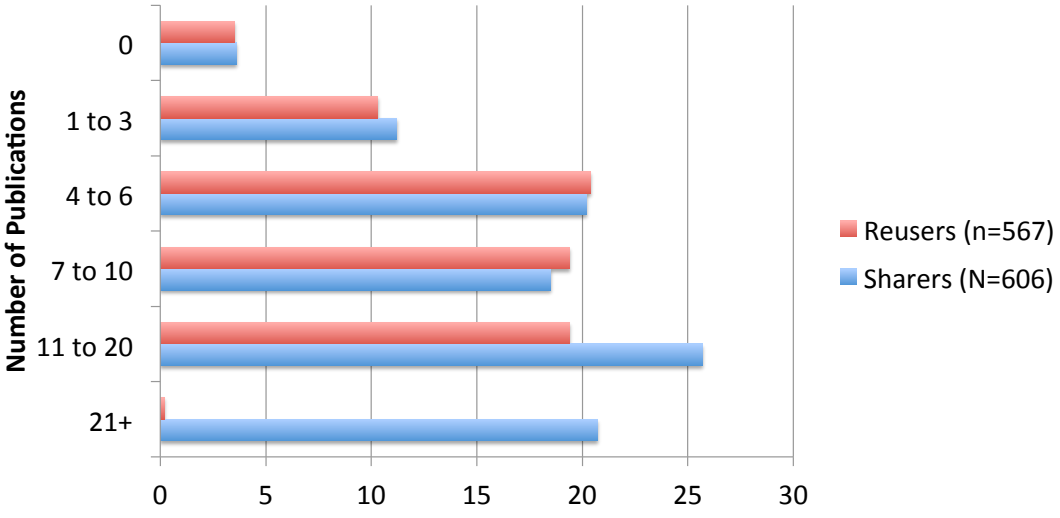


Source: Preliminary analysis by Cheryl Thompson, University of Illinois

NationalDataService.org

Publications

During the past five years, approximately how many total publications have you authored or co-authored?



Source: Preliminary analysis by Cheryl Thompson, University of Illinois

NationalDataService.org

Steps in the process

Phase I: Orienting

- **Working Group:** Establish a representative working group
- **Scale/Scope:** Define the current and potential scale and scope of the initiative(s)

Phase II: Map the Terrain

- **Stakeholders:** Specify stakeholders with appropriate granularity
- **Interests:** Identify interests on current and enduring issues
- **Instrument:** Develop and pre-test the stakeholder mapping instrument
- **Exploratory Map:** Survey initially identified stakeholders on targeted issues
- **Visualization:** Visually display the landscape of stakeholder interests

Phase II: Map the Terrain (cont.)

- **Analytics:** Highlight opportunities and risks with appropriate analytics
- **Feedback:** Provide synchronous and a-synchronous feedback to stakeholders
- **Confirmatory Map:** Survey a broad cross-section of stakeholders on many issues
- **Visualization, Analytics, Feedback, and Future Stakeholder Maps:** Repeat with an appropriate cadence

Phase III: Facilitated Dialogue and Action (optional)

- **Vision:** Construct a shared visions of success
- **Charter:** Charter public-private forums
- **Guidelines:** Establish protocols and standards
- **Barriers:** Reduce institutional barriers
- **Iteration:** Track progress over time – check and adjust
- **Improvement:** Enable continuous improvement

Collective impact model

Stanford SOCIAL INNOVATION REVIEW

- ***Common Agenda:*** . . .all participants to have a shared vision for change, one that includes a common understanding of the problem and a joint approach. . .
- ***Shared Measurement Systems:*** Collecting data and measuring results consistently on a short list of indicators at the community level and across all participating organizations not only ensures that all efforts remain aligned, it also enables the participants to hold each other accountable and learn from each other. . .
- ***Mutually Reinforcing Activities:*** . . . not by requiring that all participants do the same thing, but by encouraging each participant to undertake the specific set of activities at which it excels in a way that supports and is coordinated with the actions of others. . .
- ***Continuous Communication:*** . . . several years of regular meetings to build up enough experience with each other to recognize and appreciate the common motivation behind their different efforts. . . Even the process of creating a common vocabulary takes time. . .
- ***Backbone Support Organizations:*** *Creating and managing collective impact requires a separate organization and staff with a very specific set of skills to serve as the backbone for the entire initiative. . .*

Source: http://www.ssireview.org/articles/entry/collective_impact