The Missing Millions:
Democratizing Computation and Data to Bridge Digital Divides and Increase Access to Science for Underrepresented Communities

Alan Blatecky (RTI),
Damian Clarke (Alabama A&M),
Joel Cutcher-Gershenfeld (Brandeis),
Deborah Dent (Jackson State),
Rebecca Hipp (RTI),
Ana Hunsinger (Internet2),
Al Kuslikas (AIHEC),
Lauren Michael (Wisconsin)

We acknowledge the traditional owners of the lands where Washington, Maryland and Virginia now stand -- the Pamunkey, Chickahominy, Upper Mattaponi, Rappahannock, Monacan, and Nansemond. We pay respect to their Elders past, present and emerging.

Open Science Grid Webinar; March 15, 2022

1

The Missing Millions

For 35 years the NSF, through the Office of Advanced Cyberinfrastructure, has led the world with high-performance computing that has accelerated research and teaching.

Yet there are millions of researchers, students, staff, and citizens who do not have access or the needed capabilities to utilize this infrastructure

Image: https://www.google.com/search?q=Icon+millions+of+people&rlz=1C5GCEM_enUS950US951&sxsrf=ALeKk009OyBwwarUvHvAGuUpadnCR-8DQ:1627752031667&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjlmTfc63yAhWXElkFHd2AEYQ_AUoAXoECAEQAw&biw=838&bih=630#imgrc=8WtnWQZIpq88MM

2

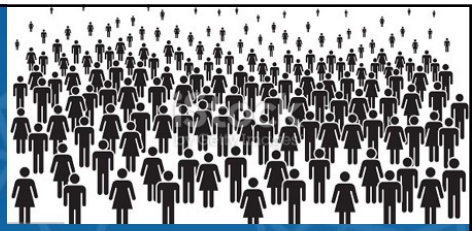
Long-term motivating question

How can NSF significantly expand, diversify, and support the development of new cohorts and communities of scientists and researchers to address pressing research, social, and global issues in 2030?

- While the focus of the study is on computing and data capabilities that support and advance the nation's research, there are broader implications as these capabilities are relevant to virtually all fields and disciplines and are essential for 21st century science and research.
- There is urgency in addressing these matters. We are in a post-industrial, digital era in which the pace of change with technologies is accelerating. While the COVID-19 pandemic has thrown into sharp relief social disparities and societal divides, it is clear that overall social institutions, specific institutional arrangements, and broader engagement with diverse faculty and students in science and engineering, are not keeping pace with technology and the gaps are widening.

3

Elements of a Vision of Success



- Broader Impacts
- Ubiquitous Access
- Innovation at the "Edge"
- Communities
- Data Leadership
- CI Careers
- Expertise at Under-Resourced Institutions
- CI Spectrum
- Collaboration
- Outreach Success Indicators
- Documented Results

Image: https://www.google.com/search?q=icon+millions+of+people&rlz=1C5GCEM_enUS950US951&sxsrf=ALeKk009OyBwwarUvHvAGuUpadnCR-8DQ:1627752031667&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjlmTfb3yAhWXEIkFHd2AEYQ_AUoAXoECAEQAw&biw=838&bih=630#imgrc=8WtnWQZIpq88MM

4

Findings

1. There are substantial barriers to access
2. Accessibility = Access + Ability
3. Racial, gender, and other forms of underrepresentation in data and computing need to be studied, socialized, and addressed
4. Insufficient engagement with underrepresented institutions
5. It is computation, *and* software, *and* data

"I am hopeful that this will create change. Too often, agencies have these conversations and still do business in the same way—with no change. There needs to be a change in who gets funded and who doesn't. Most review panels haven't heard of many of the HBCUs."

"Career paths with data and computing [need to be] clear and widely evident—especially for community colleges and HBCUs."

"Doing broader impacts is work that requires expertise. Example of an NSF panel where everyone was able to assess scientific merits but no one had expertise in broader impacts. There should be a seat on review panels for broader impacts."

5

Findings (cont.)

