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

Re: Request for Information; Equitable Data Engagement and Accountability (87 FR 54269)

Dear Members of the NSTC Subcommittee on Equitable Data,

Thank you for the opportunity to comment on the development of infrastructures and resources for public engagement with equitable and open data. In this comment, we express that meaningful community engagement with equitable data is a necessary condition for the public to hold the government accountable with and for open data. We first address why access and accountability go hand in hand, then we move to answer directly question 6 regarding useability and accessibility and then question 5 regarding accountability. The comments expressed here are my own, developed with Aishani Aatresh, a Fellow of the Harvard Kennedy School's Program on Science, Technology and Society. The comments express here are not necessarily endorsed, nor meant to imply endorsement, by the Harvard Kennedy School, Argonne National Laboratory, or the Department of Energy.

Civic knowledge and public engagement are at an all-time low, according to a 2016 Annenberg Public Policy Center report.[1] We believe that public data, especially those released by local, state, and federal governments, presents a major and unprecedented opportunity for increasing the public's access to civic engagement. Data can be used to identify harms faced by citizens and their communities, understand trends and progressive realization of development and socio-economic goals, and offer a new participatory perspective into government at various levels.

Accessibility and useability are necessary conditions for accountability because accountability is a condition for democratic discourse, at a minimum.[2] Because "*to hold government accountable to the American public*" demands a citizen providing account of the government's actions vis-a-vis data, it also requires citizens to *be able to* develop an account conceptually and practically. This can only be through the kinds of awareness and familiarity that mechanisms of accessibility and usability provide. Otherwise, citizens cannot see how to participate in accountability discourse and make it be responsive their interests—"it is like playing a game whose rules do not make any sense to one."[3] This relationship between access and accountability is currently under development in the European Union. A recent proposal in the EU aims to increase access to civil liability law in artificial intelligence (AI) and data systems cases because the current bar for admission "may make it difficult or prohibitively expensive for victims to identify the liable person and prove the requirements for a successful liability claim" due to specific characteristics of AI and data systems.[4] Given the limited access citizens currently face, usability and understanding of AI and data systems are prohibitively complex. Accountability suffers as the knowledge-gap between citizens, the government, and technology developers is extremely wide in terms of the technical information available, even if the democratic sensibilities of working towards building representative and usable technologies are not limited to those of experts.

---

[1] Sarah Shaprio and Catherine Brown, "The State of Civics Education," *Center for American Progress* (blog), February 21, 2018, https://www.americanprogress.org/article/state-civics-education/.

[2] Michael W Dowdle, "Public Accountability: Conceptual, Historical and Epistemic Mappings," *Regulatory Theory: Foundations and Applications*, 2017, 197–215.

[3] Thomas Christiano, *The Constitution of Equality: Democratic Authority and Its Limits* (Oxford ; New York: Oxford University Press, 2008), 62.

[4] Procedure 2022/0303/COD COM (2022) 496: Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). https://ec.europa.eu/info/sites/default/files/1_1_197605_prop_dir_ai_en.pdf

*RFI Question 5. What resources, programs, training, or tools can make equitable data more accessible and useable for members of the public?*

We address this question with three main points. In summary, accessibility requires a *bi-directional relationship* between the government and community organizations and citizens using the data; this relationship must to provide the necessary equipment (physical and conceptual) for developing an understanding of their usage, and for citizens to co-develop new avenues of usage; and, useability and accessibility require explicit understanding and avenues for institutional responsiveness to the kinds of civic engagement citizens undertake with the data.

