Katharine W. Ray. A Motivating Climate: An Exploratory Study of the DataRescue Movement. A Master's Paper for the M.S. in L.S. degree. May, 2018. 58 pages. Advisor: Christopher A. Lee.

This study explores motivations in the DataRescue movement of 2017 that led to the identification and archiving of at-risk federal environmental data in digital repositories. It describes interviews with three national DataRescue coordinators involved in Data Refuge and the Environmental Data and Governance Initiative (EDGI) and analyses of their organizations' websites. Particularly, it looks at the impact of federal record-keeping laws on DataRescue and whether organizers considered the data's eventual users.

Motivations for these projects included determining the scope of at-risk federal environmental data, communicating and raising awareness of the risk, and creating a long-term plan for access to and preservation of data. Environmental data justice (EDJ) was a stated principle for the DataRescue movement, and organizers utilized storytelling to demonstrate environmental data's impact on people's lives. The existence of federal record-keeping laws impacted DataRescue events only insofar as participants saw the laws protecting data as inadequate.

Headings:

       Environmental sciences -- Information services

       Records management laws

       Government websites

       End users (Information technology)

       Web archiving

       Environmentalism -- United States

A MOTIVATING CLIMATE: AN EXPLORATORY STUDY OF THE DATARESCUE
MOVEMENT

by
Katharine W. Ray

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

May 2018

Approved by

_____

Christopher A. Lee

Table of Contents

1. Introduction

"We swim in a sea of data…and the sea level is rising rapidly" (Anderson & Rainie, 2012).

Between December of 2016 and June of 2017, archivists, librarians, scientists, students, and data analysts organized a series of DataRescue events across North America. Concerned about the continued availability of federal websites and datasets after a change in American administration and the establishment of a climate change denier as head of the U.S. Environmental Protection Agency (EPA), they sought to create an independent archives of federal climate data websites by duplicating selected federally hosted datasets, archiving websites, and documenting the data's provenance. The DataRescue events initially focused on environmental and climate data. Two organizations quickly formed: the Environmental Data and Governance Initiative (EDGI) and Data Refuge.

DataRescue events were locally based but coordinated with the help of EDGI and Data Refuge. At the events, participants copied federal data to locally and distributed servers in the interest of protecting it in case the Trump administration took down data related to climate change. This meant figuring out what data federal agencies hosted, and of that data, what concerned the environment. It also meant figuring out what data was archive-able using current tools, what tools were needed, and then coordinating a process to copy that data into external repositories. Rather than having just one central data repository, federal agencies control their data and determine how to make it accessible.

Federal data also come in a large variety of content, size, and form. They include educational websites on polar ice caps and historical buoy information from 1662 to the present. But the DataRescue-d datasets also included data that was not directly concerned with temperature or sea level rise. This included census data and the National Library of Medicine's index of prescription and non-prescription drugs, among many others.

The basic process for archiving federal data at a DataRescue event was that some people nominated websites for archiving and then others "seeded" them. An indexing tool then crawled those pages, meaning it captured the websites and the relationships between them. Then participants added them to the Internet Archive. Other participants would figure out ways to archive data that could not be crawled for the Internet Archive. In part because of the speed with which the movement took place—EDGI's and Data Refuge's sites went live within two months of the presidential election, and by June 2017, over fifty DataRescue events had happened—people duplicated efforts, nominating and seeding the same websites or ones that had already been archived. Other participants developed different strategies, using archiving tools such archive.is, and then storing those files on personal or shared servers. Still, by organizers' estimation, DataRescue did not rescue all the data, but it led to a wider movement on what environmental data is and how it impacts people's lives.

Even from the beginning, DataRescue events included storytelling. EDGI and Data Refuge sought to do more than the simplified process of web archiving above. DataRescue events utilized local expertise. Across the fifty local events, coordinators and participants not only nominated and copied federal data and websites, they also hosted

data teach-ins, environmental law sessions, and how-to-archive-a-website sessions. Some participants engaged in media outreach to demonstrate that the problem was ongoing.

The first event was in Toronto ("Toronto 'guerrilla' archivists to help preserve US climate data," 2016), although it was not yet under the DataRescue title. While utilizing an extant infrastructure of data and web harvesting partnerships, Data Refuge and EDGI supported over 50 DataRescue events between them and coordinated the hosting of at least 395 datasets and 200 terabytes of archived websites and digital objects. The last DataRescue event occurred in June 2017. EDGI and Data Refuge have continued their work, with evolving missions: "EDGI documents, contextualizes, and analyzes current changes to environmental data and governance practices through multidisciplinary and cross-professional collaborative work" (Environmental Data & Governance Initiative, 2017), and Data Refuge "helps to build refuge for federal data and supports climate and environmental research and advocacy" (Data Refuge, 2017).

While the events garnered attention from the press that focused on archiving federal climate data, there has been little analysis of the motivations and cultural factors that led to the DataRescue events. This paper is an exploratory case study on the motivations of the organizers of DataRescue events concerning the preservation of federal datasets and websites. It aims to answer the following research questions:

RQ1: What were the motivations of the organizations and archivists who participated in the DataRescue movement?

RQ2: What makes the DataRescue movement different from other partnerships between public agencies and other institutions and organizations to archive federal datasets and websites?

RQ3: Did organizers and participants consider the users of archived datasets?

RQ4: How did laws and regulations concerning federal recordkeeping of datasets impact the DataRescue movement?

In this paper, the term "DataRescue" encompasses both specific events and the overarching movement of archiving perceived at-risk federal datasets and websites between December 2016 and June 2017. This research is important because it explores the motivations of archivists taking on roles as activists, particularly in relation to their perceptions of the rules, laws, and norms that regulate the preservation of federal data. While there is literature on the role of the archivist as activist, there is little in the realm of digital preservation and activism as it relates to federal websites and datasets, and even less specific to federal data about climate and the environment. It also communicates the huge array of federally funded data at risk of no longer being accessible and the limitations of tools at archivists' disposal to address these risks.

Digital curation is not a generalizable set of activities, but rather "variegated persons undertak[ing] variegated tasks in variegated settings" (Poole, 2016, p. 963). This is also true of how federal government manages environmental data. DataRescue brought light to some of the issues with the management of federal data by seeking to create an external repository. This study encompasses aspects of self-, societal-, and professional-perception by focusing on the DataRescue movement as a model of activism. Through preserving datasets and websites and thus maintaining access to resources, they developed a model for reaction and laid the groundwork for emergency back-up systems of access. Archival norms favoring ostensible neutrality can perpetuate state control, particularly when the state seeks to obstruct access to the truth (Zinn, 1977). Through

DataRescue, archivists, librarians, and others—understanding that there were already obstacles in communities obtaining environmental data—did something new. They archived, en masse, federal environmental and climate datasets in externally controlled repositories. The movement bridged data activism, archivist activism, and environmental activism.

2. Literature Review

DataRescue bridges the existing regulations that concern recordkeeping as it relates to federally managed environmental datasets, data preservation, and activism. This activism encompasses data activism, archivist activism, and environmental activism. There have been no studies on the motivations of the DataRescue organizers as they pertain to knowledge of federal recordkeeping and environmental data law. There are many possible responses to the state of government information access; the DataRescue movement, in its breadth of data and participants' perspectives, included varying actions, but the motivations were grounded in data vulnerability, data justice, and the eventual use of data. This movement happened in the context of previous archival activism, federal data preservation, and regulations that govern access to federal data, and it contributed to the emerging field of environmental data justice.

2.1 Preserving Federal Government Data and Websites

One challenge to the preservation of federal datasets is a lack of consensus on what a dataset is. In an interim report from August 2017, the Library of Congress Collection Development Office defined a dataset as "a collection of records, presented in a digital or non-digital format…a data set is considered to be a type of electronic resource that consists of machine readable data" (Library of Congress, 2017, p. 1). In contrast, some definitions include time-based criteria. Meloda, a Spanish organization that advocates for "a metric for releasing open data," defines dataset as a "group of structured

data retrievable in a link or single instruction as a whole to a single entity, with updating

frequency larger than a once a minute" (Meloda, 2017).

There are also differing understandings of which datasets are records. Agencies

have interpreted federal record-keeping laws differently. Further, the agencies' ability to

collect and preserve datasets as records depends on funding and on the utility of datasets

to serve the mission of the agency.

This study explores how archivists handled, if at all, federally managed

environmental datasets as records. Furthermore, due to the structure of the U.S.

Environmental Protection Agency (EPA), one of the foremost federal agencies dedicated

to the collection and promulgation of environmental data, varying by both media and

contaminant, there are many hurdles to accessing federal environmental data (Lamdan,

2017). These factors have impacted the regulations that control management of federal

data and websites. They have also impacted the formation of partnerships between

federal agencies and other institutions.

2.1.1   Regulations for Federal Data and Website Preservation

Federal laws, policies, and regulations govern the preservation of federal data and

websites produced by agencies. However, federal agencies interpret preservation

requirements differently. There are both laws and memoranda that influence how

agencies preserve access to federal data.

While agencies' priorities change due to a change in presidential administration,

different laws apply agencies and the presidential office; the Federal Records Act and the

Presidential Records Act impact recordkeeping requirements differently. As the federal

agency responsible for providing guidance to federal agencies on best practices for

recordkeeping and access to federal archives, the U.S. National Archives and Records Administration (NARA) offers agencies guidance on creating records retention schedules but does not have enforcement or punitive powers.

When, for example, NARA found in 1992 that some offices at the EPA "have no prescribed records creation, maintenance, or disposition procedures" which is "particularly disturbing in an agency like EPA, which creates a staggering amount of recorded information in both electronic and textual form," they were unable to enforce changes (National Archives and Records Administration, 1992, p. 3). The Government Accountability Office has also criticized the EPA for failing to fulfill its oversight role on multiple occasions (Swartz, 2008). NARA, along with its archival and recordkeeping recommendations responsibilities, also houses the Office of the Federal Register, which publishes rules and proposed rules of the federal government (National Archives and Records Administration, 2017).