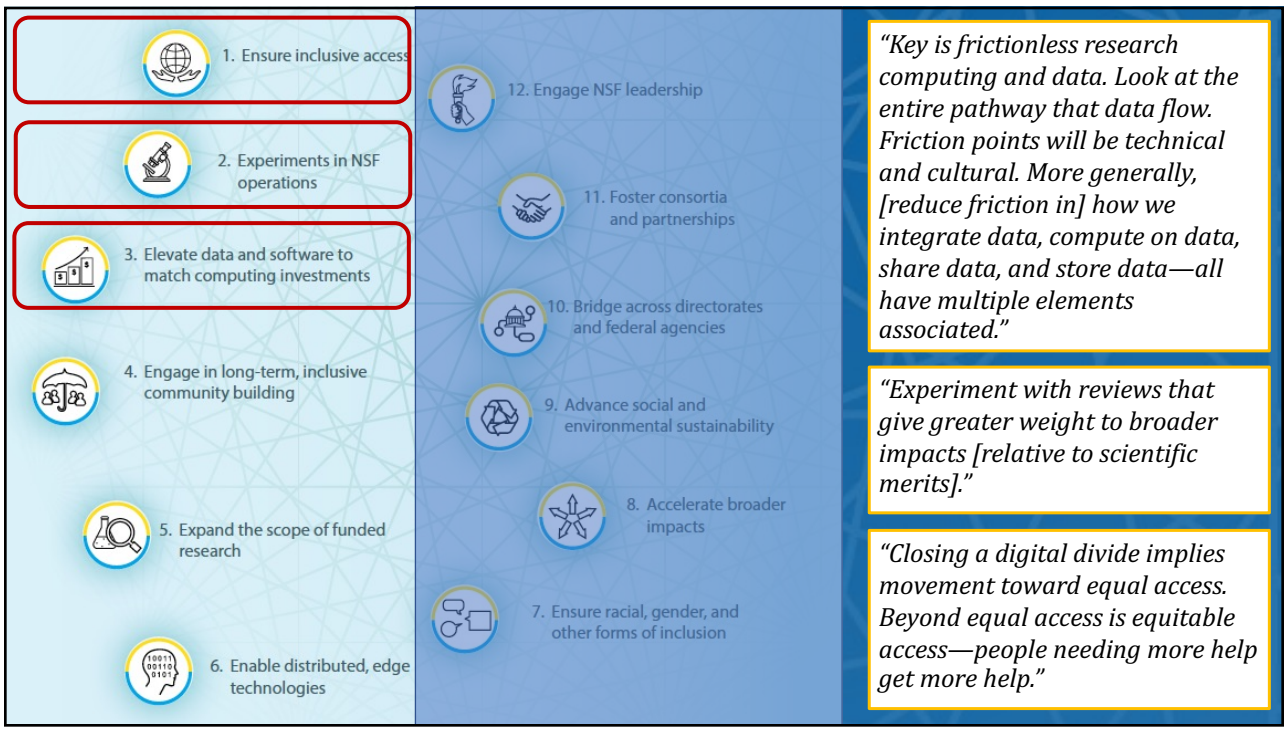
6. More diverse fields and disciplines need support
7. Facilitation successes need for expansion
8. Distributed/edge computing is expanding
9. There is an openness to science-based experimentation in how the NSF operates
10. Systemic change is needed

"I am hopeful that this will create change. Too often, agencies have these conversations and still do business in the same way—with no change. There needs to be a change in who gets funded and who doesn't. Most review panels haven't heard of many of the HBCUs."