**First, it is imperative that the resources, programs, and tools developed to increase data's accessibility and usability are understood as a relationship between the government and the citizens and community organizations which aim to use these tools.** It is a collective effort undertaken by the government and citizens where understanding is a moving target, not a one-off deposit of some tools without responsive avenues for feedback and development. Technology and data policy are steeped a tradition which follows "a paternalistic, top-down, technocratic approach where the context, values, and agendas of datafication are obscured from the citizen."[5] It communicates to citizens the technocratic, and undemocratic, idea that "modern life is too complicated to be managed by ordinary people" and "only those with specialist knowledge and skills can rise to the task."[6] A unidirectional relationship reduces "citizens and civil society…to passive but demanding 'users' to be served by the public sector" rather than partners and owners of this democratic project.[7] Therefore, access must be developed in conjugation with community stake-holders and citizens, especially if it is for fostering government accountability (question 5).

As citizens and community groups face increasing access to data tools, their conceptual understanding grows. When communities become proficient in accessing and using data tools, the horizon of use cases and activities which can be undertaken with the data also grows. This means that access for who, access for what, and usability for what purpose is a constant moving target. Therefore, the ways in which citizens want to access and use the data moves and if the government cannot be responsive to these shifts, then the government risks alienating citizens from civic engagement further. Quintessential to the ideal of democracy is that decision making "tracks the interests and the ideas of those citizens whom it affects."[8] It is through government responsivity to considered public opinion and through the government fostering the access to the tools for the formation of considered public option, that the complex and dynamic democratic process of civic engagement can function.

**Second, equitable data useability requires developing infrastructures for conceptual understanding, computing devices and public computing resources, and community-building for collective understanding and engagement.** Usability is the condition of data or code conceptually understandable by the intended audience, physically accessible, and collectively endorsed. A great challenge in fostering citizen participation in equitable data releases is the development of the conceptual tools for understanding *what are the kinds of questions and concerns one can address through data? What are the methods, tools, and resources a required to answer those questions? And who can help and guide citizens through this process?* Without addressing these questions, datasets will simply not be useable.

---

[5] Heather Broomfield and Lisa Reutter, "In Search of the Citizen in the Datafication of Public Administration," *Big Data & Society* 9, no. 1 (January 1, 2022): 20539517221089304, https://doi.org/10.1177/20539517221089302.
[6] Sheila Jasanoff, *The Ethics of Invention: Technology and the Human Future*, First edition, The Norton Global Ethics Series (New York: W.W. Norton & Company, 2016), 19.
[7] Broomfield and Reutter, "In Search of the Citizen in the Datafication of Public Administration."
[8] Philip Pettit, *Republicanism: A Theory of Freedom and Government*, Oxford Political Theory (Oxford : New York: Clarendon Press ; Oxford University Press, 1997), 184.

For example, many local governments release data sets[9]; however, there are far and few public spaces where one can interact with someone who can share how to interpret data. Traditionally, public libraries and librarians were resources for civic engagement.[10] While traditional high school civics education taught students the conceptual tools for conducting research in these public spaces, no programs currently exist which educate students on the open-data equivalent. But without many libraires prepared to handle data civics due to lacking data-trained staff, lacking computing resources, and lacking data-minded educational programs, education in schools alone is necessary, but not sufficient. In my experience teaching undergraduate computer science at the University of Chicago, students when invited to use public data to answer questions about their local communities were beyond excited and passionate. Because they were in a data-centric course and had access to computing resources, they were empowered to use their own city's data to not only verify of their own suspicions but are able to share this knowledge with their community.

Finally, we have made calls on the scientific community to develop public computing resources as demonstration, education, and activist sites where initiative can be met with the resources needed.[11] Participation relies "on different forms of public demonstration to legitimate their particular epistemologies."[12] Public computing and demonstration are further commitments to developing new modes of algorithmic and data justice, and proposals have been made in a similar vein.[13] It is necessary that resources and tools provide both the conceptual and physical means for using data systems.