The Federal Records Act, signed into law in 1950, was the first law to mandate recordkeeping maintenance and disposition at an agency level. The Presidential and Federal Records Acts Amendments, signed into law in 2014, updated the Federal Records Act in reference to electronic records. The Federal Records Act also made the Archivist the final arbiter in what constitutes a record (Ferriero, 2017). The Code of Federal Regulations states how agencies must transfer datasets to the NARA and shows consideration of the format of the dataset:

> *Data files and databases*. Data files and databases must be transferred to the National Archives of the United States as flat files or as rectangular tables; i.e., as two-dimensional arrays, lists, or tables. All ''records'' (within the context of the computer program, as opposed to a Federal record) or ''tuples,'' i.e., ordered collections of data items, within a file or table must have the same logical format (Transfer of Records to the National Archives of the United States, 2009).

This definition of data files as potential records that agencies will transfer to NARA offers evidence that datasets can be records.

However, it is a different issue if a dataset can be something other than a record, particularly if its controlling agency updates it almost constantly. In memoranda, datasets are not records of federal agency activity but are still federally funded data. The Obama-era Presidential Memorandum of 2011 ("Presidential Memorandum, Managing Government Records") led to an administration-wide recognition for better recordkeeping practices, which included identification of policies that either inhibit or encourage effective recordkeeping, designation of an official in charge of carrying out the policies, and identification of funding for policy changes. However, there are further issues with regard to datasets as records.

There is also confusion about the responsibilities of public entities or governments to keep records and archives distinct from requirements for recordkeeping in the executive office. It could be that confusion emerged from the simultaneous Presidential Records and the Federal Records Acts Amendments (2014). For datasets not treated as records, there have also been several attempts to mandate release of datasets produced from taxpayer-funded research. Members of Congress introduced the Federal Research Public Access Act (FRPAA) in 2006, 2010, and 2012, and its successor, the Fair Access to Science and Technology Research Act (FASTR), in 2013 and 2015. However, neither bill passed.

While NARA is responsible for promulgating guidance regarding best practices, individual agencies must apply and enact these federal policies. The *Records Management Policy* (2015a) of the EPA takes as its authority the Obama Memorandum,

the Federal Records Act Amendment, and agency-wide practices for preparation for

emergencies:

> Not all information created or received constitutes a record. Non-records include reference material, supplementary or convenience copies, a draft document or working paper with no substantive comments, and personal information which is unrelated to EPA business (2015).

This policy may lead to confusion about whether datasets are records. Furthermore, as

David Ferriero, Archivist of the United States, noted in testimony before the Committee

on Oversight and Government Reform:

> Ultimately, responsibility for records management will always rest, to some degree, with individual Federal employees, no matter what systems are in place. That was true in an era of exclusively paper records and it remains true in an increasingly digital age ("Presidential Records in the New Millennium", 2011, p. 9).

Agencies' capacity to fulfill archiving and recordkeeping requirements is dependent on

employees' knowledge of these laws. Similarly, archivists' knowledge of federal

recordkeeping laws informs how they choose to act on them. Archivists contend with this

in certain ways, as discussed below concerning DataRescue. Both NARA and the

Smithsonian have modified their practices to satisfy requirements of other Obama-era

orders concerning preparation for climate change, which will impact the federal

government's ability to provide access to federal records (Tansey, 2015; Executive Order

No. 13514; Executive Order No. 13653; Executive Order No. 13694).

2.1.2   End of Term Web Archive

One example of an externally held archival collection of federal sites that

emerged from a partnership between federal agencies and non-federal organizations is the

End of Term Web Archive (EOT Archive). The California Digital Libraries and the

Internet Archive jointly host the project (End of Term Web Archive, n.d.), though it does

not currently have major consistent funding sources (Bailey, Grotke, & Phillips, 2016a).

The University of North Texas Libraries began the End of Term Web Harvest in their

capacity as a depository library in the Federal Depository Libraries Program (FDLP). The

End of Term Web Archive is particularly relevant to DataRescue because (1) most of the

DataRescue events contributed to the EOT Harvest through the Internet Archive, and (2)

an administration change was the catalyst that started the DataRescue movement

(Hegstad, 2017).

Federal officials have identified many sites in the federal website realm, some of

which come from federal agencies creating websites they do not update and maintain.

Others come from offices that were later renamed or reorganized (Phillips, 2011). The

existence of these sites is not limited to turnover from one administration to the next, but

may also be present within the span of a single administration's term. Nonetheless,

changes are more significant when the executive administration changes. Information

provided to the public varies to reflect a given administration's policy priorities. The End

of Term Web Archive (EOT Archive) illustrates this with its collection of websites from

2008, 2012, and 2016. It fits into a context of other collections held by the Government

Publishing Office, the Library of Congress, and the National Archives and Records

Administration. The EOT Archive, like the Library of Congress, has thematic web

collecting. Other countries' national libraries strive to collect every site that their

governments have published, which Shveiky and Bar-Ilan (2013) called the

comprehensive or combined approach to collection development. The DataRescue

movement contributed over 100,000 websites and digital objects to the EOT Archive by May of 2017 ("Over 200 terabytes of the government web archived!", 2017).

The Government Publishing Office's stated mission is "*keeping America informed,* as the official, digital, and secure source for producing, protecting, preserving, and distributing the official publications and information products of the Federal Government" (Government Publishing Office, 2017). The Office manages the FDLP. and incorporates web archiving as part of its National Plan for preserving access to government documents and are involved in the process, though it does not host the EOT Archive (Baish, Etkin, & Walls, 2015; FDLP, 2014). For the University of North Texas, the EOT Archive grew out of its role as a Federal Documents Depository Library and a previous project, the CyberCemetery, which began in 1997.

Bailey, Grotke, and Phillips, who work at the Internet Archive, the Library of Congress, and UNT Libraries, respectively, gave two presentations before and after the harvesting of the federal web for the 2016 End of Term Archives and described the breakdown of roles as such: the Internet Archive conducted crawling, preservation, access, and full-text search development; the Library of Congress conducted crawling, preservation, and data transfers; University of North Texas conducted nomination tool development, crawling, nomination management, preservation, and access; the California Digital Library managed the web portal and metadata; and the GPO conducted outreach and URL nomination (2016a; 2016b).

The End of Term Web Archive displays the distributed responsibilities of archives organizations and federal groups. It also shows the current gaps in preserving federally funded data, publications, and websites. In March 2018, Congress introduced a

bill to modernize the FDLP act that addresses some of the concerns of preservation for digital publications.

2.1.3   Open Data

Open data is the principle that governments should make public any data they have, as long as the data does not contain sensitive information. In 2013, President Obama signed an executive order titled "Making Open and Machine Readable the New Default for Government Information" mandating the creation of a central online access point for federally managed data (Executive Order No. 13642, 2013). This led to the creation of Data.gov. The site functions as the United States' repository of federal data, with increased transparency and accountability as two of its intended impacts (Data.gov, 2017). It ranks the reusability of data on a 5-star rating system developed by Tim Berners-Lee for open data, with stars representing open license, structured data, open standards, unique resource identifier (URI), and connected sources, respectively (Berners-Lee, 2010; Abella, Ortiz de Urbina Criado, & De Pablos Heredero, 2014). However, Data.gov does not encompass all federal data. The National Oceanic and Atmospheric Administration (NOAA), the EPA, the United States Geological Survey (USGS), and the Bureau of Census all make openly accessible environmental data available through platforms they control. Not only are there multiple platforms for accessing data, there are also many challenges to continue to make open data initiatives successful. These include open lines of communication, institutional support, and ways to distinguish between current and past versions of data (Lucia et al., 2017). Due to organizational structure changes, in addition to changes in policy, a new administration can jeopardize the long-term viability of open data at the federal level.

When federal funding institutes such as the National Institutes of Health (NIH) began to mandate data availability for research-funds recipients, librarians saw the increased need to understand and implement data curation practices (Ogburn, 2010). Scientific data present particular issues for preservation, including size, a lack of uniform standards for format, difficulty documenting provenance, and the high level of expertise required for processing datasets. Scientists who try to reuse this data face similar challenges. Accessibility issues, missing data, lack of context, and unavailability of data are common concerns (Carver, 2012). These accessibility issues include format, duplication issues, and data degradation.

The Office of Science and Technology Policy (OSTP) memorandum of February 2013 (Holdren, 2013) addressed how federal agencies need to increase access to federally funded data. Kriesberg, Huller, Punzalan, and Parr found that there was a wide variation in how different departments responded to the memo (2017). Most agencies' public access plans included clearly discussed data management plans, but most lacked digitization plans or plans for legacy data, funding sources, and persistent identifiers. Some data librarians responded to the memo by crowd-sourcing a list of requirements for disclosure and access to federally funded data. (Whitmire et al., 2016). Others put this into practice to support cultural heritage institutions on reusing data, since they had experts on sharing data (Allard, Lee, McGovern, & Bishop, 2016).

These choices speak to the need for open data to include ways to ensure access. Many agencies did not have thorough plans for data accessibility or discoverability. Simply releasing public datasets does not necessarily equate to transparency or data reuse (Boyle et al., 2015). Governments that do not provide this framework risk the long-term

sustainability of Open Data projects and the loss of public faith in transparency. Failure to properly protect subjects' identities can further undermine public trust and belief in accountability.

In March of 2018, Administrator of the EPA, Scott Pruitt, announced plans to block EPA researchers from considering data that is not available for public scrutiny when creating rules. This is similar to the House Bill H.R. 1430, the HONEST Act, which passed the House of Representatives in 2017 but failed to pass the Senate, which called for the prohibition of the EPA using any scientific studies that could not be independently reproducible (Friedman, 2018).

