

Computing, Data, and Cyberinfrastructure for a Systems Approach to Studying the Earth

The National Academies of Sciences, Engineering, and Medicine

January 2021

Joel Cutcher-Gershenfeld,
Heller School for Social Policy and Management,
Brandeis University

Support from the National Science Foundation is deeply appreciated, including: 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 CI," NSF OCI 12-56163, "Envisioning Success: A Workshop for Next Generation EarthCube Scholars and Scientists," NSF I-CORPS 1313562 "Stakeholder Alignment for Public-Private Partnerships," NSF DBI 1636461 "Revitalizing the National Ecological Network Cyberinfrastructure," NSF CNS 1939224 "Collaborative Strategies for Successful Large-Scale, Distributed Science and Engineering Projects," and sub-contracts for stakeholder maps through NDS, CaRCC, XSEDE, MS-CC, West Big Data Hub, Internet2, ARL, AAU, APLU, and others. **Stakeholder Mapping enabled by WayMark Analytics**

Stakeholder Alignment Collaborative: Karen Baker, UIUC; Nick Berente, Notre Dame University; Paul Arthur Berkman, Tufts University; Pat Canavan, WayMark Analytics; Joel Cutcher-Gershenfeld, Brandeis University; F. Alex Feltus, Clemson University; Alysia Garmulewicz, University of Santiago; Michael Haberman, UIUC; Ron Hutchins, University of Virginia; John Leslie King, University of Michigan; Chris Lenhardt, RENCI; Spenser Lewis, Draper Labs/Brandeis University; Matt Mayernik, NCAR/UCAR; Michael Maffi, Penn State University; Charles McElroy, Cal Tech; Barbara Mittleman, WayMark Analytics; Beth Plale, Indiana University; Raj Sampath, Brandeis University; Namchul Shin, Pace University; Shelly Stall, AGU; Pippa Veazey, University of Alaska; Susan Winter, University of Maryland.

Overview

Interdisciplinary, multi-stakeholder consortia will be the primary vehicle for advancing a systems approach to studying the Earth.

The computing, data, and cyberinfrastructure needed for **a systems approach to studying the Earth requires diverse stakeholders to accomplish together what they can't do separately – it is a complex set of social *and* technical challenges.**

These points will be motivated with examples from:

- EarthCube
- iSamples
- Coalition on Publishing Data on the Earth and Space Sciences (COPDESS)
- National Ecological Observatory Network (NEON)
- Minority Serving Cyberinfrastructure Consortium (MS-CC)
- Campus Research Computing Consortium (CaRCC)
- Stakeholder Alignment Collaborative

First, however, a bit of geoscience humor ... using data from 129-65 million years ago let's predict the 2020 election results (and lift up minority voting in the election) ...

Cretaceous Era 129-65 Million Years Ago



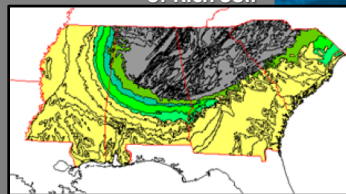
Ron Blakey/[Northern Arizona University](#)

Cretaceous Era 129-65 Million Years Ago

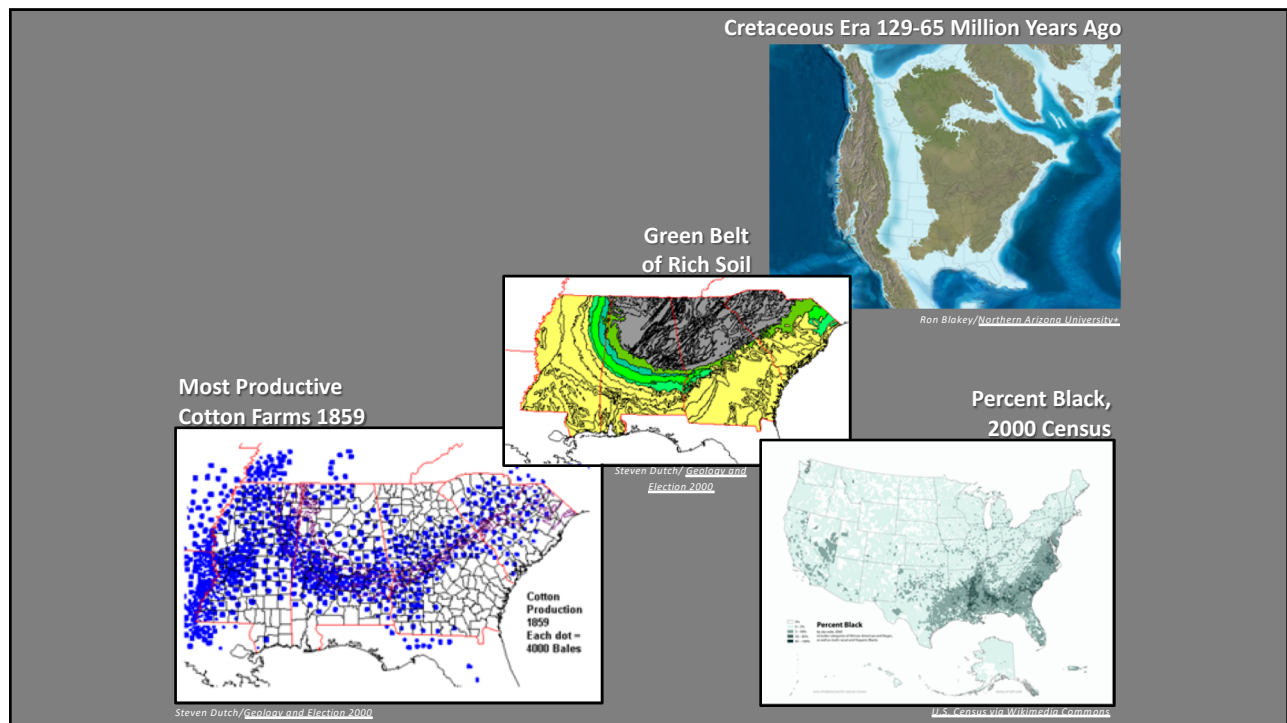
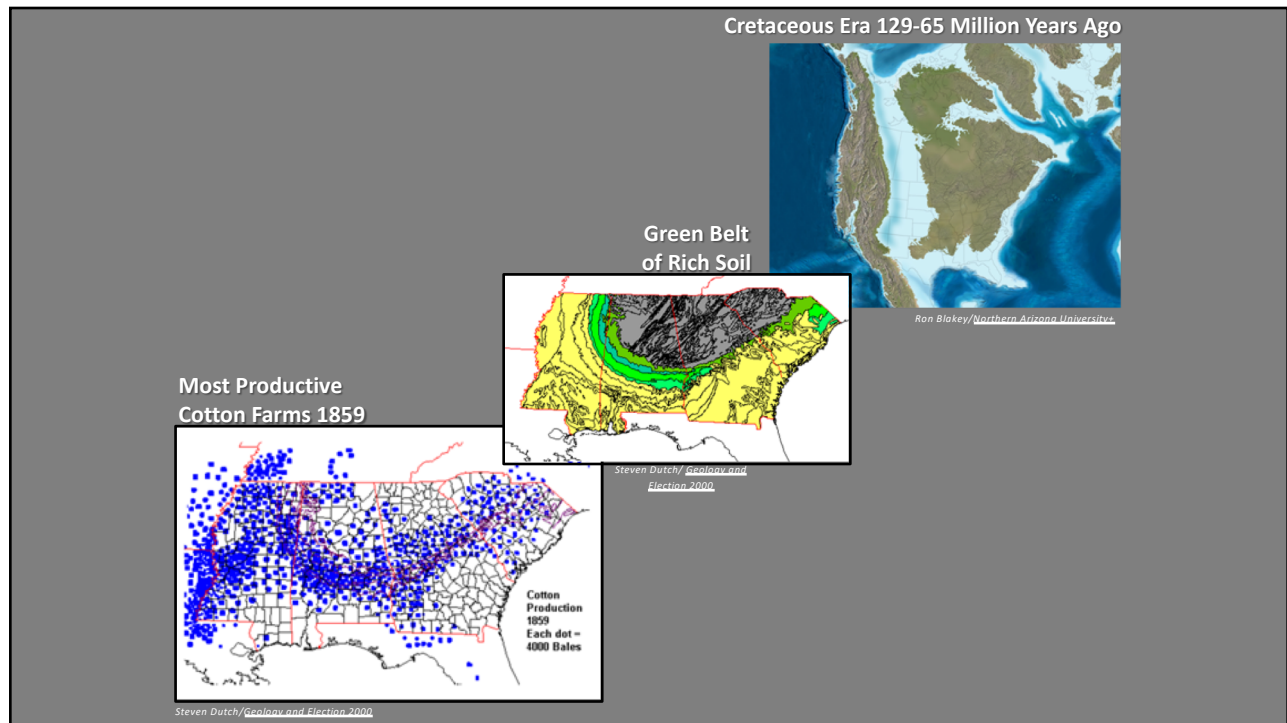


