RFI Response: Equitable Data Engagement and Accountability

To: The White House Office of Science and Technology Policy (OSTP), on behalf of the Subcommittee on Equitable Data of the National Science and Technology Council

Submitted by:
Shannon Dosemagen, Director, Open Environmental Data Project (shannon@openenvironmentaldata.org)
Katie Hoeberling, Director of Policy Initiatives, Open Environmental Data Project
Emelia Williams, Research and Policy Associate, Open Environmental Data Project

As urgency grows around climate change and other environmental crises, so too does the need for environmental datasets to be opened up to support equitable decision-making and accountability. Open data policies such as the OPEN Government Data Act enacted in 2019 and efforts like the Open Government Partnership have set a foundation for data transparency and availability in the US. Still, they often fall short of making data equitable, accessible, and usable. Incentive structures, standards and privacy, and rules and regulations related to environmental data are failing communities collecting and using data, as well as government stakeholders who could benefit from this data in their policy and decision-making structures.

Inclusive, interactive data systems can become a foundation for increasing environmental equity in communities, especially those dealing with historical or ongoing environmental injustices. Socializing government and the public to the value of community data and the concept of environmental data as a public good will be critical to building equitable and effective data systems. Open dialogue and participatory infrastructure development can help identify the political push points and necessary changes in the social behavior of Federal agency staff and researchers, to ensure adoption and use of inclusive data systems.

The Open Environmental Data Project (OEDP) builds spaces to grow the global conversation on environmental data access and use to increase the flow of usable information between communities, researchers, lawmakers, and enactors seeking to actively address environmental and climate injustices that impact the quality of life and health of our communities. Informed by our research and programming, we offer recommendations for Federal agencies to better support collaboration with other levels of government, civil society, and the research community around the production and use of equitable data. Responses to each of the RFI's questions are provided under their headings below.

Q1-2: What are examples of successful collaborations involving equitable data between the Federal government and (a) Tribal, territorial, local, and State governments, or (b) local communities? Among examples of existing Federal collaborations with (a) Tribal, territorial, local, and State governments or (b) local communities involving equitable data, what lessons or best practices have been learned from such collaborations?

In our work, we have encountered several successful collaborations involving equitable environmental data between the Federal government and other governance levels. ***Two significant examples include the Environmental Information Exchange Network and the Internet of Water.***

The Environmental Information Exchange Network (EIEN) demonstrates the social, technical, and capacity-building elements required to keep data management and sharing initiatives effective and accountable to those involved. The EIEN facilitates data sharing among the Environmental Protection Agency (EPA), states, Tribes, and territories through shared, reusable, and streamlined data collection and exchange services. This work is done largely through partnerships with the Exchange Network (EN)

and the Tribal Exchange Network Group (TXG), the maintenance of the Central Data Exchange and topic data exchanges, and grant funding for eligible states, federally-recognized tribes, and US territories.

The EIEN leverages its partnerships to focus on continuous maintenance of data infrastructure, enabling it to update its tools in response to local partner needs. The TXG comprises tribal professionals working on various aspects of the EN whose primary purpose is to ensure tribal representation and participation, promote tribal sovereignty, facilitate access to resources and tools, and communicate and collaborate with the EPA to increase understanding of unique tribal cultural values and environmental concerns. The EN assists 51 state nodes, 14 Tribes, and 3 territories in sharing data, using standardized data structures, and offering automation services. The purposeful and deliberate collaboration across governance levels allows the EIEN to support local data initiatives with technical and non-technical support, decrease data sharing errors, and cultivate a group of interested environmental professionals who use and support this data and work.

The EIEN also distributes capacity-building grants, supporting local and state projects in creating, piloting, and maintaining environmental data systems designed by those localities. For example, MassBays builds tools to support community-based monitoring and decision-making. EIEN also provides support and peer-to-peer partnership and data-sharing options throughout and after the duration of the grant cycle. This granting and support structure can serve as a model for other data collaboratives and foundations for designing capacity-building grant programs.

