Open Data
A model for proactive disclosure

SUNSHINE FEST
MARCH 16, 2026
WASHINGTON, D.C.

MODERATOR

CORI ZAREK

VP for North America, APOLITICAL

Former Deputy U.S. Chief Technology Officer

Founding staff attorney, Office of Government Information Services

What's the one dataset you wish was publicly available?

STEFANA BREITWIESER

Collections Archivist, Opioid Industry Documents Archive

JOHNS HOPKINS UNIVERSITY

LINDITA CAMAJ

Associate Professor of Journalism

UNIVERSITY OF FLORIDA

LYNDA KELLAM

Director of Research Data & Digital Scholarship

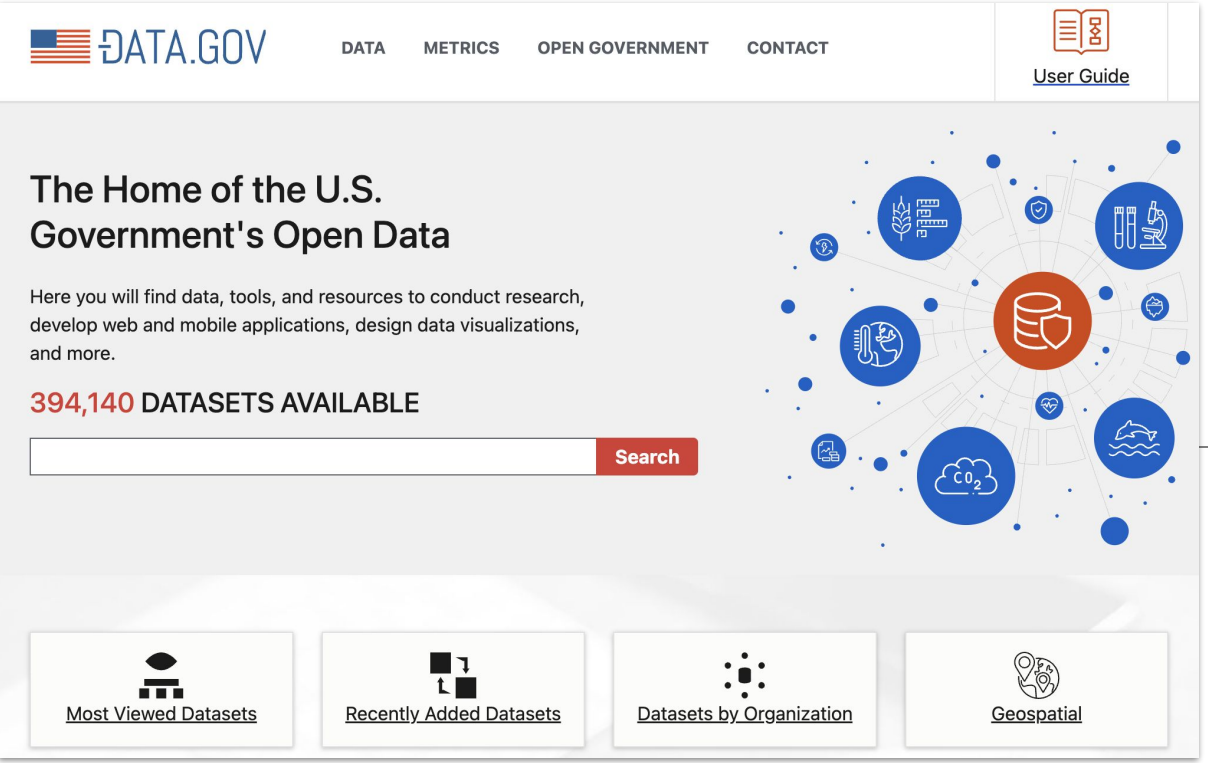
PENN LIBRARIES / DATA RESCUE PROJECT

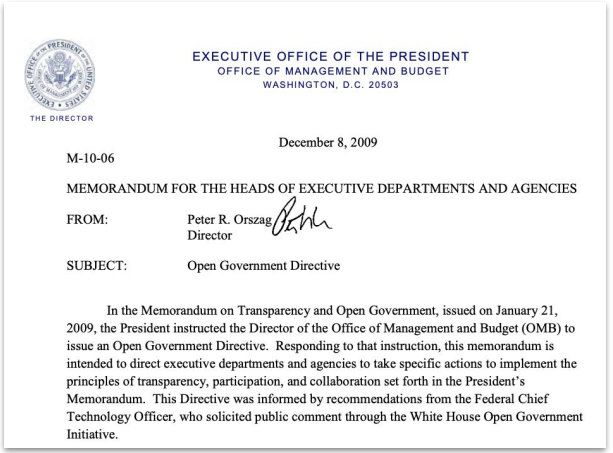
DENICE ROSS

Senior Advisor

FEDERATION OF AMERICAN SCIENTISTS

Early open data efforts





EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

THE DIRECTOR

December 8, 2009

M-10-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Peter R. Orszag
Director

SUBJECT: Open Government Directive

In the Memorandum on Transparency and Open Government, issued on January 21, 2009, the President instructed the Director of the Office of Management and Budget (OMB) to issue an Open Government Directive. Responding to that instruction, this memorandum is intended to direct executive departments and agencies to take specific actions to implement the principles of transparency, participation, and collaboration set forth in the President's Memorandum. This Directive was informed by recommendations from the Federal Chief Technology Officer, who solicited public comment through the White House Open Government Initiative.

the WHITE HOUSE PRESIDENT BARACK OBAMA

HOME · BLOG

U.S. Government, OSTP, Open New Troves of Data to the Public

JANUARY 22, 2010 AT 7:10 PM ET BY ANEESH CHOPRA

Summary: The public has a new trove of government data at its fingertips -- thanks to agencies throughout the Federal Government answering the Open Government Directive's call to make available at least three new, high-value data sets.

Cross-posted from the OSTP blog

Today marks the first milestone called for in the Administration's Open Government Directive — a call for agencies to make available at least three new, high-value data sets in machine-readable format. Thanks to the terrific responsiveness of our cabinet agencies, the public now has a new trove of government data at its fingertips.