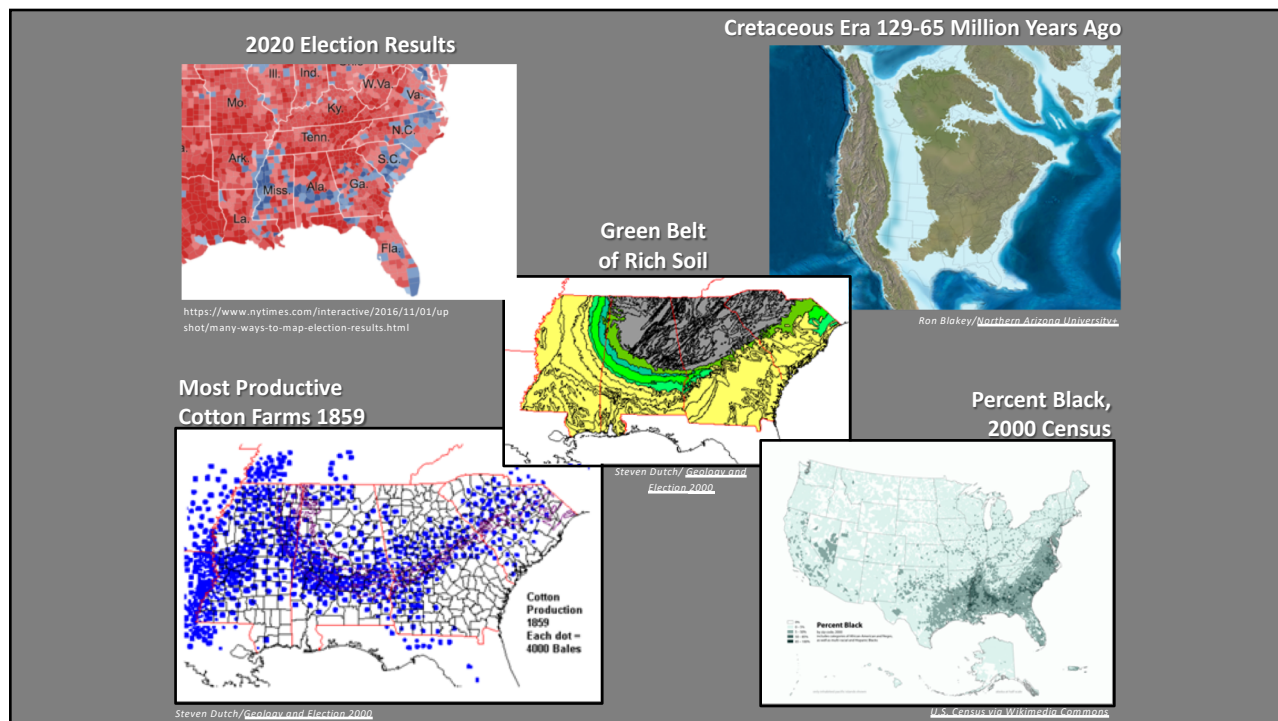
Ron Blakey/[Northern Arizona University](#)

Green Belt
of Rich Soil



Steven Dutch/[Geology and
Election 2000](#)





Cyberinfrastructure for the Geosciences

The initial 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





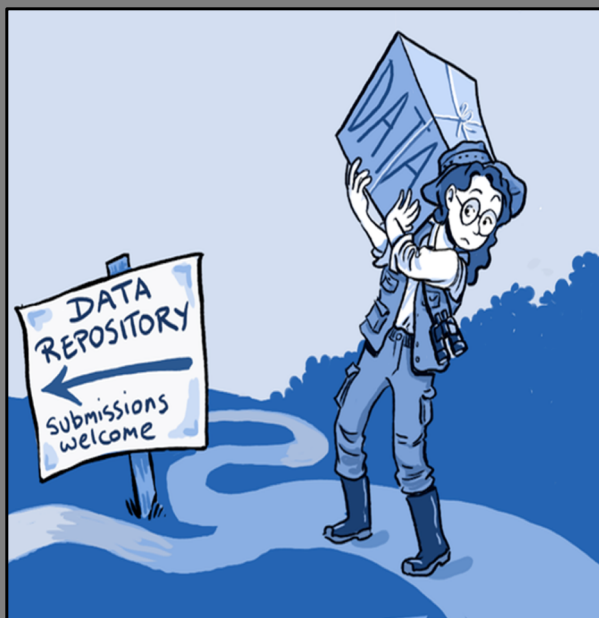
February 22, 2013
MEMORANDUM FOR THE HEADS OF EXECUTIVE
DEPARTMENTS AND AGENCIES

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law . . . the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data. . . . **These policies will accelerate scientific breakthroughs and innovation, promote entrepreneurship, and enhance economic growth and job creation.**



January 1, 2013
ARC Open Access Policy

Any publications arising from an ARC supported research Project must be deposited into an open access institutional repository within a twelve (12) month period from the date of publication. . . The Australian Government makes a major investment in research to support its essential role in improving the wellbeing of our society. To maximise the benefits from research, publications resulting from research activities must be disseminated as broadly as possible to allow access by other researchers and the wider community.



2012 Stakeholder Alignment data by Fields and disciplines

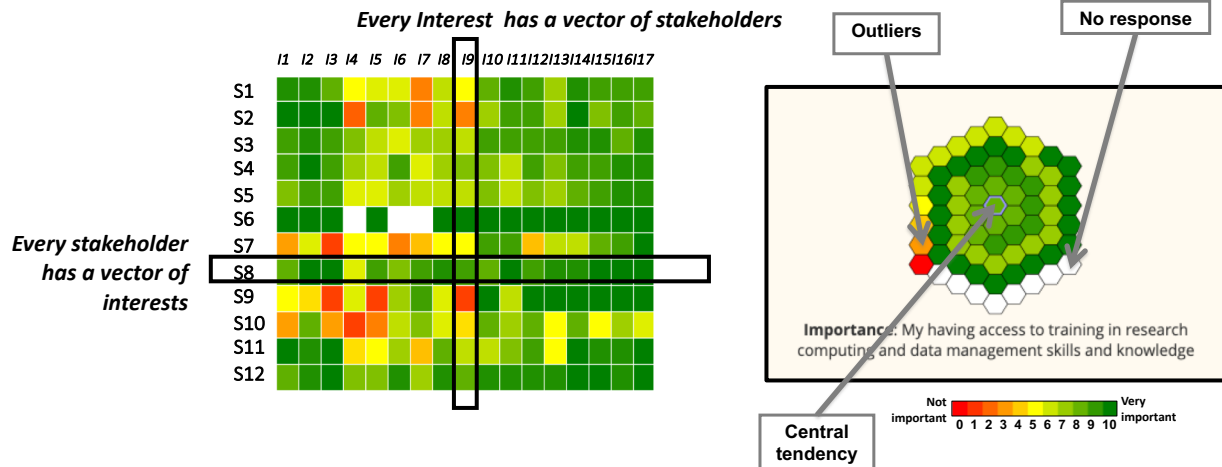
	Primary	Secondary
Atmospheric	n=175 (11.3%)	n=74 (4.8%)
Biologist/Ecosystems	n=127 (8.2%)	n=101 (6.5%)
Climate Scientists	n=78 (5.1%)	n=86 (5.6%)
Critical zone	n=31 (2%)	n=44 (2.8%)
Geographers	n=32 (2.1%)	n=34 (2.2%)
Geologists	n=358 (23.2%)	n=112 (7.3%)
Geophysicists	n=148 (9.6%)	n=73 (4.7%)
Hydrologists	n=82 (5.3%)	n=61 (4.0%)
Oceanographers	n=171 (11.3%)	n=94 (6.1%)
Computer/Cyber	n=82 (5.3%)	n=91 (5.9%)
Data managers	n=53 (3.4%)	n=86 (5.6%)
Software engineers	n=24 (1.6%)	n=50 (3.2%)

Note: additional categories included in the survey, but these are the focus here.