The Internet of Water's (IoW) work is another example of equitable data collaboration between a coalition of organizations working with Federal, state, and local government partners to build foundational water data infrastructure across the US, and create a community using water data to inform decisions. With funding support from the Infrastructure Investment and Jobs Act, the EPA will grant $15 million to projects aimed at more easily sharing information on water quality (including IoW), water infrastructure needs and water technology between state and local agencies.

The IoW's Online Resource Library and Service Provider Directory is regularly updated, providing up-to-date and high-quality training, resources, and referrals for community scientists. One standout feature is the Water Data Assessment Tool, which helps data producers improve their data's discoverability, accessibility, and usability. Other sectors of environmental data could learn from this model, in that providing actionable online and easy-to-use tools allow for greater participation and efficacy at local levels.

IoW also creates hub structures that can be customized and designed to fit diverse needs, allowing developers to compare and contrast hub types and their advantages, limitations, and barriers. Importantly, their data governance practices are based on user needs and limitations, focusing on creating standardized but customizable metadata catalogs. Good examples of IoW hubs include New Mexico Water Data and the Western States Water Council. Best practices that other sectors could replicate include IoW's clear representation of the standard technical management system with customizable models depending on the collaborative's needs and priorities.

IoW also cultivates a peer-to-peer network to connect members from across the nation. To support this network, IoW provides an online directory, regular webinars, and technical matchmaking where community scientists and small watershed monitoring programs are matched with technical service providers. By providing active participation points in network activities, IoW enhances credibility and utilization of tools that integrate community science into decision making, and helps to spread adoption and uptake of best-recommended technologies and monitoring practices.

Q3. What resources, programs, training, or other tools can facilitate increased data sharing between different levels of government (Tribal, territorial, local, State, or Federal) related to equitable data?

Coalition building, statewide data policies, and training programs that build up capacity for effective use of data can facilitate increased equitable data sharing between different levels of government. Coalition building is key to creating sustainable data-sharing programs. Coalitions of Federal agencies and programs can raise awareness, and direct spending towards improved bi-directional data flows between different levels of government and communities. The Federal Equitable Data Working Group could design a model that values community data during various decision-making processes and works for agency QA/QC needs.

Additionally, Federal agencies and states have a major role to play in managing and sharing data both with public agencies at various levels and those outside of government. Master Data Sharing Agreements, collaboratively designed with other agencies, states, and communities, can set data management standards and establish processes for safely sharing fine-grained and de-identified datasets with the public. Relatedly, Federal and statewide data policies can drive a culture of sharing by creating a presumption of openness among agencies, and by providing guidance for agencies to share data openly in usable formats and publish data dictionaries and useful metadata. States (as well as cities, counties, and universities) can also look to models such as the Western Pennsylvania Data Center, which leverages digital and legal infrastructure to support sharing between public agencies and community engagement with government data.

To increase data sharing between different levels of government, departments and agencies at every scope of government need access to relevant data literacy training to expand their capacity. Examples include programs like Georgetown's Data Labs, which offer long-term training on open and responsible data governance. In addition, governments can employ data intermediaries or ombudspeople who can liaise between communities and agencies, help users find and access data, and advocate for system improvements. At Federal and state levels, Chief Data Officers (CDOs) and agency data officers with backgrounds and expertise in library science and curation, rather than solely data analysis or IT, can be hired to improve government capacity to provide more accessible and usable data. Such officers would ensure that data practices meet standards set in Master Sharing Agreements and data policies, and establish collaborative data governance practices. They could also offer guidance and technical assistance to agency staff regarding data quality, curation, and sharing.

Q4. 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?

Communities most affected by environmental harms and climate change tend to be those already facing the burdens of poverty, structural racism, and disadvantages in academic advancement. They also have large stakes in using data for equitable environmental governance and decision-making. Making data more findable, available, understandable, and usable can support the inclusion of these groups in the scientific enterprise. A common theme in OEDP's conversations with community partners is an awareness of an abundance of data but uncertainty around where to look (see our Opportunity Brief on "Environmental Data as a Public Good"). Increased access and understanding of environmental data can spark interest in scientific research and lower barriers for scholars from frontline communities to pursue scientific careers.