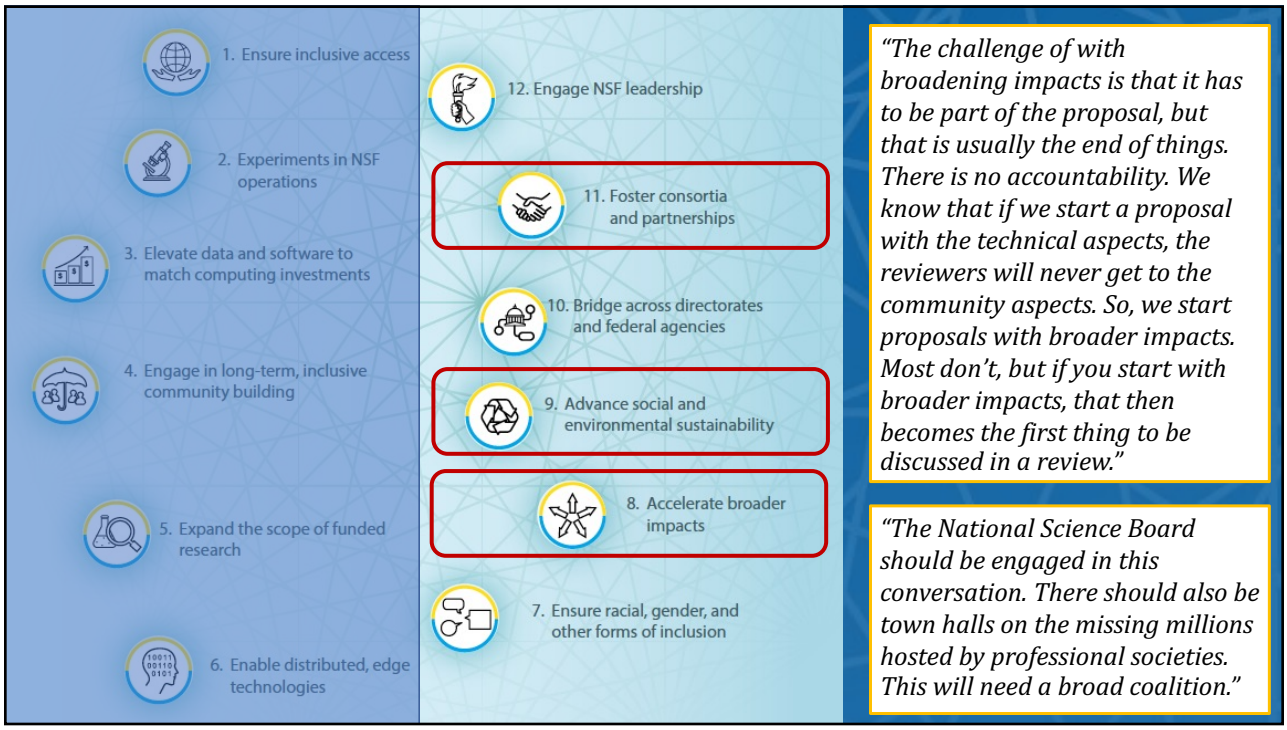
"Career paths with data and computing [need to be] clear and widely evident—especially for community colleges and HBCUs."

"Doing broader impacts is work that requires expertise. Example of an NSF panel where everyone was able to assess scientific merits but no one had expertise in broader impacts. There should be a seat on review panels for broader impacts."

6


1. Ensure inclusive access
2. Experiments in NSF operations
3. Elevate data and software to match computing investments
4. Engage in long-term, inclusive community building
5. Expand the scope of funded research
6. Enable distributed, edge technologies
12. Engage NSF leadership
11. Foster consortia and partnerships
10. Bridge across directorates and federal agencies
9. Advance social and environmental sustainability
8. Accelerate broader impacts
7. Ensure racial, gender, and other forms of inclusion
"Key is frictionless research computing and data. Look at the entire pathway that data flow. Friction points will be technical and cultural. More generally, [reduce friction in] how we integrate data, compute on data, share data, and store data—all have multiple elements associated."
"Experiment with reviews that give greater weight to broader impacts [relative to scientific merits]."
"Closing a digital divide implies movement toward equal access. Beyond equal access is equitable access—people needing more help get more help."

7


1. Ensure inclusive access
2. Experiments in NSF operations
3. Elevate data and software to match computing investments
4. Engage in long-term, inclusive community building
5. Expand the scope of funded research
6. Enable distributed, edge technologies
12. Engage NSF leadership
11. Foster consortia and partnerships
10. Bridge across directorates and federal agencies
9. Advance social and environmental sustainability
8. Accelerate broader impacts
7. Ensure racial, gender, and other forms of inclusion
"The challenge of with broadening impacts is that it has to be part of the proposal, but that is usually the end of things. There is no accountability. We know that if we start a proposal with the technical aspects, the reviewers will never get to the community aspects. So, we start proposals with broader impacts. Most don't, but if you start with broader impacts, that then becomes the first thing to be discussed in a review."
"The National Science Board should be engaged in this conversation. There should also be town halls on the missing millions hosted by professional societies. This will need a broad coalition."