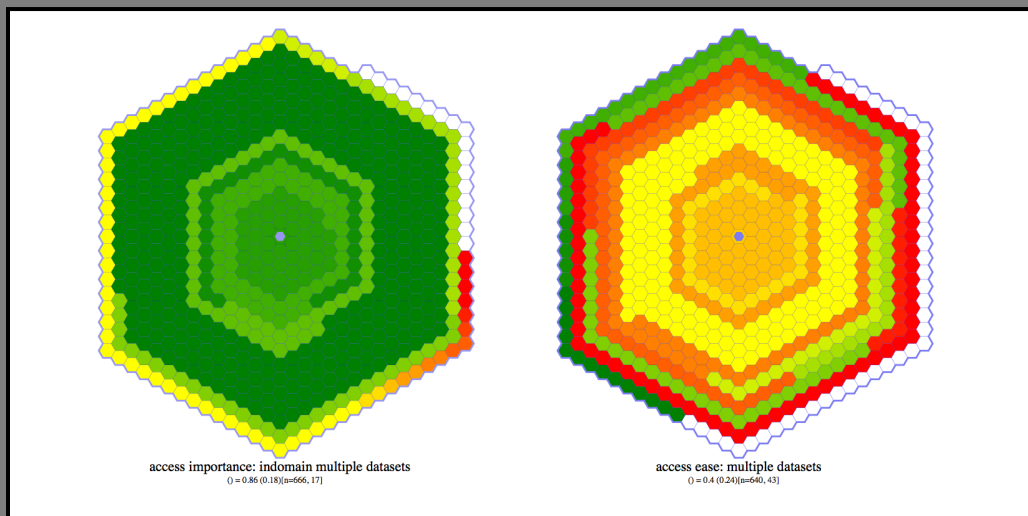
2012 Specific areas of expertise

- | | |
|--|---|
| <ul style="list-style-type: none"> • 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 | <ul style="list-style-type: none"> • 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 |
|--|---|

Stakeholders and interests

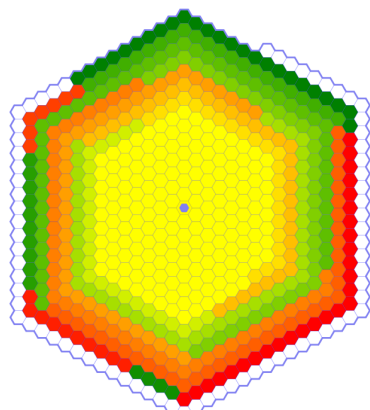


2012 Accessing data, models, and software within fields/disciplines: Importance and ease

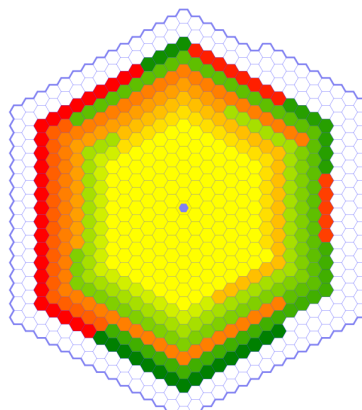


How IMPORTANT is it for you to find, access, and/or integrate multiple datasets, models, and/or software (e.g. visualization tools, middleware, etc.) in your field or discipline? (v58)
How EASY is it for you to find, access, and/or integrate multiple datasets, models, and/or software (e.g. visualization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

2012 Cooperation/sharing among geoscientists / Cooperation/sharing among cyber-developers



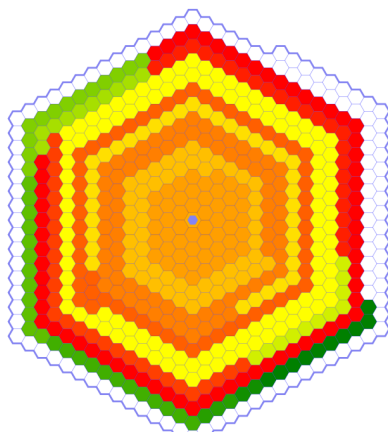
current: coop./sharing among geoscientists
() = 0.49 (0.23)[n=611, 72]



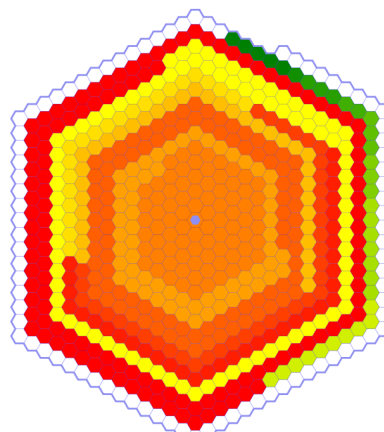
current: coop./sharing among cyber. developers
() = 0.49 (0.23)[n=489, 194]

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)

2012 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 (0.22)[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)

NSF End-User Workshops 2012-2014

Version 1 of the survey instrument:

1. Early Career	24.7% (n = 37 of 150)	Oct. 17–18, 2012
2. Structure and Tectonics	70.5% (n = 24 of 34)	Nov. 19–20, 2012
3. EarthScope	31.9% (n = 22 of 69)	Nov. 29–30, 2012
4. Experimental Stratigraphy	42.9% (n = 21 of 49)	Dec. 11–12, 2012
5. Atmospheric Modeling / Data Assimilation and Ensemble Prediction	31.2% (n = 29 of 74)	Dec. 19, 2012

Version 2 of the survey instrument:

6. OGC	28.0% (n = 14 of 50)	Jan. 13, 2013
7. Critical Zone	28.3% (n = 39 of 138)	Jan. 21–23, 2013
8. Hydrology / Envisioning a Digital Crust	48.9% (n = 23 of 47)	Jan. 29–31, 2013
9. Paleogeoscience	50.6% (n = 40 of 79)	Feb. 3–5, 2013
10. Education & Workforce Training	57.9% (n = 33 of 57)	Mar. 3–5, 2013
11. Petrology & Geochemistry	71.1% (n = 59 of 83)	Mar. 6–7, 2013
12. Sedimentary Geology	55.6% (n = 50 of 90)	Mar. 25–27, 2013
13. Community Geodynamic Modeling	46.4% (n = 45 of 97)	Apr. 22–24, 2013
14. Integrating Inland Waters, Geochemistry, Biogeochem and Fluvial Sedimentology Communities	39.0% (n = 46 of 118)	Apr. 24–26, 2013
15. Deep Sea Floor Processes and Dynamics	49.2% (n = 29 of 59)	June 5–6, 2013
16. Real-Time Data	23.4% (n = 25 of 107)	June 17–18, 2013
17. Ocean 'Omics	71.2% (n = 42 of 59)	Aug. 21–23, 2013
18. Coral Reef Systems (two workshops)	91.7% (n = 44 of 48)	Sept. 18–19/ Oct. 23–24, 2013
19. Geochronology	44.6% (n = 66 of 148)	Oct. 1–3, 2013
20. Ocean Ecosystem Dynamics	45.0% (n = 36 of 80)	Oct. 7–8, 2013
21. Clouds and Aerosols	63.9% (n = 39 of 61)	Oct. 21–22, 2013
22. Rock Deformation and Mineral Physics	44.3% (n = 37 of 79)	Nov. 12–14, 2013
23. Marine Seismic	46.2% (n = 24 of 52)	Dec. 11–12, 2014

Table 1: Response Rates and Timing of NSF EarthCube Disciplinary Domain Workshops (n=824 of 1,828).

2012 NSF Early Career Workshop



Selected "Tweets" from Day One

- EarthCube will give all scientists same chance of making major contributions regardless of institution size and funding.
- EarthCube could provide databases of minerals to give access to hard-to-obtain data to see big picture processes (tectonic, geochemical)
- Why are people hoarding data? How can we change the incentives for data sharing and reuse?
- Confession on the floor: "I just came to the realization that I am a Data Hoarder."
- Strategic planning 15 years out is *really* hard when technology changes so fast.
- Love this answer to "What is your area of expertise:" "It's complicated"
- Curious as to which data bases are used the most across the field of earth science.

Selected "Tweets" from Day Two

- EarthCube panel to early career scientists: "You are the future."
- The science leader panelists are all men, and the education panelists are women. Where are the women leaders and men educators?
- Good points about difficulties of being interdisciplinary when going up for promotion/tenure review. No easy answers.
- Maps don't have error bars. My mind is blown.
- Abductive approaches require reflection, deep listening. . . totally at odds with current academic environments
- EarthCube is just one piece of infrastructure within national and international efforts. How do we coexist?

2013 and 2014 Behavioral indications on sharing data

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

2013 Users of geospatial data	2013 All others	2014 iSamples
5.06	2.99	--
66%	57%	45%
65%	51%	31%
57%	39%	20%

Have you submitted a Data Management Plan as part of funded research (current or past)?

If you did so, did your Data Management Plan include a commitment to make data publicly available?

