
ESSAY

LESSONS FROM DATA RESCUE: THE LIMITATIONS OF GRASSROOTS CLIMATE CHANGE DATA PRESERVATION AND THE NEED FOR FEDERAL RECORDS LAW REFORM

SARAH LAMDAN†

INTRODUCTION

Shortly after Donald Trump's victory in the 2016 Presidential election, but before his inauguration, a group of concerned scholars organized in cities and college campuses across the United States, starting with the University of Pennsylvania,¹ to prevent climate change data from disappearing from government websites. The move was led by Michelle Murphy, a scholar who had previously observed the destruction of climate change data and muzzling of government employees in Canadian Prime Minister Stephen Harper's administration.² The "guerrilla archiving" project soon swept the nation, drawing media attention as its volunteers scraped and preserved terabytes³ of climate

† Professor and librarian, CUNY School of Law. Many thanks to all of the volunteers at the Environmental Data and Governance Initiative and Data Refuge, and to CUNY faculty, staff and students for their support.

¹ See *DataRescue Philly: January 13-14, 2017*, PPEH, <http://www.ppehlab.org/datarefugephilly/> [<https://perma.cc/58Q3-3ZT3>]; see also Zoë Schlanger, *Rogue Scientists Race to Save Climate Data from Trump*, WIRED (Jan. 19, 2017), <https://www.wired.com/2017/01/rogue-scientists-race-save-climate-data-trump/> [<https://perma.cc/8W9R-DGFV>].

² See Schlanger, *supra* note 1 ("Researchers watched as Harper officials threw thousands of books of aquatic data into dumpsters as federal environmental research libraries closed."); see also Lindsey Dillon et al., *Environmental Data Justice and the Trump Administration: Reflections from the Environmental Data and Governance Initiative*, 10 ENVTL. JUST. 186, 187 (2017).

³ A terabyte is a data storage measurement that equals approximately a million bytes, or 1,024 gigabytes. A terabyte of storage can hold enough words that it would take every adult in America speaking at the same time five minutes to say them all. See *Terabyte*, MERRIAM-WEBSTER (Feb. 18, 2018), <https://www.merriam-webster.com/dictionary/terabyte> [<https://perma.cc/62UY-TTGH>].

change and other environmental data and materials from .gov websites.⁴ The archiving project felt urgent and necessary, as the federal government is the largest collector and archive of U.S. environmental data and information.⁵

As it progressed, the guerrilla archiving movement became more defined: two organizations developed, the DataRefuge at the University of Pennsylvania, and the Environmental Data & Governance Initiative (EDGI), which was a national collection of academics and non-profits. These groups co-hosted data gathering sessions called DataRescue events. I joined EDGI to help members work through administrative law concepts and file Freedom of Information Act (FOIA) requests. The day-long archiving events were immensely popular and widely covered by media outlets.⁶ Each weekend, hundreds of volunteers would gather to participate in DataRescue events in U.S. cities. I helped organize the New York DataRescue event, which was held less than a month after the initial event in Pennsylvania.⁷ We had to turn people away as hundreds of local volunteers lined up to help and dozens more arrived in buses and cars, exceeding the space constraints of NYU's cavernous MakerSpace engineering facility. Despite the popularity of the project, however, DataRescue's goals seemed far-fetched: how could thousands of private citizens learn the contours of multitudes of federal

⁴ See *DataRefuge and Related Press*, PPEH, <http://www.ppehlab.org/datarefugenews/> [<https://perma.cc/HSF6-5X58>] (listing a bibliography of news articles discussing the project).

⁵ Memorandum from the Office of Mgmt. and Budget for Heads of Executive Departments and Agencies on Management of Federal Information Resources (Nov. 28, 2000), <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A130/a130trans4.pdf> [<https://perma.cc/XG2R-QLWR>] (“The Federal Government is the largest single producer, collector, consumer, and disseminator of information in the United States.”); see also ENVTL. PROTECTION AGENCY, U.S. ENVIRONMENTAL PROTECTION AGENCY OPEN GOVERNMENT PLAN 4.0 (Sept. 2016), https://www.epa.gov/sites/production/files/2016-09/documents/2016epaopengovplan4_odraft091516update1.pdf [<https://perma.cc/FC6A-4QCY>] (describing an overview of the many data and information systems within the EPA's purview).

⁶ See e.g., Alleen Brown, *A Coalition of Scientists Keeps Watch on the U.S. Government's Climate Data*, THE INTERCEPT (Jan. 27, 2017), <https://theintercept.com/2017/01/27/a-coalition-of-scientists-keeps-watch-on-the-u-s-governments-climate-data/> [<https://perma.cc/ZZN6-HKRV>]; Amina Khan, *Fearing Climate Change Databases May Be Threatened in Trump Era, UCLA Scientists Work to Protect Them*, L.A. TIMES (Jan. 21, 2017), <http://www.latimes.com/science/sciencenow/la-sci-sn-climate-change-data-20170121-story.html> [<https://perma.cc/M2EQ-VFX3>]; Dana Varinsky, *Scientists Across the US are Scrambling to Save Government Research in 'Data Rescue' Events*, BUS. INSIDER (Feb. 11, 2017), <http://www.businessinsider.com/data-rescue-government-data-preservation-efforts-2017-2> [<https://perma.cc/L2UC-Z5RY>]; Activists Work to Preserve Government Environmental Data, NPR (Mar. 11, 2017), <https://www.npr.org/2017/03/11/519807763/activists-work-to-preserve-government-environmental-data> [<https://perma.cc/8XGS-RUAB>].

⁷ Jerome Whittington, *Data Rescue NYC*, FACEBOOK (Feb. 4, 2017), <https://www.facebook.com/events/704371486406298/> [<https://perma.cc/EA6C-9C4Q>] (Facebook Event, ITP-NYU Tisch School of Arts, 721 Broadway New York, NY 10003).

environmental information warehouses, gather the data from all of them, and then re-post the materials in a publicly accessible format?⁸

Participants raised numerous concerns, as well. Computing experts feared that we were working inefficiently; there had to be better ways to crawl the websites more quickly. Information science experts were concerned with how to verify the .gov data. How could we prove that the data was what we said it was and sufficiently document the chain of custody for each bit of information? Scientists were worried that the DataRescue events were not collecting the right data, and that we were missing relevant climate change information from databases, tucked away in the lower branches of website maps. Others worried that the data we were capturing was already outdated. Federal agencies are notoriously slow to post information online,⁹ and most of the data was from 2016 or earlier. Despite these varied concerns, we pressed onwards, investing tens of thousands of working hours into DataRescue efforts.

