



October 3, 2022

NSTC Subcommittee on Equitable Data  
Office of Science and Technology Policy  
Eisenhower Executive Office Building  
1650 Pennsylvania Ave. NW  
Washington, DC 20504

Re: Engagement and Accountability RFI

On behalf of the Center for Data Innovation ([datainnovation.org](http://datainnovation.org)), I am pleased to submit this response to the Office of Science and Technology Policy's (OSTP) request for information (RFI) on the collection and use by federal agencies of equitable data—a term which OSTP uses to refer to “data that allow for rigorous assessment of the extent to which government programs and policies yield consistently fair, just, and impartial treatment of all individuals.”<sup>1</sup>

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, London, and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as important data-related technology trends. The Center is a part of the nonprofit, nonpartisan Information Technology and Innovation Foundation.

In response to the RFI, the Center offers three main recommendations: 1) OSTP should support partnerships that bolster access to high-performance computing of historically underrepresented groups in the field; 2) OSTP should promote robust data literacy curriculums in U.S. schools to ensure more Americans have opportunities to make use of data about themselves and their communities; and 3) OSTP should prioritize closing the “data divide”—the social and economic inequalities that result from a lack of collection or use of data about individuals or communities—to improve the effectiveness of data-driven services and decision making.

Please find our responses to the relevant questions in the document below.

Sincerely,

Gillian Diebold

---

<sup>1</sup> “A Vision for Equitable Data: Recommendations from the Equitable Data Working Group,” White House, April 2022, <https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf>.



Policy Analyst  
Center for Data Innovation  
[gdiebold@datainnovation.org](mailto:gdiebold@datainnovation.org)



#### ***4. What resources, programs, training, or other tools can expand opportunities for historically underrepresented scholars and research institutions to access and use equitable data across levels of government?***

High-performance computing (HPC) is essential for solving many data-intensive computational problems in a wide range of fields, including engineering, biology, and Earth science. Increasing access to HPC for traditionally underrepresented groups in science and engineering will help ensure that researchers from all backgrounds can access and use data equitably. Partnerships that coordinate the sharing of computing resources between the federal government and universities, particularly Minority-Serving Institutions (MSIs) that include Historically Black Colleges and Universities, Hispanic-Serving Institutions (HSIs), and Tribal Colleges and Universities (TCUs) will create a direct connection with traditionally underrepresented communities and newly collected data as well.<sup>2</sup>

Such partnerships existed from 1997 to 2004 with the support of the National Science Foundation (NSF) through the Education, Outreach, and Training Partnership for Advanced Computation Infrastructure (EOT-PACI), an effort that included dozens of institutions and organizations to coordinate the sharing of computing resources with MSIs.<sup>3</sup> More recently, the NSF Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science (INCLUDES) initiative works to enhance the participation of underserved communities in scientific research.<sup>4</sup> The INCLUDES program has provided more than \$7 million in funding for the Computing Alliance of Hispanic-Serving Institutions but that is the only current initiative to advance minority participation in computing.<sup>5</sup> Similar efforts need to target HBCUs and TCUs, and reestablish grants that fund HPC resources at MSIs. Industry-university partnerships, such as the partnership between the University of Florida and NVIDIA, can also expand access to AI computing capabilities.<sup>6</sup>

#### ***6. What resources, programs, training, or tools can make equitable data more accessible and useable for members of the public?***

One important step to improving the accessibility and useability of equitable data is improving data literacy for Americans of all backgrounds. Many organizations and individuals face barriers to using data,

---

<sup>2</sup> Hodan Omaar, "How the United States Can Increase Access to Supercomputing," (Center for Data Innovation, December 2020), <https://datainnovation.org/2020/12/how-the-united-states-can-increase-access-to-supercomputing/>.

<sup>3</sup> "NSF EPIC Press Release," last modified April 2005, <http://mvhs.shodor.org/epic/pressrelease.html>.

<sup>4</sup> "About Us: NSF INCLUDES," accessed November 6, 2020, <https://www.includesnetwork.org/new-a/about-us>.

<sup>5</sup> "NSF INCLUDES Alliance: Computing Alliance of Hispanic-Serving Institutions," accessed November 6, 2020, [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=1834620&Histo](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1834620&Histo).

<sup>6</sup> Hodan Omaar, "Industry-University Partnerships to Create AI Universities" (Center for Data Innovation, July 2022), <https://www2.datainnovation.org/2022-ai-universities.pdf>.



including a lack of technical skills.<sup>7</sup> Improving data literacy would not enable more Americans to develop in-demand job skills to participate in the growing data economy, but it would also allow them to better understand public data and decisions based on that data.