If so, have you now made your data publicly available? (descriptive data for physical samples)

In addition to issues of cost, time, and skills, there are underlying cultural issues of competition and proprietary views of data.

International Geo Sample Number (IGSN) DOI for physical samples

From an early career geoscientist in 2015...



Source: Jon Stelling, Lehigh University, iSamples Workshop 2015

2015 Joint statement of commitment by science publishers and data facilities

... the vast majority of data submitted along with publications are in formats and forms of storage that makes discovery and reuse difficult or impossible....

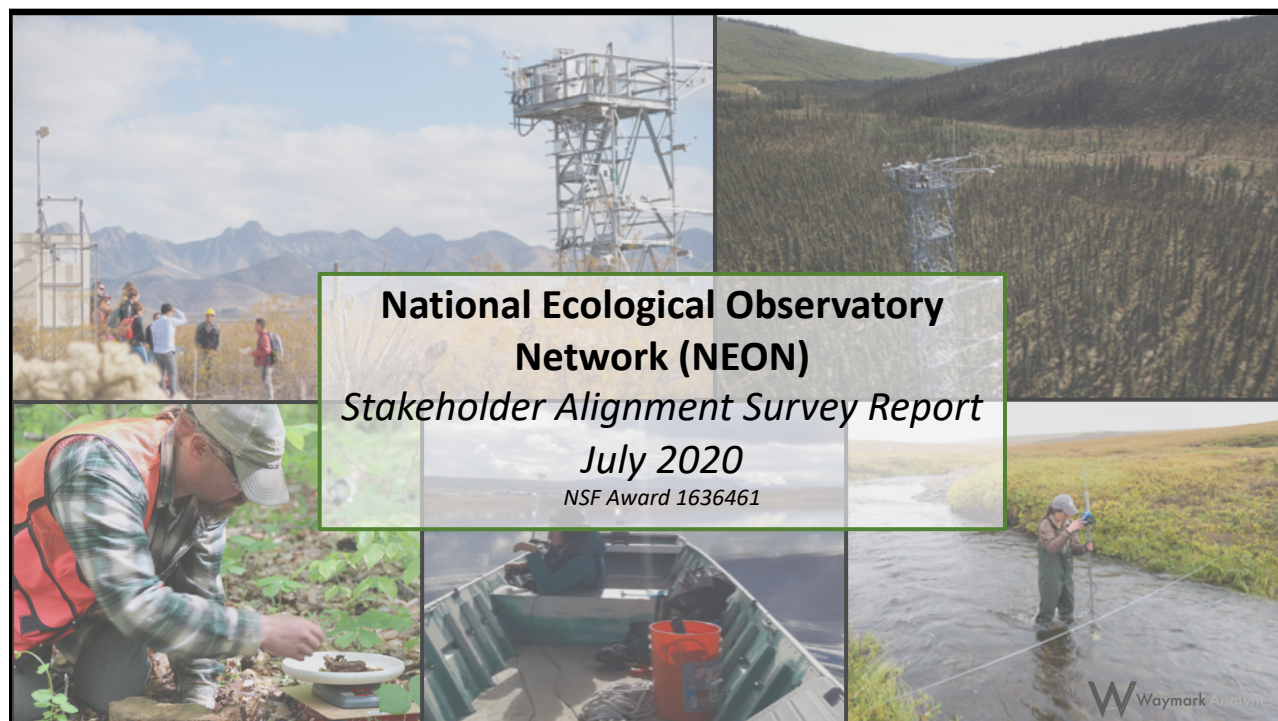
Connecting scholarly publication more firmly with data facilities... has many advantages for science in the 21st century and is essential in meeting the aspirations of open, available, and useful data envisioned in ... funder guidelines....

Source: <http://www.copdess.org/statement-of-commitment/>

American Astronomical Society (AAS)
 American Geophysical Union (AGU)
 American Meteorological Society (AMS)
 Biological and Chemical Oceanography Data Management Office, Woods Hole Oceanographic Institution (BCO-DMO)
 Center for Open Science
 CLIVAR and Carbon Hydrographic Data Office (CCHDO)
 Community Inventory of EarthCube Resources for Geosciences Interoperability (CINERGI)
 Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)
 Continental Scientific Drilling Coordination Office (CSDCO)
 Copernicus Publications
 Council of Data Facilities (CDF)
 Dryad
 Elsevier
 European Geosciences Union (EGU)
 Geochemical Society
 Geological Data Center of Scripps Institution of Oceanography
 ICSU World Data System
 Incorporated Research Institutions for Seismology
 Integrated Earth Data Applications (IEDA)
 John Wiley and Sons
 LacCore: National Lacustrine Core Facility
 Magnetism Information Consortium (MagIC)
 Mineralogical Society of America (MSA)
 Neotoma Paleocology Database
 National Snow and Ice Data Center (NSDC)
 Nature Publishing Group
 OpenTopography
 Paleontological Society
 Proceedings of the National Academy of Sciences
 Rolling Deck to Repository (R2R) Program
 Science
 UNAVCO

2020 Sample EarthCube Tools

<https://www.earthcube.org/tools-inventory>



Meet the respondents, 2020 (n=54)

Relevant Field or Discipline (all that apply)

Atmospheric Sciences	15.4% n=8
Aquatic Sciences (including limnology)	25.0% n=13
Biogeochemistry	40.4% n=21
Biology	40.4% n=21
Chemistry	3.8% n=2
Computer Sciences	5.8% n=3
Critical Zone Sciences	9.6% n=5
Ecological Sciences (including, organisms, populations, and communities)	76.9% n=40
Environmental Sciences	59.6% n=31
Evolutionary Ecology and Biology	23.1% n=12
Geosciences	13.5% n=7
Hydrology	15.4% n=8
Information/Data Sciences	17.3% n=9
Land Cover and Processes	11.5% n=6
Oceanography	3.8% n=2
Science education	11.5% n=6
Soil Science	13.5% n=7
Other	3.8% n=2

Note: For this report “fields and disciplines” are the relevant stakeholder categories.

Primary Role

Natural and/or Physical Scientist	83.6% n=46
Computer and/or Data Scientist	5.5% n=3
Other	10.9 n=6

Relation to NEON

Interested in, aware of, and knowledgeable of the NEON program (engaged)	84.6% n=44
Interested in and aware of the NEON program, but no detailed knowledge of it (keen)	9.6% n=5
Interested in, but not made aware of the program prior to this survey (piqued)	3.8% n=2
I have no interest in NEON (uninterested)	1.9% n=1

Indicator Issues – Importance (10=very important ... 0=not important) (2020)

Being able to discover what data are available overall from NEON. *mean=.91 (sd=.13)*

Interoperability of NEON data with data from external sources (combining multiple data sets/types from multiple sources). *mean=.89 (sd=.14)*

Being able to discover what data are available at a specific NEON site. *mean=.87 (sd=.18)*

Being able to use NEON data to expand the impact of ecological science in society. *mean=.84 (sd=.19)*

Being able to get user support (e.g., with data, documentation, and code) when accessing NEON data. *mean=.81 (sd=.18)*

Being able to use NEON data to transform the way ecological science is conducted. *mean=.81 (sd=.22)*

Being able to get provenance information, tracing the origins and transformations of NEON data. *mean=.80 (sd=.22)*

Having a Digital Object Identifier (DOI) or other globally persistent, unique identifiers attached to NEON data. *mean=.74 (sd=.26)*

Being able to offer ideas for improvement to NEON Cyberinfrastructure. *mean=.72 (sd=.22)*

Being able to get training in the use of NEON data. *mean=.65 (sd=.28)*

Being able to use visualization tools (provided by NEON or outside software tools) with NEON data. *mean=.63 (sd=.29)*

Being able to identify and request access to NEON physical samples (such as soil or invertebrate samples). *mean=.58 (sd=.36)*

Recognition and appreciation from my colleagues (including professional societies) for research that involves the use of NEON data. *mean=.56 (sd=.28)*

Indicator Issues – Difficulty (10=very important ... 0=not important) (2020)

Being able to use NEON data to expand the impact of ecological science in society. *mean=.30 (sd=.24)*

Interoperability of NEON data with data from external sources (combining multiple data sets/types from multiple sources). *mean=.32 (sd=.24)*

Being able to use NEON data to transform the way ecological science is conducted. *mean=.34 (sd=.24)*

Being able to get provenance information, tracing the origins and transformations of NEON data. *mean=.41 (sd=.28)*

Being able to use visualization tools (provided by NEON or outside software tools) with NEON data. *mean=.45 (sd=.22)*

Being able to offer ideas for improvement to NEON Cyberinfrastructure. *mean=.45 (sd=.27)*

