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

MARA ROJESKI BLAKE & REID BOEHM
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

Federally Funded, Distributed, Diverse

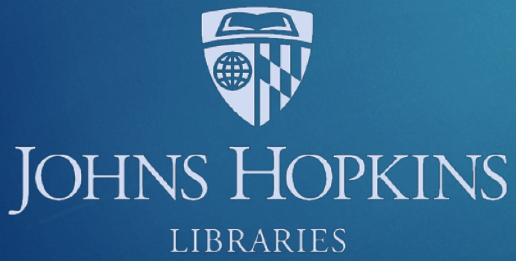
- Acknowledged efforts
  - Identify and Locate
  - Coordinate how and where

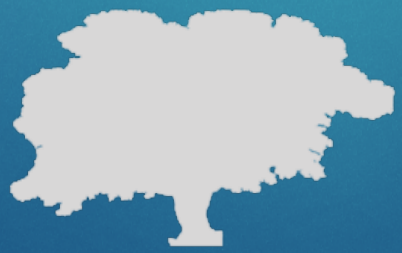
- Data Rescue is not new.

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….



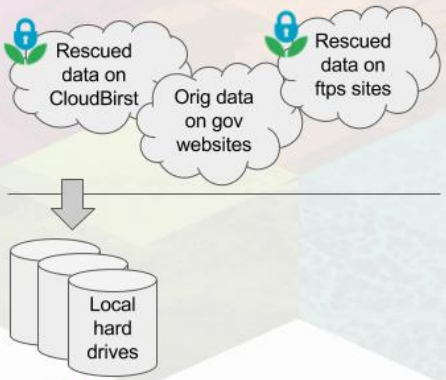


PENN PROGRAM IN
ENVIRONMENTAL HUMANITIES

Pilot Work

Data curation through OSF

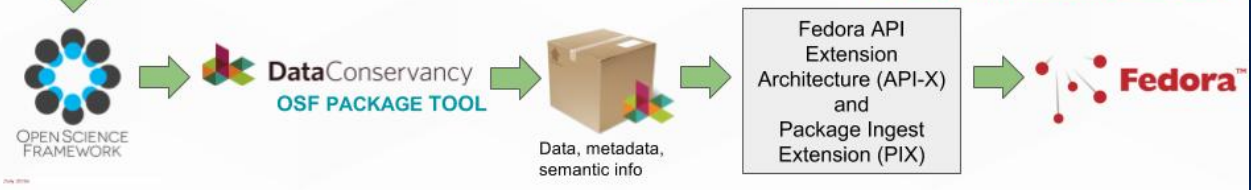
JOHNS HOPKINS LIBRARIES DataConservancy



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

Link to OSF workflow demo video



OSFHOME

NOAA 15 MIN Precipitation Data | Files | Wiki | Analytics | Re

NOAA 15 MIN Precipitation Da

Contributors: **Reid Boehm**, Karen L. Hanson, Chen Chiu, Hanh Vu, Mark Patton, Aaron

Affiliated Institutions: Johns Hopkins University

Date created: 2017-04-19 03:37 PM | Last Updated: 2017-08-18 09:48 AM

Create DOI

Category: Project

Description:

U.S. 15 Minute Precipitation Data is digital data set DSI-3260, archived at the National Centers for Environmental Information (NCEI). [Formerly NCDC] This is precipitation data. The primary source of data for this file is approximately 2,000 mostly U.S. weather stations operated or managed by the U.S. National Weather Service. Stations are primary, secondary, or cooperative observer sites that have the capability to measure precipitation at 15 minute intervals. This dataset contains 15-minute precipitation data (reported 4 times per hour, if precip occurs) for U.S. stations along with selected non-U.S. stations in U.S. territories and associated nations. It includes major city locations and many small town locations. Daily total precipitation is also included as part of the data record. NCEI has in archive data from most states as far back as 1970 or 1971, and continuing to the present day. The major parameter is precipitation amounts at 15 minute intervals, when precipitation actually occurs.

License: CC0 1.0 Universal

Wiki

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/dataset/u-s-15-minute-precipitation-data

Originators: DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce

Original Publishers: DOC/NOAA/...

Read More

Files

Click on a storage provider or drag and drop to upload

Filter

Name Modified

Citation osf.io/grhz7

Components Add Component Link Projects

NOAA Data Catalog dsi 3260

Boehm, Hanson, Chiu & 9 more

NCDC Data documentation for DSI - 3260 15 Minute Precipitation Data, created April 26, 2005

NOAA Cooperative Station Information and Map File

Boehm, Hanson, Chiu & 9 more

This information corresponds to the geographic location of the cooperative stations. The two digit number in each of the data set files represents a s...