2.2 Archives, Activism, the Environment, and Justice

Archivists have long been activists, particularly with regard to community archives (Wakimoto, Bruce, & Partridge, 2013; Findlay, 2016). Professional organizations have also increased their responsiveness to current events. For example, the Society of American Archivists (SAA) has also increased its issuance of statements with regard to current events that pertain to archival activities. In 2015, the SAA Council issued a statement on the "Strengthening of Federal Records Authority" (Society of American Archivists Council), noting that all Federal employees received mandatory annual training on information security, yet they did not receive the equivalent training on recordkeeping. In 2017, partly in response to the pushing from outside the agency, the EPA required all employees, contractors, and volunteers to have recordkeeping training. But there are a number of other ways in which archiving, data, and environmental activism have interacted in the past.

2.2.1 Data Activism

Data justice literature often focuses on the collection and control of data, individual rights of visibility and invisibility, engagement and disengagement with technology, and discrimination (Taylor, 2017; Noble, 2018). This includes examining who and whose ideas are represented in data, and whose are not.

Currie, Donovan, and Paris discuss the position of power that data curators have, and what happens when that power transfers to communities (2018). They compare the data archives of EDGI with Fatal Encounters, a volunteer-run organization that documents police killings in the U.S. Similarly, there has been significant activism concerning archiving social movements as they happen as well as embedding archivists in social movements, including #archivesforblacklives, archivists of Occupy Wall Street, and archivists similarly coordinating to document the Women's March (Drake, 2016; Erde, 2014; Russell & Vandeven, 2017), which incorporated the ethos and organization of the respective movements.

2.2.2 Environmental and Climate Data

Archivists have grappled with how sustainability and resiliency might inform the archival practice beyond disaster preparedness, including through the use of collections to identify climate data and document the present moment as both a climate occurrence and social occurrence (Davis, 2015; Tansey, 2015). Welch (1999) described the process of researchers using archival collections concerning environmental activism. There have been numerous instances of archival activism around climate change. Project ARCC, which stands for Archivists Responding to Climate Change, has taken a lead on a response to what archivy may look like in response to a changing environment. Gordon-

Clark (2010, 2012) has documented the impact of climate change on coastal and island

repositories, and the Council of State Archives (2007), Climate Change and the Historic

Environment (Cassar, 2005) have published guides for cultural heritage institutions to

respond to climate change.

There have also been considerations of the impact of digital preservation on

climate, the carbon footprint of archivists, how archives can be used as evidence of

climate change, and, generally, how libraries and archives in the Anthropocene Age can

respond to the present moment. In the last few years, there have been extensive strides

made towards incorporating climate change into how archivists may prepare for the

future, as evidenced by a topical resource page on the SAA website ("Climate Change

Resources for Archivists," 2017).

Environmental data justice (EDJ) is an emerging field of study in this vein. Dillon

et al. (2017) defines EDJ as:

> Public accessibility and continuity of environmental data and research, supported
> by networked open-source data infrastructure that can be modified, adapted, and
> supported by local communities. Environmental data justice also includes
> maintaining attention to long-standing [environmental justice] concerns about the
> politics of evidence, such as what counts as data, what data are collected, and
> whose interests they serve.

Environmental justice (EJ) has long figured into the work of communities

disproportionately impacted by environmental damage and toxicity. For example, in the

United States, 17% of the population live within three miles of a Superfund site—a toxic

waste site that requires cleanup—and that 17% is more "minority, low income,

linguistically isolated, and less likely to have a high school diploma" than the total

population (Environmental Protection Agency, 2015b). Immigrants, the poor, and people

of color have long participated in environmental activism (Gottlieb, 1993; Taylor, 2014)

and used data to change environmental policies, laws, and practices. A final understanding of environmental data not explored in this paper can be seen in how human bodies are sites of federal policies (Murphy, 2008). Petryna (2002) created the term *biocitizen* to describe a person who is inscribed with biological damage inflicted by nations. While the biological damage—in other words, the physical data—is in some sense federally funded (federally inflicted), it is outside the scope of this research project. The preservation of federally-funded environmental data is important to document and fund changes for environmental justice, but there is more to explore in how the federal government defines data worthy of collection and preservation. Environmental data, without the tools to collect, provide context, and document it, often on a vast scale, is, in Jerome Whitington's words, "insensible" since "anyone with experience trying to enforce regulations in say, the Appalachian coal fields will be familiar with the long history of legal failures in the face of insufficient evidence" (2017).

Many forces, including a lack of appropriate legislation and action for the preservation of federal data, predate the change in administration that might have led archivists to discuss how to create a separate repository for federal data. However, when Donald Trump took office, DataRescue and EDGI seemed to appear overnight. When someone who had called climate change a hoax appointed another climate change denier to head the agency in charge of protecting human health and the environment, the movement took off.

3. Methodology

      This research undertook an explanatory, retrospective, and mixed-methods case study of the Data Refuge movement. Semi-structured interviews with organizers of Data Refuge, DataRescue events, and EDGI took place via Skype. In order to offset the inherent limitations of a retrospective case study (Yin, 1984), I analyzed the websites of Data Refuge and EDGI at three points in time between January 22, 2017, and November 30, 2017. Analysis included the following: links to data, workflows, and other artifacts that had been archived or made publicly accessible; consideration of end users; reasons and motivations for the existence of these sites; and references to federal laws or policies.

      The website snapshots used were those closest to, but not preceding, the following dates: January 22, 2017, June 30, 2017, and November 30, 2017, using the Internet Archive's Wayback Machine, if unavailable on the Internet Archive. I excluded an earlier date, December 1, 2016, because no website snapshots of Data Refuge or EDGI existed for that point in time. I chose January 22 because of its proximity to the change in administration, and June 30 because it was after the last DataRescue event had occurred. November 30 was chosen at the development of this research proposal and five months after the last DataRescue event had occurred as documented on Data Refuge's site. If there was not an archived instance of a site after June 30, 2017, I archived the site using the Internet Archive. (For a full list of analyzed archived sites, see Appendix C).

I recruited participants based on their involvement in planning DataRescue events and archiving and then through snowball interview technique (Penn Program for the Humanities, 2017; Environmental Data Governance Initiative, 2017). Thirteen invitations to participate in the study were sent out via email, and five recipients responded. I corresponded briefly to present the subject of my research to the steering committee of EDGI, but they did not get back to me by the time of writing this. I then conducted three 30-minute interviews, which I transcribed and provided to the interview subjects as a method of verifying accuracy (Hagens, Dobrow, & Chafe, 2009). The unit of analysis in this case was the DataRescue movement, events, and public-facing websites.

Participants of this study have been given pseudonyms. "Ellie" is the director of a digital scholarship department at a research library. She identifies as a librarian and was instrumental in Data Refuge's formation. She created the Comprehensive Knowledge Archive Network (CKAN) instance that hosts data on the Data Refuge website, organized two DataRescue events, and created a framework for identifying at-risk data. She has been intimately involved in library and data preservation initiatives for over 15 years. "Steven" is a librarian. Before the DataRescue events, in his capacity as head of a design lab of a research university library, he organized citizen scientist events and was actively involved in library consortia. "Helen" is a scholarly communications and data curation librarian at a large research university library.

This mixed-methods approach is appropriate because, as Stake (1995) has shown, case studies often lend themselves for an understanding of a greater phenomenon. The DataRescue movement involved the copying, archiving, and curating federal datasets and

websites in a widely distributed geography. Social processes led to the creation of workflows and underpinnings of this activist response as a squarely archivist position. Using both artifact analysis and interviews supporting triangulation through use of multiple data sources.

3.1 Limitations

Since this study was undertaken as a master's paper project, the time available for data collection and analysis was limited. This limited time frame impacted my ability to provide the depth of "thick description" required to make this study fully transferable to other settings (Geertz, 1973).

In the interest of having a specific unit of analysis, this study was bounded to the organizations EDGI and Data Refuge and individuals who are affiliated with one of them. A number of institutional or organizational partners were affiliated with or emerged from the DataRescue movement. These included Penn Libraries, University of Michigan Libraries, Internet Archive, Temple University Libraries, Johns Hopkins Library, ProjectARCC, University of Toronto, Union of Concerned Scientists, End of Term Archive, Climate Mirror, and Dat Project. However, it quickly became apparent that membership or affiliation with these groups overlapped frequently. Further research could expand to a greater understanding of how these organizations intersected and identify more clearly the workflows that emerged from the groups.

4. Results and Discussion

Website snapshots contained referrals to federal policies and decisions; reasons for the existence of the organizations, mission statements, and motivations; mentions of data use; and ways to get data access. The snapshots displayed motivations that included (1) determining the scope of at-risk federal environmental data, (2) communicating and raising awareness of the risk between different groups who were already involved in the work of data preservation and those new to this area, and (3) creating a long-term plan for addressing access to and preservation of the data. These support the environmental data justice mission of the organizations.

