The Data Hoarders Resisting Trump's Purge

Can librarians and guerrilla archivists save the country's files from *DOGE*?

By [Julian Lucas](#)

March 14, 2025

Illustration by Erik Carter

From https://www.newyorker.com/news/the-lede/the-data-hoarders-resisting-trumps-purge

The deletions began shortly after Donald Trump took office. C.D.C. web pages on vaccines, H.I.V. prevention, and reproductive health went missing. Findings on bird-flu transmission vanished minutes after they appeared. The Census Bureau's public repository went offline, then returned without certain directories of geographic information. The Department of Justice expunged the [January 6th insurrection](#) from its website, and whitehouse.gov took down an explainer page about the Constitution. On February 7th, Trump sacked the head of the National Archives and Records Administration, the agency that maintains the official texts of the nation's laws, and whose motto is "the written word endures."

More than a hundred and ten thousand government pages have gone dark in a purge that one scientist likened to a "digital book burning," and which has proved as frightening in its imprecision as in its malice. Racing to comply with executive orders [banning "D.E.I."](#) and "gender ideology extremism," agencies have cut materials on everything from supporting transgender youth in school to teaching children about sickle-cell disease, which disproportionately affects people of African descent. But they have also axed records having little to do with the Administration's ideological priorities, seemingly assisted by A.I. tools that flag forbidden words without regard to context. A recently leaked list of pages marked for deletion on military websites includes references to the Enola Gay—not, as it turns out, a member of the L.G.B.T.Q. community but, rather, the B-29 bomber that nuked Hiroshima.

Oblivion menaces every scrap of information that doesn't spark joy in the Oval Office. "It's gone," Trump said of "wokeness," during his [recent address to Congress](#), in almost motherly tones. "And we feel so much better for it, don't we? Don't we feel better?" But on this front, at least, the Administration is facing well-organized resistance. It comes from a loose coalition of archivists and librarians, who are standing athwart history and yelling "Save!" They belong to organizations such as the Internet Archive, which co-created a project called the End of Term Web Archive to back up the federal web in 2008; the Environmental Data and Governance Initiative, or *EDGI*; and libraries at major universities such as M.I.T. and the University of Michigan. Like the Encyclopedists of Isaac Asimov's "[Foundation](#)"—who race to compile a collapsing empire's accumulated knowledge—they're assembling information arks to ride out the chaos.

Get the News & Politics newsletter

The latest from Washington and beyond, covering current events, the economy, and more, from our columnists and correspondents.

By signing up, you agree to our user agreement (including class action waiver and arbitration provisions), and acknowledge our privacy policy.

"This is how we know about our country," Lynda Kellam, a social scientist and data librarian, told me. "People who support the 'drain the swamp' mentality don't seem to understand how much the government does." Kellam, who's served in leadership positions at two Ivy League universities, described the vulnerable data as "irreplaceable." The U.S. government is one of the world's biggest publishers, and its research on everything from carbon emissions to infant mortality is conducted on a scale that few private institutions can match. Kellam told me that she'd participated in a small data-preservation effort during Trump's first term, but had never seen anything like the frenzy that ignited in late January, when the C.D.C. began removing information from its website. (It has since been restored by means of a temporary restraining order.) As the *DOGE*-led assault on civil society continued, her peers began saving files and hosting "datathons" at their universities.

Kellam was encouraged, but worried about a lack of coördination. What if the backups languished on private hard drives? What if archivists duplicated one another's work? She created a Google Doc to centralize information about preëxisting initiatives—an archive of archives, with detailed instructions on how to contribute to each. "It was really just meant to be a place where people could go and nominate things for the End of Term crawl," she said. Within days, though, there were more than a hundred people in the document at any given time. Kellam met with the heads of other data-librarian organizations, and together they founded the Data Rescue Project to preserve the enormous data sets that website-focussed efforts had missed. Its tracker now catalogues more than four hundred publicly accessible volunteer backups of government repositories, from the C.F.P.B.'s Consumer Complaint Database to the C.D.C.'s National Immunization Survey.