Knowledge Management in Journalism

Knowledge Management: practices in coding, storing, sharing, transmitting and leveraging knowledge assets within and between organizations

1. inward-facing approach (databases organize information gathered through reporters' own investigations)
2. outward-facing approach (databases that manage larger flow of data generated by other actors in the open data ecosystem

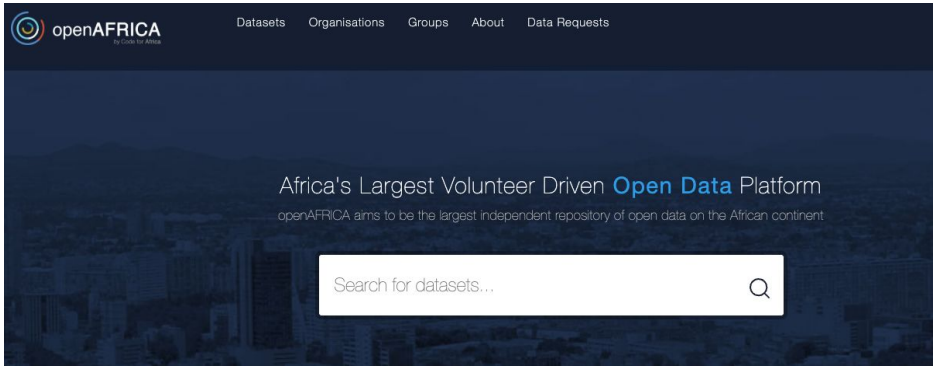
Sources: Alavi & Leidner, 2001; Kakabadse et al. 2003; Lewis & Usher, 201

Data Liberation, Preservation & Archiving

Pooling Resources and Expertise - Share technical skills, navigate more effectively local laws, languages, and cultural norms. Methods include web scraping, FOIa requests, offline sources, and a network of coordinators.

- ✔ e.g. **Environmental Data Projects**: Journalists in countries with strong FOI laws can obtain pollution data and share it with colleagues in countries where such data is hidden.
- ✔ e.g. **ARIJ Archival Initiatives** - ARIJ has been collecting and archiving datasets since 2009 from 18 Arab countries, transfer their archives to the OCCPR Aleph Global database

Collaborative Databases

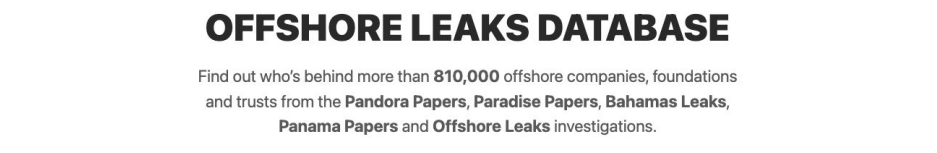


Open Data Repositories

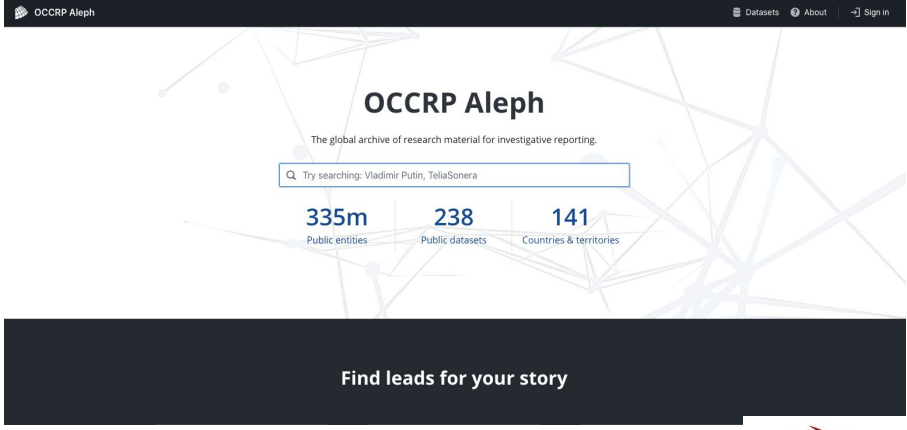
ADVOCATING FOR TRANSPARENCY

e.g. openAfrica *by Code for Africa*; **Offshore Leaks Dabatase** by *International Consortium of Investigative Journalists (ICIJ)*

- ✔ broad civic open data repositories oriented towards transparency and public use
- ✔ Built as a data warehouses & data preservation
- ✔ Designed for public accessibility and usability
- ✔ Designed as knowledge repository & retrieval system



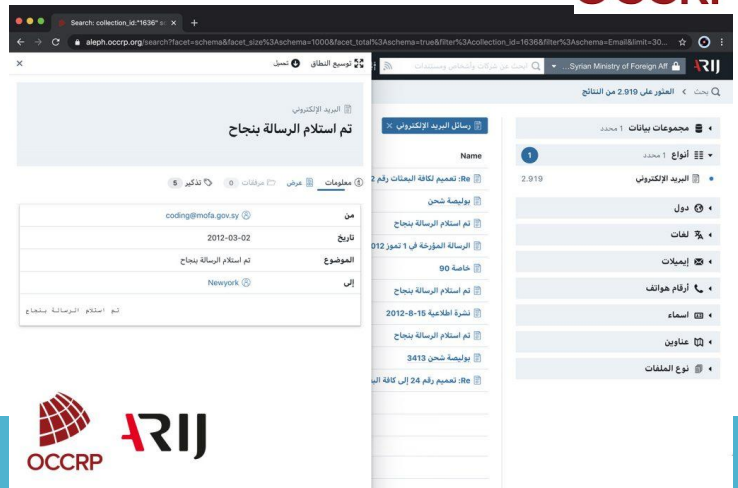
Data Standardization



HARMONIZING DATA ACROSS JURISDICTIONS

E.g. Aleph by *Organized Crime and Corruption Reporting Project (OCCRP)*; **ARIJ Aleph** by *Arab Reporters for Investigative Journalism*