Months into the DataRescue project, its volunteers had to reckon with the project's infeasibility. The data collection failed to yield a searchable, publicly accessible, complete archive of the federal government's climate change data. The DataRescue project revealed that, no matter how talented and ambitious, private citizens should and could not bear the burden of major .gov data preservation efforts. The project confirmed the need for updated federal records preservation laws and processes.

Preserving federal records is the duty and responsibility of the government itself, and the federal government is uniquely equipped to properly maintain and preserve its data collections. It is nearly impossible for the public to rebuild government data infrastructure, recreating the work of building thousands of .gov web pages, databases, and weblinks. Rather than compelling volunteers to do this official work, federal laws should require government entities to maintain and preserve online access to environmental information and data no matter which direction the political pendulum sways.

⁸ For a discussion of the goals of the Data Project by DataRefuge founder, see *How Data Refuge Works, and How YOU Can Get Involved* (Feb. 6, 2017), <https://sunlightfoundation.com/2017/02/06/how-data-refuge-works-and-how-you-can-help-save-federal-open-data/> [<https://perma.cc/S33D-RNG8>].

⁹ A 2015 study of www.data.gov, the major publishing platform for the "most important federal datasets," cited outdated data as a major issue, finding that most agencies do not commit to regularly updating online datasets, and in the rare cases where agencies commit to updating datasets, they fail to do so 32% of the time. Karine Nahon & Alon Peled, *Data Ships: An Empirical Study of Open (Closed) Government Data*, 2015 48th Hawaii International Conference on System Sciences, <https://pdfs.semanticscholar.org/d642/8cof9caeo26do6f7ac43e61e4c33668ad620.pdf>. [<https://perma.cc/YY73-R6WE>]. FOIA processes are also notoriously slow. See STAFF OF H. COMM. ON OVERSIGHT AND GOV'T REFORM, 114TH CONG., *FOIA IS BROKEN: A REPORT 25* (Comm. Print 2016) (describing the FOIA process as "frustratingly slow"). This could be in part because the federal government's IT infrastructure is riddled with technical issues. See Sean Gallagher, *Why US Government IT Fails So Hard, So Often*, ARS TECHNICA (Oct. 10, 2013), <https://arstechnica.com/information-technology/2013/10/why-us-government-it-fails-so-hard-so-often/> [<https://perma.cc/G9S5-4L2R>].

Federal laws confirm that government records preservation and access are the government's responsibility. The Federal Records Act (FRA) was drafted to ensure the proper maintenance, preservation, and notification before destruction, of records, designed to reassure people like those running the DataRescue project against fears that climate change data could just disappear.¹⁰ However, the FRA was drafted in the 1950s, long before the Internet existed. Legislators at the time did not contemplate a world where most federal records would be "born digital,"¹¹ or created in electronic, rather than print, format and only available online. In a world where information access has shifted to a paperless, search-engine driven forum, the FRA has not been updated to properly address the online accessibility of federal records.

Federal records laws and policies in the U.S., like the FRA, the Presidential Records Act (PRA) of 1978¹² and the Freedom of Information Act (FOIA),¹³ were created to prevent government administrators from engaging in censorship, lack of transparency, and the destruction of relevant government records. The DataRescue project proves that these laws, in their current iterations, are insufficient in the digital information age, and must be reformed to ensure proper records preservation and access. This Essay describes DataRescue's efforts and limitations, and examines the FRA, prescribing statutory updates that would safeguard online access to government records, allowing volunteers across the nation to rest easy knowing that climate change data and other information will be preserved through changing presidential administrations.

THE DATA RESCUE PROJECT

DataRescue was the best of grassroots efforts. It was organic, innovative, and optimistic, driven by the idealistic notions of saving and valuing government data.¹⁴ The project shared the communal and progressive energy of

¹⁰ See Federal Records Act, 44 U.S.C. § 3105 (2012); see also Elizabeth Williams, Annotation, *Validity, construction, and Application of Federal Records Act of 1950*, 44 U.S.C.A. §§ 2101 et seq., with Respect to Presidential Records and Records of Executive Office of President, 24 A.L.R. Fed. 3d Art. 8 (2017) ("The Federal Records Act . . . is a collection of statutes that sets forth federal agencies' duties regarding the creation, management, and disposal of records, and is intended to assure, among other things, accurate and complete documentation of the policies and transactions of the federal government, control of the quantity and quality of records produced by federal government, and judicious preservation and disposal of records.").

¹¹ "Born digital" is defined by the Society of American Archivists as any "information created in electronic format." This definition is necessary to differentiate information "born" online from print information that is later digitized. See *Born Digital*, SOCIETY OF AM. ARCHIVISTS, <https://www2.archivists.org/glossary/terms/b/born-digital> [<https://perma.cc/27VZ-8J7D>].

¹² See 44 U.S.C. §§ 2201–2209 (1981) (regulating preservation of presidential materials).

¹³ See 5 U.S.C. § 552 (2011) (requiring agency information to be made public in specified ways).

¹⁴ Both DataRescue host organizations post similar goals and missions on their websites. See *DataRefuge*, PPEH, <http://www.ppehlab.org/datarefuge/> [<https://perma.cc/7MBM-7HJS>]; see also *Archiving Data*, EDGL, <https://envirodatagov.org/archiving/> [<https://perma.cc/4SKQ-P8L8>].

the women's marches and other events that sprung from the 2016 election.¹⁵ DataRescue offered hope to people concerned with climate change denial and deprioritization, prevalent in the new U.S. administration,¹⁶ as they warily awaited President Trump's decision to withdraw from the Paris Climate Agreement.¹⁷ The initial Guerilla Archiving event that spurred the DataRescue movement "focused on preserving information and data from the Environmental Protection Agency, which has programs and data at high risk of being removed from online public access or even deleted."¹⁸

DataRescue's organizers coalesced into two subgroups: DataRefuge and the Environmental Data and Governance Initiative (EDGI). Both organizations were led by volunteers. Despite time and energy limitations, DataRescue members hosted 49 data rescue events, and almost 1500 people attended those events. Their efforts captured over 200 terabytes of government website materials, including over 70 million web pages and 40 million PDFs.¹⁹ The Internet Archive, host of the larger 2016 "End of Term" (EOT) archiving project,²⁰ tallied 63,076 web pages gathered by DataRescue, making EPA websites the most comprehensively archived of all federal agency web pages in the EOT collection.²¹

¹⁵ See Tanvi Misra, *The Local Resistance to President Trump Begins*, CITYLAB (Jan. 20, 2017), <https://www.citylab.com/equity/2017/01/the-local-resistance-to-president-trump-begins/513191/> [<https://perma.cc/4JFB-2ZGS>] (describing the upswing in grassroots movements following the 2016 Presidential election).

