

Disappearing Data: *Concerns, Implications, & Resiliencies*

Sara Curran (scurran@uw.edu): Research Resilience Panel (5/19/25)

Overview & Purpose

Part 1.

- Sketch of the issues as they relate to research (at this time)
- Provide some current tools and support for sustaining/recovering/re-using/re-purposing data
- Discussion with panelists and audience: Additional concerns? Additional support?

Part 2.

- Faculty Leadership Panel Reflects on the Challenges and Approaches for Resiliencies
- Moderated Discussion with Audience

Concerns

Known concerns (on again/off again caveat):

- Federal (tax-payer funded) data removed from platforms or access eliminated (e.g. NCES 6/25)
- Federal data no longer collected
 - surveys stopped e.g. CDC PRAMS, NSDUH, NOAA)
 - classifications of categories changed or eliminated (e.g. gender measures)
- Scientific input on Federal data and research restricted or eliminated (e.g. CSAC cancelled)
- Research cancelled and resulting data and research not generated for public information (e.g. cancellation of SSA research networks)

Future concerns:

- Changes to Civil Service Code (OPM) politicizes federal statistical management
- Federal data systems no longer able to protect privacy of individuals and communities
- Federal data changed for non-scientific reasons
- Lack of scientific review, oversight, input significantly reduces data quality



Public Implications



Loss of critical information hampers good government and business planning



Loss of taxpayer-funded information harms health, disaster preparedness, etc.



Distributed, digital archive system of Federal libraries program reveals weaknesses in research repositories system for all

← → ↻ guides.lib.uw.edu/research/federaldataaccess ☆

Google Calendar KNOW University of Wash... Get the Proxy Boo...

W UNIVERSITY LIBRARIES
UNIVERSITY of WASHINGTON

Research Guides

HOME / START YOUR RESEARCH / RESEARCH GUIDES / FEDERAL DATA / ALTERNATIVE ACCESS

Federal data: alternative access

This guide provides alternative access points for federal data that has been removed from official sources. Please **reach out** to suggest additional sources.

Archived government data

- **End of Term Web Archive**
Captures U.S. government websites at the end of Presidential terms.
- **Wayback Machine**
General internet archiving, including government sites.
- **GovWayback**
Archived versions of government websites prior to January 20, 2025
- **Library Innovation Lab**
Mirror of data files from data.gov.
- **UCSB Public Data**
Several archived public datasets, including CALMA data rescue, and more.
- **Find Lost Data**
Searches for datasets across various sites.

<https://guides.lib.uw.edu/research/federaldataaccess>

Some.... Research Implications

Current scramble to rescue data or save data collection efforts, pointing towards alt. research resources

- Office of Research & Libraries Created Research Guide
- Libraries & eScience considering scraping sites of major repositories (e.g. Pub Med)
- Local efforts, interest groups, distributed efforts
 - E.g. [UW CSDE's links](#); [Proton](#)

But...

- Haphazard, unsystematic and not coordinated



Data Rescue Project

[iAssist](#)

[RDAP – Research Data Access and Preservation](#)

[Data Curation Network](#)



Photo by [Clark Tibbs](#) / [Unsplash](#)

On April 24, 2025, [Research Data Access and Preservation Association](#) (RDAP) hosted a Data Rescue Hackathon. Facilitated by members from both the Data Rescue Project (DRP) and RDAP, 147 attendees (including data professionals and members of the public) joined in on the event.



Data Rescue Project

URL:

https://www.datarescueproject.org/data-rescue-tracker/?utm_source=substack&utm_medium=email

The Data Rescue Tracker

	ID	Dataset	Website	Status	Maintainer
1	f9927cab-...	Billion-Dollar Weather and Climate Disasters	ncei.noaa.gov	In Progress	GTDC
2	c7787744-...	BLS Downloads	download.bls.gov	Finished	DRP
4	f1e011bf-d...	CDC FTP	ftp.cdc.gov	Finished	DRP
5	26ba71f4-f...	US Census Bureau FTP	census.gov	Finished	DRP
6	7cb43fc4-...	National Hurricane Center	nhc.noaa.gov	Finished	DRP
7	2c78d527-...	United States Geological Survey	usgs.gov	Finished	DRP
8	98666e77-...	Office of General Counsel Documents	noaa.gov	Finished	DRP
9	a575ee2e-...	National Hurricane Center FTP	ftp.nhc.noaa.gov	Finished	DRP
10	50606b34-...	Enforcement and Compliance History Online	echo.epa.gov	Finished	DRP
11	0c0e50d0-...	System for Award Management	sam.gov	In Progress	GTDC
12	ca207fa3-...	American Communities Survey	census.gov	In Progress	GTDC
13	e1f1b49b-...	Consumer Expenditure Survey	bls.gov	In Progress	GTDC
14	16cd1ff5-f...	USA Spending	www.usaspending.gov	In Progress	GTDC
15	36977438-...	Consumer Finance Bureau Website	consumerfinance.gov	Finished	DRP
16	7adcdfa4-...	College Scorecard	collegescorecard.ed.gov	Finished	DRP DL
17	913ae562-...	Campus Safety and Security	ope.ed.gov	Finished	DRP DL
18	02fda5bc-...	Equity in Athletics Data Analysis Cutting Tool	ope.ed.gov	Finished	DRP DL

