What if you could save the data?

Endangered Data Week & how libraries can protect public data

Joseph Koivisto, MSLIS

Systems Librarian
University of Maryland

https://go.umd.edu/mla2019

Some background

- **02 January 2017**: HR 05 - Remove requirement for long-term data analysis for ACA costs [1]
- **24 January 2017**: EPA cuts climate change page from website [2]
- **03 February 2017**: USDA blacks out animal welfare data [3]
- **08 March 2017**: HR 1430 -- Honest and Open New EPA Science Treatment Act [4]

[1] https://docs.house.gov/billsthisweek/20170102/BILLS-115hres5-PIH-FINAL.pdf
[2] https://www.reuters.com/article/us-usa-trump-epa-climatechange-idUSKBN15906G
[3] http://www.sciencemag.org/news/2017/02/usda-blacks-out-animal-welfare-information
[4] https://www.congress.gov/bill/115th-congress/house-bill/1430

Some background

- **July 2018**: Urgent Investigation of Destruction of Records on Family Separation Crisis [1]
- **October 2018**: Holes in History: The Dept of Interior request to destroy records [2]

[1] Jesse Franzblau. https://www.openthegovernment.org/coalition-calls-for-urgent-investigation-of-destruction-of-records-on-family-separation-crisis/
[2] James R. Jacobs. https://freegovinfo.info/node/13099

Endangered Data Week

Endangered Data Week

WAIT A SEC...

What is "Endangered Data"?

- **Deleted**
- **Repressed**
- **Mishandled**
- **Lost**

- **Deleted**
- **Repressed**
- **Mishandled**
- **Lost**
- *Never Collected*

https://endangereddataweek.org/

Endangered Data Week

Promote care for endangered collections by:

- publicizing the availability of datasets
- increasing critical engagement with them, including through visualization and analysis
- encouraging political activism for open data policies
- fostering of data skills through workshops on curation, documentation and discovery, improved access, and preservation

Why should we care about Endangered Data

It's <u>YOUR</u> data

Threats aren't new

But they aren't going away

Data continuity is on the line

Lives are on the line

What we can do about Endangered Data

Awareness

- Consciousness Raising

- Public awareness increases sense of ownership and understanding of value

- Serves as the foundational "Why" that guides conversations on what to save and how to save it

Awareness

- Tailored to your community need

  - What data is important to you and your patrons
    - Can be alienating, confusing if data is too big or too removed from audience needs

  - What important details can you convey about the data and its value

  - Consider spectrum of abilities necessary for accessing and understanding data

Awareness — Maryland

Maryland Open Data Portal

- https://opendata.maryland.gov/

Maryland Open Data Council

- https://opendata.maryland.gov/stories/s/Council-on-Open-Data/vyny-sdaq/

Awareness — Delaware

Delaware Open Data

- https://data.delaware.gov/

Delaware Data Council

- https://opendatacouncil.delaware.gov/

Awareness — Federal

Data.gov - Open data portal for federal government agencies

- https://www.data.gov/

Federal Register - official resource on federal Laws; presidential documents; administrative regulations and notices; and descriptions of federal organizations, programs and activities

- https://www.federalregister.gov/
- https://www.federalregister.gov/developers/api/v1 (API)

Skills-based literacy

- A lot of users need specialized instruction to work with and understand data

- Not all librarians have in-depth data skills

- But most of us work with data in some capacity or another
  - Genifer Snipes, "Everyone's a data librarian now"

[1] https://www.newlibs.org/article/3240

Skills-based literacy

- Data concepts & languages (CSV, TSV, XML, JSON)

- Programming languages & skills (Python, Git)

- Data visualization (Tableau, R)

- Webscraping, data collection & data preservation (ad hoc approaches & workflows, BagIt[1], checksums)

- Data assessment & interpretation (Open Refine, Excel)

[1] https://github.com/LibraryOfCongress/bagit-python

Skills-based literacy

- Helpful to present skills in task-oriented context

  - "I want to read this NOAA data so I can understand what it is saying about carbon emissions"

  - "I want to make graphs from this NASA data that displays ocean temperatures over time"

- Be familiar enough with the technologies you are teaching to be able to answer a wide array of questions

- If you are providing users with computers or applications, *make sure they work!*

Data preservation initiatives

- Initiatives can get you started without having to develop a full program from scratch

- Contributing helps to increase the scope and diversity of preserved data

- Can accommodate a wide range of skills (zero technical skills ←→ highly technical)

- Plugs you in to larger discourse of endangered data issues
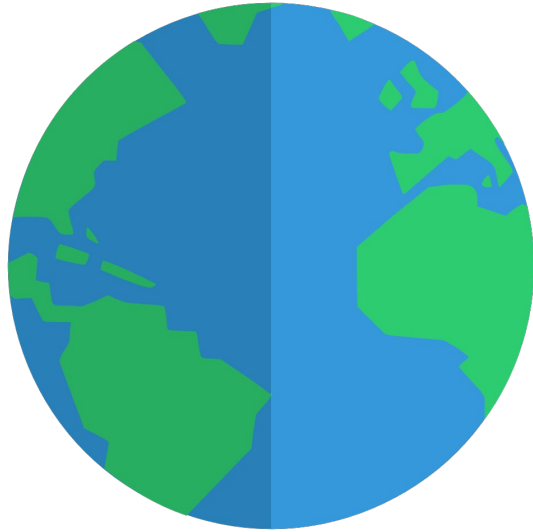
Data preservation initiatives

Data preservation initiatives

https://envirodatagov.org/

Data preservation initiatives

CLIMATE MIRROR

http://climatemirror.org/

Data preservation initiatives

https://www.datalumos.org/

Data preservation initiatives

https://archive.org/

Organizing & advocacy

- While data preservation is very important, proactive development of data policies and regulations is equally important

- Regulatory infrastructure defines much of governmental practice around data collection, preservation, and dissemination

- Engaging with legislators and the community on governmental data practices, regulations can motivate change

Organizing & advocacy

- Open Data Advocacy & Letter Writing[1]

- Teach-ins on data policy, data activism

- Legislative outreach & advocacy day

- Digital Library Federation's Government Records Transparency &

    Accountability group

    - https://www.diglib.org/groups/transparency-accountability/

[1] https://github.com/endangereddataweek/resources/blob/master/resource-advocacy.md

What can I do?

- Start planning early

- Review existing EDW resources
  https://github.com/endangereddataweek/resources

- Add events to EDW web page
  https://endangereddataweek.org/

Endangered Data Week

Less of an event, more of a lifestyle