¹⁶ See Michael Greshko, *The Global Dangers of Trump's Climate Denial*, NAT'L GEOGRAPHIC (Nov. 9, 2016), <https://news.nationalgeographic.com/2016/11/president-trump-global-climate-change-denial-environment/> [<https://perma.cc/5XM6-NNEL>].

¹⁷ The President eventually announced the U.S. withdrawal from the Agreement in June of 2017, making the U.S. one of two nations to eschew membership. See *President Trump Announces U.S. Withdrawal From the Paris Climate Accord*, WHITE HOUSE (June 1, 2017), <https://www.whitehouse.gov/articles/president-trump-announces-u-s-withdrawal-paris-climate-accord/> [<https://perma.cc/EV4Y-U969>]. Syria, the other nation, eventually also signed on to the Agreement, leaving the U.S. as the singular nation to skip out on the international agreement to lower greenhouse gases. See Fiona Harvey, *Syria Signs Paris Climate Agreement and Leaves US Isolated*, THE GUARDIAN (Nov. 7, 2017), <https://www.theguardian.com/environment/2017/nov/07/syria-signs-paris-climate-agreement-and-leaves-us-isolated> [<https://perma.cc/EUF4-KJAU>].

¹⁸ Technoscience Salon and Research Unit (@technosciencesalon), *Guerilla Archiving, Saving Environmental Data from Trump*, FACEBOOK (Dec. 17, 2017), <https://www.facebook.com/events/1828129627464671/> [<https://perma.cc/TW5C-JEKS>] (Facebook Event, University of Toronto Faculty of Information, 140 St. George Street, Toronto, Ontario M5S 3G6).

¹⁹ *Archiving Data*, EDGI, <https://envirodatagov.org/archiving/> [<https://perma.cc/4SKQ-P8L8>].

²⁰ The EOT archiving project is held at the end of each Presidential administration to capture administrative webpages because there are no statutory requirements that the federal government archive government web pages before they are changed by a new Presidential administration. See *END OF TERM WEB ARCHIVE: US FEDERAL WEB DOMAIN AT PRESIDENTIAL TRANSITIONS*, <http://eotarchive.cdlib.org/> [<https://perma.cc/R2ZG-E5RP>].

²¹ Dillon et al., *supra* note 2, at 188.

As the project developed, EOT contributions became DataRescue's main goal. While the Internet Archive and its Wayback Machine search engine helps people looking for old versions of websites, it is not the ideal form or format for viewing stores of climate change data. The Wayback Machine does not perform keyword searches or sift through data. Instead, visitors must come to the Wayback Machine search equipped with the URLs they wish to see. To make a useful archive of .gov climate change data, DataRefuge would need to spool up government data from the Internet and then unspool it in a searchable format in an online archive very different from the URL-freeze-frame focus of the Internet Archive. What's more, most of DataRescue's work entailed crawling and capturing the "uncrawlable" databases and materials that the Internet Archive bots could not reach or provide online. Much of the data captured in DataRescue events are stored away on servers and are not searchable or publicly accessible.

DATE RESCUE'S SHORTCOMINGS

While projects that sprung up around DataRescue proved successful, like tracking changes on federal webpages, interviewing EPA employees, and elevating the discussion around environmental data justice,²² DataRescue's ultimate goal—to create continued access to all of the data and records that were at risk of being removed from government websites—was never fully realized. Congressman Mike Quigley, co-chair of the Congressional Transparency Caucus, even acknowledged the futility of archiving government data onto one platform, agreeing that "compiling all the government's data into one archive or platform, even if just on one topic, is an immensely difficult task—for a group of volunteers or even for the government itself." DataRescue's attempt to do just that proved Quigley right.²³ Despite the hundreds of hours of work by thousands of individuals, DataRefuge efforts failed to yield an easy-to-use alternate source of climate change data in a political climate that devalued, altered, and hid that crucial information.

DataRescue is hardly the first effort to store and preserve government records, but it is the first major public attempt to create a comprehensive archive of all the U.S. government's environmental data and information. Government records preservation has been carried out for years by

²² *Id.* at 188-191 (describing the numerous programs that sprung from the DataRescue efforts, including a sustained Web Tracking project, scholarly writing, and timely response to legislative activities).

²³ See Varinsky, *supra* note 6.

individuals²⁴ and organizations.²⁵ However, while DataRescue focused on huge swaths of data, most traditional efforts to digitally preserve government data focus on small-scale records collections and use FOIA as the primary records retrieval mechanism rather than bots and hackers. These small-scale efforts yield fewer, but better curated results, focusing on select bits of fruitful information rather than on an entire field of data. For instance, the National Security Archive focuses on declassified U.S. documents and records related to national security and MuckRock publishes the FOIA correspondence of people who opt in to the MuckRock filing system.

DataRescue's ambitious goals were hampered, in large part, by social and technological issues. These organizational failures illustrate the problematic aspects of large-scale private efforts to preserve .gov web content. Socially, the grassroots enterprise could not possibly form an organization capable of planning for proper data storage and use. There were no blueprints, clear directives, or singular leaders at the helm. DataRescue events generally followed the format of the inaugural archiving event, dividing people into various "paths" and having them do assorted, unrelated tasks rather than working cohesively or discussing big-picture issues associated with the project,²⁶ as there was little time or capacity to evaluate and correct course as events unfolded. Additionally, the voluntary nature of the undertaking meant that there was a lack of expertise at most DataRescue events. While people were interested in the one-time-only DataRescue events, it was hard to gather sustained personal investment in the project. Many specialists who would have been instrumental—government librarians, agency experts, etc.—did not have time to oversee the project. The people that did volunteer scrambled to learn new and unfamiliar skills. Scientists trained themselves to scrape Internet data, and computing experts learned about federal agencies and environmental law.