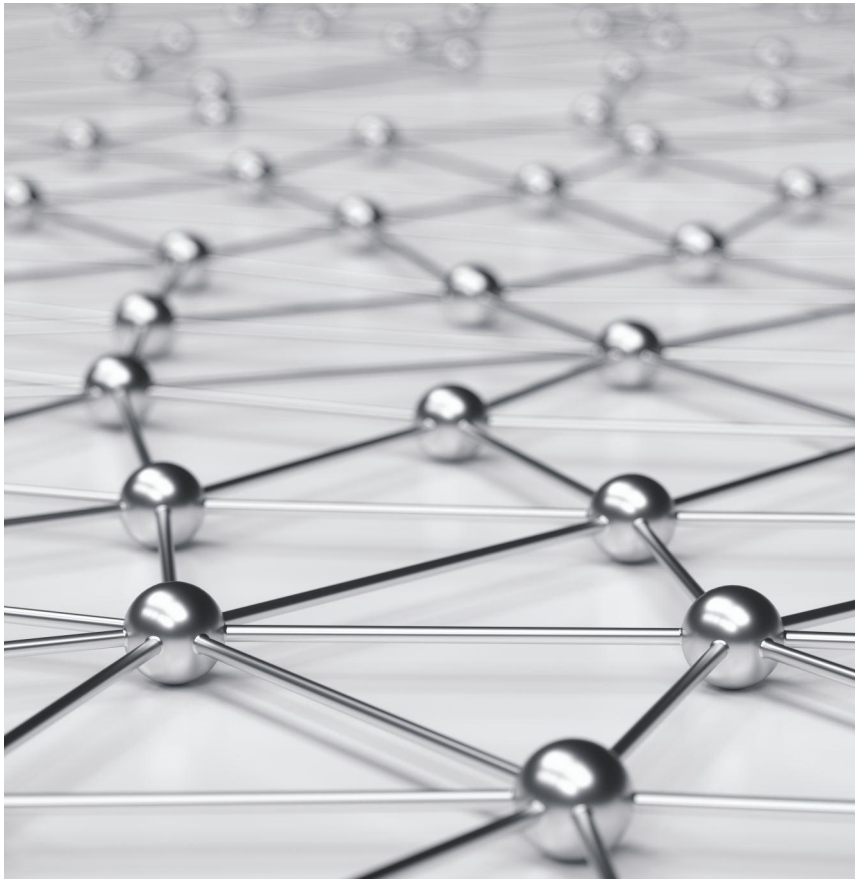
- ✔ investigative-first platforms tailored to journalists
- ✔ Partially open to external actors
- ✔ Designed as a search, discovery, and verification engine for large datasets
- ✔ Data standardization aligns government records, corporate registries, procurement data, and court documents from various jurisdictions.



Data Connectors

e.g. NINA by *Latin American Center for Investigative Journalism (CLIP)*

- ✔ a decentralized data commons that connects public and private databases from more than 21 countries, enabling users to search and cross-reference datasets in a single unified interface
- ✔ Rather than storing documents and other information, NINA's architecture aggregates and harmonizes heterogeneous data sources
- ✔ translates diverse API languages and structures into a unified schema that allows seamless querying and analysis across multiple national systems



SOUTH-SOUTH NETWORKS

DATA ECOSYSTEMS FOR REGIONAL AUTONOMY

South-South networks promote knowledge exchange and co-development of tools tailored to local needs. Collaborating within these contexts fosters mutual understanding, shared strategies, and resilience.

- ✔ e.g. NINA by CLIP works with structured and unstructured data, connecting to various databases focusing on public information data from Latin America. Mini NINAs (Guatemala, Honduras, El Salvador), specific versions of NINA catering to specific needs of those countries

Reframing Journalism's Professional Identity

Knowledge Management Systems – a novel practice of epistemic culture

- *from reporting events → to co-producing civic data*
  - ✔ Cross-border collaboration hubs in journalism have become both producers and custodians of open data infrastructures, blurring the line between news production and knowledge brokers

Open-Source Culture & Data Activism

- *from objectivity → to civic accountability practice*
  - ✔ cross-border journalism infrastructures straddle the line between professional investigative tools and civic data commons
  - ✔ by creating shared databases and advocating for open government, journalists contribute to broader transparency and open-source data ecosystems.
  - ✔ journalist frames data activism through the lens of information disclosure, watchdog roles, and public accountability – *"*creating open datasets for social good"

Opioids Industry Documents Archive

Our vision: to serve a groundbreaking digital archive of documents arising from the opioid industry that advances understanding of the **root causes** of the **opioid epidemic** and helps address **corporate behavior** that is **harmful** to **public health**.

Investigations from 2022

Hopkins Bloomberg
PUBLIC HEALTH

Documenting Pharma Role in the Opioid Cris

Opioid indu
cynical stra
deaths. Can

'Eat what you kill': How a fentanyl drugmaker bribed doctors, harmed patients and collected millions

Insys Therapeutics is the only drug marketer whose top executives were jailed as part of the opioid epidemic. The fallout from the company's cash-for-scrips strategy that made it a Wall Street favorite reverberates today.

Ken Alltucker
USA TODAY

Burlakoff dashed off an urgent
ondre.

HUFFPOST

How The CEO Of A Leading LGBTQ Rights Group Played A Role During The Opioid Crisis

Amit Paley, CEO of The Trevor Project, was part of a team advising Purdue Pharma on how to boost sales of opioids during his time at McKinsey & Co., according to a trove of documents reviewed by HuffPost.

REUTERS® INVESTIGATES | Going for Broke

How corporate chiefs dodge lawsuits over sexual abuse and deadly products

A REUTERS SPECIAL REPORT

Scandals b
studio and
their wealth
granted life
uptc
in r

The Washington Post
Democracy Dies in Darkness

Sections | Subscribe | Sign in

THE OPIOID FILES

Inside the sales machine of the 'kingpin' of opioid makers

A cache of more than 1.4 million newly released records exposes the inner workings of the nation's largest opioid manufacturer.

McKinsey Opened a Door in Its Firewall Between Pharma Clients and Regulators

The firm let consultants advise bot
government overseers, internal rec
what we know" was part of their pi

The New York Times

Behind the Scenes, McKinsey Guided Companies at the Center of the Opioid Crisis

464

ael Forsy

By Meryl
May 10, 2

STAT+

A STAT INVESTIGATION

Documents detail how pharmacy giants Walgreens, CVS, and Walmart failed patients in the opioid crisis

By Lev Facher, Kate Sheridan and Ed Silverman Oct. 14, 2022 Reprints

Investigations from 2023

THE WALL STREET JOURNAL.