8

Questions or comments?

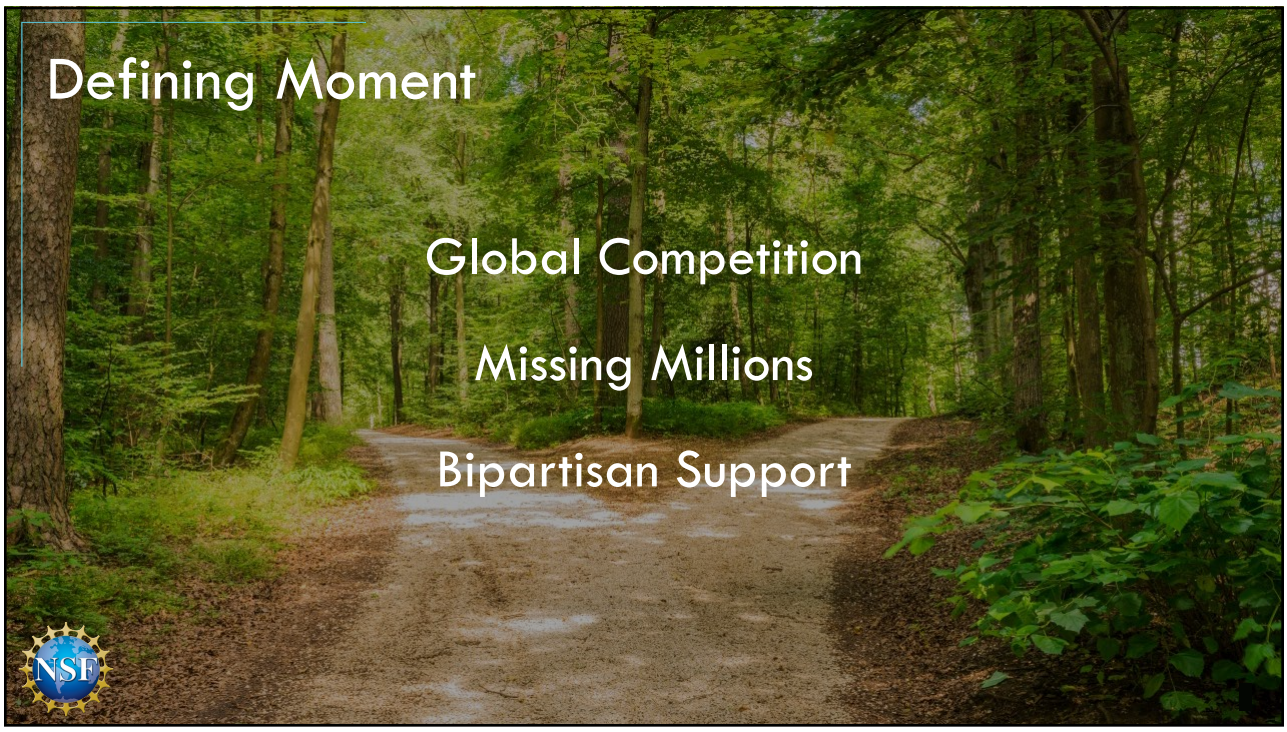
Full Report can be downloaded from
https://www.rti.org/publication/missing-millions/fulltext.pdf

Image Source: https://www.cio.com/article/3296703/how-to-find-and-implement-emerging-technologies-as-a-cio.html

9

Appendix A

10



11

Methods

15 Focus groups, 88 individuals in mixed groups combining:

- **Researchers/scholars and administrative leaders in math, science, social science, humanities, and other domains**
- **Early career scholars**
- **HBCU Leaders, Faculty and CI Professionals**
- **TCU Leaders, Faculty and CI Professionals**
- **University CI professionals**
- **Government and FFRDC CI Professionals**
- **Digital fabrication ecosystems**
- **Industry experts**