The main social challenge, however, was the overwhelming magnitude of the necessary tasks. Nobody anticipated the scale of the data nor the number of people who would want to participate. It was like planning a dinner party in your New York City studio apartment only to find out that the whole

²⁴ Individuals like Russ Kick and Michael Ravnitsky file FOIA requests to obtain government records and then share those records online. See GOVERNMENTATTIC.ORG, <http://www.governmentattic.org/> [<https://perma.cc/G42W-KZ2B>] (providing "electronic copies of thousands of interesting Federal Government documents obtained under the Freedom of Information Act"); THE MEMORY HOLE, <http://themoryhole2.org/> [<https://perma.cc/RVJ7-YDZK>] (detailing important documents saved by writer and anthropologist Russ Kick).

²⁵ Organizations like MuckRock and the NSA Archive also collect certain federal records using FOIA requests, providing their curated collections online with annotations and related news stories. See MUCKROCK, <https://www.muckrock.com/> [<https://perma.cc/Y73C-W8MR>]; NAT'L SECURITY ARCHIVE, <https://nsarchive.gwu.edu/> [<https://perma.cc/NG74-YZ6U>].

²⁶ See *Data Refuge Paths*, PPEH, <http://www.ppehlab.org/datarefugepaths> [<https://perma.cc/79W3-LSJT>] (describing the various paths for volunteers).

neighborhood planned on attending. The organizers quickly took on a deluge of participants and found themselves leading a quite large, undefined group of people. As one of the group leaders summarized, “We’re having a heart attack... In the last four days I think we’ve been working 22 hours a day” to react to news reports of changes to climate change data access.²⁷

Technologically, the volume of data collected was enormous, and nobody was sure what to do with it or how it would eventually be used. Many well-intentioned volunteers lacked the requisite training to properly scrape the data. Instead, data was often collected by people who did not know how to properly “bag” it (preserve the data’s integrity as it entered storage) or “tag” it (label the metadata to identify the contents of each retrieved item) to ensure that the information being spooled up could be unspooled in a useful way later. Proper bagging and tagging was imperative to prevent the retrieved data from becoming an incomprehensible, unverifiable jumble.

The grassroots, decentralized nature of the project meant that volunteers were doing things in different ways across the various DataRescue events. There was no way to ensure uniformity in procedures and processes, so inconsistencies abounded throughout the project. For example, while librarians working within DataRefuge meticulously broke apart and labeled the metadata to ensure data integrity, automated webcrawlers were collecting exponentially more data without doing the thought-intensive labeling work. The data collection was also inefficient. “Bagging” was very labor intensive, which turned out to be wasted effort when the bagged data never actually disappeared from .gov websites. Because DataRescue participants did not have a way to intuit which web pages and what data were actually at risk, a lot of the project’s efforts ended up merely preserving readily available .gov resources that never went offline.

Finally, there was no consensus on how to store the data long term. DataRescue did not employ a comprehensive archival organization system that would make the data easy to sort and search. Much of the stored data was not clearly tagged, and the rudimentary primers that EDGI drafted to sort the various federal agencies and programs could not compare to the expertise of agency insiders who built the systems.²⁸ There is no easy way to organize materials from, or even to draw up a sitemap for, existing agency websites without understanding the back-end of the webpages and databases involved. Beyond the cataloging issues, no one source

²⁷ Zoe Schlanger, *Hackers Downloaded US Government Climate Data and Stored it on European Servers as Trump was Being Inaugurated* (Jan. 21, 2017), <https://qz.com/891201/hackers-were-downloading-government-climate-data-and-storing-it-on-european-servers-as-trump-was-being-inaugurated/> [<https://perma.cc/XT8V-FFDQ>].

²⁸ For an example of the primers, see *Primers Database*, EDGI <https://envirodatagov.org/datarescue/primers-database/> [<https://perma.cc/2XVC-6LKC>]. The primers are in essence a brief overview of the administrative office as well as all of the links on that office’s webpage.

could provide or maintain an online repository of the magnitude necessary to make terabytes of data available to the public in a searchable format.

Ultimately, the EPA, whose web pages were at the heart of DataRescue efforts, offered the final salvo, punchline, and proof that the federal government is best suited to provide and preserve its online materials. In mid-February 2017 as DataRescue scrambled to scrape every last corner of the EPA's website, the EPA posted a mirror website that preserved a snapshot of all of the EPA's web pages in its January 19, 2017 form, saving everything preceding Trump's inauguration.²⁹ With the flip of a switch, the federal government created the ideal, perfectly accessible archive that DataRescue had deployed hundreds of individuals to try to duplicate.

Future grassroots efforts to save government data will confront the same social and technological limitations that hindered DataRescue. What would be a straightforward, streamlined preservation project for federal agency experts becomes a Tower of Babel for volunteers spread across the continent. Group leaders juggle creating a harmonious DataRescue workflow with predicting which datasets were most at risk. At best, the work takes a toll on its most dedicated volunteers. One EDGI leader and full-time professor described how it felt to squeeze family, an academic career, and data rescue work into 24 hours: "I'm tired . . . I haven't read a book for pleasure since before Christmas."³⁰ At worst, trying to reach the DataRescue goal was an exercise in futility, as nobody short of a magician could recreate the maze of .gov websites and data collections.

IMPROVING FEDERAL RECORDS LAW: ACCOUNTING FOR ONLINE ACCESSIBILITY AND ADDING SOME TEETH

Changing presidential administrations sometimes lead Congress to update laws to prevent the loss or destruction of federal records. For instance, the Nixon presidency ended in the Watergate scandal, where incriminating Presidential records were destroyed, prompting policymakers to draft new transparency provisions to prevent the erosion of American information access in future presidencies.³¹ Similarly, an administration that threatens

²⁹ See Matt Novak, *The EPA Just Posted a Mirror Website of the One Trump Plans to Censor*, GIZMODO (Feb. 16, 2017), <https://gizmodo.com/the-epa-just-posted-a-mirror-website-of-the-one-trump-p-1792430343> [<https://perma.cc/693V-U9W9>].

³⁰ Emily Atkin, *Inside the Scientists' Quiet Resistance to Trump*, NEW REPUBLIC (Feb. 24, 2017), <https://newrepublic.com/article/140731/inside-scientists-quiet-resistance-trump> [<https://perma.cc/GZ8V-2TAS>].

