Parasitic Knowledge Infrastructures: Data Reuse by Anthropogenic Climate Change Skeptics

Wofford, Morgan University of Michigan, United States | mwofford@umich.edu

ABSTRACT

Stakeholders from academia, industry, funding agencies, and scholarly publishing are increasingly investing in open data partially in the hope that it will democratize science and promote more diverse data reuse. However, fewer studies examine how unconventional communities outside academia and industry use open data. Through an investigative digital ethnography, I observed the data practices of anthropogenic climate change (ACC) skeptics, specifically how they discuss, evaluate, and reuse open data. This poster focuses on the knowledge infrastructure that affords the data practices of ACC skeptics. I argue that ACC skeptics are building a parasitic knowledge infrastructure on the back of the climate science knowledge infrastructure it often seeks to discredit. Understanding the infrastructure that supports skeptics' data reuse can inform how we design policies and infrastructure to actualize open data's promises while minimizing its perils.

KEYWORDS

Data practices; knowledge infrastructures; open science; climate science; climate change skepticism

INTRODUCTION

One of the purported promises of the open data movement is its democratizing potential, affording people outside of orthodox research communities the ability to take advantage of data (Baack, 2015; Cavalier & Kennedy, 2016; Espinosa et al., 2014; Nielsen, 2011; Ricker et al., 2020; Zuiderwijk & Janssen, 2014). Open data activists envision data allowing individuals to generate their own knowledge representing a "democratization of information" (Baack, 2015). Citizen groups like the Anti-Eviction Mapping Project and the Mapping Police Violence Database use open government data to support the needs of the traditionally marginalized (*Anti-Eviction Mapping Project*, 2021; *Mapping Police Violence*, 2021). Tech-savvy unconventional data reusers view their activities in their communities as philanthropic endeavors (Kassen, 2021).

However, open data supporting the traditionally marginalized is not guaranteed, nor is it the only eventuality. Those with the most privilege and power are more likely to take advantage of open data, reinforcing extant hegemonic structures (Mirowski, 2018). Even if open data facilitates broader and more diverse data reuse, this can sometimes result in misuse or misinterpretation. For instance, local governments and private companies frequently misuse climate data in determining their financial climate-related risk (Fiedler et al., 2021). In another example, Lee et al. showed that anti-maskers use orthodox visualization techniques on open government data to support their unorthodox beliefs of removing mask mandates (Lee et al., 2021).

Broadly, unconventional data reusers could include citizen scientists, students, conspiracy theorists, community organizers, activists, teachers, and many more. Understanding how these data reusers leverage open data is essential to inform infrastructural development and support appropriate unconventional data reuse. The poster is based on an investigative digital ethnography of a single group of unconventional data reusers—ACC skeptics—examining their data practices, specifically how they discuss, evaluate, and reuse open climate data. The artifacts are the knowledge they construct from the data in the trimmings of scholarly discourse, including statistics, figures, graphics, and other computational models. Following conventional digital ethnography methods, I account for the infrastructures that afford this data reuse (Pink et al., 2015). I contextualize these as knowledge infrastructures that are "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010).

METHODS

Through an investigative digital ethnography, I qualitatively traced how open data travels through the ACC skepticism community and becomes "knowledge," following cascades of documents. Investigative digital ethnography combines the search for specific information or actors with longer-term observation (Friedberg, 2020). This methodology draws on various works on digital ethnography to understand how artifacts move through online communities (Donovan, 2019; Lewis & Marwick, 2017; Pink et al., 2015).

I used "deep lurking," which draws on Clifford Geertz's concept of "deep hanging out," to participate by observing and systematically documenting the data practices of this community (Geertz, 1998; Lee et al., 2021). The units of analysis are data from repositories and the "knowledge artifacts" skeptics produce from those data. These knowledge artifacts include graphs, statistics, tables, charts, maps, computational models, code, and more. Because

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

of the nature of digital media, I collected data with an added focus on the knowledge infrastructure that supports the data reuse (Pink et al., 2015).

I began by identifying relevant publications, social media platforms, organizations, and influencers to define an ACC skepticism knowledge infrastructure. The preliminary list seeded my initial monitoring environment for various social media platforms. I then engaged with platform affordances, viewing, liking, and reposting content to influence recommendation algorithms. The recommendation algorithms offered new accounts to follow through algorithmic-assisted snowball sampling.

For my monitoring strategy, I began each observation period deep lurking on social media platforms. I archived these web pages if open data were discussed or reused. If these posts linked to blogs or organizations that mentioned open data, those web pages were archived. I observed the community for 75 hours from September 2021 through November 2021, archiving the relevant web pages and taking 15 pages of memos. All documents were uploaded into NVivo for qualitative coding. I synthesized themes across my data using grounded theory (Glaser & Strauss, 1967) adapted for social media analysis (Postill & Pink, 2012), inductively coding the data for emergent themes. These themes were then coded into higher-level concepts, constructing theories grounded in the data (Charmaz, 2014). I analyzed 125 files consisting of 2,867 pages of text over 67 hours with 19 pages of memos.

FINDINGS

I propose that ACC skeptics are building a parasitic knowledge infrastructure upon the back of the mainstream climate science knowledge infrastructure. A parasitic knowledge infrastructure generates, shares, and maintains its knowledge using components of another knowledge infrastructure while simultaneously weakening that infrastructure it relies upon. The "hypertransparency [of] open data, open code, commodity software tools, and alternative publication venues" allows skeptics to selectively use these components where feasible and advantageous (Edwards, 2019, p. 21). The digital ethnography shows that the parasitic knowledge infrastructure relies on open climate science data and tools to produce much of their "knowledge." For instance, ACC skeptics use global temperature data from NASA and NOAA to attempt to disprove the very temperature trends these agencies publish.

