

Beyond Data Refuge: Leveraging Libraries

The Problem: Scope, Scale, and Timing

Archiving federal web sites, gray literature, and data is a fast-moving, complex, and urgent challenge. Given our mission to preserve and make knowledge accessible, academic research libraries and librarians are particularly well-suited to develop collaborative solutions. Toward this end, there are already a few strategies in place including (1) seeding the [End-of-Term Archive](#) to harvest government web site pages and (2) a growing number of [Data Refuge](#) events to extract and make accessible climate and environmental data that can not be web archived.

What Libraries & Librarians Can Do

In response to the overwhelming number of requests that we've received from colleagues at universities across the US who want to help, we propose an additional scaling strategy: Leveraging existing capacity within libraries -- in staff and expertise -- to archive web sites *immediately*. Many academic research libraries have web archiving systems in place with knowledgeable librarians engaged in this work. If a fraction of those skills and systems are directed to address a small slice of this challenge, together we can make substantive and critical progress toward preserving federal web sites.

One Example: Penn's Experience

Penn Libraries is web archiving NOAA pages. Over the course of three days more than 3,700 URLs were crawled accounting for more than 400,000 web pages and 600GB of data storage.

Choose One Way to Get Involved Today

1. *Start Web Archiving Federal Sites*

- Determine how much time, effort, and web archiving space you are willing to commit to preserving federal government web sites.
- Identify a staff member to use Archive-It and allocate 2-5 days to perform the task.
- Use the [US Agency Coordination Spreadsheet](#) to claim an agency and one or more programs. Penn Libraries staff will provide light coordination to assist **one librarian** from each participating library to provide access to this and other support resources to help us all work efficiently together. Please send names to Kim Eke (kimeke@upenn.edu).

2. [Organize a Data Rescue event](#) on your campus or participate in [upcoming events](#).

3. *Dedicate a set of staff hours to provide metadata and quality assurance assistance for climate & environmental data that has already been extracted.*

4. *Dedicate a set of staff hours to help us complete [US Agency Coordination Spreadsheet](#).*

Questions

1. Who might be good national partners to help coordinate and communicate around all of the different US agencies?
2. What networks or consortia might be interested partners?
3. Are there upcoming library conferences or events where we can hold discussions about web archiving, extracting federal data, and related challenges?
4. Are there grants that might be available to support these efforts in the form of term positions or other needed resources?

Recent Press Coverage

VICE News (1/26/17) *Saving climate data: Scientists are scrambling to safeguard climate data from the Trump administration* <https://news.vice.com/story/scientists-are-scrambling-to-safeguard-climate-data-from-the-trump-administration>

CBS News (1/24/17) *Trump admin orders EPA media blackout and contract freeze* <http://www.cbsnews.com/news/president-trump-administration-orders-epa-media-blackout-contract-freeze/>

Wired (1/19/17) *Rogue Scientists Race to Save Climate Data from Trump* <https://www.wired.com/2017/01/rogue-scientists-race-save-climate-data-trump/>

Gizmodo (1/19/17) *Trump Is Launching a War on Science— These People Are Fighting Back* gizmodo.com/meet-the-people-racing-to-stop-trump-from-purging-feder-1791317464

URLs

End of Term Archive
<http://eotarchive.cdlib.org/>

DataRefuge for Climate & Environmental Data
<http://www.datarefuge.org>

US Agency Spreadsheet (in progress):
https://docs.google.com/spreadsheets/d/1yIrhFrZkv2Yhdk48W_P5bd-C-jxLtbzJFcm2E5oq-ec/edit#gid=1007301026

Organize a Data Rescue Event
<http://www.ppehlab.org/datarescue/>

Upcoming Data Refuge Events
<http://www.ppehlab.org/datarefuge>