However, EDGI and Data Refuge developed two separate approaches to moving forward with environmental data justice, but there was a great overlap in DataRescue events. One challenge encountered in this research was identifying the motivations of individuals based on collective action:

> I was working so, so, so closely with a group of people that it's hard to pick out my own values—I mean, I had my own values, absolutely, and I absolutely brought them. But I spent a lot of time on trying to just get a handle on the scope of the problem. (Ellie)

Because of the collaborative nature of the organizations and the speed with which actions occurred, it was difficult to extricate organizers' values and motivations from those of their colleagues. However, mission statements, group definitions, goals, and calls to

Place	Title/Hashtag	Date Start	Date End	Called DataRescue?	City	State
Toronto		2016-12-17	2016-12-17	No	Toronto	Ontario (Ca.)
Philadelphia	#DataRescuePhilly	2017-01-13	2017-01-14	Yes	Philadelphia	PA
Chicago	#DataRescueChicago	2017-01-17	2017-01-17	Yes	Chicago	IL
Indianapolis	#DataRescueIndy	2017-01-19	2017-01-19	Yes	Indianapolis	IN
Los Angeles	#ProtectClimateData	2017-01-20	2017-01-20	Yes	Los Angeles	LA
Philadelphia	Philadelphia Mini Rescue	2017-01-25	2017-01-25	Yes	Philadelphia	PA
Ann Arbor	#A2DataRescue	2017-01-27	2017-01-28	Yes	Ann Arbor	MI
Boston @ Harvard	#DataRescueBoston	2017-02-01	2017-02-01	Yes	Cambridge	MA
UC Davis	#DataRescueDavis	2017-02-02	2017-02-02	Yes	Davis	CA
Portland	#DataRescuePDX	2017-02-03	2017-02-03	Yes	Portland	OR
New York	#DataRescueNYC	2017-02-04	2017-02-04	Yes	New York	NY
Austin	#DataRescueAustin	2017-02-10	2017-02-11	Yes	Austin	TX
SF Bay Area	#DataRescueSFBay	2017-02-11	2017-02-11	Yes	San Francisco	CA
Durham	#DataRescueNH	2017-02-17	2017-02-17	Yes	Durham	NH
Boston @ MIT	#DataRescueBoston	2017-02-18	2017-02-18	Yes	Cambridge	MA
Haverford	#DataRescueHC	2017-02-18	2017-02-18	Yes	Haverford	PA
Washington, DC	#DataRescueDC	2017-02-18	2017-02-19	Yes	Washington	DC
Boulder		2017-02-18	2017-02-19	Yes	Boulder	CO
Twin Cities	#DataRescueTC	2017-02-24	2017-02-25	Yes	Minneapolis	MN
Seattle	#DataRescueSeattle	2017-02-25	2017-02-25	Yes	Seattle	WA
Austin @ UT	#DataRescueATX	2017-03-03	2017-03-03	Yes	Austin	TX
New Haven @ Yale	#DataRescueNHV	2017-03-04	2017-03-04	Yes	New Haven	CT
Houston, Texas @ Rice University	#DataRescueHouston	2017-03-04	2017-03-04	Yes	Houston	TX
Oxford, Ohio @ Miami University		2017-03-04	2017-03-04	Yes	Oxford	OH
Chapel Hill		2017-03-04	2017-03-04	Yes	Chapel Hill	NC
Madison	#DataRescueMadison	2017-03-05	2017-03-05	Yes	Madison	WI
Indianapolis		2017-03-08	2017-03-08	Yes	Indianapolis	IN
Dover	#DataRescueNH	2017-03-10	2017-03-10	Yes	Dover	NH
Chicago	#DataRescueCHI	2017-03-10	2017-03-10	Yes	Chicago	IL
Los Angeles @ Indie Desk	#DataRescueLA	2017-03-11	2017-03-11	Yes	Los Angeles	CA
Eugene, OR		2017-03-11	2017-03-11	Yes	Eugene	OR
Corvallis, OR	#DataRescueOregon	2017-03-17	2017-03-18	Yes	Corvallis	OR
Boston @ Northeastern	#DataRescueBoston	2017-03-24	2017-03-24	Yes	Boston	MA
New York	#DataRescueNYC2	2017-03-25	2017-03-25	Yes	New York	NY
Easton, PA	#DataRescueEaston	2017-04-02	2017-04-02	Yes	Easton	PA
Pittsburgh	#DataRescuePGH	2017-04-02	2017-04-02	Yes	Pittsburgh	PA
Alburquerque		2017-04-05	2017-04-05	Yes	Alburquerque	NM
New Haven @ Yale		2017-04-08	2017-04-08	Yes	New Haven	CT
St. Louis	#DataRescueWUSTL	2017-04-14	2017-04-14	Yes	St. Louis	MO
Endangered Data Week Events		2017-04-17	2017-04-21	Some	Multiple	Multiple
Raleigh		2017-04-20	2017-04-20	Yes	Raleigh	NC
Reno	#DataRescueReno	2017-04-22	2017-04-22	Yes	Reno	NV
Chicago	#DataRescueCHI	2017-04-22	2017-04-22	Yes	Chicago	IL
Austin	#DataRescueATX	2017-04-26	2017-04-26	Yes	Austin	TX
Des Plaines, IL	#DataRescueCHI	2017-04-27	2017-04-27	Yes	Des Plaines	IL
Washington, DC	Sustainable Development Goals Data Archive-a-thon	2017-05-06	2017-05-06	Yes	Washington	DC
Detroit		2017-05-12	2017-05-12	Yes	Detroit	MI
Research Triangle Park	Primer Writing Event	2017-05-15	2017-05-15	Yes	Research Triangle Park	NC
Princeton	#DataRescuePton	2017-05-19	2017-05-19	Yes	Princeton	NJ
Denton	#DataRescueDenton	2017-05-20	2017-05-20	Yes	Denton	TX
Research Triangle Park	#DataRescueRTP	2017-06-10	2017-06-11	Yes	Research Triangle Park	NC

Table 1. DataRescue Events, chronologically. Penn Program for Environmental Humanities (2017).

action displayed motivations of the groups. These sources conveyed explicit and implicit values such as the importance of cite-ability of data for eventual use.

The organizations' websites showed differing levels of change across time. Data Refuge's home page remained relatively unchanged, with slight changes to the shades of green on the page. As of March 30, 2018, it had a major change with the addition of a link to the Data Refuge Stories project.

Data Refuge hosted complex datasets that the Internet Archive could not host due to their complexity. This front-facing data repository at Data Refuge ("Datasets") was not captured between January 28, 2017, and March 30, 2018, in the Internet Archive, meaning I could not tell what it looked like between those dates. In January 2017, there were 18 datasets in the repository, including one test bag. Bags are a method for gathering and transferring datasets and websites. Metadata was sparse, with little descriptive information. Each dataset had a JSON file, which contained preservation metadata, and a ZIP file of the data package itself. When archived on March 30, 2018, many datasets had not only JSON files but also PDFs that contained notes from the data harvest, information produced by the archival "bags" that data were transferred in, a file list of the types of files (by the file extension), and checksums.

EDGI's home page changed greatly, from a static landing page to a rolling visual display highlighting recent publications. As 2017 progressed, EDGI's website grew in scope and its goals became more specific. In 2017, on January 26, EDGI's "About" page describes who were involved, which then changed at a later date:

> This project brings together an international network of social and natural scientists, lawyers, and other information and environmental professionals that compose the diverse range of skills needed to document and advocate for the vital importan[ce] of evidence-based environmental policy.

On September 5, the "About" page communicated the diverse array of organizations and no longer mentioned specific professions: "We also aim to create an open, collaborative network of individuals, non-profits, universities and companies who believe that science and data are vital for environmental governance."

EDGI's "Mission, Vision, Values" had not been captured before March 30, 2018, nor had its page devoted to "Environmental Data Justice." The group's values are justice; environmental and human health; intersectionality; anti-fascism, anti-racism, and anti-oppression; accessibility; responsivity and proactivity; and participatory knowledge-making. This specificity of values, and the formation of the Environmental Data Justice Working Group, show that EDJ was shaping as the focus. It also showed that the group had a larger set of activities than archiving websites. They interviewed long-term employees at the EPA, tracked and drew attention to websites' changes.

4.1 Determining Scope

The DataRescue struggled in defining the scope of what participants should archive. All the organizers I spoke to agreed that the scope of DataRescue's archiving goals was "capacious" (Ellie). Data Refuge's first workflow ("DataRescue Workflow: An Overview", snapshot taken February 23, 2017) described four paths for having an event. Path I, "Website Archiving," involved identifying websites for archiving in Internet Archive. This used EDGI-developed primers of federal agencies, which are identifications of data, links to that data, and both short descriptions. Path II, "Archiving More Complex Datasets," worked on the data that was uncrawlable using tools like Archive-It. In Path II, researchers investigated the uncrawlable URLs, identified by Path I

participants. Harvesters figured out how to capture the data, checkers inspected data for completeness of data, baggers assured quality and bagged the data, and describers described the contents of the bag. Data Refuge, rather than the Internet Archive, then hosted these more complex datasets, those that could not be crawled.

Both groups have always been interdisciplinary at core. Penn Libraries and the Penn Program in Environmental Humanities (PPEH) jointly sponsored Data Refuge, which has as part of its manifesto, "The history of sustainability has been the history of fragmentation: science and the humanities, their generation and our generation, abstraction and action" ("The Manifesto," n.d.). EDGI involved people from a wide range of disciplinary backgrounds.

The interdisciplinary perspectives influenced the breadth of the project and archived data. Ellie expressed how broad the scope of the project was. She also described how the definition of "environmental data" meant that the organizers wanted to capture even more than historical temperature data: "It also includes data about human populations and animal populations. Environmental and climate data to some degree is any data that's relevant to the earth, which is to say arts data could be environmental and climate data." The breadth of the definition contributed to why the workflow changed to encompass more roles. The second iteration of the workflow ("DataRescue Events," snapshot taken July 23, 2017) widened from seeding and identifying complex datasets, storytelling, and strategizing about the future, to include Wikipedia Edit-a-thons, web archiving how-to, more storytelling, and citizen science.

Agencies provide access to federal datasets on many sites with architectures that are not consistent from agency to agency. Partly, this is due to the fact that agencies have

different organizational structures. As the DataRescue organizers found out, some federal

data curators data creators in the government did not know what they had. Helen, the

scholarly communications and data curation librarian, found that DataRescue participants

thought there should be a list of data that was at risk for deletion. However, the creation

of a list was predicated on an even greater and complete list of federal environmental

data, which they did not have.