PRO BANKRUPTCY

McKinsey Consulted VA While Advising Opioid Makers to Target Agency for Sales

Newly released documents show the firm advised companies such as Purdue Pharma and Endo on how to sell opioids to Veterans Affairs, which was also a client; McKinsey denies a conflict of interest

THE GLOBE AND MAIL

SUBSCRIBE LOG IN

CANADA WORLD BUSINESS INVESTING WATCHLIST PERSONAL FINANCE OPINION POLITICS SPORTS

McKinsey pitched Purdue Pharma Canada on plan to boost opioid sales in 2014, memo reveals

Purdue Pharma Case | Supreme Court Appears Split | Understand the Case | Who Gets Legal Immunity? | Addressing the Opioid Crisis

Sacklers Gave Millions to Institution That Advises on Opioid Policy

Even as the nation's drug crisis mounted, the National Academies of Sciences, Engineering and Medicine continued to accept funds from some members of the Sackler family, including those

UCSF MAGAZINE

Corporate Strategy, National Tragedy

UCSF's industry archives expose the marketing tactics that fueled the opioid epidemic.

Investigations from 2024

Slideshow: Opioid Advertisements From the Height Of the Crisis

The opioid crisis didn't happen by accident.

1 HR AGO • BEN WESTHOFF

Peer-reviewed papers

AJPH OPINIONS, IDEAS, & PRACTICE

The Opioid Industry Documents Archive: A Living Digital Repository

G. Caleb Alexander, MD, MS, Lisa A. Mix, MLA, Sayeed Choudhury, MSE, Rachel Taketa, MLIS, Cecilia Tomori, PhD, Maryam Mooghali, MD, MSc, Anni Fan, Sarah Mars, PhD, Dan Ciccarone, MD, MPH, Mark Patton, MS, Dorie E. Apollonio, PhD, MPP, Laura Schmidt, PhD, Michael A. Steinman, MD, Jeremy Greene, MD, PhD, Pamela M. Ling, MD, MPH, Anne K. Seymour, MS, Stanton Glantz, PhD, and Kate Tasker, MLIS

ELSEVIER

Social Science & Medicine

Volume 356, September 2024, 117158

The role of metaphor in the corporate political strategies of health harming industries: Comparing the concept of balance in the gambling and opioid industry discourses

May CI. van Schalkwyk [a], Benjamin Hawkins [b], Daniel Eisenkraft Klein [c], Adam D. Koon [d e]

Journal of Health Politics, Policy and Law

RESEARCH ARTICLE | FEBRUARY 07 2024

Inside "Operation Change Agent": Mallinckrodt's Plan for Capturing the Opioid Market FREE

Daniel Eisenkraft Klein; Ross Mackenzie; Ben Hawkins; Adam Koon

Journal of Health Politics, Policy and Law

RESEARCH ARTICLE | AUGUST 01 2022

Marketing Opioids to Veterans and Older Adults: A Content Analysis of Internal Industry Documents Released from *State of Oklahoma v. Purdue Pharma LP, et al.*

Hanna Yakubi; Brian Gac; Dorie E. Apollonio

J Health Polit Policy Law (2022) 47 (4): 453–472.

ELSEVIER

Drug and Alcohol Dependence Reports

Available online 2 November 2023, 100199
In Press, Journal Pre-proof What's this?

Retail Chain Pharmacy Opioid Dispensing Practices from 1997 to 2020: A Content Analysis of Internal Industry Documents

THE JOURNAL OF LAW, MEDICINE & ETHICS

Diverting Data and Drugs: A Narrative Review of the Mallinckrodt Documents

Antoine Lentacker[1], Kelly Pham[1], and Jason M. Chernesky[1]

Health Affairs Scholar, 2024, 2(10), qxae119
https://doi.org/10.1093/haschl/qxae119
Advance access publication: October 24, 2024
Research Article

Health Affairs Scholar
EMERGING & GLOBAL HEALTH POLICY
OXFORD

The opioid industry's use of scientific evidence to advance claims about prescription opioid safety and effectiveness

Ravi Gupta[1,2,*], Jason Chernesky[3], Anna Lembke[4], David Michaels[5], Cecilia Tomori[6], Jeremy A. Greene[1,3,7], G. Caleb Alexander[7,8], Adam D. Koon[9]

How an opioid giant deployed a playbook for moulding doctors' minds

BMJ 2024 ; 385 doi: https://doi.org/10.1136/bmj.q1208 (Published 10 June 2024)
Cite this as: *BMJ* 2024;385:q1208

Article | Related content | Metrics | Responses

Sergio Sismondo [1 2], Maud Bernisson [2]

OIDA Addresses Many Important Questions

1. How have **pharmaceutical companies** driven the current opioid epidemic?
2. What was **known within these companies, at what time, by whom**, regarding potential risks of addiction and overdose?
3. How have **marketing and promotional practices deviated** from FDA standards?
4. How have **intermediary organizations, such as advocacy organizations,** been used by the industry to promote messaging that misconstrues the risks and benefits of opioids?
5. What role have **other stakeholders (e.g., wholesalers, pharmacies)** played in driving opioid oversupply?