Being able to discover what data are available overall from NEON. *mean=.46 (sd=.29)*

Being able to get user support (e.g., with data, documentation, and code) when accessing NEON data. *mean=.46 (sd=.25)*

Being able to identify and request access to NEON physical samples (such as soil or invertebrate samples). *mean=.48 (sd=.25)*

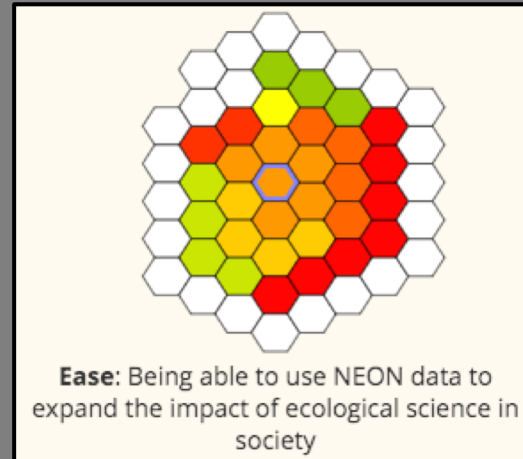
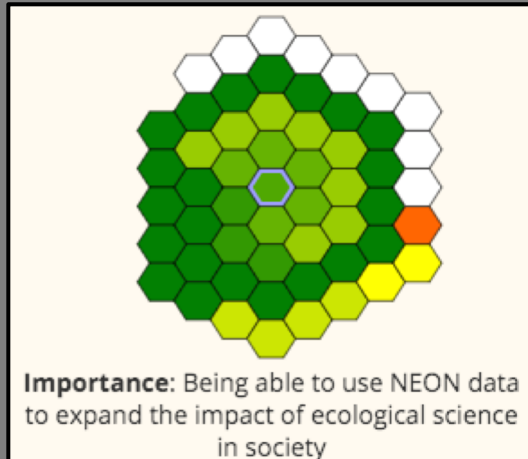
Being able to discover what data are available at a specific NEON site. *mean=.49 (sd=.30)*

Being able to get training in the use of NEON data. *mean=.52 (sd=.25)*

Having a Digital Object Identifier (DOI) or other globally persistent, unique identifiers attached to NEON data. *mean=.56 (sd=.26)*

Recognition and appreciation from my colleagues (including professional societies) for research that involves the use of NEON data. *mean=.59 (sd=.24)*

Being able to use NEON data to expand the impact of ecological science in society.

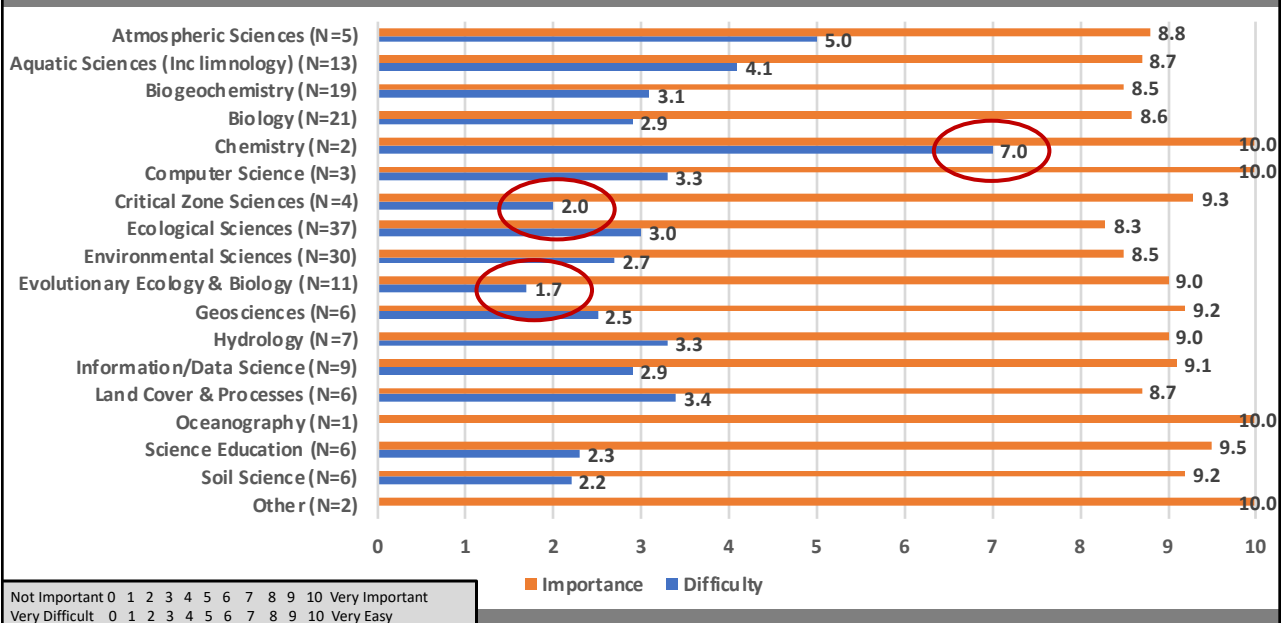


Not important 0 1 2 3 4 5 6 7 8 9 10 Very important

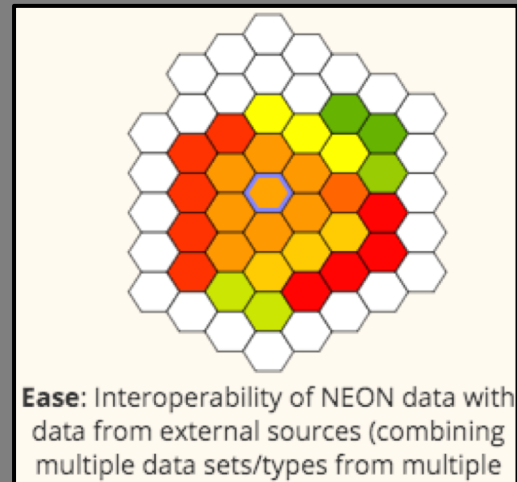
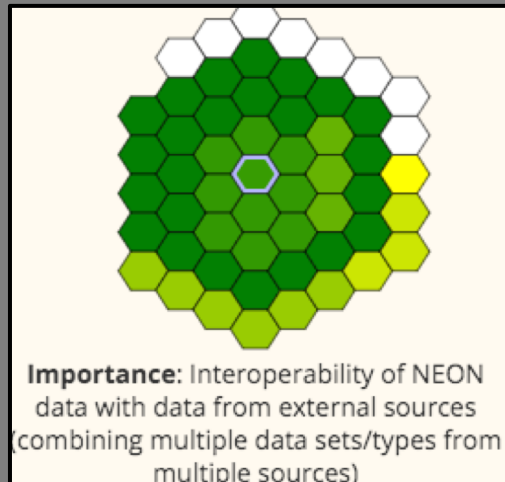
Very Difficult 0 1 2 3 4 5 6 7 8 9 10 Very Easy

27

Being able to use NEON data to expand the impact of ecological science in society.
(cont.)



Interoperability of NEON data with data from external sources (combining multiple data sets/types from multiple sources).



Not important 0 1 2 3 4 5 6 7 8 9 10 Very important

Very Difficult 0 1 2 3 4 5 6 7 8 9 10 Easy

29

Interoperability of NEON data with data from external sources (combining multiple data sets/types from multiple sources). (cont.)



MINORITY SERVING- CYBERINFRASTRUCTURE CONSORTIUM

in partnership with

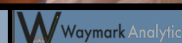


Data and Computing Infrastructure for Historically Black Colleges and Universities (HBCUs), Hispanic Serving Institutions (HSIs), Tribal Colleges and Universities (TCUs), and other Minority Serving Institutions (MSIs)

Stakeholder Alignment Survey Report Summary – January 2021



WayMark Analytics acknowledges the traditional owners of the lands across the North American Continent.
We pay respect to Elders past, present and emerging.



MINORITY-SERVING CYBERINFRASTRUCTURE CONSORTIUM

Meet the respondents, 2020 (n=291)

Institutional Type

HBCU (44 of 101 HBCUs represented)	43% n=125
HIS (63 of 539 HSIs represented)	28% n=82
TCU (32 of 38 TCUs represented)	28% n=83
Other MSI (not in the above categories)	1% n=4

Note: A few HBCUs are also HSIs.

Primary Role

Educator and/or researcher (faculty or staff)	43% n=126
Cyberinfrastructure professional (or professional in development) supporting research computing, facilitation, data curation, etc.	7% n=21
Senior administrative leader (president, CIO, CTO, VPR, Dean, etc.)	42% n=181
Other (please specify):	8% n=22

Note: "Other" includes students, leaders or staff in not-for-profit or private CI or STEM support organizations, representatives of government agencies, and others.