While the lack of a list made DataRescue events made it challenging to identify

all vulnerable data at DataRescue events, some participants thought this was not

necessarily a bad thing. The lack of a list meant that the new administration could not

systematically removed climate data from public view. The labyrinthine organization of

federal data led, according to Steven, to its protection: "I mean, it was so distributed,

decentralized, and disorganized that it made it really hard to figure out if we're actually

doing a comprehensive job. But it also meant that you couldn't just take a bunch of stuff

down now."

The organizers viewed all environmental data and data that hosted on federal

websites as being at risk, and Ellie worked with others to establish one kind of framework

for understanding that risk. The at-risk data, though, included as wide a definition of

"data" and "environmental" as anyone participating in a DataRescue event thought of

those terms. The definition of "record" was not relevant in determining scope, since

participants viewed all environmental data, record or not, as at risk.

Professional data curators were less inclined to be involved because they knew

more about this scale, and Steven believed that if he and organizers like him had known

as much, the movement might never have taken off. The scale of the data did not

incapacitate them because they did not know how great it was until they were already waist-deep in data.

According to a document written by Justin Schell (2018), EDGI's guiding principle is Environmental data justice. This principle draws on what data is collected and how. I did not find evidence that Data Refuge used this term directly; however, the PPEH Schuylkill River Research Seminar demonstrated many core principles of environmental justice and data collection.

4.2 Opening Channels of Communication

Perspectives on how participants could facilitate discussions between differing perspectives played a huge role in the movement. DataRescue was a partnership in that there were competing interests and widely differing perspectives on how things should be saved. Danielle Robinson in her guide for the Portland DataRescue (see Table 1) mentioned, "Usually academia does not hack! We form subcommittees." (2017, as cited by Allen, Stewart, & Wright, 2017). While this may be a simplification, there is some truth to how communal norms influenced the expectations and motivations of DataRescue events. Hack-a-thons tend to be one-off swarms of people devoted to a task for a short time to accomplish a lot. DataRescue bridged the hack-a-thon model to the subcommittee model in the midst of both concerned scientists and civic data advocates. The event also took place across many different types of venues and cities (see figure 1), most occurring in colleges and universities but at least one happened in a mixed workspace/shopping center.
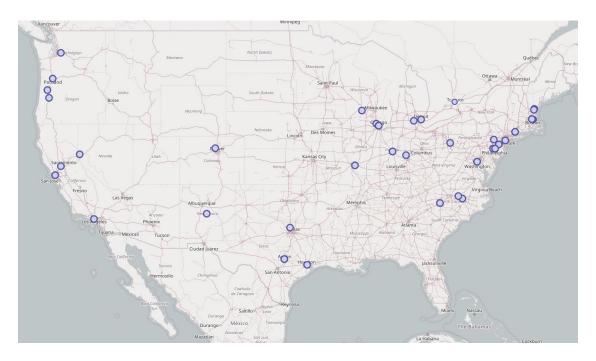
Figure 1. Geographic locations of the 51 DataRescue events.

On "Archiving" (snapshot March 16, 2018), EDGI had begun to promote Data Together, which are intended to be local, communal, and frequent metadata editing sessions. This draws on the collective strength of a hack-a-thon as well as the sustaining slowness of community development.

"A huge part of what we were trying very hard to do was getting different communities to talk to one another," Ellie said. The three communities she described were (1) the data producers / scientific community, (2) the Civic Tech community, and (3) the librarian and archivist community. These groups had different ideas on what one thing must happen to solve the problem of vulnerable data, and each felt ownership in contributing to the solution. However, they had different strategies for how to do so. As it turned out, one perspective could not address it, since the problem is still not fully understood:

> To some degree, everyone who was insisting, "my community knows exactly how to solve this, we just haven't built the tool," from the civic tech community, and

"we don't have funding or institutional mandate from the library and archival community..." It was sort of like, hey if you already knew and had all the answers, this probably—we wouldn't be here. (Ellie)

Communication between groups also led to partnerships with large research libraries through the Libraries+ Network. The Libraries+ Network builds on existing consortia to plan for future data grabs, in a kind of ex officio FDLP capacity, although these data are not publications as the FDLP would consider them. Numerous projects have arisen out of this huge network, such as the Dat Project, which is building a new "web of commons" (Dat Project, 2017). Like both Data Refuge and EDGI, web archiving and open data are only part of what the Dat Project hopes to achieve.

4.3 Impact of Federal Record-Keeping Laws

There was a great deal of fear that environmental data would be taken offline. However, it was rarely seen as a records or legal issue. EDGI portrayed the availability of federal environmental data as part of an infrastructure: "Dismantling this infrastructure—which ranges from databases to satellites to models for climate, air, and water—could imperil the public's right to know, the United States' standing as a scientific leader, corporate accountability, and environmental protection" ("About," snapshot taken September 5, 2017). Ellie developed a framework for understanding vulnerability of data that looked at various threats. Federal laws that govern the preservation of federal data were among several considerations, including technical, social, political, and institutional threats. She also worked to identify not only what legal requirements were in place to ensure that data was collected, stored, and made available, but also the enforcement mechanisms that ensure agencies follow the legal requirements.

DataRescue events formed around these laws reflected an understanding of the persistent distinction between law, policy, and practice. The definition of a federal record mattered less than the practices of the federal government: "It's less about what the definition is for me and [more about] what that definition means in practice for how much work this would take" (Steven).

There were mixed perspectives on how government regulation could support the long-term accessibility of data. Helen stated that the OSTP memorandum and the Obama-era "Managing Government Records" memorandum were not as desirable as federal statutory requirements might be; however, she also saw those as indicative of how complicated ensuring public access to data can be. The OSTP memo applied mostly to the recipients of federal grants, rather than the agencies themselves.

Sarah Lamdan, an expert in environmental legal research and a member of EDGI, concluded that the federal government should be required to preserve the digital media it creates, since federal agencies have the ability to do so. She pointed to the EPA Archive the EPA made public in February 2017 as evidence of the agency's ability to preserve and make accessible the data (Lamdan, 2018). Lamdan called for an update to law around federal record-keeping to address and enforce federal data access. However, the EPA archive was far from perfect and presented some of the same access challenges often found in open data.

Links in the EPA Archive often directed to live links that did not hold the needed information; for example, a link that was supposed to lead to a Spanish-language educational handout instead directed users to a live EPA site that did not contain Spanish-language information (Environmental Protection Agency, 2004; Environmental

Protection Agency Archive, n.d.). This lack of versioning—identifying what was current

data and what was archived data, even on the EPA Archive site—made it difficult for

users to find information. The searchability of the EPA Archive was also severely

limited.

4.4 Consideration of the User

The main consideration of the user was in supporting the cite-ability of the data.

Other considerations included the selection of datasets, findability, accessibility, or ease

of use. Cite-ability was how to ensure provenance. This influenced both where data went

and how DataRescue participants have described the data. The bulk of the data in a place,

the Internet Archive, that does not have very good discoverability affordances. Even

though the Internet Archive does not have strong discoverability properties—it is

searchable by URL—organizers felt that it had strong cite-ability properties. In order for

the data to be used as they intended, it needed this cite-ability: "Scientists don't cite data

in their—you don't build a bridge because someone on the internet said they got data that

says this is what the flood plain is going to look like" (Ellie). Cite-ability in this case,

rather than referring only to the presence of a persistent identifier, refers to the basis for

provenance, for how data moved from its original federal platform to the Internet

Archive, and its original context.

Still, the user was at the heart of what DataRescue archiving was trying to

accomplish. If data was not discoverable and as close to research quality as possible,

there would be little point to doing what they were doing. The Internet Archive was

determined to be a good platform for supporting the cite-ability of the archived websites,

even if it is possible to manipulate archived sites (Nelson, 2018). The greater challenge was designing the Data Refuge repository to support cite-able data.

The metadata hosted on Data Refuge were never meant to support discoverability, as Ellie described. The Data Refuge repository is an instance of the Comprehensive Knowledge Archive Network (CKAN), which is an open source data management system; data.gov is also an instance of CKAN. Data Refuge's site includes archived complex data objects that could not be crawled using Archive-It. The metadata on the Data Refuge repository was, according to Ellie, "more about how the data was moved through the system and who had it" rather than what the datasets actually are.

Snapshots of the Data Refuge "Datasets" site lacked much of the metadata recommended by groups to support discoverability. For example, when compared against EPA's recommendations for metadata documentation for data in the Environmental Dataset Gateway (Environmental Protection Agency, 2014), most of the datasets, particularly early in the project, had very little descriptive information. The Data Refuge site was a storage mechanism for data rather than a site for discovery. Metadata included information on which DataRescue event bagged the data. For example, a snapshot of the Data Refuge site on January 26, 2017, had two Organization categories, "Holding Area Philly" and "test org." The only group was "Metadata," and included language from the CKAN template. ("Groups"). Tags were uncontrolled headings and included both "EPA" and "epa."

One challenge to creating metadata up front was the perceived urgency. Steven described how this was the "smash-and-grab" phase of data archiving: "We were thinking the basics, the basic metadata schema pulled from different standards…but this was

mostly like, 'Okay, if this goes away tomorrow, we have it here, at the very least.'" The

concerns that went into building a repository, and the attention to the development of a

system to ensure continued access, was less at the forefront for the groups in their data

archiving. However, Steven saw this as Phase I of an ongoing move to make data more

findable. With another organization, Steven was planning to assess what is vulnerable

and then determine a method for reconstitution of EDGI-archived data. He could see the

existence of a repository like one proposed to enhance access to Environmental Impact

Statements, which are created by various agencies at both the state and federal level

through the creation of a centralized database (Wentz, 2016).

Students at Ellie's institution are currently adding descriptive metadata to the

datasets in the Data Refuge repository. She hopes the data will find an eventual home in

the MIT Libraries Dataverse. This may also increase the long-term sustainability of the

data, once it has more description.

At some DataRescue events, organizers asked subject area specialists what data

they relied on. Through a partnership with the Union of Concerned Scientists, Data

Refuge organizers sent out surveys to determine similar dependencies on datasets.