6 Individual interviews

12

Interview Protocol

- **Success Vision:** If OAC and the NSF more broadly were fantastically successful over the next decade in democratizing research computing and achieving broader impacts, what would be key elements of that success? Focus less on the specific technologies and more on what the technologies are enabling. Consider "what if" and "why not" statements. Don't limit your thinking to current NSF funding and priorities – think out of the box on this.
- **Specific Technologies:** What new computing and data technologies will be integral to this vision? What technologies don't exist but might be compelling and impactful? What are new or non-traditional research applications that we might consider?
- **Communities of Practice:** How can we build diverse and inclusive communities of practice associated with advanced computing technologies? How can we make more than incremental advances in terms of democratizing computation? What will it take for innovation to be community-driven on a continuing basis?
- **Barriers:** What do you see as organizational, institutional, and societal barriers to the success vision and the technologies? What would be failure modes, deep problems, risk or unanticipated aspects of potential future computing technologies in society that we should worry about?
- **Anything Else:** Are there any other comments and questions that you would like to share?

13

Paths
Access and Experiments

1. **Ensure Inclusive access**
   - 1A Expand investments in apprenticeships, internships and training grants
   - 1B Document and make visible full spectrum of research data and computing resources
   - 1C Expand investments in easy usable software, data and services
   - 1D Ensure investments integrate for frictionless workflow across data, software and compute
   - 1E Explore investments in compute and data that don't require command line expertise
   - 1F Develop entry-point mechanisms for K-12, undergraduates, and graduates to use data

"Key is frictionless research computing and data. Look at the entire pathway that data flow. Friction points will be technical and cultural. More generally, [reduce friction in] how we integrate data, compute on data, share data, and store data—all have multiple elements associated."

Shorter-Term

Longer-Term

14

Paths
Access and Experiments

2. Innovate with experiments in NSF operations

- 2A "Block grants" to HBCU, TCUs, HSIs and other minority-serving institutions and consortia
- 2B Experiment with solicitations and review processes that give greater weight to broader impact
- 2C Increase flexibility and broader leeway to program officers for RAPIDs, EAGERs, RCNs, CC*
- 2D Foster an "investment portfolio mindset" to include more "high-risk, high reward" initiatives aimed at reaching the "missing millions"
- 2E Support mechanisms and integrative adaptation to incorporate lessons learned

"Experiment with reviews that give greater weight to broader impacts [relative to scientific merits]."

Shorter-Term

Longer-Term

15

Paths:
Data/Software and Long-Term

3. Elevate data and software investments to the level of computing investments

- 3A Support broader experiments and impact in usability of services with a focus on reproducibility, interoperability, extensibility and sustainability of these capabilities
- 3B Pioneer innovations in open and consistent sharing and reuse of data and software, especially for the next generation of students, researchers and scientists
- 3C Balance investments in new technology innovations with investments in sustained services
- 3D Ensure that useful and usable data can be found and utilized by anyone, anywhere

"Decisions about what is funded and prioritized should reflect interests in a democracy—not limited to the views of academics at the front of their field. It is not the newest thing—shiny objects—what about sustainability of technologies, usability of technologies?"

Shorter-Term

Longer-Term

16

Paths:
Data/Software and Long-Term

4. Engage in long-term, inclusive community building

- 4A Define community broadly to include underrepresented voices so that their research priorities reflect their interest
- 4B Foster dialogue on the data and computing portfolio that would benefit from a longer-term horizon
- 4C Experiment with longer-term missing million grant formats like early career awards (investing in the person or institution with more flexibility on the research focus)
- 4D Learn from long-term thinking found in communities such as LTER, TCUs, etc.
- 4E Build expertise in participatory research methods

"Focus on research that makes a difference in people's lives."

Shorter-Term

Longer-Term

17

Paths:
Expanded Scope and "Edge" Innovation

5. Expand the scope of funded research in computing and data

- 5A Adopt a "clean sheet" approach to identifying promising new research computing and data investments
- 5B Use social impact language to bring in broader and more diverse proposals
- 5C Expand scope of solicitations to take into account social impacts of compute and data

"Increases in computing power have led to expanded capabilities but no change in how we think about computing—how might we rethink this if we were designing computing today?"