Microsoft

Research Repositories: *Not yet a long-term solution*

- **Reuse and Replication Resource=Re3**
- **re3data** is currently the **most** comprehensive source of information on research data repositories, with more than 3000 records
- D Strecker, H Pampel, R Schabinger, NL Weisweiler (2023) “Disappearing repositories: Taking an infrastructure perspective on the long-term availability of research data.” *Quantitative Science Studies*; 4 (4): 839–856. https://doi.org/10.1162/qss_a_00277
 - 6.2% of all repositories listed in re3data were shut down. The median age of a repository when shutting down was 12 years
 - Shutdown strategies (47% have no strategy):
 - Maintain limited access to data (for example via a simple FTP interface)
 - data migration (transferring data custody to another repository).
- Burden of infrastructure maintenance and data preservation is currently placed on individual repositories alone

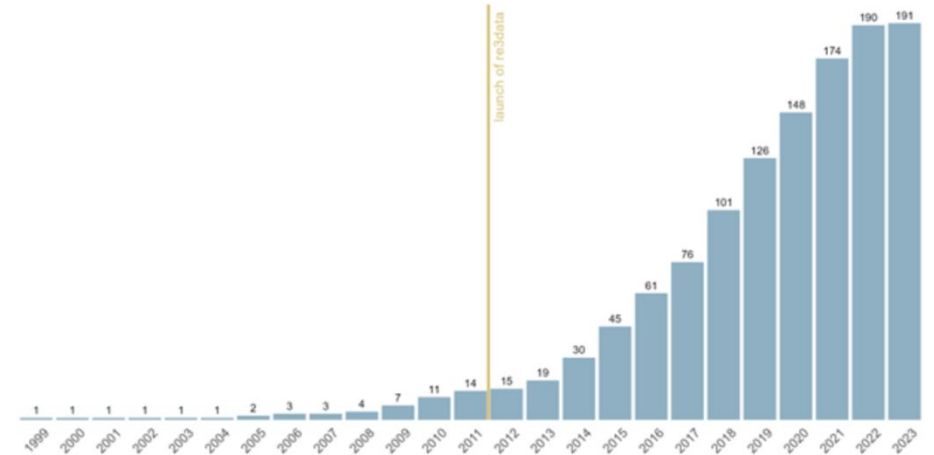


Fig. 1: Number of closed repositories indexed in re3data per year (cumulative)



Research Implications & Resiliency Efforts

- **Disruption** of critical lines of scientific inquiry
- **Reliance** on **replication** and **reuse**, as opposed to relying on recent or future data availability
 - scientific reviews may not appreciate
- **More infrastructure** required for **independent and/or shared repositories** for stability and reliability
- **Renewed** efforts to reclaim public trust in data and research
- **Need for new sources of data** outside of Federal government
 - Very expensive and complex infrastructure required
- New **extension of innovative data collection** (used for hard-to-reach pop or rare species) for broader purposes
 - RDS, mobile app surveys, remote sensing satellite info, other surveillance



Additional References

<https://www.404media.co/archivists-work-to-identify-and-save-the-thousands-of-datasets-disappearing-from-data-gov/>

No single organization monitoring and some things disappear and then reappear. There is one major court case regarding public health data (Doctors for America v. OMB et al.), which has so far resulted in a TRO that required the administration to maintain some websites

(<https://www.justsecurity.org/107087/tracker-litigation-legal-challenges-trump-administration/>; <https://www.cdc.gov/hiv-data/about/index.html>).

Additional References

[U.S. Government Data Purge \(Brookings\)](https://www.brookings.edu/articles/the-us-government-data-purge-is-a-loss-for-policymaking-and-research/?utm_source=substack&utm_medium=email)

https://www.brookings.edu/articles/the-us-government-data-purge-is-a-loss-for-policymaking-and-research/?utm_source=substack&utm_medium=email

UW Libraries' Initiatives for Resilience

Negeen Aghassibake, Data Visualization Librarian

tinyurl.com/uwlib25

Best Practices for Data Recovery

Naomi Alterman, eScience Institute

<https://staff.uw.edu/naomila/slides/2025-data-recovery-panel>

Discussion

Questions for you....

- Are there any realities or consequences of broad relevance that we haven't yet thought about?
- Are there any information pieces that can be shared?
- Are there broader professional associations that are working on this and should / could we connect with them?
- Is anyone interested in becoming part of a campus-wide or conference-wide effort to coordinate around?
- Are there more ways to think about research resilience in this domain?
- Other thoughts?

Faculty Leadership Reflections: Supporting Research in the Context of Disappearing Data

Stephanie Farquhar School of Public Health

David Pigott, School of Medicine

Luanne Thompson, School of Oceanography