³¹ See Daniel P. Finney, *Watergate Scandal Changed the Political Landscape Forever*, USA TODAY (June 16, 2012), <http://usatoday30.usatoday.com/news/nation/story/2012-06-16/watergate-scandal-changed-political-landscape/55639974/1> [<https://perma.cc/BBT4-BFV3>] (identifying three major federal open records reforms that passed after Watergate: the Presidential Records Act, the Government in Sunshine Act in 1976, and the Ethics in Government Act).

online access to environmental records should spark updates to records laws that more explicitly protect access to records on government webpages.

While a few provisions in federal records laws have been updated to account for technological advances (for instance, the Electronic Freedom of Information Act Amendments of 1996 mandated online FOIA reading rooms and explicitly incorporating electronic records into FOIA's scope³²), specifically modifying federal records laws to account for the online accessibility of modern government information and provide real consequences for violating records management requirements would better protect online federal records. Congress can mandate the preservation of digital records, including non-government webpages, like social media accounts and links to outside data, and it can guarantee accessible online archives like the EPA mirror website to ensure that rescuing government data is a completely unnecessary task.

In an era where government information is often born digital,³³ government transparency is only ensured when online accessibility to government records is guaranteed.³⁴ Our current federal records laws were drafted pre-Internet,³⁵ when getting government records necessarily involved visiting a library or waiting for a letter to arrive in the mail. The laws prescribe outdated data access practices and they do not require online publication or webpage preservation.

Despite these outdated provisions, the reality is that today, the majority of government-created electronic content is never printed and is only available on the Internet. Online information research and access is the norm, as paper archives and records have largely become relics of the past. New records access laws should acknowledge this reality and treat removing information from .gov websites as akin to burning books in a bonfire.

Congress recognizes the waning value of paper records in libraries. Recent bills have sought to eliminate complimentary print versions of major federal publications, including the Federal Register, in favor of online access to

³² Electronic Freedom of Information Act Amendments of 1996, Pub. L. No. 104-231, 110 Stat. 3048 (codified at 5 U.S.C. § 552(a)(2) (2012)).

³³ See *supra* note 11.

³⁴ For example, over 97% of the U.S. Government Publishing Office (GPO) content is born digital. Lisa Rein, *Government Printing Office Has New Strategies to Keep Presses Rolling*, WASH. POST (Oct. 26, 2012), https://www.washingtonpost.com/lifestyle/style/government-printing-office-has-new-strategies-to-keep-presses-rolling/2012/10/26/d76dcof4-1ef8-11e2-9746-908f727990d8_story.html?utm_term=.59ae17e44d20 [https://perma.cc/YRJ6-MSPC].

³⁵ The FRA, for example, was passed in 1950 and the Freedom of Information Act was enacted in the 1960s. Both have been amended with more online-friendly provisions—the eFOIA Act and the 2014 FRA amendments—but neither have been thoroughly reformed to bring them into the digital information era. See Electronic Freedom of Information Act Amendments of 1996, *supra* note 32; Presidential and Federal Records Act Amendments of 2014, Pub. L. No. 113-187, 128 Stat. 2003.

electronic versions.³⁶ Some laws have even required open access portals for certain types of federal data to be shared online.³⁷ Even the FOIA shifted agency reading rooms online twenty years ago.³⁸

In the FRA—the main federal records preservation law—“records” is defined broadly to include digital media.³⁹ But beyond that expansive definition, the law does little else to ensure proper digital records management, ignoring modern archiving norms and practices for online materials.⁴⁰ For instance, while the National Archives and Records Administration (NARA) digital records strategy says that the agency will store digital content in a trusted Digital Object Repository and properly document digital preservation actions to ensure authenticity,⁴¹ there is no statutory mandate to do so in federal records preservation law.

Even as the legislature fails to bring federal records laws into the Internet age, federal agencies are adjusting programs in recognition of the shift to online access of government records. The Federal Depository Library Program (FDLP), the government program that ensures access to government documents through local libraries, is moving away from sending print materials to libraries and moving toward a digital platform called FDLP eXchange.⁴² Despite administrative changes moving federal records online, federal records laws remain mostly in the pre-digital age, failing to ensure adequate online access preservation. As the government offers more information resources online in lieu of print records, it should update its federal records laws to safeguard online access to those resources.

Federal records laws also lack effective enforcement mechanisms. The records laws are largely discretionary, leaving the NARA Archivist in charge of decisions about what will be preserved and what preservation entails.⁴³

36 *See, e.g.*, Federal Register Printing Savings Act of 2017, Pub. L. No. 115-120.

37 *See, e.g.*, Digital Accountability and Transparency Act of 2014, Pub. L. No. 113-101, 128 Stat. 1146-1153 (requiring the U.S. federal government to transform its spending information into open, accessible data).

38 Electronic Freedom of Information Act Amendments of 1996, *supra* note 32.

39 *See* 44 U.S.C. § 3301(a) (2012).

40 While the National Archives, the federal agency charged with overseeing and managing federal records preservation programs, recently published a strategy for digital archiving, this strategy is not incorporated or discussed in the Federal Records Act, and does not carry the force of law. *See, Strategy for Preserving Digital Archival Materials*, NAT'L ARCHIVES (June 8, 2017), <https://www.archives.gov/preservation/electronic-records.html> [<https://perma.cc/6K8S-YAN5>].

41 *See id.* Several NARA regulations do allude to electronic records, but they hardly comprise a legislative scheme dedicated to full governmental compliance with the newest archival strategies for preserving and providing access to born digital records. *See, e.g.*, Electronic Records Management, 36 C.F.R. § 1236 (2017).

42 *See* FDLP EXCHANGE, <https://www.fdlp.gov/project-list/fdlp-exchange> [<https://perma.cc/64D7-VJLV>] (describing its goal of building an application to be a one-stop-shop for libraries in the program).

43 *See, e.g.*, 44 U.S.C. § 3301(b) (2012) (“The Archivist’s determination whether recorded information, regardless of whether it exists in physical, digital, or electronic form, is a record as defined in subsection (a) shall be binding on all Federal agencies.”)

There are little, if any, consequences for administrators who violate records management policies.⁴⁴ The FOIA makes federal agency records access a right. Part of that right is the proactive disclosure of records of public importance.⁴⁵ All of our federal records management rules, and especially the FRA, should be crafted with this right in mind, creating substantive enforcement measures to protect the right to government transparency.