Sample Titles Associated with Primary Roles:

President	Provost
Vice President	Chancellor
CIO	Vice-Chancellor
COO	Dean
CFO	Librarian
Math/Science Division Head	Chair, Arts & Humanities
Chair, Computer Science	Chair, Social Sciences
Director, Institutional Research	Director, Research Center
Director, Library Services	MBA Program Manager
Institutional Data Manager	Chief Security Officer
IT Director	IT Manager
IT Systems Administrator	IT Technician
Network Engineer	Systems Administrator
Agriculture Faculty	Biology Faculty
Computer Science Faculty	Economics Faculty
English Faculty	Engineering Faculty
Geoscience Faculty	Math Faculty
Music Faculty	Regional Planning Faculty
Science Faculty	Social Work Faculty
Information Science Instructor	Social Science Instructor
Environmental Specialist	Public Health Professional

Overall highlights

- **Basic needs:**

Across HBCUs, HSIs, and TCUs there is a deep need for basic infrastructure support, such as broadband WIFI on campus and at home for students, staff, and faculty (heightened by the pandemic).

- **Potential action:** Submit separate and joint proposals for infrastructure investments. *(timing: Immediate)*

- **Consistency across institutions:** Although there are some unique considerations for certain types of institutions (such as data sovereignty with TCUs), the vast majority HBCUs, HSIs, and TCUs have similar responses.

- **Potential action:** Develop and build alignment around a shared vision and joint goals. *(timing: 3-6 months)*

- **Workforce development:**

Students need literacy and advanced skills with data and computing (the educational mission), while faculty and staff need training and support for a robust cyberinfrastructure (the research mission).

- **Potential action:** Share curricula (faculty and students) and develop career paths (cyber professionals). *(timing: 12-24 months and ongoing)*



33

Overall highlights

- **Collaboration:** There is strong support for collaboration across institutions to accomplish together what they can't do separately (with little support for each acting on their own).

- **Potential action:** Expand engagement in MS-CC and other relevant consortia and initiatives. *(Timing: 3-6 months and ongoing)*

- **Institutional operations:** Administrators need a more accessible and responsive data infrastructure for campus operations, surfaced when asked about research data and computing.

- **Potential action:** Foster alliances among campus leaders regarding infrastructure for institutional data. *(Timing: 6-12 months and ongoing)*

- **Societal impact:** There is a strong potential for data and computing to advance research on issues central to community culture and disparities in society, with infrastructure as a constraint on achieving these broader impacts.

- **Potential action:** Identify and track impact measures for investments and impacts for HBCUs, HSIs, TCUs, and other MSIs. *(Timing: 6-12 months and ongoing)*



34

MINORITY-SERVING CYBERINFRASTRUCTURE CONSORTIUM



If access to data and research computing could deliver one thing that you value, "a must have," what would it be? (categories & illustrative examples, 2020)

Connectivity, Equipment, Security, and Access to Technology (n=101, 34%)

- Access for all, on and off campus (HSI)
- Cyber infrastructure and big data enabling technologies (HBCU)
- Best in Class IT Security Team & Systems (HSI)
- Internet access for students in rural communities (TCU)

Data Storage, Data Management, and Data Analytics (n=49, 16%)

- Secure storage of and access to unstructured data in customizable formats (HBCU)
- A data governance structure (HSI)
- Track our first gen/Hispanic students and aggregating their data in a more efficient way – our data tracking is a huge issue on campus – any report gathering becomes a significant event (HSI)
- Dedicated computer resources for data (TCU)

User Features and Resources for teaching and research (n=39, 13%)

- Software and hardware for teaching virtually and in class (HBCU)
- Single sign on (TCU)
- Simulation of real-world problems (Other MSI)
- A platform and dedicated resources for running python notebooks, docker images, and other "platform" independent software packages for basic training (HSI)

Impacts on Students and Society (n=30; 10%)

- At this point, a small HPC environment to allow our students to develop HPC skills to help them with admission to Graduate Schools and compete for Jobs (HBCU)
- High quality data sets for student use (HSI)
- Social data on Native Americans that is equivalent to the data for Blacks and Hispanics (TCU)
- Our students will acquire the skills to use big data (TCU)

35

MINORITY-SERVING CYBERINFRASTRUCTURE CONSORTIUM



If access to data and research computing could deliver one thing that you value, "a must have," what would it be? (categories & illustrative examples, 2020)

Training, Continuing Education, and Access to Expertise (n=28, 9%)

- Mentoring (HBCU)
- Access for acquiring industry certifications (HSI)
- 1 FTE in Institutional Research trained in quantitative AND qualitative research, SPSS, MAXQDA (TCU)

Strategic Planning, Assessment, Administrative Support, and Funding (n=8, 3%)

- Data Management Plan; User Friendly System to enhance data stewardship (HBCU)
- Adequate budget and stakeholder Understanding (HBCU)
- A short term and long term IT Plan (TCU)
- Data-driven decision making (TCU)
- To have faculty and senior administration understand and embrace the importance of having a cyberinfrastructure that protects, expands and supports faculty and campus research (HSI)

Inter-Institutional Collaboration (n=7, 2%)

- A strong alliance with an existing program to assist us in gaining an established program (HBCU)

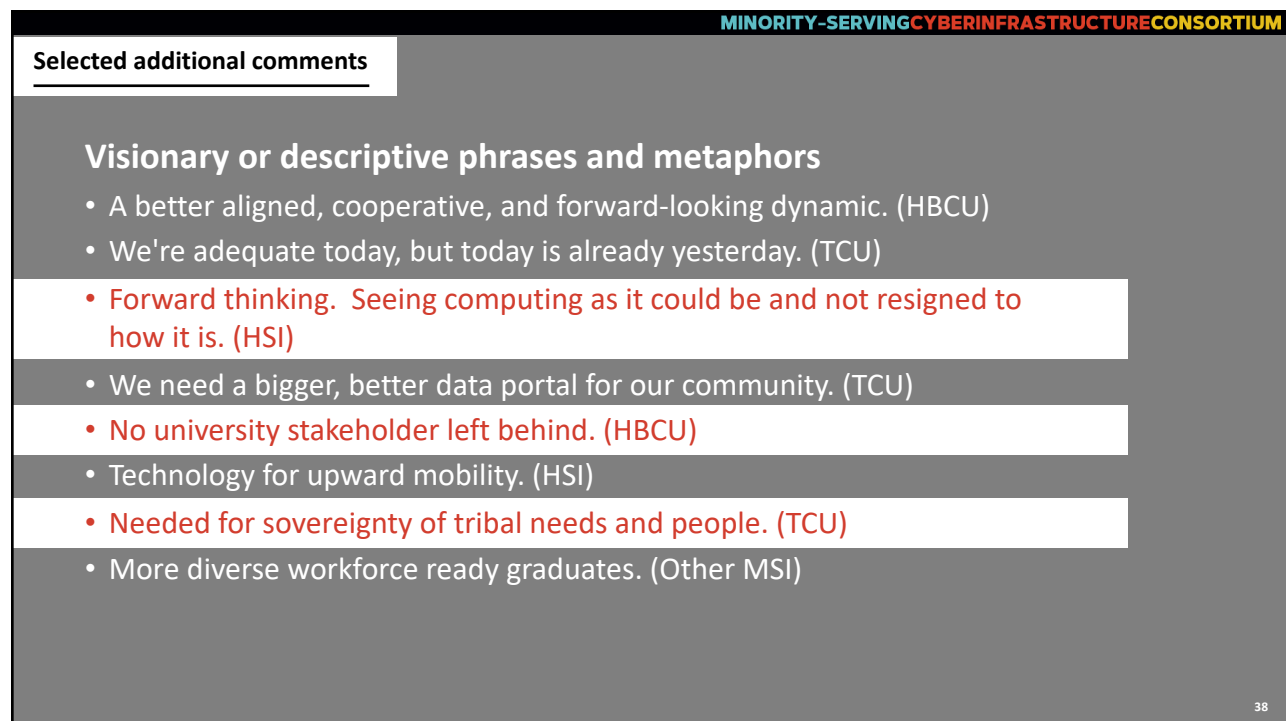
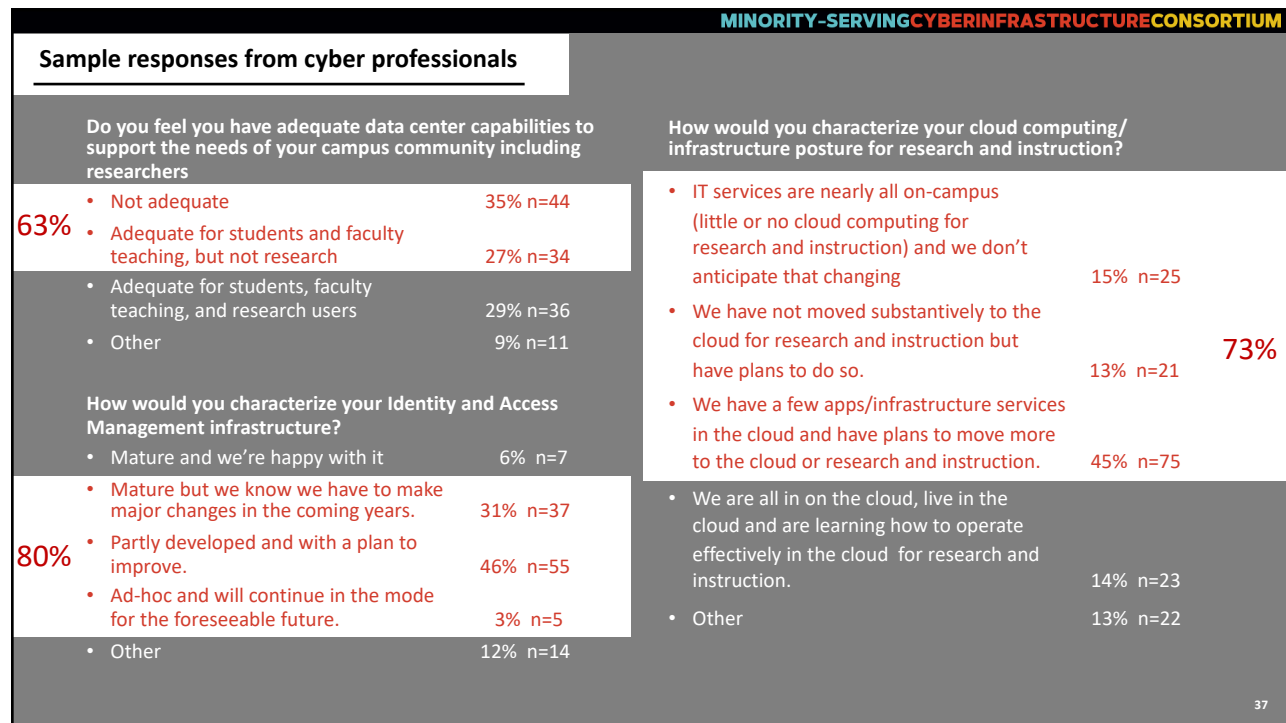
Access to Library Resources (n=7, 2%)

