

Commonplace

Collaborative climate futures? Envisioning the role of open data infrastructures for collaborative socio-environmental research

Luis Felipe R. Murillo¹, Shannon Dosemagen²

¹University of Notre Dame, ²Open Environmental Data Project

Published on: Apr 28, 2022

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Environmental research has been at the forefront of the debate concerning the present and the future of planetary health. As the question of climate change and crisis looms large and the struggles around environmental injustice become more visible, both private and public climate research funders and institutions are emphasizing the urgent need for knowledge cross-pollination and co-production. This is happening not only across technoscientific projects but across communities that are often included as "support personnel," but not *bona fide* research partners with voice and decision-making power. Geophysical characterizations of climate change are not enough to understand and respond to the challenges that vulnerable communities face worldwide. It is in this context that we seek to articulate a future in which technologists, researchers, and community members have the basic conditions to create common data infrastructures to enable collaborative environmental research. In this article, we discuss the main motivation for creating a stack for socio-environmental research based on the need for infrastructures that can be understood and governed by the communities on whose land climate research is conducted. In the conclusion, we revisit open questions of community stewardship as they relate to pressing challenges of data sovereignty based on open technologies for climate research.

Introduction

Environmental research has been at the forefront of the debate concerning the present and the future of planetary health. As the question of climate change and crisis looms large and the struggles around environmental injustice become more visible, both private and public climate research funders and institutions are increasingly emphasizing the urgent need for knowledge cross-pollination and co-production. This is happening not only in technoscientific projects but in communities that are often included as "support personnel," yet not considered *bona fide* research partners with voice and decision-making power ([Kawerak et al. 2020](#); [Nanda and Mohamed 2021](#); [Ellam Yua et al. 2022](#)).

Community data stewardship has become of increasing interest in the socio-environmental space, highlighting the need for robust measures and safeguards for data sovereignty. In a practical sense, community data stewardship means having data in the hands of community organizations that can lead the design, production, governance, and distribution of environmental knowledge for their own purposes.

As practitioners with decades of experience in community science and open technologies, we contend there is a need for a different approach to what counts as "infrastructure" for socio-environmental research. After years of studying and experiencing walls of inter-incomprehension across domains of knowledge, we understand that the difficulties in realization of open and relational infrastructures primarily have to do with the (mis)recognition of sociotechnical aspects that support scientific practices. It has become clear in the public debate that geophysical characterizations of climate change are not enough for understanding and responding to the challenges that vulnerable communities face worldwide. It is in this context that we seek to articulate a potential future in which community members as technologists and researchers have the conditions to devise common infrastructures for collaborative environmental research.

What, and whose, infrastructure for open climate stewardship?

The public debate on digital infrastructures has become increasingly concentrated, for good reasons, on the question of power imbalance. What has been identified in Europe with the acronym "GAFAM" (Google, Apple, Facebook, Amazon, and Microsoft) is now perceived worldwide as the source of the problem of large-scale data collection, mining, and brokerage, where an unprecedented degree of infrastructural monopoly and opacity (at once, institutional, infrastructural, technical) has been reached ([Burrell 2016](#); [Annany and Crawford 2018](#); [Zuboff 2019](#)). It is in this domain that community-oriented climate research proposals appear as "moonshots." If big government institutions and universities have now turned to GAFAM's "cloud-based" services, relinquishing control over their own infrastructures, and collaborative environmental data efforts are being organized and funded by monolithic philanthropic initiatives, how can we seriously entertain the idea of designing, implementing, and sustaining our own digital tools, infrastructures, and data services for community-driven action research?

The answer is, of course, far from straight-forward. Yet one thing has become very clear in face of the urgency of the issue of community data stewardship: it is no longer possible to exercise any form of meaningful autonomy when constant data is being mined on users of GAFAM services. For this reason, our work concerns primarily the importance of action research infrastructures that attend, first and foremost, to the needs of most affected communities by environmental injustice and the climate crisis.

In order to realize this vision of common data infrastructures, the first step is to reimagine what infrastructures are and what they can be with an expanded understanding of the responsibility that our proposal entails. “Infrastructures,” we suggest, based on key contributions from Science and Technology Studies, are better understood as the human and non-human support work that, once combined, provide the foundations for the work of climate research and activism that is to be conducted ([Star and Ruhleder 1996](#); [Star 1999](#); [Star and Bowker 2010](#)). In this relational sense, infrastructures are specific assemblages of humans and technical objects that are mobilized to make collaborative research possible. Applied to our conservational context, environmental data needs to be produced with clearer attention to the contextual conditions of its production. It should, in other words, be produced at a micro and regional scale to address the needs of affected communities. It is in this expanded, relational sense of what counts as “data” (i.e. numeric data from environmental sensors, ethnographic data in the form of oral histories) that open data technologies are to be understood as well in their infrastructural capabilities.