The parasitic knowledge infrastructure utilizes mainstream tools when they provide affordances that skeptics cannot replicate. For instance, the World Meteorological Organization's KNMI Climate Explorer allows users to investigate various climate data, including time series, model scenario runs, and more. KNMI is only mentioned as an available tool to visualize climate model scenario runs. This is an affordance that skeptical and agnostic tools do not offer. When suggesting tools to explore time series data, skeptics mention their alternatives rather than the KNMI Climate Explorer. Their devices more easily show data trends that support skeptical conclusions than mainstream tools. Many sites provide resource lists of mainstream climate science data, tools, and organizations alongside their skeptical counterparts, presented as equally legitimate.

Where infrastructural components are not available but could prove beneficial for their goals, skeptics critique the infrastructure for its lack of transparency. One skeptic explains that while National Centers for Environmental Information published a paper on how the Global Historical Climatology Network (GHCN) dataset is processed, the code used for data processing is held "specifically by the US government that we cannot test externally [or] even replicate it." He concludes that the code must be made public to understand how the GHCN is calculated. In other instances, alternative infrastructural components are created to mimic mainstream climate science to maximize its perceived legitimacy. For example, one skeptical organization calls its lists of world temperature datasets, a data repository, and its weekly newsletter a journal.

This parasitic knowledge infrastructure masquerades as a trustworthy knowledge infrastructure trying to spread its knowledge in the trappings of scholarly discourse. As we open more components of knowledge infrastructures to the public, parasitic infrastructures are more likely to arise. How do we deal with the largely unavoidable misuse of open data? Is the misuse significant enough to address through changes in open science policies or the design of infrastructural components, such as data repositories?

CONCLUSION

The ACC skeptics' data practices are made possible through their parasitic knowledge infrastructure built on the back of the mainstream climate science knowledge infrastructure. The parasitic knowledge infrastructure gains strength from the mainstream knowledge infrastructure's increased openness. ACC skeptics' data practices call into question the often uninterrogated assumption that open data is a universal and democratizing social good. We need a more comprehensive picture of what open data leads to in practice and how various groups, including unconventional communities, reuse data. This picture can inform how we should design open data policy and infrastructure to actualize open data's promises while minimizing its perils.

ACKNOWLEDGMENTS

I am grateful to Andrea Thomer, Libby Hemphill, and Irene Pasquetto for their guidance and support on this research for my pre-candidacy project.

REFERENCES

Anti-Eviction Mapping Project. (2021). Anti-Eviction Mapping Project. https://antievictionmap.com

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 2053951715594634. https://doi.org/10.1177/2053951715594634

Cavalier, D., & Kennedy, E. B. (2016). *The rightful place of science: Citizen science*. Consortium for Science, Policy & Outcomes.

Charmaz, K. (2014). *Constructing Grounded Theory*. SAGE.

Donovan, J. (2019). Toward a Militant Ethnography of Infrastructure: Cybercartographies of Order, Scale, and Scope across the Occupy Movement. *Journal of Contemporary Ethnography*, *48*(4), 482–509. https://doi.org/10.1177/0891241618792311

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.

Edwards, P. N. (2019). Knowledge infrastructures under siege: Climate data as memory, truce, and target. In D. Bigo, E. F. Isin, & E. S. Ruppert (Eds.), *Data politics: Worlds, subjects, rights* (pp. 21–42). Routledge, Taylor & Francis Group.

Espinosa, R., Garriga, L., Zubcoff, J. J., & Mazón, J.-N. (2014). Linked Open Data mining for democratization of big data. *2014 IEEE International Conference on Big Data (Big Data)*, 17–19. https://doi.org/10.1109/BigData.2014.7004479

Fiedler, T., Pitman, A. J., Mackenzie, K., Wood, N., Jakob, C., & Perkins-Kirkpatrick, S. E. (2021). Business risk and the emergence of climate analytics. *Nature Climate Change*, *11*(2), 87–94. https://doi.org/10.1038/s41558-020-00984-6

Friedberg, B. (2020). Investigative Digital Ethnography: Methods for Environmental Modeling. In *Media Manipulation Casebook*. https://mediamanipulation.org/research/investigative-digital-ethnography-methods-environmental-modeling

Geertz, C. (1998). Deep hanging out. *The New York Review of Books*, *45*(16), 69–72.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing.

Kassen, M. (2021). Understanding motivations of citizens to reuse open data: Open government data as a philanthropic movement. *Innovation*, *23*(1), 44–70. https://doi.org/10.1080/14479338.2020.1738940

Lee, C., Yang, T., Inchoco, G. D., Jones, G. M., & Satyanarayan, A. (2021). Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3411764.3445211

Lewis, B., & Marwick, A. E. (2017). Media Manipulation and Disinformation Online. *Data & Society*. https://datasociety.net/library/media-manipulation-and-disinfo-online

Mapping Police Violence. (2021). Mapping Police Violence. https://mappingpoliceviolence.org

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

Nielsen, M. A. (2011). *Reinventing discovery: The new era of networked science*. Princeton University Press.

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2015). *Digital Ethnography: Principles and Practice*. SAGE.

Postill, J., & Pink, S. (2012). Social Media Ethnography: The Digital Researcher in a Messy Web. *Media International Australia*, *145*(1), 123–134. https://doi.org/10.1177/1329878X1214500114

Ricker, B., Cinnamon, J., & Dierwechter, Y. (2020). When open data and data activism meet: An analysis of civic participation in Cape Town, South Africa. *The Canadian Geographer / Le Géographe Canadien*, *64*(3), 359–373. https://doi.org/10.1111/cag.12608

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003