NOAA 15 Minute Precipitation Data Documentation February 2016

Boehm, Chiu, Vu & 8 more

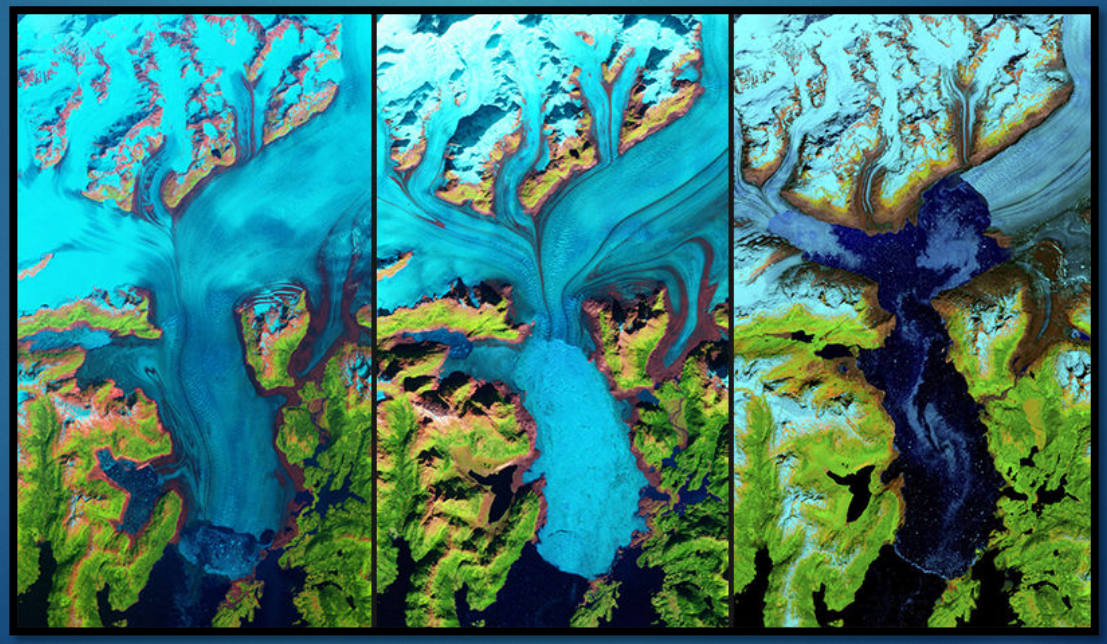
Documentation related to NOAA Precipitation data, this file covers the data that is included in the online system. Please note that output format will...

Where did our energy from 2017 go?

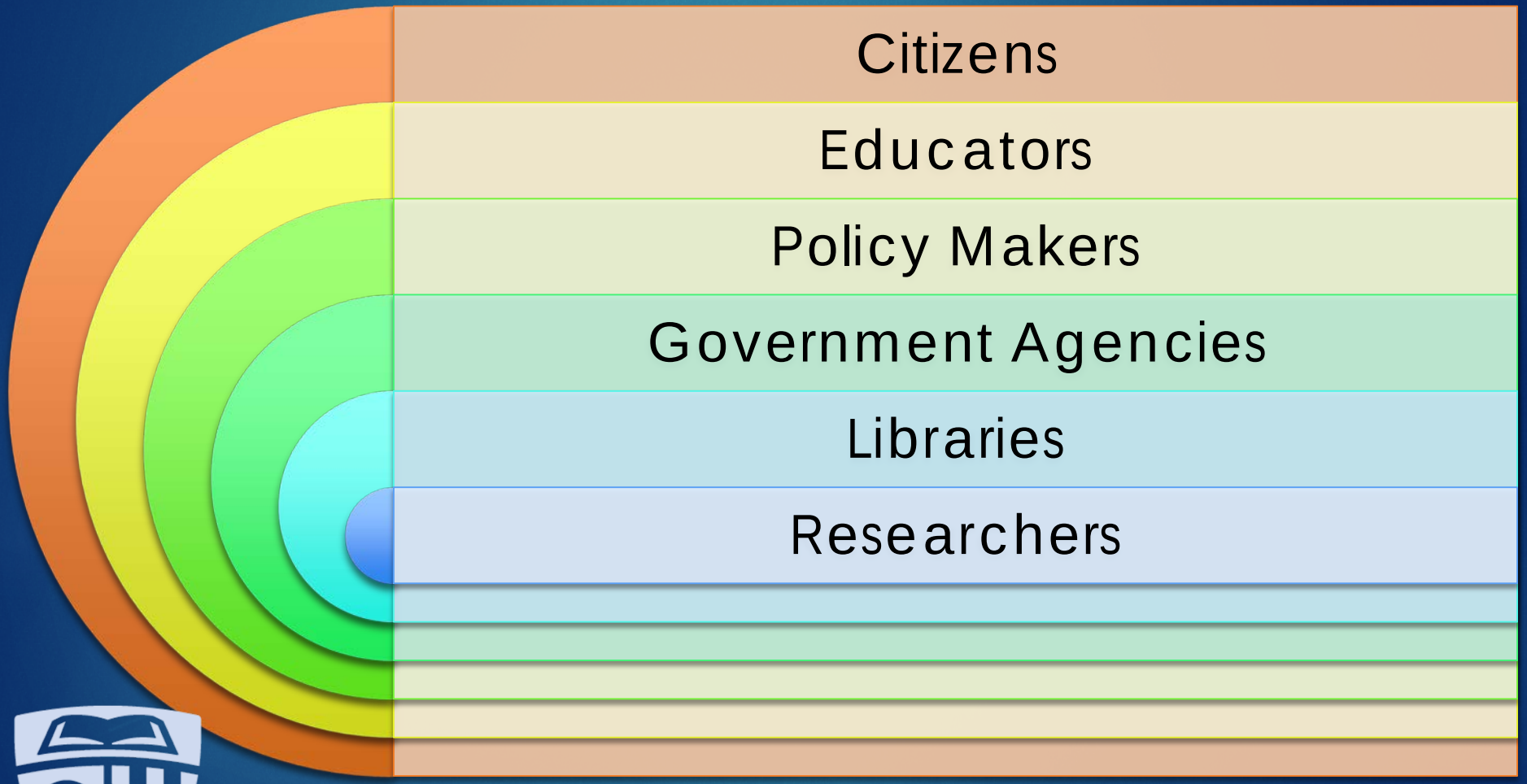
And how do we proceed?