Organizers encouraged groups to include local expertise at DataRescue events on

identifying data, but also as part of the wider motivation, to draw attention to the

vulnerability of data. Ellie described how they "wanted the workflow to accommodate as

much local expertise as possible, but that was … more difficult than expected.

4.5 Differences Between This and Other Archiving Movements

The scale of data, the challenge in defining at-risk environmental data, and the

development of environmental data justice made this project different from those that

came before it. Similarities included the challenges in dealing with a wide array of format and size to the data. Since there was an existing infrastructure, the End of Term Harvest's model was used for part of the data collection. However, the sheer scale of the data meant that things had to be different than in open data initiatives with defined scope or with a narrower-subject-focused scope.

Non-government and non-institutional data initiatives typically have one repository and are smaller in scope. The Free Law Project gives people who use PACER—the federal district courts and appeals records system—a browser extension to add the court records they download to a RECAP Archive. Like DataRescue, it is a distributed nominating and voluntary data archiving system; however, unlike DataRescue, it has one nominating or pointing tool that is directed at one set and system of records.

It is also different because of how many people have participated. Attendances at events ranged from twenty to several hundred (Penn Program for Environmental Humanities, 2017). The DataRescue movement generated more attention than any other previous web archiving project (Environmental Data & Governance Initiative, 2018) and sparked the interests of many different groups of people. As Lamdan wrote, "It was like planning a dinner party in your studio New York City apartment only to find out that the whole neighborhood wanted to attend" (Lamdan, 2018, p. 7). The actual count of individuals who have contributed to the DataRescue is unknown; organizers felt like it would be unethical to track people's involvement, since participants did not consent to that upon participation.

Owing to its basis in environmental data justice, DataRescue also questioned the presumption that large-scale data collection and preservation are the best ways to rescue data. Instead, DataRescue sought to balance the value of transparency with a value for protecting individual privacy by not collecting certain data. When information transfers from a private sphere to a publicly accessible one without the consent of information creator, it can put the information creator's well-being in jeopardy (Robertson, 2016; Noble, 2018). Organizers considered who would be able to use this data and who controlled access to it. They chose not to track participants' information in order to protect their privacy. One of the main channels of communication for participants, the Slack messaging service, required email registration to set up the accounts, but after participants requested to be added, organizers deleted the records they had of the email addresses. Storytellers at DataRescue events followed guidance not to take photographs of participants who did not want their likenesses circulated. As "at-risk environmental data" came to include information that was never on federal websites, this principle of stewardship and respect for individual privacy led to a more nuanced interpretation of transparency.

5. Implications for Future Study

DataRescue has implications for environmental data justice, storytelling in digital archiving and data preservation, and the development of other archiving communities working together. Archivists and data curators could explore how the legacy of DataRescue affects the emerging field of environmental data justice and other interdisciplinary fields of study. Another study could explore, rather than archivists perceiving themselves as activists, whether activists moved towards being archivists.

Allen, Stewart, and Wright (2017) documented the importance of storytelling in the DataRescue movement and the commitment to storytelling in the first workflows. Storytelling as another form of metadata can have implications for future archiving occurrences and could become a standard documenting procedure.

This also leads to a greater understanding of how practice, policy, and law play out in the federal record-keeping setting. As this study found, archivists and librarians brought a unique perspective on the understanding that institutional culture and support can have on the viability of projects. Further research can be conducted on how this relates to the success of the DataRescue movement. Government documents librarians have long pointed out that there are websites that fall in between differing responsibilities or through the cracks of one responsible party's container.

6. Bibliography

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2014). *El profesional de la información, 23*(6):582-588. http://www.elprofesionaldelainformacion.com/contenidos/2014/nov/04_eng.pdf

Allard, S., Lee, C., McGovern, N., & Bishop, A. (2016). *The open data imperative: How the cultural heritage community can address the federal mandate*. Washington, D.C.: Council on Library and Information Resources. https://www.clir.org/wp-content/uploads/sites/6/2016/07/pub171.pdf

Allen, L., Stewart, C., & Wright, S. (2017). Strategic open data preservation: Roles and opportunities for broader engagement by librarians and the public. *College & Research Libraries News 78*(9):482. https://doi.org/10.5860/crln.78.9.482.

Anderson, J., & Rainie, L. (2012, July 20). The future of big data. *Pew Research Center, Internet & Technology.* Retrieved 3 December 2017 from http://www.pewinternet.org/2012/07/20/the-future-of-big-data/

Bailey, J., Grotke, A., & Phillips, M. (2016a). After the Harvest: Preservation, Access, and Researcher Services for the 2016 End of Term Archive (Presentation). CNI. Retrieved from https://docs.google.com/presentation/d/1_eBnTjaTeFnD5diGhIWDy00Vgousffl5kCF-ErZ0wZ0/edit#slide=id.g153e4cbb38_0_0

Bailey, J., Grotke, A., & Phillips, M. (2016b). Harvesting Democracy: Archiving Federal Government Web Content at End of Term (Presentation). AALL. Retrieved from https://docs.google.com/presentation/d/17-VoM_8ykrWdX-ys-L66xedPwTcQBVsdcW-kQdfVoDc/edit#slide=id.g153e4cbb38_0_0.

Baish, M. A., Etkin, C., & Walls, D. (2015). National Plan for Access to U.S. Government Information. FDLP Academy. Accessed 24 November 2017. Retrieved from https://www.fdlp.gov/file-repository/about-the-fdlp/gpo-projects/national-plan-for-access-to-u-s-government-information/2700-national-plan-for-access-to-u-s-government-information-a-framework-for-a-user-centric-service-approach-to-permanent-public-access

Bellardo, L. J. (2001). Memorandum to agency records officers and information resource managers: Snapshot of agency public web sites (NWM 05.2001). National Archives and Records Administration. Accessed 21 November 2017. Retrieved from https://www.archives.gov/records-mgmt/basics/snapshot-public-web-sites.html

Berners-Lee, T. (2010). Is your Linked Open Data 5 Star? Linked data (personal website). Accessed 2 December 2017. https://www.w3.org/DesignIssues/LinkedData.html

Bower, D. & Walls, D. (2014). Bringing order to chaos: Capturing and preserving the federal web for permanent public access [Webinar]. FDLP Academy. Retrieved from https://www.fdlp.gov/about-the-fdlp/fdlp-academy

Brügger, N. (2005). *Archiving websites: General considerations and strategies*. Aarhus, Denmark: The Centre for Internet Research.

http://cfi.au.dk/fileadmin/www.cfi.au.dk/publikationer/archiving_underside/archiving.pdf

Boyle, E., Chapman, J., Featherstone, S., Frantela, A., Krane, L., Shears, B. (2015).
Empowering the public through open data. Los Angeles: USC Annenberg.
Retrieved from
https://communicationleadership.usc.edu/files/2015/10/CivicTechUSCOpenDataLACounty.pdf

Carver, N. B. (2012). *Scientists' Perception of Endangered Data and Data Reuse*
(Master's Thesis). University of North Carolina at Chapel Hill, Chapel Hill, North
Carolina. https://cdr.lib.unc.edu/indexablecontent/uuid:b2235b8b-2093-467d-9aea-00f705f70763.

Climate Change Resources for Archivists (2017). Accessed 3 December 2017 from
https://www2.archivists.org/groups/regional-archival-associations-consortium-raac/climate-change-resources-for-archivists

Collection Development Office, Library of Congress (2017). Data sets - interim
guidance. Library of Congress collections policy statements supplementary
guidelines. Retrieved from https://www.loc.gov/acq/devpol/datasets.pdf

Costa, M., Gomes, D. & Silva, M.J. (2017). The evolution of web archiving.
International Journal of Digital Libraries, 18(3):191-205. https://doi-org.libproxy.lib.unc.edu/10.1007/s00799-016-0171-9

Currie, M., Donovan, J., & Paris, B. (2018). Preserving for a More Just Future: Tactics of
Activist Data Archiving. In U. M. Munshi & N. Verma (Eds.), *Data science
landscape* (67-78). Singapore: Springer Nature Singapore.

Dat Project (2017, October 11). A distributed data community: Introducing the new Dat website. [Blog post]. Retrieved from https://blog.datproject.org

Data Refuge (2017). Building refuge for federal climate & environmental data. Accessed 2 December 2017. https://www.Data Refuge.org/

Penn Program in the Environmental Humanities (2017). Data rescue events. Accessed 30 November 2017. http://www.ppehlab.org/datarescue-events/

Data.gov (2017). Impact. Accessed 2 December 2017. https://www.data.gov/impact/

Davis, C. (2015, August 23). Preserving our future: Understanding and acting on the implications of climate change on the archival profession. [Presentation]. Retrieved from http://www.slideshare.net/caseyedavis/climate-change-caseydavis

Dillon, L., Walker, D., Shapiro, N., Underhill, V., Martenyi, M., Wylie, S., Lave, R., Murphy, M., Brown, P. (2017). Environmental data justice and the Trump administration: Reflections from the Environmental Data and Governance Initiative. *Environmental Justice 10*(6). https://doi.org/10.1089/env.2017.0020

Drake, J. (2016). Expanding #ArchivesForBlackLives to traditional archival repositories. Accessed 3 December 2017 from https://medium.com/on-archivy/expanding-archivesforblacklives-to-traditional-archival-repositories-b88641e2daf6

End of Term Archive Web Archive (n.d.). US federal web domain at presidential transitions. California Digital Libraries and the Internet Archive. Accessed 24 November 2017. Retrieved from http://eotarchive.cdlib.org/

Environmental Data & Governance Initiative (2017). Mission, vision, & values. Accessed 2 December 2017. https://envirodatagov.org/about/mission-vision-values/

Environmental Data & Governance Initiative (n.d.). Member bios. Accessed 22 March

    2018. https://envirodatagov.org/about/members/

Environmental Data & Governance Initiative (2018). Press coverage. Accessed 30 March

    2018. Environmental Data & Governance Initiative.

Environmental Protection Agency (2014, October 22). Recommendations for metadata

    documentation. Retrieved from

    https://edg.epa.gov/metadata/webhelp/en/gptlv10/inno/EDG_Metadata_Recomme

    ndations.pdf

Environmental Protection Agency (2015a). *Records Management Policy* (EPA

    Classification No. CIO 2155.3). Washington, DC: EPA Office of Environmental

    Information, Office of Information Collection.