Shorter-Term

Longer-Term

18

Paths:
Expanded Scope and "Edge" Innovation

6. Enable technologies and communities that enhance capabilities at the "edge"

- 6A Invest in CI and community laboratories at the "edge"
- 6B Explore investments in data and compute that are easily accessible (GUI, Apps)
- 6C Build ubiquitous compute and data, widely distributed and accessible

"Closing a digital divide implies movement toward equal access. Beyond equal access is equitable access—people needing more help get more help."

Shorter-Term

Longer-Term

19

Paths:
Racial, Gender, Underrepresentation and Broader Impacts

7 . Foster dialogue, learning, and action on racial, gender, and other forms of underrepresentation

- 7A Experiment with language in initiatives that will draw In underrepresented entities
- 7B Experiment with cross-institutional partnerships to support underrepresented, next-generation scholars along professional development pathways
- 7C Explore investments that will deepen understanding AI)and related technologies in generating *and* ameliorating racial, gender, and other disparities

"It is really important to support people on the journey so they don't quit. It is easy to think 'I don't belong here' if you don't see people like you, and that is when people are more likely to quit. It is not just about getting them in the door but also about giving them support along the way."

Shorter-Term

Longer-Term

20

Paths:
Racial, Gender, Underrepresentation and Broader Impacts

8. Accelerate the broader impacts of CI investments

- 8A Document past NSF awards that have achieved broader impacts with technology adoption
- 8B Provide technical assistance to prospective PIs on effective ways to address broader impacts
- 8C Approach broader impacts with the same level of methodological rigor given to scientific merits

"The challenge of with broadening impacts is that it has to be part of the proposal, but that is usually the end of things. There is no accountability. We know that if we start a proposal with the technical aspects, the reviewers will never get to the community aspects. So, we start proposals with broader impacts. Most don't, but if you start with broader impacts, that then becomes the first thing to be discussed in a review."

Shorter-Term

Longer-Term

21

Paths:
Sustainability and Bridging Across

9 . Advance social and environmental sustainability

- 9A Identify and support sharing of innovations in reducing the carbon footprint of research computing
- 9B Develop long-term sustainability models for community-building initiatives
- 9C Pioneer distributed, environmentally sustainable approaches to research computing

"The research community should take into account the environmental impact of computing—can we be on the cutting edge on the impacts of computing?"

Shorter-Term

Longer-Term

22

Paths:
Sustainability and Bridging Across

10. Bridge across directorates and federal agencies

- 10A Review cross-directorate initiatives to identify implications for solicitations, budgeting, leadership, and other matters relevant to cross-cutting research data and computing g initiatives elevating cyberinfrastructure (social and technical) as a foundational agency-wide resource
- 10B Identify policy enablers associated with interagency collaboration, educating NSF program officers and PIs on what is and is not possible.

"The problems in higher education are very siloed—you have to get rid of the silos. The same is true with government agencies. NSF needs to collaborate more with FAA, NOAA, NASA, NIH, DOD, OSTP—cyberinfrastructure is cross-cutting. Otherwise, you will not have the resources you need—you will only be talking to yourself."

Shorter-Term

Longer-Term

23

Paths:
Consortia and NSF Leadership

11. Foster Consortia and Partnerships

- 11A Map the ecosystem of consortia associated with research computing and data, and then develop mechanisms for sustaining and coordinating among them.
- 11B Identify and support innovations in bridging research computing and data services across fields and disciplines

"There are many consortia and networks: CaRCC, Campus Champions, etc.—but it takes hard work to sustain."

Shorter-Term

Longer-Term

24

Paths:
Consortia and NSF Leadership

12. Engage with NSF Leadership

- 12A Foster dialogue and action with the OAC Advisory Committee, the Computer and Information Science and Engineering Advisory Committee (CISE AC), and NSB

"The National Science Board should be engaged in this conversation. There should also be town halls on the missing millions hosted by professional societies. This will need a broad coalition."

Shorter-Term

Longer-Term

25