If we adopt a relational understanding of “data infrastructures,” we can place ourselves at a vantage to better situate the problem of open climate data stewardship: no longer as a problem to be addressed solely by cadres of software and data engineers, but by a distributed collective of community organizations that share common tools and protocols for advancing climate activism (organizations such as indigenous organizations, youth-led groups, and activist networks that have regional and global expression). This relational understanding is particularly important for disentangling the most difficult issues we confront today with respect to collaborative work around open technologies and projects (such as open scientific software, hardware, and data). These issues include:

1. The lack of meaningful, collaborative ties among affected communities, researchers, and government representatives;
2. The challenges of socializing newcomers—researchers and communities alike— with the intricate tools and protocols of environmental data management, analysis, and stewardship;
3. And the dearth of acknowledgment of and appreciation for different forms of knowledge that cannot be reduced to numerical data formats (such as qualitative data in various forms: oral histories, narratives of natural-social change, community-driven cartography and collective memory archives, among many others).

As much as any other infrastructural affordance, open technologies that make up the ensemble (often identified among technologists as a "stack") of open climate research tools are relational objects: they depend on a set of sociotechnical conditions to become meaningfully "open," that is, they are partial objects in the partial view of everyone and everything—technologists, policy makers, activists, environmental science students, and artificial agents such as automated data management processes—that is involved in climate crisis mitigation.

The question of openness, therefore, speaks directly to the issue of connection of partial views, but also, and more fundamentally, to the question of trust which is slowly cultivated: *learning to give more than to take* is key when approaching any collective for collaborative work. It is extremely difficult to pursue meaningful bonds without socializing the knowledge of tools and infrastructures in a way that does not alienate those who need them for community purposes on a much more urgent timeline and at a different level of priority than a scientific project. Once we start disentangling the multiple threads of expertise, community work, and open technologies for climate research, collaborative work on data stewardship starts looking more feasible through the multiple activities that any open data management plan needs to perform, which include:

1. Documenting data provenance and acquisition tools and techniques with contextual information that is respectful of collaborative protocols and mindful of power imbalances;
2. Selecting a set of existing infrastructural tools that can be used and improved to support data workflows. This ranges from the open formats and protocols for data storage to the interfaces that we will need to query, analyze and visualize data, but also coordinate access with proper stewardship concerns since data (as relational object) is to be understood as a collective production;
3. And situating sociotechnically and institutionally our procedures for protecting and preserving the data, according to the needs of a specific group to whose land (and ecological urgency) the data may belong to.

All of these dimensions of (open) data management must be conceived for the purposes of community-based stewardship if we are serious about creating alternatives to the infrastructures that are impossible to govern because we are not allowed to understand what goes on inside the gated, smart-badge controlled walls of the R&D sections of GAFAM.

Examples include, but are not limited to, the creation of protocols that respect community self-determination by fostering community-based data stewardship, such as The First Nations Principles of OCAP® ("Ownership, Control, Access, and Possession") and CARE principles ("Collective benefit, Authority to control, Responsibility, Ethics") ([FIGC 2018](#); [Caroll et al 2020](#); [2021](#)); creating open science hardware-based instruments for bottom-up approaches to community science in the context of the "Gathering for Open Science Hardware" (GOSH) network or community science spaces such as Public Lab ([Arancio and Dosemagen 2022](#); [Rey-Mazon et al. 2018](#)); using and disseminating open tools to support environmental data pipelines and workflows ([Goodman et al. 2014](#); [Gentemann et al. 2021](#); [Gabrys et al. 2016](#)); and supporting community data centers in autonomous spaces, such as Casa Tainã in Brazil which hosts its own server infrastructure for preserving Quilombola multimedia collections and organizing online gatherings ([Silva and Murillo 2021](#)). Once we are allowed to familiarize ourselves with the infrastructure that supports our data workflows, we start to understand data problems otherwise. It all starts by recognizing the alternatives that are already in movement as an exercise in expanding our technopolitical imaginations.

Conclusion

Community-led data stewardship, we suggest, demands thinking about infrastructures in more relational ways. Openness, as an aspect of our infrastructural work, implicates framing the problem of data stewardship as a matter of building meaningful and trustworthy sociotechnical ties. No longer a topic of expert control, infrastructures can be better devised as means for placing community questions, approaches, and urgencies front and center in our discussion about the present and future of collaborative environmental research.

This reckoning of the need for infrastructural alternatives has to do with the halting paradox that we face today: technical work on data management is situated at a polar opposite to the community work that is often organized for the purposes of environmental justice. One cannot, of course, go anywhere without the other when it comes to the creation of a shared understanding of our common planetary condition. It is with a sense of urgency that we suggest open infrastructures as an alternative interface between technical, scientific, and community approaches to environmental research and activism: an interface that necessarily implicates understanding infrastructures as a sociotechnical matter that needs to be, as Michael Fischer ([2007](#)).

puts it, “peopled.” We see the face and recognize the labor of our colleagues across various fractal dimensions in community, research, and technical work as they “infrastructure” our work and we, in turn, participate in “infrastructuring” theirs for open climate research ([Bowker et al. 2010](#); [Edwards 2010](#); [Dosemagen et al. 2021](#)). Once we re-situate the question of stewardship as a problem that involves research instruments, infrastructures, and data that are directly implicated in people’s lives, we start reorienting ourselves for collaborative climate futures that are not yet recognized as such, but are urgently needed for true co-production of knowledge.