Environmental Protection Agency (2015b). Population surrounding 1,388 Superfund

    remedial sites. Office of Communications, Partnerships, and Analysis. Retrieved

    from https://www.epa.gov/sites/production/files/2015-

    09/documents/webpopulationrsuperfundsites9.28.15.pdf

Environmental Protection Agency (2004). Dusty la carpa dorada del asma y sus

    provocadores del asma - revista de muñequitos. Indoor Environments Division

    (6609J), Office of Air and Radiation. 402-F-00-009. Retrieved from

    https://espanol.epa.gov/sites/production-es/files/2015-

    08/documents/dustythegoldfish_sp.pdf

Environmental Protection Agency Archive (n.d.) Homework resources.

    https://archive.epa.gov/students/web/html/homework.html

Erde, J. (2014). Constructing archives of the Occupy movement. *Archives and Records, 35*(2):77-92.

Exec. Order No. 13514 (2009). 3 C.F.R. (revoked by EO 13693).

Exec. Order No. 13642, 3 C.F.R. 244-246 (2013).

Exec. Order No. 13653, 3 C.F.R. 330-336 (2013).

Exec. Order No. 13694, 3 C.F.R. 15871-15884 (2015).

Federal Depository Library Program (2017, February 3). Web Archiving.

https://www.fdlp.gov/project-list/web-archiving

Ferriero, D. S. (2017). NARA's role under the Presidential Records Act and the Federal

Records Act. *Prologue Magazine 49*(2). Retrieved from

https://www.archives.gov/publications/prologue/2017/summer/archivist-pra-fra

Findlay, C. (2016). Archival activism. *Archives and Manuscripts, 44*(3): 155-159.

http://dx.doi.org/10.1080/01576895.2016.1263964

Friedman, L. (2018, March 26). The E.P.A. says it wants research transparency.

Scientists see an attack on science. *The New York Times.*

https://www.nytimes.com/2018/03/26/climate/epa-scientific-transparency-honest-

act.html

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic Books.

Gottlieb, R. (1993). *Forcing the spring: The transformation of the American

environmental movement.* Washington, D.C: Island Press.

Government Publishing Office (2017). Mission, vision, and goals. Accessed 2 December

2017. https://www.gpo.gov/who-we-are/our-agency/mission-vision-and-goals

Hartman, C.N., Murray, K., Phillips, M. (2013). *Classification of the End of Term Archive: Extending collection development practices to web archives. Final report*. https://digital.library.unt.edu/ark:/67531/metadc152437/m2/1/high_res_d/LG-06-09-0174-09_UNT_Feb2013_FINAL.pdf

Hegstad, M. (2017, January 30). Networks seek to archive, protect environmental information from Trump. *Inside EPA: An online news service from the publishers of Inside EPA*. Accessed 29 November 2017. https://insideepa.com/daily-news/networks-seek-archive-protect-environmental-information-trump

Holdren, J. P. (2013, February 22). Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research. Office of Science and Technology Policy, Executive Office of the President. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

International Internet Preservation Consortium (2017). Tools and software: Awesome web archiving. Accessed 24 November 2017. Retrieved from http://netpreserve.org/web-archiving/tools-and-software/

Kriesberg, A., Huller, K., Punzalan, R., Parr, C. (2017). An analysis of federal policy access to scientific research data. *Data Science Journal 16*: 27. http://doi.org/10.5334/dsj-2017-027

Lacey, A. (2017, November 29). CPR Fears Pruitt's 'sophisticated sabotage' could hinder EPA for years. *Inside EPA: An online news service from the publishers of Inside*

EPA. Accessed 29 November 2017. https://insideepa.com/interview/cpr-fears-pruitts-sophisticated-sabotage-could-hinder-epa-years

Lachapelle, E., Borick, C. P., & Rabe, B. (2012). Public attitudes toward climate science and climate policy in federal systems: Canada and the United States compared. *Review of Policy Research, 29*(3): 334-357.

Lamdan, S. (2017). Environmental information: Research, access & environmental decisionmaking. Washington, D.C.: Environmental Law Institute.

Lamdan, S. (2018). *Lessons from DataRescue: The limitations of grassroots climate change data preservation and the need for federal records law reform*. [Manuscript].

Liboiron, M. (2015). Disaster data, data activism: Grassroots responses to representations of Superstorm Sandy. In J. Leyda & D. Negra (Eds.), *Extreme Weather and Global Media*. New York: Routledge.

Library of Congress (n.d.). Web sites and pages: Quality and functionality factors. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. Accessed 21 November 2017. https://www.loc.gov/preservation/resources/rfs/websites.html

Littman, J., Kerchner, D., Wrubel, L. (2017). End of Term 2016 U.S. Government Twitter Archive. Harvard Dataverse, V1. doi:10.7910/DVN/TQBLWZ

Lucia, J., Appel, R., Khanna, D., Nelson, C., Sly, M., Sneff, G. (2017, September 29). Future proofing civic data: Exploring the challenges of preserving open civic data for the long term [White Paper]. Temple University Libraries. Retrieved from https://sites.temple.edu/librarynews/2018/01/31/future-proofing-civic-data/

Meloda (2017). Dataset definition, version 3. Accessed 29 November 2017. Retrieved from http://www.meloda.org/dataset-definition/

Milan, S. (2016). Data Activism as the New Frontier of Media Activism (SSRN Scholarly Paper No. ID 2882030). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2882030

Millar, L. (2017). On the crest of a wave: Transforming the archival future. *Archives and Manuscripts, 45*(2):59-76. https://doi.org/10.1080/01576895.2017.1328696

Murphy, M. (2008). Chemical regimes of living. *Environmental history, 13*(4):695-703. http://www.jstor.org/stable/25473297

National Archives and Records Administration (1992, February). *Records Management in the Environmental Protection Agency: A NARA Evaluation.* Washington, DC: National Archives and Records Administration. https://hdl-handle-net.libproxy.lib.unc.edu/2027/mdp.39015025245401

National Archives and Records Administration (2017, February 10). Federal Laws, Policy & Regulations. Retrieved from https://www.archives.gov/records-mgmt/laws

Nelson, M.L. (2018, March 23). Weaponized web archives: provenance laundering of short order evidence. [Presentation]. National forum on ethics and archiving the web. Slides retrieved from https://www.slideshare.net/phonedude/weaponized-web-archives-provenance-laundering-of-short-order-evidence

Noble, S.U. (2018). *Algorithms of oppression*. New York: New York University Press.

Ntoulas, A., Cho, J., Olston, C. (2004). What's new on the web? The evolution of the web from a search engine perspective. In *Proceedings of the 13th International*

Conference on World Wide Web: 1–12. Accessed 21 November 2017.

http://www2004.wwwconference.org/docs/1p1.pdf

Ogburn, J. L. (2010). The imperative for data curation. *portal: Libraries and the Academy, 10*(2): 241-246. https://doi.org/10.1353/pla.0.0100

"Over 200 Terabytes of the Government Web Archived!" (2017, May 9). The Internet Archive (Blog). Accessed 2 December 2017. Retrieved from http://blog.archive.org/2017/05/09/over-200-terabytes-of-the-government-web-archived/

Petryna, A. (2002). *Life exposed: Biological citizens after Chernobyl.* Princeton: Princeton University Press.

Peyton, J. & Franks, A. (2016). The new nature of things? Canada's conservative government and the design of the new environmental subject. *Antipode 48*(2): 453–473. https://doi.org/10.1111/anti.12179

Poole, A. H. (2015). How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science, 15*(2): 101-139. https://doi-org.libproxy.lib.unc.edu/10.1007/s10502-014-9236-y

Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation, 72*(5): 961-986. https://doi-org.libproxy.lib.unc.edu/10.1108/JD-10-2015-0123

Phillips, M., Chudnov, D., & Jacobs, J. (2016). Exploratory analysis of the End of Term Web Archive: Comparing two collections. *Web Archiving and Digital Libraries, JCDL 2016*.

https://digital.library.unt.edu/ark:/67531/metadc854115/m2/1/high_res_d/wadl_2016_eot.pdf

Phillips, M. (2011). TooManyWebsites.gov. The White House (Blog). Obama White House Archives. Retrieved from

https://obamawhitehouse.archives.gov/blog/2011/06/13/toomanywebsitesgov

"Presidential Memorandum, Managing Government Records" (2011, November 28). Obama White House Archives. Retrieved from

https://obamawhitehouse.archives.gov/the-press-office/2011/11/28/presidential-memorandum-managing-government-records

Presidential and Federal Records Act Amendments of 2014, Pub. L. 113-187, 128 Stat. 2003, codified as amended at 44 U.S.C. 101 §2101.

Presidential records in the new millennium: Updating the Presidential Records Act and other federal recordkeeping statutes to improve electronic records preservation: Hearing before the Committee on Oversight and Government Reform, House of Representatives, 112th Congress, 9 (2011) (Testimony of David S. Ferriero).

Robertson, T. (2016, March 20). Digitization: Just because you can, doesn't mean you should. [Blog]. Retrieved from www.tararobertson.ca

Robinson, D. (2017). Data-Rescue-PDX: Volunteer guide, and other materials for DATA RESCUE PDX. Retrieved from https://github.com/daniellecrobinson/Data-Rescue-PDX.