- More seamless integration between the LMS and Publisher-based online systems (BB and Pearson Labs, BB and Cengage, etc.) (HBCU)
- Online high quality LMS system (HSI)

Additional Comments (n=31, 11%)

- A welcoming, inclusive, and growth oriented mindset (Other MSI)
- To see the future, i.e. what faculty and students will need in the future. What technologies will be needed in short and long term. (HBCU)

36





The Campus Research Computing Consortium

2017 top interests (n=255) (*not important=0; very important=1; very difficult=0; very easy=1*)

Rank by importance:

1. **Workforce development** for cyberinfrastructure administrators and staff (mean=.84)
2. **Supporting facilitators** (broadly defined) on campus, bridging between research teams and research computing resources (mean=.84)
3. Research computing **expertise sharing** among universities (mean=.84)

Rank by difficulty:

1. Influencing **state and federal policies** impacting research cyberinfrastructure (mean=.18)
2. Research computing **resource sharing** among universities (mean=.26)
3. Effective models for demonstrating **return on investment (ROI)** in research computing resources (mean=.26)

Gaps between importance and difficulty:

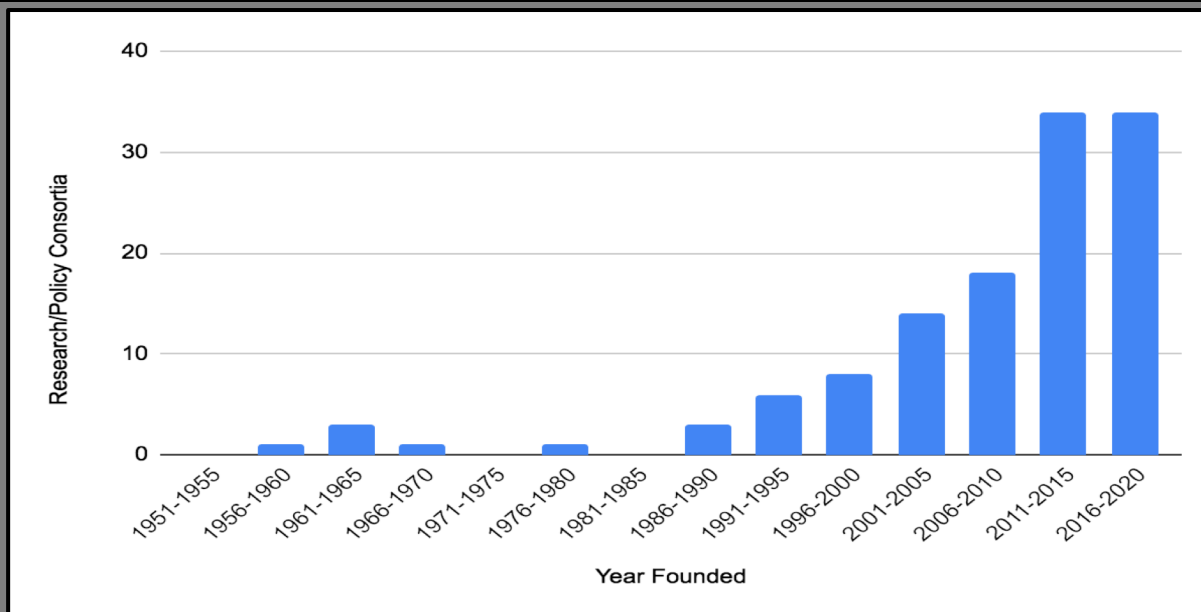
1. Influencing **state and federal policies** impacting research cyberinfrastructure (gap=.59)
2. **Workforce development** for cyberinfrastructure administrators and staff (gap=.56)
3. **Supporting facilitators** on campus, bridging between research teams and research computing resources (gap=.56)

2019 CaRCC ecosystem workshop lightening talks

- Association of Research Libraries (ARL)
- Big Data Hubs
- Campus Champions (CC)
- Campus Research Computing Consortium (CaRCC)
- Carpentries
- Coalition for Academic Scientific Computation (CASC)
- *Coalition for Networked Information (CNI)*
- Education Opportunities (HPC University, SIG HPC Education)
- EDUCAUSE
- EPOC/CI Engineers
- Extreme Science and Engineering Discovery Environment (XSEDE)
- Global Environment for Network Innovations (GENI)
- GPN/Regional Network
- HPC Systems Professionals
- Midscale Experimental Research Infrastructure Forum (MERIF)
- Minority Serving Institutions / Historically Black Colleges & Universities
- Open Science Grid (OSG)
- Research Data Access & Preservation Association (RDAP)
- Women in HPC (WHPC)

Launching and Sustaining Multi-stakeholder Consortia

Multi-Stakeholder Consortia Supporting Research Computing and Open Data: Year Founded 1950-2020

