Selected Papers of #AoIR2017:
The 18th Annual Conference of the
Association of Internet Researchers
Tartu, Estonia / 18-21 October 2017

GRAND INTENTION, SMALL INTERVENTIONS: CLIMATE DATA RESCUE AS COUNTER-DATA ACTION

Britt S. Paris
Department of Information Studies, UCLA

Morgan E. Currie
Stanford University Center for Philanthropy and Civil Society

Joan Donovan
Data and Society Research Institute

Introduction

In this scholarly literature on data activism (Milan, 2016), counter-data action (Kitchin and Lauriault, 2014) or statactivism (Bruno 2014), the work of data activists involves the *generation* of data to create new statistical representations or to challenge official ones. However, much less attention has been focused on activity at the other end of the data lifecycle spectrum**:** data *archiving* and preservation as an act of political critique. In fact, literature on data activism has widely ignored issues of stewardship and long-term preservation, which raises the important question of whether largely ad hoc, diverse groups of data activists can manage data with sufficient integrity and sustainability over the long term. This paper focuses on framing the Environmental Data & Governance Initiative's (EDGI) data archiving and preservation as an act of political critique, and a concrete type of counter-data action.

The Environmental Data & Governance Initiative (EDGI) is a mostly-volunteer, collaborative network that investigates potential threats to the scientific research infrastructure necessary for environmental and energy policy in the United States. EDGI began in November 2016, soon after Donald Trump was elected to the U.S. presidency. Internationally, scholars shared similar concerns that Trump's ideological position on climate change would result in the removal of this information from public access. EDGI formed with a few goals in mind: to design web-based tools, foster a network of

Suggested Citation (APA): Paris, B., Currie, M., Donovan, J. (2017, October 18-21). *Grand Intention, Small Interventions: Climate Data Rescue As Counter-Data Action*. Paper presented at AoIR 2017: The 18th Annual Conference of the Association of Internet Researchers. Tartu, Estonia: AoIR. Retrieved from http://spir.aoir.org.

researchers, and host public events in an effort to keep public environmental data accessible.[1] To better understand this phenomenon, I will briefly examine data activism's questioning the power of official statistics with relation to social justice issues.

Data Collection as Activism

Building archives has already been an integral part of struggles for social justice (Flinn & Alexander, 2015). There are many examples of archiving as an inherently political activity from Jimerson (2009), Caswell (2010) and countless others. This archival work challenges the dominant narratives of the past and makes us rethink how we preserve memories for the next generation. Autonomy is key to the success of these archives, which are often maintained, owned and used by the very people who generate them (Sellie et al, 2015). By remaining independent from formal institutions, archivists are making a statement about how entrenched organizations play a role in their political necessity in the first place. Past and present marginalization of and violence to particular minority communities and pervasive ecological exploitation remain central to institutions of American democracy – whether universities or federally funded historical archives. For this reason, Johnson, Drake, and Caswell (forthcoming) note that we cannot always count on such institutions to meaningfully memorialize on behalf of these voices and Autonomy from central institutions can also protect valuable materials within politically volatile environments.

Data rescues and their archival work strive not to challenge a dominant scientific narrative, but to protect it from a "post-truth" mentality that makes climate change denial seem a viable social act, one in which facts pertain only to individual perspectives. In this respect it may be different from past radical archive of the pasts. But the work is autonomous and comprised of heterogeneous actors interested in resisting the actions of those in power to cast aside empiricism and our future progress on climate change.

Archiving for the Future

To counter any possible loss, EDGI members devote themselves to several tasks: website monitoring for any changes to content on a subset of federal websites; archiving datasets and web pages alongside local organizers and volunteers at archiving events, called DataRescues, in collaboration with other Internet archiving efforts, including Internet Archive, the University of Pennsylvania's DataRefuge, and Climate Mirror. They have moved to focus on creating their own data infrastructure to maintain the data, its trust and accessibility in collaboration with Protocol Labs. This discussion will focus on EDGI's data archiving activity in particular.

At the time of this writing, Multiple EPA climate change subdomains have been removed and now redirect to an "update" page (Mooney and Eilperin, 2017). Links to the State department's 'Climate Change' pages also stopped working after the Obama-Trump whitehouse.gov changeover (Varinsky, 2017). To be clear, federal data is not so much in danger of being deleted than of becoming difficult to access thanks to policy

[1] See EDGI's Report, "Introducing the Environmental Data and Governance Initiative," published February 1, 2017. https://envirodatagov.org/publication/introducing-edgi/.

changes and related funding cuts. EDGI's concern is that data that previously circulated as "open" will disappear from the web in an effort to reduce the capacity for scientific research and calls for reform and regulation.

One vital part of EDGI's archiving work involved the coordination of 'DataRescue' events around the United States; over 30 of these occurred from December 2016 until July 2017. Often hosted by higher educational institutions, DataRescues invited volunteer participants together to copy federal scientific datasets, documents, and webpages using an array of archiving tactics that EDGI participants call "seeding," "harvesting," and "bagging." "Seeding" is EDGI's term for nominating federal webpage URLs to the Internet Archive's End of Term project. End of Term routinely archives .gov websites during periods of a presidential transitions. The project uses an automated web crawler to "crawl," or map out relations between federal web pages, as well as archive web snapshots; it also accepts nominations for URLs from external parties who want to ensure particular web pages are replicated. EDGI volunteers contributed to the End of Term project by nominating the pages of agencies devoted to energy and environmental research in particular.