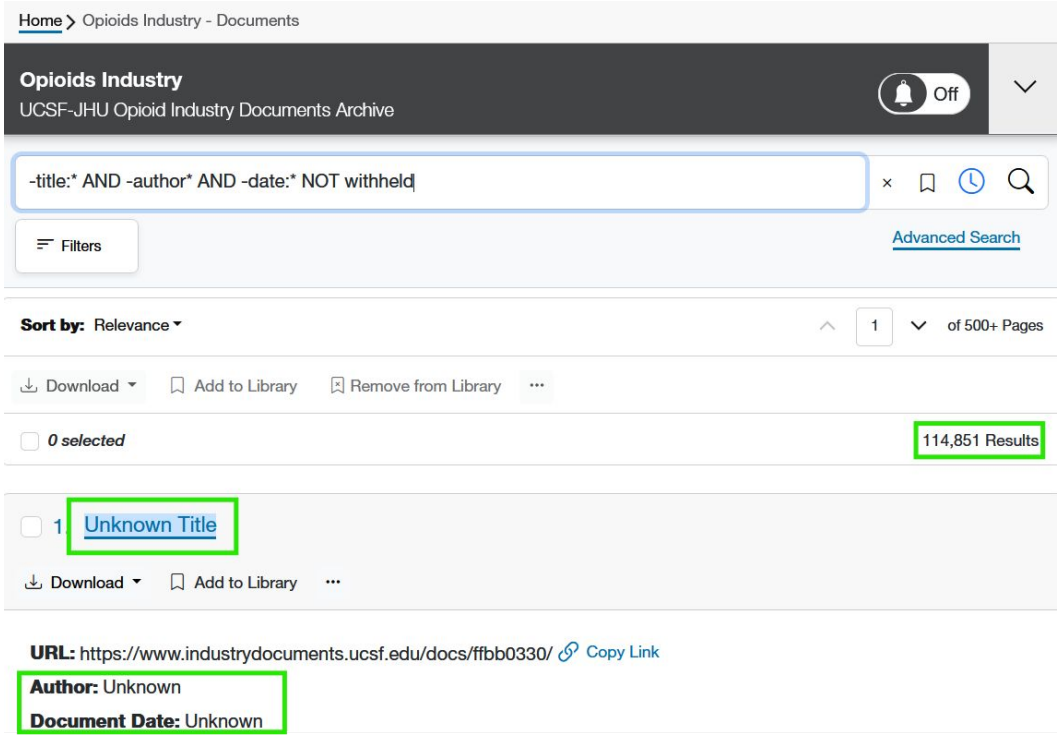
OIDA Toolbox for computational research methods

- **Social network analysis**: Which employees within a company communicated with each other most frequently, and during which time periods? What does this social network tell us about how communication and decision-making work in a large organization?
- **Sentiment analysis**: Did employees in certain roles use language (euphemisms, tone, commands, etc.) in certain ways, or did it change over time? What does this tell us about how language is used in the modern workplace and in relation to specific steps in the wrongdoing of these companies?
- Other forms of **text mining**:
  - At what point were key decisions made, and by whom? How did these patterns differ from one company to another?
  - After manually locating a few instances of deception or misconstrual of evidence in the OIDA corpus, can we detect more instances automatically using machine-learning models?