REFORMING THE FRA IN THE DIGITAL INFORMATION AGE

The FRA is the main federal law governing records maintenance and preservation.⁴⁶ At a glance, the FRA appears well-situated to ensure the maintenance and preservation of the government's records, but a closer look reveals an outdated and largely unenforceable records management scheme.

A 2014 FRA amendment attempted to pull the law into the digital era by expanding the definition of "federal records" to include "all recorded information," thereby including electronic records.⁴⁷ Changing the definition, however, did not inspire innovation in federal records management. Although the amendment tasked the NARA Archivist with identifying what recorded information falls within the FRA's scope,⁴⁸ the Archivist has not created a definitive guide on what electronic records must be preserved, and the statute is silent regarding non-government-hosted records, including social media postings and government-hosted links to non-governmental content.⁴⁹ There is

⁴⁴ The only enforcement action permitted by the Federal Records Act for unlawfully destroyed or removed records is an Attorney General action to recover the records. Under 44 U.S.C. § 3106 the Attorney General can initiate an action to recover records that have been unlawfully destroyed or removed if a federal agency tells the National Archivist about unlawfully removed or destroyed records. In situations where agencies fail to report, the statute empowers the National Archivist itself to report the unlawful removal or destruction of federal records. Beyond that statutory provision, there are no citizen suit provisions or other ways for the public to ensure that records are not unlawfully removed or destroyed.

⁴⁵ 5 U.S.C. § 552(a)(1-2)(2012) (requiring proactive disclosure of many types of public records).

⁴⁶ 44 U.S.C. §§ 2101-2118 (2012). The law covers federal agency records and is enforced by NARA. See David S. Ferriero, *NARA's Role Under the Presidential Records Act and the Federal Records Act*, NAT'L ARCHIVES: PROLOGUE MAG. (2017), <https://www.archives.gov/publications/prologue/2017/summer/archivist-pra-fra> [<https://perma.cc/42X4-8ZQ9>]. Another records management statute, the Presidential Records Act (PRA), governs the management of non-agency executive office records, including the records of the Executive Office of the President and the National Security Council. See *Armstrong v. Exec. Office of the President, Office of Admin.*, 1 F.3d 1274, 1293 (D.C. Cir. 1993) (holding that National Security Council records fall under the PRA, which is distinct and mutually exclusive from the FRA, and less rigorously protected).

⁴⁷ Presidential and Federal Records Act Amendments of 2014, Pub. L. 113-187, 128 Stat. 2003, amending 44 U.S.C. 3301(a) (2012).

⁴⁸ Presidential and Federal Records Act Amendments of 2014, Pub. L. No. 113-187, 128 Stat. 2003, 209 (codified at 44 U.S.C. 3301(b) (2012)) ("The Archivist's determination whether recorded information, regardless of whether it exists in physical, digital, or electronic form, is a record as defined in subsection (a) shall be binding on all Federal agencies").

⁴⁹ Notably, the NARA told the Trump administration to save tweets in order to comply with the PRA, which indicates that they may fall under the scope of the law, although the National Archivist is

no policy consensus on whether all types of webpage materials or social media posts are “federal records” or what their place is in the preservation scheme.

An updated version of the FRA and PRA would specify that presidential and executive social media accounts and other online platforms used by the administration to discuss federal policies and plans, or as an administrative mouthpiece, must be preserved. Similarly, online items that federal officials link to through their webpages or social media accounts would be included in the definition of federal or presidential records. Incorporating privately owned digital content created or referenced by government administration into federal records preservation laws could be accomplished simply by adding official content on external online platforms to the definitions of “records” in federal records laws rather than leaving the task of choosing what counts as a federal record to the National Archivist.

PRESERVING ONLINE ACCESS TO GOVERNMENT RESOURCES

Most importantly, the FRA does not mandate the preservation of online access to government records. Currently, if records are removed from .gov websites, they are not considered to be destroyed or non-accessible.⁵⁰ Under the current FRA provisions, agencies may remove or scrub agency webpages and links. As long as paper copies or offline electronic iterations of federal records exist and could ostensibly be handed over as a responsive record in a FOIA request, there is no FRA violation associated with taking records offline or modifying online records to remove words or links. Instead, people must file FOIA requests to retrieve the records.

The FOIA process is notoriously time consuming, complicated, and fraught with potential record-blocking exemptions and procedural obstacles.⁵¹ Even

clear that the power to determine what constitutes a presidential record is vested in the President and not the NARA. *See* Letter from David S. Ferriero, Archivist of the U.S., to The Honorable Claire McCaskill & The Honorable Tom Carper, Comm. on Homeland Sec. & Governmental Affairs (Mar. 30, 2017), <https://www.archives.gov/files/press/press-releases/aotus-to-sens-mccaskill-carper.pdf> [<https://perma.cc/Q2H8-WKPL>] (“[R]ecords management authority is vested in the President, and NARA does not make ‘determinations’ with respect to whether something is or is not a Presidential record”).

⁵⁰ While the NARA has created guidelines for webpage maintenance and “trustworthiness” (maintaining reliable and authentic webpages), there is no penalty for changing or removing online content. *See NARA Guidance on Managing Web Records*, NAT’L ARCHIVES (Jan. 2005), <https://www.archives.gov/records-mgmt/policy/managing-web-records.html#1.1>. [<https://perma.cc/2D85-PDRE>]. The NARA describes agency webpages as agency information repositories, linking the public to records that are also available in print format. So long as the online content can be accessed through FOIA, there is no FRA violation associated with removing the content from the online repository.

⁵¹ Electronic Freedom of Information Act Amendments of 1996, Pub. L. No. 104-231, 110 Stat. 3048 (1996), amending 5 U.S.C. 552(a)(2). For a description of some of the obstacles and complications, see STAFF OF H. COMM. ON OVERSIGHT AND GOV’T REFORM, 114TH CONG., FOIA IS BROKEN: A REPORT, *supra* note 9.

FOIA itself acknowledges the need for proactive records release with online access, envisioning electronic reading rooms filled with volumes of agency records.⁵² Yet the FRA fails to recognize the value of click-of-a-button access as opposed to weeks, months, or even an eternity of tug of war with an agency under a FOIA request for access.⁵³ Even worse, if a researcher does not know what records they seek, they will not know what to records to request via FOIA.