The Internet Archive's basic crawler, however, encounters many intractable webpages that it cannot crawl. In order to archive this uncrawlable data, EDGI participants designed an open source web application called Archivers.space, a project management tool that tracks a dataset's process from its uploading to a server through multiple stages of research and vetting by participants. In one step participants "bag" the data - EDGI's term for running a python script that generates checksums. Another step entails pulling the data into a zip file along with information that pertains to its web context, metadata, and any other important descriptive information. In a final step the dataset is fully described, including its chain of custody, context, and provenance. Once a dataset is ready, it is pushed to DataRefuge.org (maintained by DataRefuge), where scientists and activists can use it.

Archiving data does not necessarily preserve it, and exigent dissemination of data may struggle without custodial publics. The recent rise of autocratic tendencies in various jurisdictions (Turkey, the Philippines, Russia, the United States) is a salutary reminder of the fragility of any centralized data storage system. In a centralized model, however, the problem of data authority is simple -- the central warehouse authorizes a piece of data simply by storing it. In the decentralized web, that problem is more difficult, as are the problems of locating data, versioning it, and storing it reliably. EDGI's Data Together collaboration with Protocol Labs and qir.io exists to decentralize the rescued data. More broadly this collaboration exemplifies Kitchin and Lauriault's notion of counter-data action as it serves as a juncture at which communities work cooperatively, share responsibilities, and reinforce each other so that data is accessible to all, immediately discoverable, easily verifiable, and robustly preserved.

The *verification* problem can be solved by cryptographic signing provided by Protocol Labs' Interplanetary File Sharing System (IPFS). An application layer protocol that allows data o take advantage of many efficiencies lost via the de facto file retrieval system, HTTP, making the latter work more like a peer-to-peer system.
The *authority* problem can be solved by webs of trust provided by EDGI and Data

Together. The *access* problem remains a complex problem for which scientific and activist communities must be polled to determine how to meet their needs within the existing IFPS structure.

Conclusion

EDGI archives data created by U.S.-funded scientists to document evidence of climate change and human-induced ecological violence; it replicates and preserves scientific evidence to protect it from a vicious relativism, not to augment, challenge, or reinterpret it. EDGI's web archiving and data mirroring pluralizes and distributes the material context of the data; its political work lies in decentralizing information by removing it from institutional control, not re-presenting it.

It does so by building upon data archives that typically fall somewhere within the civic sphere involving grassroots political activism, who commune voluntarily to collect evidence around shared matters of concern. To be trustworthy, the archivist will want to show that a dataset has not been tampered with or degraded in any way; she will want to use professional standards for digital data preservation: metadata that documents chain of custody and provenance; metadata that preserves context; analog and automated checksums; and multiple, mirrored copies to prevent loss. In the case of EDGI volunteers designed bespoke platforms for data ingestion and documentation in order to ensure the integrity of the data.

By treating federal scientific data as a public utility, data rescues create an occasion for community and political resistance and constitute an emergent form of counter-data action. In fact, we might find that the importance of mirroring federal climate data lies less in rescuing data sets for the scientific community – since it's too soon to tell whether more information will vanish or be defunded – but instead in creating spaces for community dialogue and wider public awareness of the vulnerabilities of politically contentious scientific work.

References

Bruno, I., Didier, E., & Vitale, T. (2014). Statactivism: forms of action between disclosure and affirmation. *Bruno Isabelle, Didier Emmanuel and Vitale Tommaso (2014)," Statactivism: Forms of Action between Disclosure and Affirmation", in Partecipazione e Conflitto. The Open Journal of Sociopolitical Studies*, *7*(2), 198–220.

Caswell, M. (2010). Khmer Rouge archives: accountability, truth, and memory in Cambodia. *Archival Science* 10 (1), 25-44

Doria D. Johnson, J. M. Drake, and M. Caswell. (Forthcoming). "From Cape Town to Chicago to Colombo and Back Again: Towards a Liberation Theology for Memory Work," forthcoming.

Eilperin, J. (2017, May 6). The EPA just buried its climate change website for kids. Retrieved June 24, 2017, from https://www.washingtonpost.com/news/energy-environment/wp/2017/05/06/epa-buries-climate-change-site-for-kids/

Flinn A. and B. Alexander. (2015). "'Humanizing an inevitability political craft': Introduction to the special issue on archiving activism and activist archiving," *Arch Sci*, vol. 15, no. 4, pp. 329–335.

Jimerson, R. (2009) *Archives Power*, Society of American Archivists.

Kitchin, R., & Lauriault, T. P. (2014). *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work* (SSRN Scholarly Paper No. ID 2474112). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2474112

Mooney, C., & Eilperin, J. (2017, April 29). EPA website removes climate science site from public view after two decades. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/energy-environment/wp/2017/04/28/epa-website-removes-climate-science-site-from-public-view-after-two-decades/

Varinsky, D. (n.d.). Data on climate change progress is disappearing from the US State Department website. Retrieved June 27, 2017, from http://www.businessinsider.com/climate-action-reports-disappeared-2017-1