Access to data collected and maintained by Federal agencies such as the Census Bureau and NASA has historically been restricted. Often there are good reasons for this—in the case of Census data, significant

risks to respondent privacy currently outweigh the benefits of providing open access to the data. However, policies that require researchers to pay access fees or travel to data centers effectively restrict data's use to well-resourced scholars and research institutions, especially since centers are often near "top" universities (Tranchero & Nagaraj, 2021).

Lifting or easing these restrictions can expand opportunities for a more diverse group of scholars to access government datasets and contribute scientific knowledge. A 2020 study on NASA's Landsat data, for example, demonstrated that lifting restrictions and reducing the costs of accessing the data led to significant increases in associated, highly cited publications by scientists from "lower-ranked institutions," as well as those earlier in their careers. And if in situ access is deemed necessary for protecting privacy, funding to support the establishment of more data centers would enable scholars with fewer financial resources or those unable to travel to use the data.

Publication metrics, in conjunction with author demographics, for outputs produced using federally funded data can be used to measure changes in who is accessing and using such data. Agencies providing the data might also collect demographic data from users who download their available datasets to assess who is using them.

Federal funding agencies like the National Science Foundation should also strengthen and enforce open data policies within the confines of consent-based protection. While the NSF, which maintains a large environmental and climate research portfolio, encourages data sharing and actively funds the development of data infrastructure, many of its datasets are never uploaded to public repositories. An enforcement mechanism lies in tying future award decisions to prior data sharing for those who have previously received NSF support. A precedent for this lies in NSF's proposal requirement that investigators report on the results of prior NSF awards. The National Institutes of Health's new data sharing policy, which will go into full effect in 2023, can be used as a model for other agencies like the NSF. ***These agencies can further support researchers in understanding and using this data by requiring data uploaders or stewards to include standardized and detailed metadata.***

Q5 & 7. 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? 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?

Government-led and funded research as a whole has not holistically represented the experiences of communities facing the worst impacts of environmental degradation and climate change—even when data are collected in a participatory manner, limited funding and avenues for follow-through restrict the usefulness of findings to the communities in closest proximity to the data sources. Furthermore, privatization of the instruments and digital platforms used for collecting or processing data, as well as inefficiencies in Federal data infrastructure, prevents many communities from collecting, accessing, or using data. These barriers threaten to undermine progress in the work of environmental justice, and community and citizen science, as well as erode trust in public institutions, and isolate participating communities, many of whom already bear the brunt of climate change impacts and environmental degradation.

Public agencies should prioritize open data infrastructure that meets the needs of community-based organizations and researchers, as well as government staff. Government and industry datasets could be powerful tools for environmental action at many levels. However, environmental monitoring and compliance data collected by agencies like the EPA are fragmented, stored in different data systems, managed by different entities, and shared in different ways. To ensure broad

findability, access, and use of such data, data infrastructure must be designed with multiple uses and stakeholders in mind, and to accommodate and promote open practices.

Programs such as OEDP's Beyond Compliance Network (in partnership with Fair Tech Collective and Intertidal Strategies) aim to modernize environmental data systems and democratize knowledge creation and use by investigating and re-thinking approaches to the management and sharing of environmental compliance data. The obstacles faced in this process include a common reliance on personal relationships to find or understand data, gaps in metadata and data dictionaries, and large variations in data's scale and granularity that make it difficult to integrate (see our recent report for more discussion on these challenges).

Agencies can also look to tools such as NEPAccess and Fair Tech Collective's benzene emissions database as models. The latter consolidates data reported by oil refineries to the EPA to help communities understand their exposure and to support research on benzene's environmental health impacts. Equitable data infrastructure should be measured by how efficiently it handles multi-directional flows of information while (i) serving multiple types of users with widely varying expertise, including concerned constituents and communities, environmental regulators, and private sector actors capable of reducing pollution; (ii) incorporating quantitative and qualitative data from diverse sources and temporalities; and (iii) establishing performance on existing environmental metrics while fostering the development of new metrics that are meaningful, and actionable, for diverse constituencies and geographies.