At the local, state, and federal levels, data literacy can be bolstered by improving educational programs in data science and related disciplines, beginning in K-12 schools.<sup>8</sup> Strong programs in math, statistics, and computer science can equip students with the skills necessary for advanced data literacy. This training should continue through higher education, where degrees in technical fields also can provide the highly skilled workforce needed to participate in the data economy. Programs that reduce disparities in data literacy will also reduce the digital use divide, or the gap between those that use technologies in ways that augment and transform their daily lives and those that use technology passively. These divides go hand-in-hand and increasing data literacy will necessarily increase digital literacy.

**7. In which agencies, programs, regions, or communities are there unmet needs, broken processes, or problems related to participation and accountability that could be remedied through stronger collaborations and transparency around equitable data?**

The data divide refers to the social and economic inequalities that result from a lack of collection or use of data about individuals or communities.<sup>9</sup> Many individuals and communities across the United States lack the data necessary to benefit from data-driven innovation. The data divide has emerged between the data-haves and the data have-nots, and these inequities can significantly impact how individuals and communities participate in the data economy.<sup>10</sup> Opportunities to benefit from data vary based on different demographic and geographic factors. Whether or not data-driven services and data-driven decision-making works for someone often depends on where they live and their background. Historically underrepresented communities typically face the widest data gaps resulting from this lack of equitable data collection.

Instances of inequitable data systems pose the biggest threat to participation in the data economy as they can encompass communities of all sizes, characteristics, and locations. These systems collect and store data needed for key services, such as education, financial services, public health, and more. But, disparities occur when these systems have weak or nonexistent infrastructure. Moreover, some system data might exist but have steep barriers to access, like data siloes. For example, in education, schools cannot take advantage of data-driven technologies when they lack the systems necessary to collect and

---

<sup>7</sup> Gillian Diebold, "Closing the Data Divide for a More Equitable U.S. Digital Economy," (Center for Data Innovation, August 2022), <https://datainnovation.org/2022/08/closing-the-data-divide-for-a-more-equitable-u-s-digital-economy/>.

<sup>8</sup> Daniel Castro, Joshua New, John Wu, "The Best States for Data Innovation," (Center for Data Innovation, July 2017), <https://www2.datainnovation.org/2017-best-states-data.pdf>.

<sup>9</sup> Gillian Diebold, "Closing the Data Divide for a More Equitable U.S. Digital Economy," (Center for Data Innovation, August 2022), <https://datainnovation.org/2022/08/closing-the-data-divide-for-a-more-equitable-u-s-digital-economy/>.

<sup>10</sup> Ibid.



utilize high-quality data. As a result, students, families, and administrators are forced to make decisions about enrollment or interventions based on incomplete or inaccurate data. Some states have Statewide Longitudinal Data Systems (SLDSs) that store data on education from early childhood to the workforce (P-20W), but other states still lack this type of specialized system. This leaves data siloed and often inaccessible to key stakeholders. While some students and families have access to data about their entire educational journey, others lack this type of information, meaning they must make consequential decisions with unequal knowledge. As of 2017, only 17 states and the District of Columbia had fully linked SLDS systems.<sup>11</sup>

A similar situation arises with financial services. Credit scores determine everything from whether someone qualifies for a mortgage or car loan, but also whether they can obtain certain services without a deposit, or even apply for a lease or a job. Credit agencies often lack the necessary data infrastructure to collect and score individuals based on “alternative” data sources, such as rent or utility payments, cell phone bills, or cash-flow data in a bank account. As a result, many consumers have limited or no credit history.<sup>12</sup> As of 2020, 21 percent of U.S. adults do not have a credit card and 19 million are considered “unscorable” due to insufficient or outdated data.<sup>13</sup> Agencies governing financial services need to expand the use of alternative credit to include more consumers in the financial system.<sup>14</sup>

---

<sup>11</sup> “50-State Comparison: Statewide Longitudinal Data Systems,” Education Commission of the States, last modified December 14, 2021, <https://www.ecs.org/state-longitudinal-data-systems/>.

<sup>12</sup> “Data Point: Credit Invisibles,” Consumer Financial Protection Bureau, May 2015, [https://files.consumerfinance.gov/f/201505\\_cfpb\\_data-point-credit-invisibles.pdf](https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf).

<sup>13</sup> Ibid.

<sup>14</sup> Gillian Diebold, “Congress Should Expand Use of Alternative Credit Data,” *Center for Data Innovation*, March 8, 2022, <https://datainnovation.org/2022/03/congress-should-expand-use-of-alternative-credit-data/>.