Citations

1. Kawerak, Inc.; Association of Village Council Presidents; Bering Sea Elders Group; and the Aleut Community of St. Paul. 2020. Source: <https://kawerak.org/download/navigating-the-new-arctic-program-comment-letter>, accessed June, 15, 2021. [↵](#)
2. Nanda, A. and Mohamed. S. 2021. Environmental Data is Critical to Fighting the Climate Crisis: Data Stewardship Can Put It In the Right Hands. *The Bastion*. Source: <https://thebastion.co.in/politics-and/environmental-data-is-critical-to-fighting-the-climate-crisis-data-stewardship-can-put-it-in-the-right-hands>, accessed January 13, 2022. [↵](#)
3. Ellam Yua, J. Raymond-Yakoubian, R. Aluaq Daniel. and C. Behe. 2022. A framework for co-production of knowledge in the context of Arctic research. *Ecology and Society* 27(1):34. <https://doi.org/10.5751/ES-12960-270134> [↵](#)
4. Burrell, J. 2016. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512. <https://doi.org/10.1177/2053951715622512> [↵](#)
5. Ananny, M., & Crawford, K. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645> [↵](#)
6. Zuboff, S. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs. [↵](#)

7. Star, S. L., and Ruhleder, K. 1996. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, 7(1), 111-134. <https://doi.org/10.1287/isre.7.1.111> ↵
8. Star, S. L. 1999. *The Ethnography of Infrastructure*. *American Behavioral Scientist*, 43(3), 377-391. <https://doi.org/10.1177/00027649921955326> ↵
9. Star, S. L., and Bowker, G. C. 2010. How to Infrastructure. In *Handbook of New Media: Social Shaping and Social Consequences of ICTs* (pp. 230-245). London: SAGE Publications Ltd. <https://doi.org/10.4135/9781446211304> ↵
10. First Nations Information Governance Centre (FNIGC). 2018. The First Nations Principles of OCAP®. *First Nations Information Governance Centre*. Source: <https://fnigc.ca/what-we-do/ocap-and-information-governance>, accessed on Nov. 21, 2020. ↵
11. Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., and Hudson, M. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), 43. <https://doi.org/10.5334/dsj-2020-043> ↵
12. Carroll, S. R., Herczog, E., Hudson, M., Russell, K., and Stall, S. 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Science: Scientific Data*, 8(1), 108. <https://doi.org/10.1038/s41597-021-00892-0> ↵
13. Arancio, J., and Dosemagen, S. 2022. Bringing Open Source to the Global Lab Bench. In: *Issues in Science and Technology*. Source: <https://issues.org/open-source-science-hardware-gosh-arancio-dosemagen>. Accessed on April 15, 2022. ↵
14. Rey-Mazon, P., Keysar, H., Dosemagen, S., D'Ignazio, C., Blair, D. 2018. Public Lab: Community-based approaches to urban environmental health and justice. *Science and Engineering Ethics*, 24(3), 971-997. <https://doi.org/10.1007/s11948-018-0059-8> ↵
15. Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., Stefano, R. D., Gil, Y., Groth, P., Hedstrom, M., Hogg, D. W., Kashyap, V., Mahabal, A., Siemiginowska, A., & Slavkovic, A. (2014). Ten Simple Rules for the Care and

Feeding of Scientific Data. *PLOS Computational Biology*, 10(4), e1003542.

<https://doi.org/10.1371/journal.pcbi.1003542>

16. Gentemann, C., Crawford, S., Martin, P., Perez, F. 2021. Open Science in Action. In: American Geophysical Union, Fall meeting, New Orleans, Louisiana. Source:

<https://agu.confex.com/agu/fm21/meetingapp.cgi/Session/122142>, accessed January 13, 2022. ↵

17. Gabrys, J., Pritchard, H., and Barratt, B. 2016. Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society*, 3(2), 2053951716679677. <https://doi.org/10.1177/2053951716679677> ↵

18. Silva, T. C., Murillo, L. F. R. 2021. Computing from the Global South. In: *Data & Society: Points*. Source: <https://points.datasociety.net/computing-from-the-global-south-reflections-from-within-dd569fd0df9>, accessed September 2, 2021. ↵

19. Fischer, M. M. J. 2007. Four Genealogies for a Recombinant Anthropology of Science and Technology. *Cultural Anthropology*, 22(4), 539–615.

<https://doi.org/10.1525/can.2007.22.4.539> ↵

20. Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. 2010. Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: J.

Hunsinger, L. Klastrup, & M. Allen (eds.), *International Handbook of Internet Research* (pp. 97–117). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_5 ↵

21. Edwards, P. N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press. ↵

22. Dosemagen, S., Heidel, E. Murillo, L. F. R., Velis, E., Stinson, A. and Thorne, M. 2021. Open Climate Now! In: *Branch*. Issue 2, June 14, 2021. Source:

<https://branch.climateaction.tech/issues/issue-2/open-climate-now>, accessed on May 1, 2021. ↵