Russell, D., & Vandeven, K. (2017, January 10). Project spotlight: Women's March on Washington archives project. SAA Women Archivists Section. Accessed 3 December 2017 from

https://womenarchivistsroundtable.wordpress.com/2017/01/10/project-spotlight-womens-march-on-washington-archives-project/

Schell, J. (2018). *One-year review of data rescue efforts*. [Manuscript]. Accessed 25 March 2018 from https://docs.google.com/document/d/1vCUALNoGvHE4F2AkcbvXvjLmMGrxx4CP1gYayc5RZdw/edit?usp=sharing

Sellie, A., Goldstein, J., Fair, M., & Hoyer, J. (2015). Interference Archive: a free space for social movement culture. *Archival Science, 15*(4): 453–472. https://doi.org/10.1007/s10502-015-9245-5

Shveiky, R., & Bar-Ilan J. (2013). National libraries' traditional collection policy facing web archiving. *Alexandria, 24*(3):37-72. https://doi-org.libproxy.lib.unc.edu/10.7227/ALX.0001

Society of American Archivists Council (2015, November). Issue brief: Strengthening of federal records authority. Society of American Archivists. Retrieved from https://www2.archivists.org/statements/issue-brief-strengthening-of-federal-records-authority

Stempler, A. (2013). The use and availability of environmental activism collections in academic archives. *CUNY Academic Works*. Retrieved from https://academicworks.cuny.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1012&context=si_pubs

Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Summers, E., & Punzalan, R. (2017). Bots, seeds and people: Web archives as infrastructure. *ACM Conference on Computer-Supported Cooperative Work and Social Computing Proceedings*. https://arxiv.org/pdf/1611.02493.pdf

Swartz, N. (2008). GAO: NARA not doing its job. *Information Management Journal, 42*(5), 10. Retrieved from http://libproxy.lib.unc.edu/login?url=https://search-proquest-com.libproxy.lib.unc.edu/docview/227758665?accountid=14244

Tansey, E. (2015). Archival adaptation to climate change. *Sustainability: Science, Practice and Policy, 11*(2):45-56. https://doi.org/10.1080/15487733.2015.11908146

Tansey, E. (2016). Archives without archivists. *Reconstruction: Studies in Contemporary Culture, 16*(1):11.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society,* July-December 2017: 1-14.

Taylor, D. (2014). *Toxic communities: Environmental racism, industrial pollution, and residential mobility*. New York: NYU Press.

"The Manifesto" (n.d.). Penn Program in Environmental Humanities. http://www.ppehlab.org/manifesto/

"Toronto 'guerrilla' archivists to help preserve US climate data" (2016, December 15). *BBC*. Retrieved from http://wayback.archive-it.org/8311/20161229232250/http://www.bbc.com/news/world-us-canada-38324045

Transfer of Records to the National Archives of the United States, 36 C.F.R. §1235 (2009).

Wentz, J. (2016). *Using online databasing to unlock the full value of environmental impact assessments*. New York: Sabin Center for Climate Change Law, Columbia Law School.

Welch, T. (1999). 'Green' archivism: The archival response to environmental research. *The American Archivist, 62*(1):74-94.

Whitington, J. (2017, January 17). Environmental data, guerrilla archiving, and the Trump transition. Dispatches. *Cultural Anthropology* https://culanth.org/fieldsights/1048-environmental-data-guerrilla-archiving-and-the-trump-transition

Whitmire, A., Briney, K., Nurnberger, A., Henderson, M., Atwood, T., Janz, M., Kozlowski, W., Lake, S., Vandegrift, M., Zilinski, L. (2016). A table summarizing the federal public access policies resulting from the US Office of Science and Technology memorandum of February 2013. Version 5. figshare Fileset. https://doi.org/10.6084/m9.figshare.1372041.v4

Zinn, H. (1977). Secrecy, archives, and the public interest. *Midwestern Archivist, 2*(2): 14-26. http://digital.library.wisc.edu/1793/44118

Appendix A: Recruitment Email

Subject: Master's Paper Study Interview on Motivations Involved in DataRescue

Dear (Participant's Name),

I'm writing today to invite you to consider participating in a research study I am conducting as part of my MSLS degree at UNC – Chapel Hill.

The purpose of this study is to learn more about the motivations of archivists, librarians, and others that led to their participation in DataRescue events. I am focusing on EDGI and Data Refuge due to the organizations' initial involvement in DataRescue events and their coordination with dataset mirroring and archiving. This study will allow for greater understanding of the archival community with respect to federal record-keeping and the considered use of archived federal data. It considers the motivations of participants of DataRescue events.

I am recruiting participants who have been actively involved in the organization of DataRescue events and either EDGI or Data Refuge. Your participation would be limited to an interview of approximately 30 minutes that can take at an agreed upon location or via video conference.

If you'd like to participate, please respond, and I will send you a consent form to consider. All participation is voluntary, and you can withdraw at any time. There is no compensation.

If you have questions or concerns about your rights as a research subject, you can contact the Institutional Review Board at UNC at 919-966-3113 or by email at IRB_subjects@unc.edu.

Thank you for your time, and hope to hear from you soon!

Best regards,

Whitney Ray
IRB Study #: 18-0012

K. Whitney Ray
Research Assistant, Penn State University Libraries
Intern (Contractor), The U.S. Environmental Protection Agency Research Triangle Park Library
M.S.L.S. Candidate 2018, University of North Carolina - Chapel Hill

Appendix B: Semi-Structured Interview Questions

1. What is your profession/role at your institution?
2. Could you describe your involvement in Data Refuge/EDGI and DataRescue events?
3. What led you to being involved in EDGI/Data Refuge/organizing events?
4. When archiving federal datasets and websites, did you consider the eventual user? In what ways?
5. Did existing federal government recordkeeping and publishing laws affect your involvement in DataRescue events? In what ways?

Appendix C: Websites

Website Title	Date	Organization	URL to Archived Site
(Environmental Data Governance Initiative) Get Involved	2017-01-25	EDGI	https://web.archive.org/web/20170125154526/https://envirodatagov.org/
(Home) The EPA Under Siege	2017-06-30	EDGI	https://web.archive.org/web/20170630235402/https://envirodatagov.org/
(Home) The First 100 Days and Counting Part 2: Pursuing a Toxic Agenda	2017-12-03	EDGI	https://web.archive.org/web/20171203003555/https://envirodatagov.org/
About	2017-01-26	EDGI	https://web.archive.org/web/20170126172924/https://envirodatagov.org/about/
About	2017-09-05	EDGI	https://web.archive.org/web/20170905141443/https://envirodatagov.org/about/
About	2017-12-23	EDGI	https://web.archive.org/web/20171223091231/https://envirodatagov.org/about/
Agency Forecasts (Later, redirected to "Archiving")	2017-01-26	EDGI	https://web.archive.org/web/20170126141421/https://envirodatagov.org/agency-forecasts/
Agency Primers	2017-03-23	EDGI	https://web.archive.org/web/20170323070055/https://envirodatagov.org/agencyprimers/
Agncy Primers (Oops! That page can't be found)	2018-03-30	EDGI	https://web.archive.org/web/20180330015651/https://envirodatagov.org/agencyprimers/
Archiving	2017-09-05	EDGI	https://web.archive.org/web/20170905141512/https://envirodatagov.org/archiving/
Archiving	2018-03-16	EDGI	https://web.archive.org/web/20180316145411/https://envirodatagov.org/archiving/
Data Event Toolkit (Redirects to DataRescue)	2017-01-24	EDGI	http://archive.today/2017.01.24-145547/https://envirodatagov.org/event-toolkit/
Data Refuge Stories	2018-02-27	DataRefuge	https://web.archive.org/web/20180227144612/http://stories.datarefuge.org/
DataRefuge: Building refuge for federal climate & environmental data	2017-01-22	DataRefuge	https://web.archive.org/web/20170122201347/https://www.datarefuge.org/
DataRefuge: Building refuge for federal climate & environmental data	2017-07-01	DataRefuge	https://web.archive.org/web/20170701091308/https://www.datarefuge.org/
DataRefuge: Building refuge for federal climate & environmental data	2017-12-11	DataRefuge	https://web.archive.org/web/20171211214447/https://www.datarefuge.org/
DataRescue	2017-09-05	EDGI	https://web.archive.org/web/20170905141618/https://envirodatagov.org/datarescue/
DataRescue	2017-03-30	EDGI	https://web.archive.org/web/20180330015133/https://envirodatagov.org/datarescue/
DataRescue Events	2017-07-23	DataRefuge	https://web.archive.org/web/20170723201446/http://www.ppehlab.org/datarescueworkflow
DataRescue Workflow: An Overview	2017-02-23	DataRefuge	https://web.archive.org/web/20170223162229/http://www.ppehlab.org/datarescueworkflow

Website Title	Date	Organization	URL to Archived Site
Datasets	2017-01-28	DataRefuge	https://web.archive.org/web/20170128021627/https://www.datarefuge.org/dataset
Datasets	2018-03-30	DataRefuge	https://web.archive.org/web/20180330001203/https://www.datarefuge.org/dataset
Groups	2017-01-26	DataRefuge	https://web.archive.org/web/20170126060831/https://www.datarefuge.org/group
Historical Page: DataRescue Events	2018-03-30	DataRefuge	https://web.archive.org/web/20180330002501/http://www.ppehlab.org/datarescueworkflow
Mission, Vision, Values	2018-03-30	EDGI	https://web.archive.org/web/20180330012248/https://envirodatagov.org/about/mission-vision-values/
Next Phase of Archiving: Data Together	2017-09-05	EDGI	https://web.archive.org/web/20170905141847/https://envirodatagov.org/archiving-data-together/
NOAA: National Weather Service (NWS)	2018-03-30 (attempted)	Google (EDGI-linked)	Error
Partners	2017-09-05	EDGI	https://web.archive.org/web/20170905141506/https://envirodatagov.org/about/partners/