**Third, resources and tools for access and use of equitable data systems should address communities and collectives rather than individuals.** While empowering individuals is an important endeavor, a particular focus should be placed on privileging communities and collective understanding. In this comment, I am mainly addressing uses of government data for the public to gain an understanding of their communities and governance for civic engagement. While democratic engagement can occur at the individual level, true empowerment for underprivileged communities comes through joining the forces of the community through data, building community through data efforts, and leveraging community partners to amplify their voices. The incorporation of algorithmic and data knowledge into civil life opens new modes of collective resistance against oppression. Collective resistance can "enact not only a making-visible of algorithms, but a making-affective of their role in society, providing voices through which their uneven impact can be felt by others."[14]

We provide two examples. The first example regarding the Census Bureau's differential privacy transition is a *negative* of failed public consultation, where accessibility and usability to the public suffered due to lack of a bidirectional communication channel between the public and the data sources. The second example of the COVID-19 Tracking Project is a positive one that demonstrates how data can itself be a source of bidirectionality and create modes of community involvement that counters pressing challenges of our time.

The Census Bureau is tasked with proving a count of everyone in the United States, among others such as protecting the privacy of individuals in their collection. The move to differential privacy

---

[9] Beth Simone Noveck, "Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency," *Yale Hum. Rts. & Dev. LJ* 19 (2017): 1.

[10] Rashad Young, "More than Just Books: The Role of Public Libraries in Building Community and Promoting Civic Engagement," *National Civic Review* 101, no. 4 (2012): 30–33; Rachel Scott, "The Role of Public Libraries in Community Building," *Public Library Quarterly* 30, no. 3 (2011): 191–227.

[11] Austin Clyde, "AI for Science and Global Citizens," *Patterns* 3, no. 2 (February 11, 2022): 100446, https://doi.org/10.1016/j.patter.2022.100446.

[12] Sheila Jasanoff, "Restoring Reason: Causal Narratives and Political Culture," in *Organizational Encounters with Risk*, ed. B. Hutter and M. Power (Cambridge University Press, 2005), 228, https://books.google.com/books?id=wIfAU5exEPQC.

[13] Maayan Perel and Niva Elkin-Koren, "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement," *Fla. L. Rev.* 69 (2017): 181.

[14] Garfield Benjamin, "Algorithmic Imaginaries and Political Resistance," in *ACM Conference on Fairness, Accountability, and Transparency 2022*, 2022.

for the 2020 data release meant the application of a new mathematical framework which "provable" protects privacy; however, as a result, it introduces targeted amounts of statistical noise into database queries which introduced inconsistencies when looking at aggregate and local data. Previously, noise was also added to the data but in a simplified procedural way outlined in the data releases. The way the noise is added with differential privacy now introduces some unknowability, leading to controversy by researchers and the public—even becoming center in a court case.[15] The sudden introduction of this change lead to controversy and the "'statistical imaginary' of the data— that the data is truly just a count of people without any inference involved.[16] Researchers now are "unable to evaluate the quality of the data" as the "introduced technical and procedural uncertainties" are a black-box to them and they were unprepared for the change. This example illustrates a failure of communication, access, and useability due to the lack of public consultation, preparation of the community through engagement and listening, and developing and fostering the epistemic tools to work with the shift.

The COVID-19 Tracking Project was a volunteer initiative launched by employees at *The Atlantic* in an attempt to better compile, interpret, and disseminate information about the pandemic in the United States, particularly given limited information and great uncertainty in the early stages of the outbreak about testing and cases. The Project was also one of the first to focus explicitly on racial data in the pandemic, in addition to the impact on long-term care facilities, bolstered by its entirely volunteer base of hundreds of data-collectors, software developers, writers, designers, and more dedicated to make more legible the impact of COVID-19 — not only in relation to individuals' daily lives, but to enable community efforts to hold leaders to account to *care* for citizens, to shine a light on problems that tend to be shrouded in darkness when looked upon from above, and to offer alternative avenues through which a problem as large and unwieldy as a pandemic could be tackled. It was in bringing together people from across communities to create yet another form of a collective that The COVID-19 Tracking Project was able to reimagine what responsible and usable data might look like as a means to hold power to account; federal data over time improved partially in response to these efforts such that the emergent, urgent need for the project waned over time.[17]

