



As the Trump admin deletes online data, scientists and digital librarians rush to save it

The race is on to save public information about health and science before Trump makes it disappear

By **CARLYN ZWARENSTEIN**

PUBLISHED FEBRUARY 4, 2025 11:46AM (EST)

UPDATED FEBRUARY 4, 2025 1:54PM (EST)



Donald Trump VS Scientific Data (Photo illustration by Salon/Getty Images)

Orders from the Trump administration affecting science and health in the United States — and from there, the world — are coming thick and fast, affecting a myriad of institutional and personal decisions that depend on accurate information provided by the U.S. government. This ranges from websites disappearing to the prohibition of dictionary words from federal scientist research papers. Now researchers and data nerds are rushing to preserve this vital information before it's lost.

Meanwhile, many public communications have been paused as well. For the first time in sixty years, the federal health agency, the Centers for Disease Control (CDC), has stopped its own publications, including the Morbidity and Mortality Weekly Report (MMWR). This **comes on top** of a communications gag order preventing its scientists from sharing any

new findings — from new insights in cancer treatment to [potential new pandemics like Ebola](#) — with the public.

Additionally, it [orders that a list of specific terms be removed](#) from any CDC research manuscript being submitted to, already being considered, or already in press by any scientific or medical journal, with publication paused or retracted until the terms are scrubbed from the work. The terms in question are: Gender, transgender, pregnant person, pregnant people, LGBT, transsexual, non-binary, nonbinary, assigned male at birth, assigned female at birth, biologically male, biologically female.

Related

[As Trump escalates war on facts, scientists warn “we are going to get screwed”](#)

A quick search by Salon of [PubMed](#), the National Institutes of Health-run database of academic biomedical and health publications as well as related disciplines like life sciences and chemical sciences, shows that it currently contains 145,340 pages’ worth of results (1,453,391 individual publications) featuring the term “gender,” and 5,613 publications with the term “transsexual,” with papers dating back to 1903.

Since most medical science papers report demographic details, excluding papers that report this information or use this language would mean un-publishing exciting [new findings on cancer treatment](#), or, say, [vital information about H5N1 transmission](#) among American farmworkers or the ongoing tuberculosis outbreak in the Kansas City area. The exact implications of President Trump’s orders are not entirely clear, so it’s hard to say exactly what may not be published as a result, but pre-emptive self-censorship is also likely. On Monday, some of the pages had already been restored, according to the [New York Times](#), underscoring the unpredictable status of some federal information.

"Deleting gobs of public data/resources as well is the digital equivalent of book burning."

In the meantime, a general communications gag order bans any CDC scientist from submitting any new scientific findings to the public. As federally-funded health websites and webpages disappear in real time from the Internet, the race is on to preserve vital datasets and formerly public information.

Charles Gaba, a health care policy data analyst and web developer, created links on his website to every mirrored copy of the CDC’s public facing web pages as they appear on the [Internet Archive](#), a nearly 30-year-old non-partisan, non-profit organization dedicated to preserving the internet from censorship and data decay.

“These pages and related files are both funded by taxpayers and specifically intended to be for the general public, after all,” Gaba told Salon. In quickly indexing every public-facing page on the CDC site in the nick of time, he anticipated what has happened over the past few days, although perhaps not the specifics of Trump’s Jan. 29 memo to all federal departments and agencies, outlining an executive order called “Defending Women”, that seems to have triggered the hasty scrubbing of federally-funded sites.

Want more health and science stories in your inbox? Subscribe to Salon’s weekly newsletter [Lab Notes](#).

Still, Gaba notes that both the former president's campaign statements, and language in the policy playbook [Project 2025](#), had promised to purge the federal government of "anyone or anything" that they see as related to diversity, equity and inclusion, and of purging federal agencies of references and resources to the same.

"Anyone who was paying attention (far too few, sadly) should have known that this would also include deleting gobs of public data/resources as well, which is the digital equivalent of book burning," Gaba said. "Given Trump/MAGA's disdain for science in general and given the recent/ongoing brouhahas over both vaccinations and transgender rights/treatment, CDC data seemed like a likely initial target."

Gaba notes that the actual archiving of datasets isn't his work but that of researchers and health care professionals collaborating to save what can be saved. The Association of Healthcare Journalists [suggests that journalists back up data](#) they use for ongoing stories and that is still visible on federal health websites. By last Friday, the CDC had removed data on health disparities and health equity, including datasets on HIV. But it's not only the CDC that seems to be scrubbing websites.

NIH publications and data are also at risk: the organization's Office of Research on Women's Health had also disappeared from the Internet by Friday. So had federal information on climate change.

"The change in the web since the inauguration has been more than we expected."

Another initiative attempting to save the data is the [End of Term Archive](#), a collaboration of institutions that set their automated web harvesting tools to work in September to preserve government website data, performing a second webcrawl that started just after the inauguration so that before and afters can be compared.