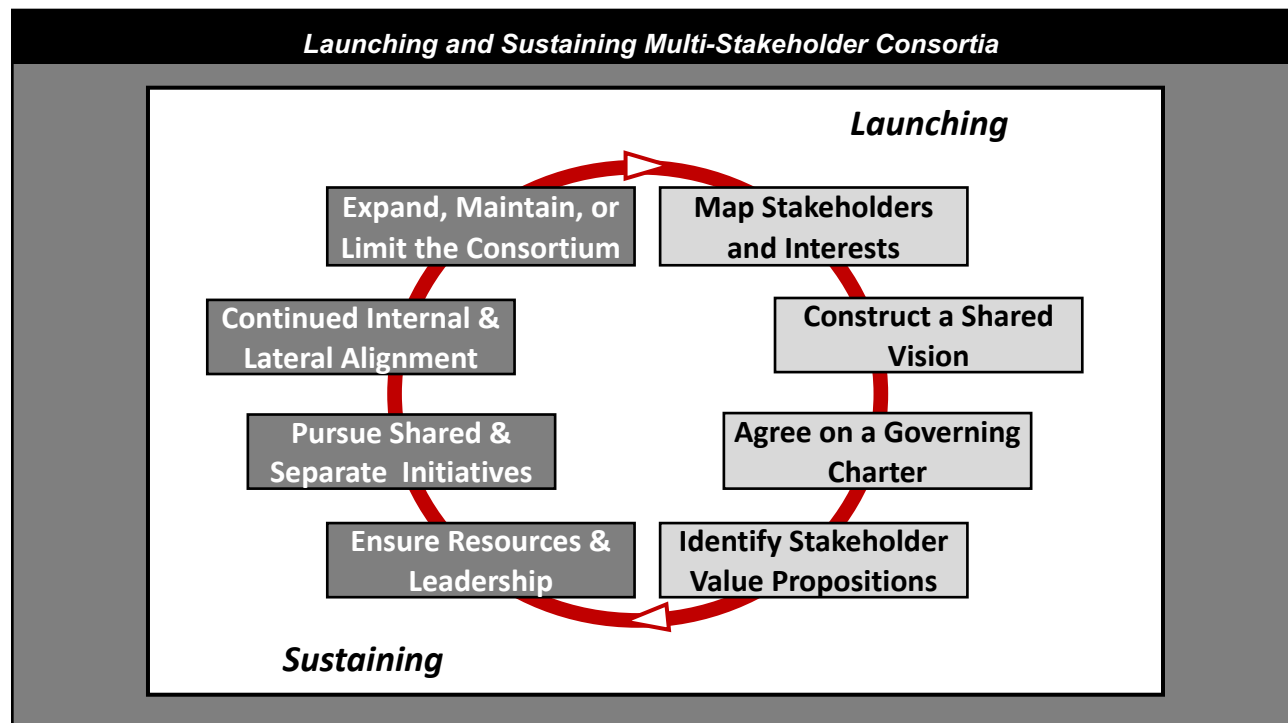


Source: The Stakeholder Alignment Collaborative, Minimum Viable Consortia (working paper)

Sample Research Computing and Open Data Consortia Founded from 1950-1999 and from 2000-2020

Domains	1950-1999	2000-2019
Cyberinfrastructure Consortia	total n=6; Examples: - Coalition for Academic Scientific Computation (CASC) - 1989 - Coalition for Networked Information (CNI) - 1990 - Internet2 - 1997	Total n=20; Examples: - The Quilt - 2000 - DataCite - 2009 - Extreme Science and Engineering Discovery Environment (XSEDE) - 2010 - Campus Research Computing Consortium (CaRCC) - 2017 - Minority Serving Cyberinfrastructure Consortium (MS-CC) - 2018
Earth, Space Science and Environmental Consortia	Total n=4; Examples - International Oceanographic Data and Information Exchange (IODE) - 1961 - Long-Term Ecological Research Network (LTER) - 1980 - Earth Science Information Partners (ESIP) - 1998	Total n=15; Examples: - EarthScope - 2003 - Interdisciplinary Earth Data Alliance (IEDA) - 2010 - EarthCube - 2013 - Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) - 2015
Multi-Domain Data Curation and Sharing Consortia	Total n=3; Examples: - World Data Center - 1957 - Committee on Data of the International Science Council (ISC) (CODATA) - 1966	Total n=16; Examples: - Dataverse - 2007 - Open Science Grid - 2012 - Research Data Alliance - 2013 - Center for Open Science - 2014 - FAIR Data Initiative - 2017
Total, these and other Consortia Studied	Total 1950-1999 = 24	Total 2000-2020 = 97

Source: The Stakeholder Alignment Collaborative, Minimum Viable Consortia (working paper)



MINORITY-SERVING CYBERINFRASTRUCTURE CONSORTIUM

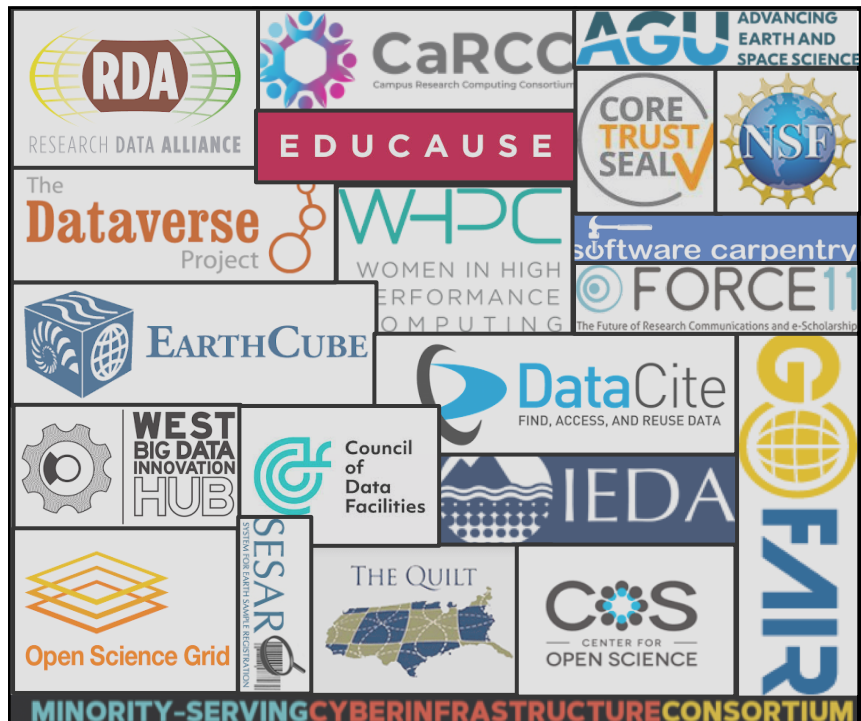
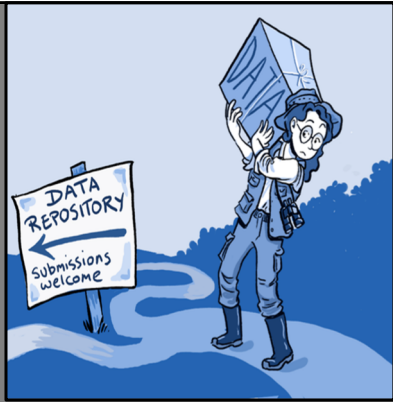
MS-CC Vision Statement

MS-CC envisions a transformational partnership to promote advanced cyberinfrastructure (CI) capabilities on HBCU, HSI, TCU, and MSI campuses, with data; research computing; teaching; curriculum development and implementation; collaboration; and capacity-building connections among institutions. MS-CC will learn and grow as a consortium. We are dedicated to lifting all participating institutions by advancing cyberinfrastructure for research and education across diverse fields, disciplines, and communities. We will engage as full contributors to the global research and education community.

MINORITY-SERVING CYBERINFRASTRUCTURE CONSORTIUM

MS-CC Stakeholder Value Propositions

- **Researchers and Educators:** MS-CC enables researchers and educators to harness the power of data and computing resources to advance the frontiers of knowledge in ways that are aligned with the mission of HBCUs, TCUs, HSIs, and other minority serving colleges and universities – spanning science, engineering, social science, humanities, arts, and other domains.
- **Students:** MS-CC enables students at HBCUs, TCUs, HSIs, and other minority serving colleges and universities to address issues of importance to them with data and computing capabilities, as well as to prepare the next generation workforce – including future cyberinfrastructure professionals.
- **Cyberinfrastructure Professionals:** MS-CC connects cyberinfrastructure professionals across HBCU, TCU, HSI, and other minority serving colleges and universities so we can accomplish together what we can't do separately – including building capability, bringing in funding, establishing career paths, advancing knowledge, and pioneering new technologies.
- **Campus Leaders:** MS-CC helps campus leaders make wise investments in the capabilities needed for a post-industrial, digital world – advancing the mission and impact of HBCUs, TCUs, HSIs, and other minority serving colleges and universities.
- **Industry Partners:** MS-CC enables industry leaders to coordinate engagement with HBCU, TCU, HSI, and other minority serving colleges and universities around new technologies, services, resources, and next-generation talent relevant to research and educational cyberinfrastructure.
- **Foundations and Funding Agencies:** MS-CC enables foundations and funding agencies to coordinate engagement with HBCU, TCU, HSI, and other minority serving colleges and universities around research priorities community development relevant to research and educational cyberinfrastructure.

Three Observations:

1. Multi-stakeholder consortia are central to understanding the Earth as a system.
2. There are considerable barriers to success – social and technical.
3. Next generation geoscientists / cyber professionals / social scientists will shape it – if we support them.

The fate of the Earth itself hangs in the balance.

MINORITY-SERVING CYBERINFRASTRUCTURE CONSORTIUM

Appendix