Public offices and agencies should support the collection and integration of community data in decision-making through funding, signaling of its value, and focusing on the infrastructure necessary to ensure environmental data is a public good. Communities and community-based organizations are creating new data streams using non-standardized sensors and processes. In addition to connecting science with people's lived experiences and empowering them to connect with and explore their environment, these datasets offer hyper-local evidence that is already informing the decisions of individuals and communities. Volunteer water monitors, for instance, have been taking water pollution samples for decades, contributing localized data to help communities, scientists, and governments monitor short- and long-term changes in water quality.

Evolving Federal data infrastructures must prioritize community data to build more equitable approaches to governance that encourage and utilize community input, and strengthen confidence in environmental decisions. Such integration would improve community representation, strengthen data's relevance to communities, and allow for a variety of potential uses by different actors. It would also expand opportunities for community engagement in science and strengthen feelings of civic trust, and efficacy and trust felt by communities often excluded from predominant scientific processes.

Critically, agencies must not stop at supporting the collection of community data and follow through by acting on the information it offers. In recent years, frontline communities have been able to access Federal and state funding to install environmental monitors at nearby polluting facilities. Too often, however, public agencies fail to hold these facilities accountable even when data show they have violated environmental regulations, further eroding trust among already skeptical communities.

Agencies should build and resource community and organizational capacity to participate in planning processes. Many communities and community organizations may be unaware of funding opportunities or lack the resources to access such support. The complexities of navigating agency processes, strategizing, and preparing proposals present barriers to access for less-resourced organizations and communities—often those most impacted by climate change and pollution. Agencies can reduce these barriers by funneling resources to capacity-building programs.

For example, California State Bill 1072 established Regional Climate Collaboratives to serve as local hubs where communities could come together, strategize, develop partnerships, and apply for state funding. It also led to creation of the Partners Advancing Climate Equity (PACE) program to provide training, mentorship, and technical assistance to frontline community leaders. Depending on communities' existing capacity and access to resources, these programs may need to focus on building a range of skills, from digital literacy to proposal writing and budgeting. Where they exist and are useful, open data portals can be leveraged as sources of baseline information and evidence (or lack thereof) to be cited in proposals. Data ombudspersons and intermediaries can support these capacity-building programs by helping organizations find and use data to support their ideas.

Q6. What resources, programs, training, or tools can make equitable data more accessible and useable for members of the public?

To make equitable environmental data accessible and usable for the public, institutions, agencies, and organizations must prioritize and instrumentalize environmental data as a public good. Two ways to do this are to implement data literacy training and localized data governance models.

Data literacy training must be issue-based and place-tailored to effectively lower barriers to data access and interaction. Communities often require access to specific information about particular aspects of their environment to make a case for local governments or grantmaking organizations. Data literacy is foundational for communities to be able to uncover useful data, and for government agencies to be able to integrate evidence into policy making. Programs catered to specific geographies and content areas will be most effective in assisting communities in finding answers to questions and creating evidence for policy making.

Localized and equitable data governance models, including collaborative governance structures, can make data more accessible, usable, and reusable for the public. New technology for data collection, storage, and management does not always need to be built from scratch; investments can be made in critical digital infrastructure and features that will make environmental data usable and useful (Dosemagen & Williams, 2022). Community ownership, direct input routes, and checks and balances can foster stronger community representation in the data. Appropriate models of collaborative governance will consider how new community data fits into current data systems while also being proactively designed to anticipate future needs.

Data collaboratives present a promising opportunity for civil society to conscientiously build and maintain spaces for environmental and climate data that intersect many issues and are shared by different contributors. More and more localities are imagining spaces where governments and communities can share and manage data collaboratively. Open data governance models should be co-created with communities, building intentional space to foster shared metadata curation practices. These processes can bridge gaps in knowledge, and ensure that collaboratives embed community-derived values from the beginning and the bottom-up. Critically, these governance models require sustained support from a team of diverse data stewards, including sociocultural, legal, policy, and technical roles.