Because there are no FRA digital access requirements, it is easy for federal agencies to strip away access to data by underfunding access efforts. For a modern-day records preservation scheme to work, agencies must have federal records management mandates and be given adequate funding to properly carry out the mandates. In the current records management scheme, each administration determines how robustly to support database development and maintenance. For instance, the current administration is proposing to strip the Toxics Release Inventory, one of the largest data collection programs governed by the EPA, of funding and support.⁵⁴ Agencies can effectively starve online data stores by failing to provide the resources required to fill them.

A statutory solution to the accessibility issues associated with born digital government records would be to include explicit language requiring continued access to and online archives for electronic government records. An improved FRA would create separate provisions for born digital and online records to ensure their continued access. The current law does not define electronic records as separate from paper records or specify that “continued records access” means actually maintaining online access to online records, including databases and webpages.

The FRA’s failure to delineate between online and offline records results in a system where scrubbing government webpages and moving government

⁵² See ADMINISTRATION OF WILLIAM J. CLINTON, STATEMENT ON SIGNING THE ELECTRONIC FREEDOM OF INFORMATION ACT AMENDMENTS OF 1996 (Oct. 2, 1996) (“The legislation that I sign today brings FOIA into the information and electronic age by clarifying that it applies to records maintained in electronic format. This law also broadens public access to government information by placing more material on-line and expanding the role of the agency reading room.”).

⁵³ Journalists lament the obstructions they face when they try to access federal records by FOIA request. See The Editorial Board, *Federal Agencies Stiff-Arm FOIA Requests: Our View*, USA TODAY (Mar. 15, 2015), <https://www.usatoday.com/story/opinion/2015/03/15/sunshine-week-foia-government-transparency-editorials-debates/24823085/> [<https://perma.cc/U8XQ-BD8G>] (describing how the federal government “slow-walk[s]” requests and explaining that delays can last for years); see also David T. Barstow, *The Freedom of Information Act and the Press: Obstruction or Transparency?*, 77 SOC. RES. 805, 805 (2010) (asserting that the FOIA makes the author’s “blood boil” because federal agencies use it like a “cloak of invisibility” for government records). Even the U.S. House of Representatives Committee on Oversight and Government Reform published a report titled “FOIA is Broken.” See STAFF OF H. COMM. ON OVERSIGHT AND GOV’T REFORM, 114TH CONG., FOIA IS BROKEN: A REPORT, *supra* note 9.

⁵⁴ See Francie Diep, *The Future of Data Justice Under Trump*, PAC. STANDARD (Nov. 2, 2017), <https://psmag.com/news/future-of-data-justice-under-trump> [<https://perma.cc/5Q4X-65QZ>] (noting the suggestion to cut the Inventory’s budget by more than \$5 million).

records offline is permissible, so long as the records are maintained in some FOIA-able format. The FRA should be amended to include specific provisions that treat removing access to digital records as a form of records destruction, so that destruction schedules and public notices are required before online government materials are removed or online links are broken.⁵⁵ Most people seek government information online and it is imperative that our federal records laws recognize this shift by protecting access to online government records.

In addition to failing to account for modern online records access, the FRA has no real “teeth,” or punitive enforcement mechanism. The FRA requires records preservation⁵⁶ and that agencies notify the public before destroying records.⁵⁷ However, there is little chance that agencies or officers will face consequences for destroying online records if they do not follow FRA procedures, and in fact, there are no major penalties that can be assessed against agencies or administrators who break FRA rules. In contrast to the transparency laws like FOIA, the FRA provides no private right of action to recover agency records wrongfully removed from government custody, and the U.S. Supreme Court has determined that no private right of action should be implied under the law.⁵⁸ Agency heads or the NARA Archivist may choose to pursue judicial action to recover or protect records through the attorney general if any records in their custody face “actual, impending, or threatened unlawful removal, defacing, alteration, corruption, deletion, erasure, or other destruction,”⁵⁹ but this is the only way to enforce FRA provisions. Recent lawsuits against NARA indicate that leaving enforcement to the National Archivist is insufficient. In 2018, NARA and the Archivist allegedly failed to investigate or enforce the FRA despite problematic recordkeeping at the EPA. Two government watchdog groups, Citizens for Responsibility and Ethics in Washington and Public Employees for Environmental Responsibility allege that:

Despite knowledge of the numerous ways in which Administrator Pruitt and the EPA have violated the FRA, the Archivist has failed to fully investigate these violations and effectively inform Administrator Pruitt of NARA’s investigative findings, has failed to make recommendations to Administrator

⁵⁵ These provisions could be appended to the current statutory provisions governing records disposal. *See* 44 U.S.C. §§ 3302–3303 (2012).

⁵⁶ 44 U.S.C. § 3101–3107 (2012).

⁵⁷ 44 U.S.C. § 3303a (2012) (requiring the National Archivist to publish notice in the Federal Register before disposing of federal records).

⁵⁸ *See* *Kissinger v. Reporters Comm. for Freedom of Press*, 445 U.S. 136, 149–50 (1980).

⁵⁹ *See* 44 U.S.C. § 3106(a) (2012).

Pruitt and the EPA for their correction, and has taken no other action despite the failure of the EPA to commence satisfactory corrective measures....⁶⁰

Despite the often ineffective enforcement of federal records laws within agencies and by the National Archivist, lower courts rarely recognize a right to file suit to recover records under the APA,⁶¹ and this cause of action is acutely limited by the deference granted to agencies under the Administrative Procedure Act (APA).⁶²

Like the FRA's scope of records and records destruction requirements, the FRA's enforcement mechanisms should also be updated for the modern age by including a right of action in the statutory language like the FOIA's citizen suit provision. Before the Internet existed, the FOIA was the optimal method for records retrieval and the FRA was the companion law that ensured the behind-the-scenes records maintenance to make FOIA records retrievals possible. When records were only created in paper format, accessing documents and materials necessarily required travelling to a library or reading room; making a phone call and scheduling a meeting; sending a letter and then waiting by the mail; or some other form of physical delivery. The FOIA's drafters did not contemplate the transparency implications of online access or a lack thereof.

The FOIA contains citizen suit provisions because its drafters recognized that government transparency and access to government records should be publicly enforceable.⁶³ Although the citizen suit provision is located in the FOIA, the FRA shares the FOIA's terms for defining a federal record, confirming that records subject to the FRA warrant the same transparency guarantees as those governed by the FOIA.

