Data Rescue
...after the smoke has cleared...

MARA ROJ ESKI BLAKE & REID BO EHM
JOHNS HOPKINS UNIVERSITY LIBRARIES
DATA SERVICES
Today’s Discussion

What is Data Rescue?

Current Efforts

Perspectives, Stakeholders

Things we’ve learned.

What’s next?
Defining Data Rescue

- Identified risks
  - Deprecation of files
  - Loss of documentation
  - Loss of public access

- Acknowledged efforts
  - Identify and Locate
  - Coordinate how and where

- Data Rescue is **not** new.

Federally Funded, Distributed, Diverse
Current, Ongoing Efforts

- Data Refuge
- Data Rescue Boulder
- Environmental Data and Governance
- Data Together
- Federation of Earth Science Information Partners
- Research Data Alliance

And More....

Johns Hopkins Libraries

EDGI Environmental Data & Governance Initiative
Data curation through OSF

Open Science Framework (OSF):
- Provides an online platform for both researchers and data curators to upload, organize and describe data collaboratively
- Is user friendly and easily accessible through free OSF accounts and OSF For Institution
- Data on OSF can be made available to the public
- Allows for data within its storage to be associated with other materials hosted outside of OSF through its various Storage Provider. This helps keep the data connected to relevant research/work

Fedora API Extension Architecture (API-X) and Package Ingest Extension (PIX)

U.S. 15 Minute Precipitation Data

This NOAA data collection, rescued by Data Rescue Boulder, is accessible to the public from the OSF. To see the original NOAA catalog entry visit: https://data.noaa.gov/datasets/U.S-15-Minute-Precipitation-Data

Originators: DOC/NOAA/NESSIS/NECI - National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce

Original Publishers: DOC/NOAA/...
Where did our energy from 2017 go?
And how do we proceed?
Current Related Issues – Spring 2018

- EPA admin move to limit data used in agency regulations
- Blocks long-standing landmark studies
- US Government considers charging for LandSat data
Stakeholders and Perspectives

- Citizens
- Educators
- Policy Makers
- Government Agencies
- Libraries
- Researchers
Relationships Among Stakeholder Communities (DataONE) Michener et al. 2012
ESIP Guidance

“Stronger Together – the case for cross sector collaboration in identifying and preserving at risk data.”

- Confirm risk level
- Collaborate with Data Centers
- Keep metadata with data
- Preserve Legacy data too
- Sync Efforts
What have we learned?
1. This is hard!

- Differences between Social and Technological parts
- What does “At Risk” really mean?
- What is the Scope?
- Formal measurements?
- The data is in diverse formats, multiple places, managed differently
2. The data is often BIG.
3. Metadata is a BEAST.

- Our interfaces may not support the previous documentation and accompanying visual tools.
- Accuracy and consistency is essential.
- Documentation influences use and interpretation.
4. Time and money are scarce.

- Volunteer Based
- Limited Infrastructure
- Distributed responsibilities?
- Large File-sharing solutions?
5. Politically charged = Difficult to discuss.

- Our institutions have specific loyalties, mindsets, cultures, funders.
- Careful word choices mean missing the full dialogue.
- Is it possible to be objective?
6. We need a shift in focus...

- Asking Questions
- Taking Action
- Creating Guides
- Offering Support
- Voicing Opinions
- Listening to Needs
What’s next?
Coordination and Communication

- Research Data Librarians as bridges between perspectives?
- Engage diverse voices across disciplines
- Materials can speak multiple dialects
- Simplicity
Focus on Sustainability

- Making space for different priorities and types of rescue.
- Resources
- Current Actions
- Participants
- Open materials need to be discoverable and adaptable over time.

Future Needs and Outcomes
Keeping Relevant, Maintaining Momentum

- Timely conversations in strategic places and spaces.
- Engage in projects with clear, simple, achievable objectives.
IMLS Planning Grant

June 2018 – May 2019
Community-created Data Rescue Tool-kit

September Meeting at Johns Hopkins University
• Coordinate distributed efforts
• Create Tool-kit Blueprint
• Document and Share Process model for similar community-based endeavors

After May 2019 – Making the toolkit a reality…
Conclusion

- Our efforts are not over
- This is a wicked problem
- We can play an important role in representing views
- Taking action = taking a service driven approach
- IMLS Grant is an exciting development to include more voices in one space through time.
References


Questions?