This isn't just a response to Donald Trump. The EOT Archive updates every four years after U.S. elections to preserve government websites at the end of a presidential administration, after which new administrations typically like to make some changes.

"I will say that the change in the web since the inauguration has been more than we expected," Mark Phillips, a librarian at the [University of North Texas Libraries](#), said in a video interview with Salon. UNT Libraries is a partner in the EOT Archive, along with the [Internet Archive](#), the [Common Crawl Foundation](#), [Webrecorder](#) and others.

"This year has been a bit more volatile in just the amount of content that's changing, the domains that are going away, or the content that is no longer available that was [there] previous to the inauguration, and so we're trying our best to either make sure that we collected it beforehand," Phillips said. Vanished content gets noted as part of the crawling process, which is done with archival practices in mind.

"And so it's information that in the future, as we go through and do analysis, or as a researcher does analysis of this content, they'd be able to say what wasn't available that had been [before]," Phillips explained.

Related

Aviation expert: "I am embarrassed" that Trump blamed DEI for D.C. air disaster

For example, the EOT Archive has records that the University of North Texas requested a [usaid.gov](#) URL on a particular date, and received a 404 error message or another indication that content was no longer available. The United States Agency for International Development, [USAID](#), was issued a stop-work order last month, and on Monday, it was reported that staffers had been told to stay home, and that Trump plans to merge USAID into the U.S. State Department, reducing its workforce and budget in the process. A Salon search for [usaid.gov](#) on the same day resulted in a “can’t find server” error message.

Phillips suggests that journalists, researchers, community groups or independent citizens should try to safeguard content that’s important to them, downloading copies of important information or datasets.

“The Internet Archive has [a utility which you can use](#) to go through and immediately do a capture so that it ends up in the Wayback Machine,” Phillips said, referring to the archive’s 28-year record of over 916 billion web pages captured in time. “It’s a free service, and they encourage you to do that.”

And of course the Internet Archive is where you can now see the entire public-facing CDC web as it was just before Jan. 27th. So far the EOT Archive has preserved sites after elections in 2008, 2012, 2016 and 2020. Now it’s open for public nominations of important sites to preserve from 2024. As well as nominations, which [anyone can make](#), the partners identify large bulk lists in the form of huge spreadsheets of domain names or specific URLs to preserve. Ultimately, they know they will run out of time.

We need your help to stay independent

[Subscribe today to support Salon's progressive journalism](#)

“Generally, the size and scale of the federal Web is very large, and so we try to get as much as we can within the kind of time constraints we work under,” Phillips said. He noted that they’ve been doing this since 2008 but this year was, well, something else.

“There are some things that we definitely didn’t have on our bingo cards,” Phillips explained. “I expected policies [and] initiatives to change because they did that in 2016, they did that in the Biden transition from Trump, this sort of thing happens. But it has been a very unexpected amount of content changing and either going away or being moved around.”

Gaba drew parallels with the memory holes in George Orwell’s “1984,” which are incinerator chutes that burn references to the past, allowing the government to rewrite history without leaving a trace of the deception.

“The Trump/Musk Administration,” Gaba told Salon, referring to the president’s [close alliance with multi-billionaire Elon Musk](#), “is attempting to rewrite history on the fly. The danger isn’t just that they’ll purge accurate data from the past but that if and when that data is ever reposted that some of it will be modified with false information. There’s also no way of knowing whether future data [and] reports will ever be published, and if it is, whether it will reflect the true state of affairs.”

Gaba says he wants “people to understand that this is likely just the beginning. ‘1984’ is today.”

Read more

[about science in the Trump era](#)

-
- **Trump declared a “national energy emergency.” Experts say it's a "farce"**
 - **Scientists debate fleeing America because of Trump — or risk their research being censored**
 - **RFK Jr.'s drug policy under Trump could worsen the overdose crisis, experts warn**

By **CARLYN ZWARENSTEIN**

Carlyn Zwarenstein is the author of *On Opium: Pain, Pleasure, and Other Matters of Substance*. She's on Twitter at @CarlynZwaren.

MORE FROM CARLYN ZWARENSTEIN

Related Topics -----

[Censorship](#)

[Internet Archive](#)

[Public Health](#)

[Tech](#)

[Trump Admin](#)

Related Articles

[Home](#) [About](#) [Staff](#) [Contact](#) [Privacy](#) [Terms of Service](#) [Archive](#) [Go Ad Free](#)

Copyright © 2025 Salon.com, LLC. Reproduction of material from any Salon pages without written permission is strictly prohibited. SALON ® is registered in the U.S. Patent and Trademark Office as a trademark of Salon.com, LLC. Associated Press articles: Copyright © 2016 The Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed.

[DMCA Policy](#)