"Everything is at risk," Sebastian Majstorovic, who administers the tracker, told me. A digital historian based in Cologne, he previously worked with Saving Ukrainian Cultural Heritage Online, or *SUCHO*, which arose in response to the Russian invasion of Ukraine in 2022. "My dad's from Bosnia, and that's informed my outlook on what can be irretrievably lost," he said, citing the destruction of nearly two million books when Sarajevo's National Library was shelled in 1992. Now he's teaching dozens of Americans to save data from their own rogue government. In early February, he began downloading census records from ftp.census.gov just before it went down—and ended up with only around two hundred gigabytes of data. But strangers online had grabbed other directories, and together they were able to complete the backup. "I think we'll be surprised by how many things have been saved by people we don't yet know, because they haven't had a chance to give it to someone," he told me. When I asked who they were, Majstorovic had a simple answer: "Nerds who care."

They came, in many cases, from r/DataHoarder, a subreddit with nearly a million members devoted to preserving files. The data hoarders collect zines, manuals, family photos, old television shows, and defunct websites—just about everything digital or digitizable at risk of disappearance. Their tastes run a wide gamut. Among the hoards cited on periodic show-and-tell threads are "1,500 90 minute recordings of church services," "15000+ hentai mangas and growing," "a digital collection of Occitan and Piedmontese books," and "someone's grandma's recipes." But the hoarders speak the same language on the subject of digital permanence, swapping tips on storage and sharing glamour shots of their elaborate server "rigs." The subreddit's banner image is a stack of hard drives emblazoned with the words "What do you mean DELETE?!"

If the community has a politics, it's defying the corporate stranglehold on the ownership of media—even personal data, which have been increasingly corralled into proprietary clouds. I was weaned on data-hoarder values by my father, a songwriter and a producer who spent the last decade of his life digitizing his vinyl collection and saving it to a music server that he dubbed "soulbro." The server, a capricious machine with a battery of hard drives, occupied a soundproof box in his home studio. It routinely malfunctioned at family dinners, prompting groans, wisecracks, and the eternal question: Why not just use Spotify?

My father countered that tech companies couldn't be trusted to keep music available, and that in the twenty-first century those who didn't own their media stood to lose it—a view vindicated by the data disasters of the past decade. Platforms such as MySpace and GeoCities have been erased in an instant, destroying vast swaths of online history. Digital outlets—most notoriously, MTV News—have shuttered without warning, leaving even the journalists who wrote for them without records of their work. Customers of e-book stores and streaming services have discovered that they don't really own their media, which can be altered or deleted by rights holders at will. "They do not want you to have or own anything, they want you to rent access to it forever," one DataHoarder wrote of such changes. "The walled gardens are being built all around us, and every year, with every new update or service change or company merger or OS 'security' feature, the walls get a bit higher."

The subreddit has spawned archiving efforts in response to terms-of-service changes at Kindle, the recent near-ban of TikTok, and [Elon Musk's threat](#) to buy Wikipedia. In November, several weeks after [Trump's election](#), a few members began worrying about scientific reference data at federal agencies. They were roundly mocked by right-wing users. "The TDS"—Trump Derangement Syndrome—"is real," one commented. Another: "What is this? Foundation?" Then, in late January, a Bluesky account purportedly run by C.D.C. staffers warned of impending deletions, and the data hoarders spun up their drives.

Video From The New Yorker

[Friendship and Hard Work Amid Italy's Anti-Immigrant Rhetoric in "Fratelli Carbonai"](#)

"That was basically total chaos—posts all day, every day, half of them thanking us and half of them bitching that there wasn't a megathread," Nicholas Serra, one of the subreddit's

moderators, said of the busy first weeks; he created a pinned post to "direct traffic." A software developer in Youngstown, Ohio, he usually archives rock-concert footage, and tries to keep the forum nonpartisan. A few right-wing users have objected to the recent campaign, Serra told me, but he sees data preservation as beyond politics: "I often find myself reminding them of the purpose of the sub, because they seemingly have forgotten."

By mid-February, the Data Rescue Project was recruiting from r/DataHoarder and a few related networks. Majstorovic and others began teaching the less experienced members how to back up government data with ArchiveTeam Warrior—an app whose creators have launched a data-rescue campaign—and to upload it to a secure public repository called DataLumos. Kellam had never even heard of data hoarders before this year. But in the first week, she told me, just one of them contributed an estimated forty per cent of the uploads. They're a largely anonymous bunch, but those willing to speak to me were all young, male I.T. professionals—normie counterparts of the *DOGE* tech tyros joyriding through civil society's back end.