*RFI Question 5. What resources, programs, training, or tools can increase opportunities for community-based organizations to use equitable data to hold government accountable to the American public?*

Withstanding the theoretical comments from question 6, that accountability requires access, I offer a concrete institutional activity—local government hackathons—that can be supported by OSTP to hold the government accountable by the public through equitable data. The idea of local government hackathons is to host community-sponsored events in public institutions such as libraries and schools where data scientists, librarians, and members of civil organizations educate the community on the tools and resources for using data, share the kinds of questions members can bring to data and computing, and offer assistance and mentorship to citizens desiring to prove cases. Not only does this activity support communities, but it represents a bi-directional relationship between expertise and communities where experts offer their knowledge to the questions and demands of citizens, and where citizens can be informed of the basic horizon to which they can have questions and demands of.

There is great emancipatory power in data, and in particular the publicity and demonstration aspect of data is extremely exciting. Demonstration is a core way the public accept and come to

---

[15] https://www.washingtonpost.com/dc-md-va/2021/09/09/alabama-drops-census-lawsuit-privacy/
[16] boyd, danah, & Sarathy, J. (2022). Differential Perspectives: Epistemic Disconnects Surrounding the U.S. Census Bureau's Use of Differential Privacy. Harvard Data Science Review, (Special Issue 2).
[17] The COVID Tracking Project. "About Us." Accessed October 3, 2022. https://covidtracking.com/about.

understand claims. This is particularly true in scientific communities and demonstrated through Professor Sheila Jasanoff's comparative study of the relationship between science and politics across different political communities.[18] Computing offers new avenues for demonstration because data projects can be reproduced, shared, and shown to many parties. While a few community members may suspect that their water quality is less than that of a wealthier neighborhood, statistical analysis which demonstrates this claim in a reproducible way is very strong evidence—evidence which can even be used to meet requirements for engaging in litigation activates which otherwise may have been prohibitively expensive or inaccessible.

To offer a concrete example how computing and data can create narratives demonstrating oppression, we can look to Amazon's attempt at creating a resume screener.[19] Utilizing data from its hiring practices department, Amazon created an algorithm which learned to screen resumes versus paying employees to do the task. A year later, the engineers reportedly noticed something troubling about their engine — it did not accept women's resumes. They examined the algorithm in the following way after a year. They looked at counterfactual cases where resumes would include signifiers common to women, without making gender explicit. For example, the inclusion of Barnard College, an all-women's liberal art college, would cause the algorithm to reject an application that otherwise would be highly desirable. Unlike a bureaucratic equivalent, a human-staffed HR office, the discrimination against women was demonstrable across time, locations, and parties. In fact, the use of data here more explanatory to those facing oppression than a traditional office would be.[20] The algorithm and data exposed itself clearly, showing its own bias printed on a screen through the right research question. Contrast this with people, who work quite hard to hide their implicit biases.

We hope you found these comments helpful and pointed. Many of the points expressed here are active areas of my research, and we are keenly working towards developing local community data hackathons in conjugation with non-profit organizations such as SEED AI. Please do not hesitate to follow up on any points that are not clear or for further information. We look forward to reading new policies and initiatives that come out of this request for information.

Austin Clyde
Assistant Computational Scientist
Argonne National Laboratory
aclyde@anl.gov

with, Aishani Aatresh
aishaniaatresh@college.harvard.edu

---

[18] Sheila Jasanoff, *Designs on Nature: Science and Democracy in Europe and the United States* (Princeton, N.J: Princeton University Press, 2007).

[19] Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women," in *Ethics of Data and Analytics* (Auerbach Publications, 2018), 296–99.

[20] Rob Reich, Mehran Sahami, and Jeremy M. Weinstein, *System Error: Where Big Tech Went Wrong and How We Can Reboot*, First edition (New York, NY: Harper, an imprint of HarperCollinsPublishers, 2021), 109.