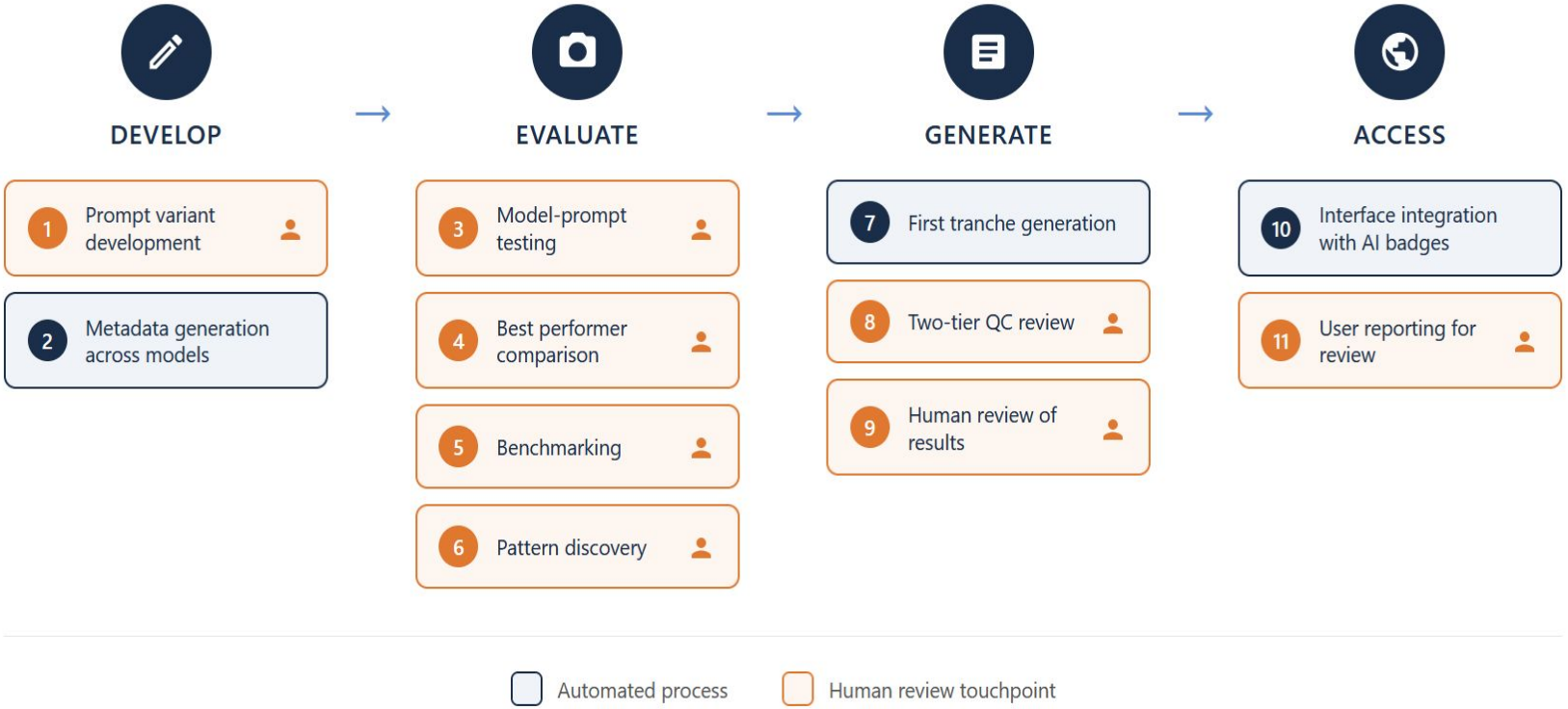
AI at OIDA

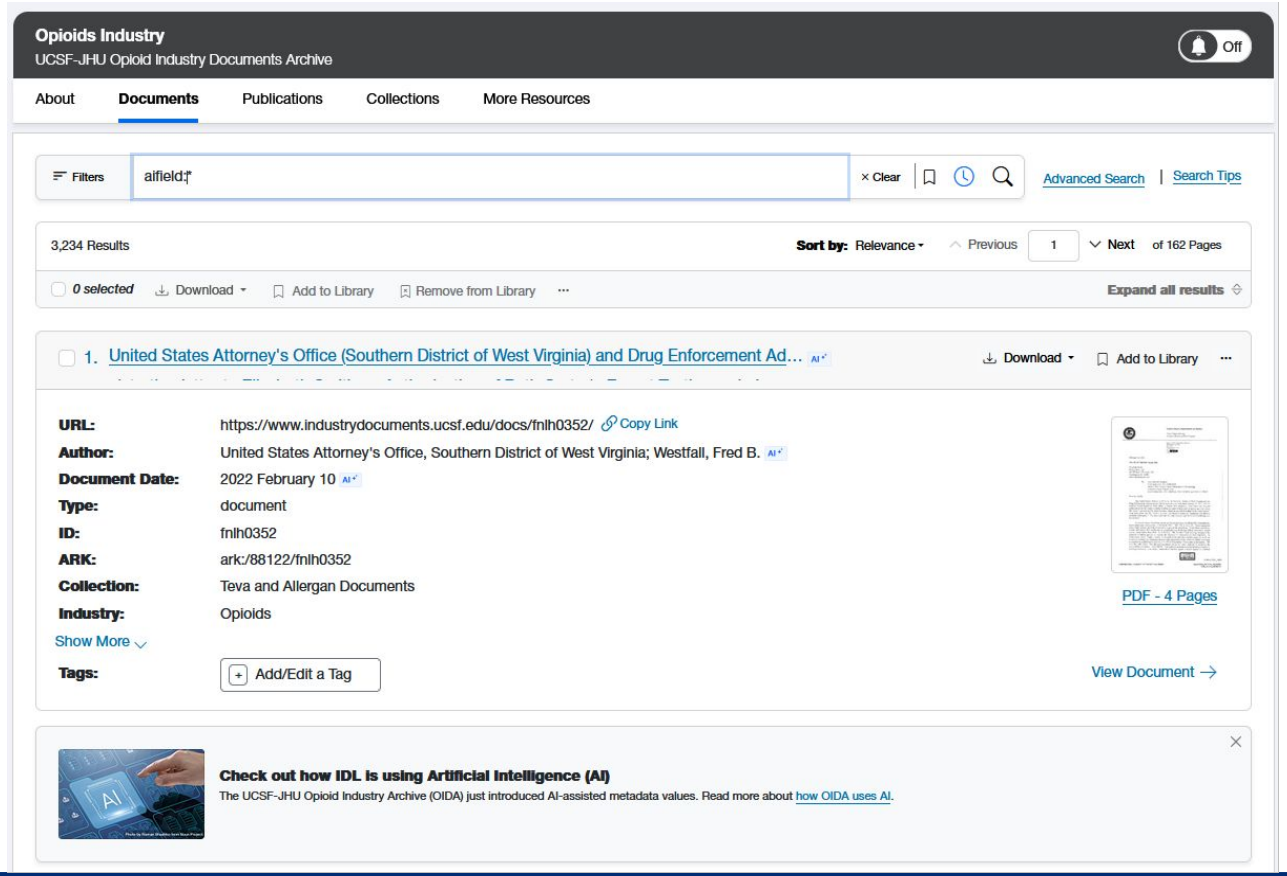
AI metadata project – Why AI?

- Almost 7M documents in OIDA
  - ~115k "triple-blanks"
  - Increasing volume of under-described metadata at receipt
- Metadata is a critical access point
- Challenges with AI



AI-Generated Metadata Workflow




Opioids Industry
UCSF-JHU Opioid Industry Documents Archive
Off
About
Documents
Publications
Collections
More Resources
Filters
aifield:*
× Clear
Advanced Search
Search Tips
3,234 Results
Sort by: Relevance
Previous
1
Next
of 162 Pages
0 selected
Download
Add to Library
Remove from Library
Expand all results
1. United States Attorney's Office (Southern District of West Virginia) and Drug Enforcement Ad...
Download
Add to Library
URL: https://www.industrydocuments.ucsf.edu/docs/fnlh0352/ Copy Link
Author: United States Attorney's Office, Southern District of West Virginia; Westfall, Fred B.
Document Date: 2022 February 10
Type: document
ID: fnlh0352
ARK: ark:/88122/fnlh0352
Collection: Teva and Allergan Documents