A volunteer named Sze, who works as a product manager at a health-care startup in Richmond, started preserving federal data as an alternative to doomscrolling. "I've been trying to get out to some of the protests, but, with a job and family, that's just difficult," he told me. Every night, after work, he downloads files to an old Dell laptop that he uses as an external hard drive; when we spoke, he was pulling data sets from *FEMA* and the C.F.P.B. Sze was born in Hong Kong. In 2019, he watched helplessly as [the city's protest movement](#) was crushed by mainland authorities; when censorship came to the United States, he was determined not to sit idly by.

Andrew, who goes by the nom de guerre Grumpy—a reference to the dwarf in "Snow White"—is a tech-support professional in Kansas. A self-described centrist, he joined r/DataHoarder for tips on downloading episodes of "[The Office](#)" after it was dropped by Netflix. He specializes in hobbyist videos about old telephone switches, which he stores on a server in his basement. His wife, a family doctor, relies on C.D.C. apps in her practice. Once the agency began removing educational materials from its YouTube channel—such as H.I.V.-prevention tips and Spanish-language tutorials on caring for newborns—he was shocked into preserving as many as he could.

Andrew has since started tracking [disappeared](#) government videos, which range from "Rwanda Ebola Preparation" to a Kennedy Center drag performance called "The Wig Party: Pussy Noir." He's struggled to find explanations for some of the deletions other than spite. One clip documented the anodyne swearing-in ceremony of Dr. Rachel Levine, an admiral and a transgender woman, as an Assistant Health Secretary in the previous Administration. "They're paying a government employee to sit in a room and screen these videos," he said, emphasizing the hypocrisy of spending taxpayer money on spite in the name of efficiency. "None of these people are going away," he said of the marginalized groups targeted. "It's petty to get rid of the videos addressing them."

Thanks, in part, to these volunteer efforts, the archivists I spoke to were confident that much less government data will be permanently lost than was initially feared. But they also saw little

reason for complacency. "What we don't know is how much material has been *changed*," Mark Graham, the director of the Internet Archive's Wayback Machine, told me. His team is tabulating how many dot-gov pages with certain keywords have been modified or deleted; in the lead are "health policy," "World Health Organization," and "systemic racism." Their backups are foundational to many of the more recent efforts to archive the federal web. But they're also closer to "snapshots" than functional substitutes: What use is an archived F.D.A. finding aid if it's been disconnected from back-end data, and doctors without coding skills can't use it to research clinical trials?

"It's a lot easier for the archival community to say, 'Yeah, we have a bunch of data,' than it is to say, 'Yeah, we're hosting a bunch of server-side applications that will help you navigate the data,' " Jack Cushman, the director of Harvard Law School's Library Innovation Lab, told me. Last month, his organization released a [backup](#) of the more than three hundred thousand data sets hosted by [data.gov](#). (At least three thousand of the originals have been removed.) They're also working on open-source tools to make all this data navigable.

"I think of the role that governments have had in building lighthouses," Cushman said, framing deleted data not just as a matter of censorship or knowledge preservation but also of civic freedom. "In the data sets we've collected, there are ones on the leading causes of death, what cities and aquifers are growing or shrinking, which crops are growing, which vehicles are safe, and which schools are succeeding," he said. "It's all this stuff that lets us coördinate but doesn't tell us what to do."

Last week, the guerrilla archiving movement reached an important milestone, when [restoredCDC.org](#) went online. It's a replica of the health agency's pre-Trump website based on backups from r/DataHoarder—one that's fully functional, with a reconstructed back end and interactive tools. But fresh challenges loom. Librarians and data hoarders have been able to save only publicly available records; restricted ones, such as the D.O.J.'s National Database of Police Misconduct—or the internal records being shredded by employees of [U.S.A.I.D.](#)—may be gone for good. Some publicly available data sets have proved unmanageably large, such as N.O.A.A. weather data, which is generated more quickly than volunteers can pull it down.

The Data Rescue Project's next priority is finding a decentralized storage solution for the data it already has. Majstorovic is working on a way to break up hundreds of terabytes into chunks small enough to share via BitTorrent, which stores files distributively among users. The result might be less vulnerable to censorship than central servers. But it would also require even more people to donate their time and terabytes. He's encouraged by the commitment shown by volunteers who ran out of hard-drive space on a previous campaign. "They started uninstalling their games," he told me. "I thought that was the ultimate nerd sacrifice." ♦