As we move toward a world where born digital records and online access are the norm, we must ensure that the FOIA's records access provisions also apply to online resources access. Including a FOIA-like citizen suit provision in the FRA would prevent the government from scrubbing its webpages and cutting its webpage and database links. Rather than leaving records preservation enforcement to NARA, the public should be able to intervene to obtain records when the government takes them offline or otherwise hinders access to electronic materials. 44 U.S.C. § 3106, governing the unlawful removal and destruction of

⁶⁰ Complaint at 19, *Citizens for Responsibility and Ethics in Washington v. Pruitt*, 1:18-cv-00406 (D.D.C. 2018).

⁶¹ *See, e.g., Armstrong v. Bush*, 924 F.2d 282, 291-93 (D.C. Cir. 1991) ("In sum, neither the statutory scheme nor the legislative history evinces a congressional intent to preclude judicial review of the adequacy of the [National Security Council's] recordkeeping guidelines and directives.").

⁶² *See, e.g., Competitive Enter. Inst. v. EPA*, 67 F. Supp. 3d 23, 32 (D.D.C. 2014) (finding that the *Armstrong* holding applies to "APA review of an agency's record-keeping guidelines," and does not allow "specific challenges to an agency's disposal determinations").

⁶³ *See* 5 U.S.C. § 552(a)(4)(B) (2012) (granting federal jurisdiction over complaints filed pursuant to the statute).

records, should be modified to provide citizens with a cause of action when the government obstructs online access to government records or destroys online materials without creating an accessible historical archive (like the EPA mirror website). Ensuring the enforcement of federal records laws would relieve public efforts like DataRescue and the EOT from taking on the full and unachievable burden of online federal records preservation.

CONCLUSION

DataRescue was an ambitious grassroots effort that brought together brilliant and motivated academics, scientists, and computer experts. EDGI coined the phrase “data justice”⁶⁴ and brought worldwide attention to the need for better environmental data preservation. The DataRescue events bridged usually siloed schools of thought and training, creating interdisciplinary bonds that will endure in the years to come. Overall, it was an empowering, positive movement.

From a legal perspective, however, DataRescue illuminated the need for updated federal records laws. The task before the volunteers was overwhelming, and DataRescue work was a frenetic attempt by members of the public to protect data that should have been legally protected by the federal government. At the end of the day, it is up to the federal government to ensure that members of the public do not shoulder these efforts. The international community recognizes a right to information access as part of its human rights scheme, especially in the environmental context.⁶⁵ To protect this right, the federal government should ensure that records, especially those containing information and data about environmental safety issues, will not be taken offline on political whims.

Government website observers have noticed that climate change data is already being removed and scrubbed from some government webpages.⁶⁶ The private sector has demonstrated the ephemeral nature of online information and the destruction caused by the simple act of removing online access to information. When reporters voted to unionize, Joe Ricketts, the CEO of major online news outlets DNAinfo and Gothamist, shut down the webpages, instantly removing over a decade’s news stories from the Internet.⁶⁷ As tens of thousands of news stories

⁶⁴ See Diep, *supra* note 54.

⁶⁵ Rebecca Bratspies & Sarah Lamdan, *Human Rights and Environmental Information*, in *HUMAN RIGHTS AND THE ENVIRONMENT: INDIVISIBILITY, DIGNITY, AND LEGALITY* (Encyclopedia of Environmental Law Series, Edward Elgar Press) (forthcoming 2018).

⁶⁶ See Leila Miller, *As “Climate Change” Fades from Government Sites, a Struggle to Archive Data*, FRONTLINE (Dec. 8, 2017), <https://www.pbs.org/wgbh/frontline/article/as-climate-change-fades-from-government-sites-a-struggle-to-archive-data/> [[https://perma.cc/SJ2\]-RA2R](https://perma.cc/SJ2]-RA2R)].

⁶⁷ See Robert Channick, *DNAinfo, Gothamist News Sites Shut Down by Billionaire Joe Ricketts After Union Vote*, CHI. TRIB. (Nov. 3, 2017), <http://www.chicagotribune.com/business/ct-biz-dnainfo-shutting-down-20171102-story.html> [<https://perma.cc/8XVE-XFKN>].

disappeared, people marveled that such a wide swath of news could instantly go dark and reporters scrambled to capture their work from data archives.⁶⁸

Being able to flip the online access switch is intensely powerful, and while it may be permissible to turn information access on and off in the private sector, it is well established that government information belongs to the public. As federal agencies scrub climate change data from government webpages,⁶⁹ they demonstrate the need for an FRA update that incorporates online information access measures and stronger enforcement mechanisms to keep federal website switch flippers and website scrubbers at bay.

Preferred Citation: Sarah Lamdan, *Lessons from DataRescue: the Limitations of Grassroots Climate Change Data Preservation and the Need for Federal Records Law Reform*, 166 U. PA. L. REV. ONLINE 231 (2018), <http://www.pennlawreview.com/online/166-U-Pa-L-Rev-Online-231.pdf>.

⁶⁸ Matthew Flamm, *A Relaunch for Gothamist, but not for DNInfo*, CRAIN'S N.Y. BUS. (Feb. 23, 2018), <http://www.crainsnewyork.com/article/20180223/ENTERTAINMENT/180229941/wnyc-to-relaunch-gothamist-acquires-dnainfo-archive-in-the-deal> [<https://perma.cc/C4UQ-YVT2>]; Andy Newman & John Leland, *DNInfo and Gothamist are Shut Down After Vote to Unionize*, N.Y. TIMES (Nov. 2, 2017), <https://www.nytimes.com/2017/11/02/nyregion/dnainfo-gothamist-shutting-down.html> [<https://perma.cc/WBT6-V7KG>]; David Yanofsky, *DNInfo's and Gothamist's Archives Still Exist and are Likely to be Resurrected*, QUARTZ (Nov. 2, 2017), <https://qz.com/1119220/the-archives-of-dnainfo-gothamist-chicagoist-laist-and-shanghaiist-are-not-deleted/> [<https://perma.cc/YGH2-UEXN>].

⁶⁹ See Coral Davenport, *How Much Has 'Climate Change' Been Scrubbed From Federal Websites? A Lot.*, N.Y. TIMES (Jan. 10, 2018), <https://www.nytimes.com/2018/01/10/climate/climate-change-trump.html> [<https://perma.cc/M56Z-XQHJ>].