Industry: Opioids
Show More
Tags: Add/Edit a Tag
PDF - 4 Pages
View Document
Check out how IDL is using Artificial Intelligence (AI)
The UCSF-JHU Opioid Industry Archive (OIDA) just introduced AI-assisted metadata values. Read more about how OIDA uses AI.

OIDA Image Collection

Highlights images extracted primarily from presentations—many for internal company audiences and board members, others targeted to prescribers.

The story of how industry has created markets for their drugs, then promoted them using established playbooks, is brought to life in these images.

6000+ images extracted from documents at launch, more documents, more images coming online.

Metadata for each image: title, description, type, category, OCR text, keywords, drug, topic

Image Classification

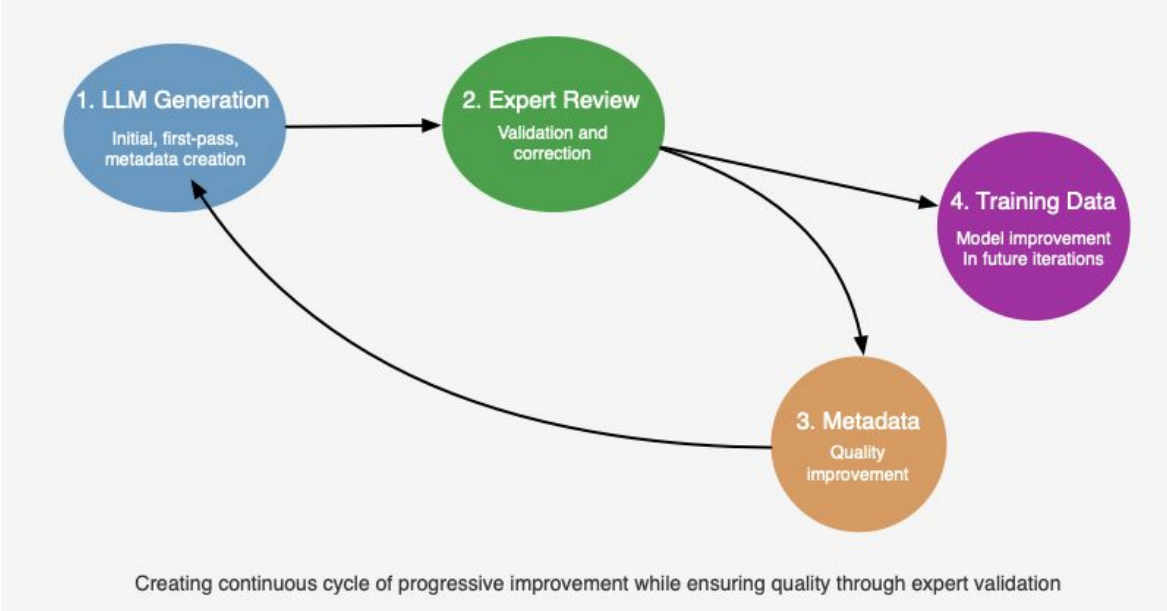
- Category assigned by zero-shot CLIP model from OpenAI
- Images sorted into six classes:
  - advertisement
  - photograph
  - table
  - Illustration
  - graph or chart
  - map
- ~58% agreement w/ expert annotator

Comparing Captioning Models

microsoft/florence-2-large

The image is an illustration of a house with a garden in front of it. The garden is filled with colorful flowers and plants, and there is a wooden planter box on the right side of the image. On the left side, there are two red garden rakes and a watering can. The house has a white door and a window on the wall. The text on the image reads "When their pain medications clock out, they do too.

Human-in-the-Loop Framework



Creating continuous cycle of progressive improvement while ensuring quality through expert validation

To receive updates on OIDA, including additional support

purl.org/oida/subscribe

stefana.breitwieser@jhu.edu

opioidarchive@jh.edu

Data Rescue Project

Presented by Lynda Kellam
on behalf of the DRP

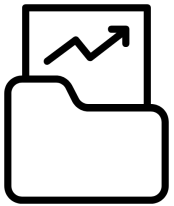
www.datarescueproject.org/ bsky.app/profile/datarescueproject.org

Our mission

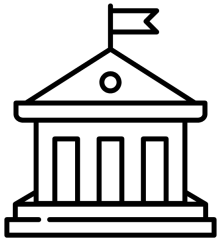
The Data Rescue Project is a grassroots volunteer community dedicated to access to public data for the public good.

Data Rescue Project

The DRP at a Glance

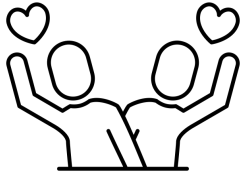


2000+ rescued datasets



80+ at-risk gov't offices covered

Hundreds of Terabytes of mirrored data



500+ volunteers

Data Rescue Project

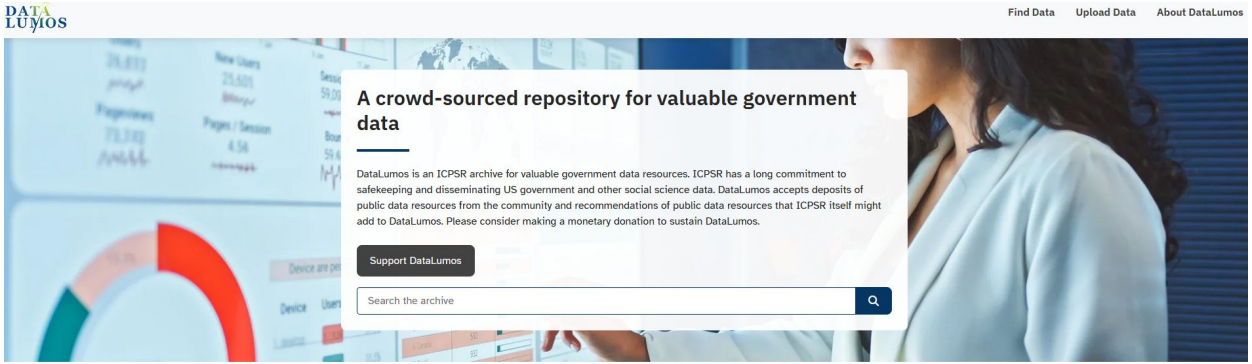
www.datarescueproject.org/

bsky.app/profile/datarescueproject.org

From Friday, February 7, 2025 - Today

Over 2500 datasets uploaded to ICPSR's DataLumos with metadata

Through February 2026, 15,000+ downloads from Data Lumos



Be the First – *Verify that the data have not already been preserved to ICPSR*

Describe Data – *Learn how to make the data more findable and useful*

Add Data Now – *Start uploading to DataLumos*

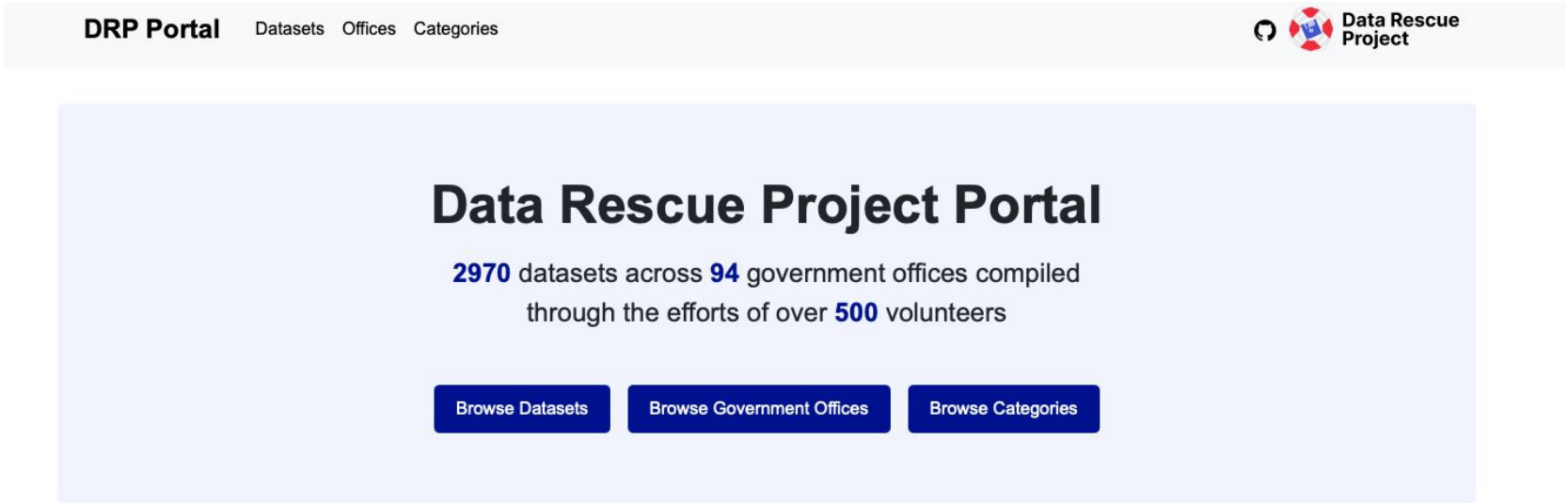
Data Rescue Project

www.datarescueproject.org/

bsky.app/profile/datarescueproject.org

Data Rescue Project Portal

portal.datarescueproject.org



Data Rescue Project

www.datarescueproject.org/

bsky.app/profile/datarescueproject.org

Partner Projects

Data Rescue Project

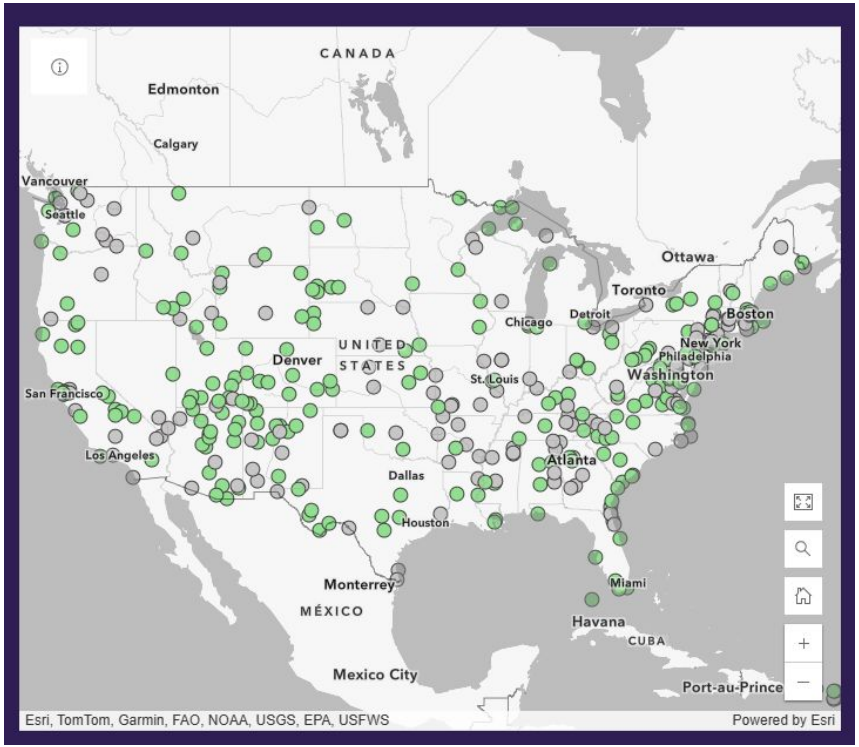
www.datarescueproject.org/

bsky.app/profile/datarescueproject.org

Save Our Signs

saveoursigns.org

- Crowdsourced effort to archive National Parks Interpretive signs
- Partnership with University of Minnesota and Safeguarding Research and Culture
- 10000+ photos submitted from around the country



Data Rescue Project

www.datarescueproject.org/

bsky.app/profile/datarescueproject.org

Community-Driven Strategic Plan to Build a Resilient Ecosystem for Federally Funded Research Data

How do we build a more resilient research data ecosystem that is resistant to single points of failure?

ASSESS, MONITOR, AND TRACK THE REPOSITORY LANDSCAPE

- Monitoring and inventorying at-risk repositories
- Developing a data repository maturity model that accounts for risk, resilience, and FAIRness
- Creating a shared repository health tool for continued monitoring

LAY A STRONG, RESILIENT, AND COORDINATED FOUNDATION

- Characterizing the core elements of a robust ecosystem
- Charting a community-informed roadmap for a consortium to coordinate on key elements
- Developing attributes of business models to support the ecosystem

DEVELOP A SHARED OUTREACH AND ADVOCACY STRATEGY

- Developing a framework for ongoing identification of and engagement with related initiatives
- Developing public communications and advocacy toolkits
- Building community capacity

We want to hear from you! data-resilience@cos.io

datarescueproject@
protonmail.com

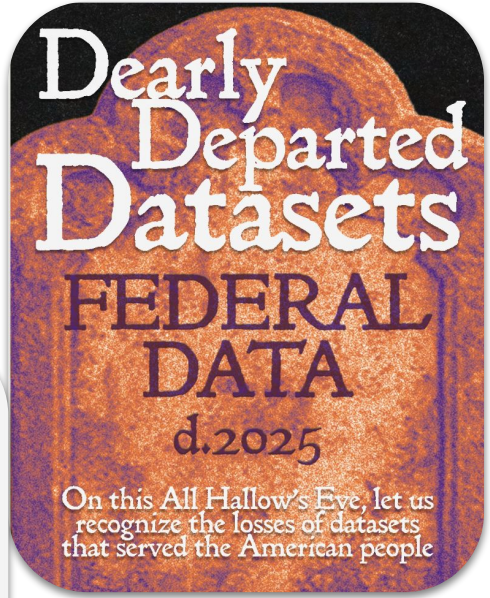
Data Rescue
Project

www.datarescueproject.org/

bsky.app/profile/datarescueproject.org

TIME
SUBSCRI
CLIMATE
POLICY
The Women Saving America's Climate Data
ADD TIME ON GOOGLE

ederal Data We're Thankful For · 7 pages
federal data
we're
thankful
for
FAS
America's Essential Data
apdu


Dearly Departed Datasets
FEDERAL DATA
d.2025
On this All Hallow's Eve, let us recognize the losses of datasets that served the American people

Thank you!