Current Related Issues – Spring 2018

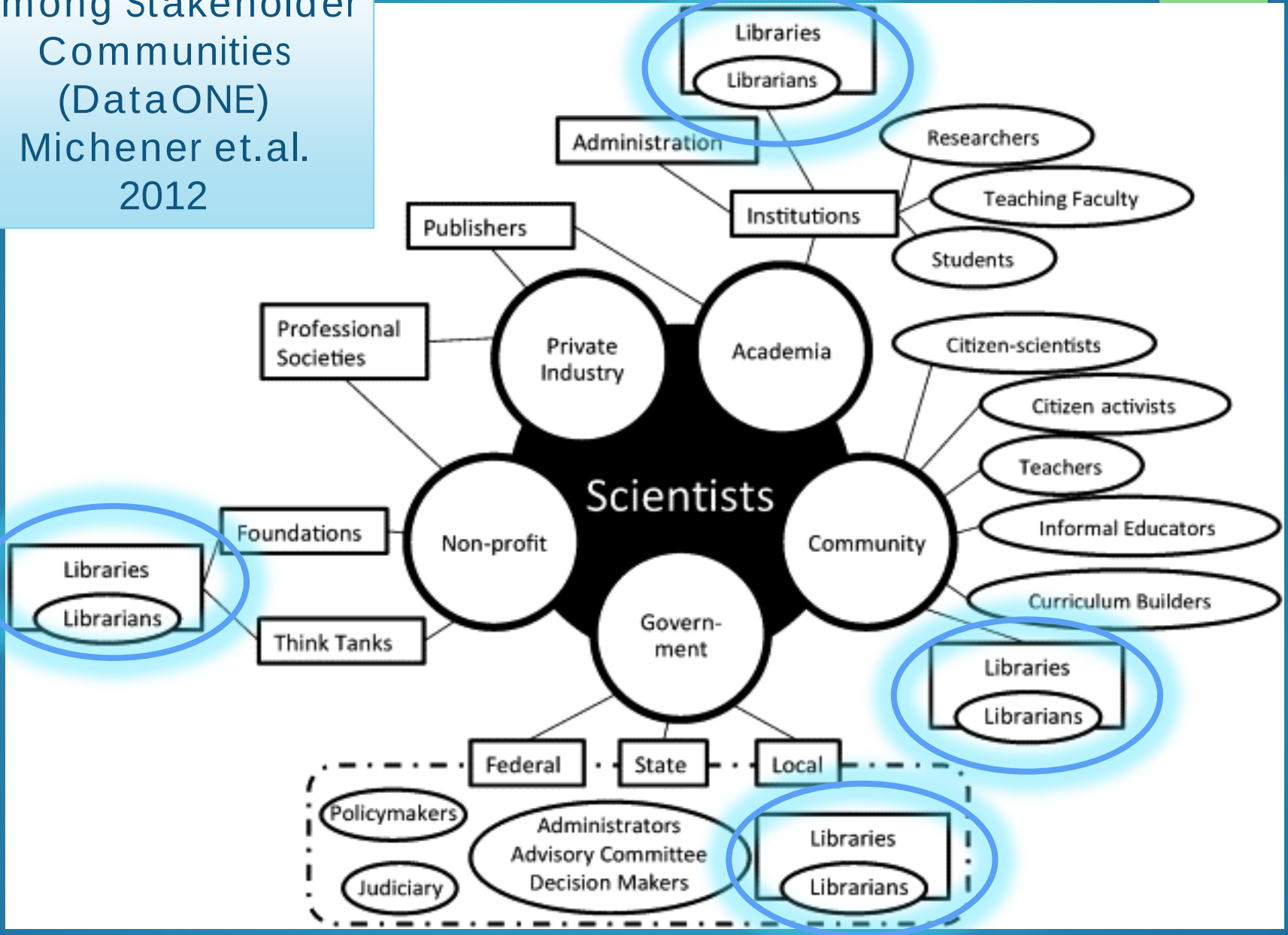
- EPA admin move to limit data used in agency regulations
  - Blocks long-standing landmark studies
- US Government considers charging for LandSat data



Stakeholders and Perspectives



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."

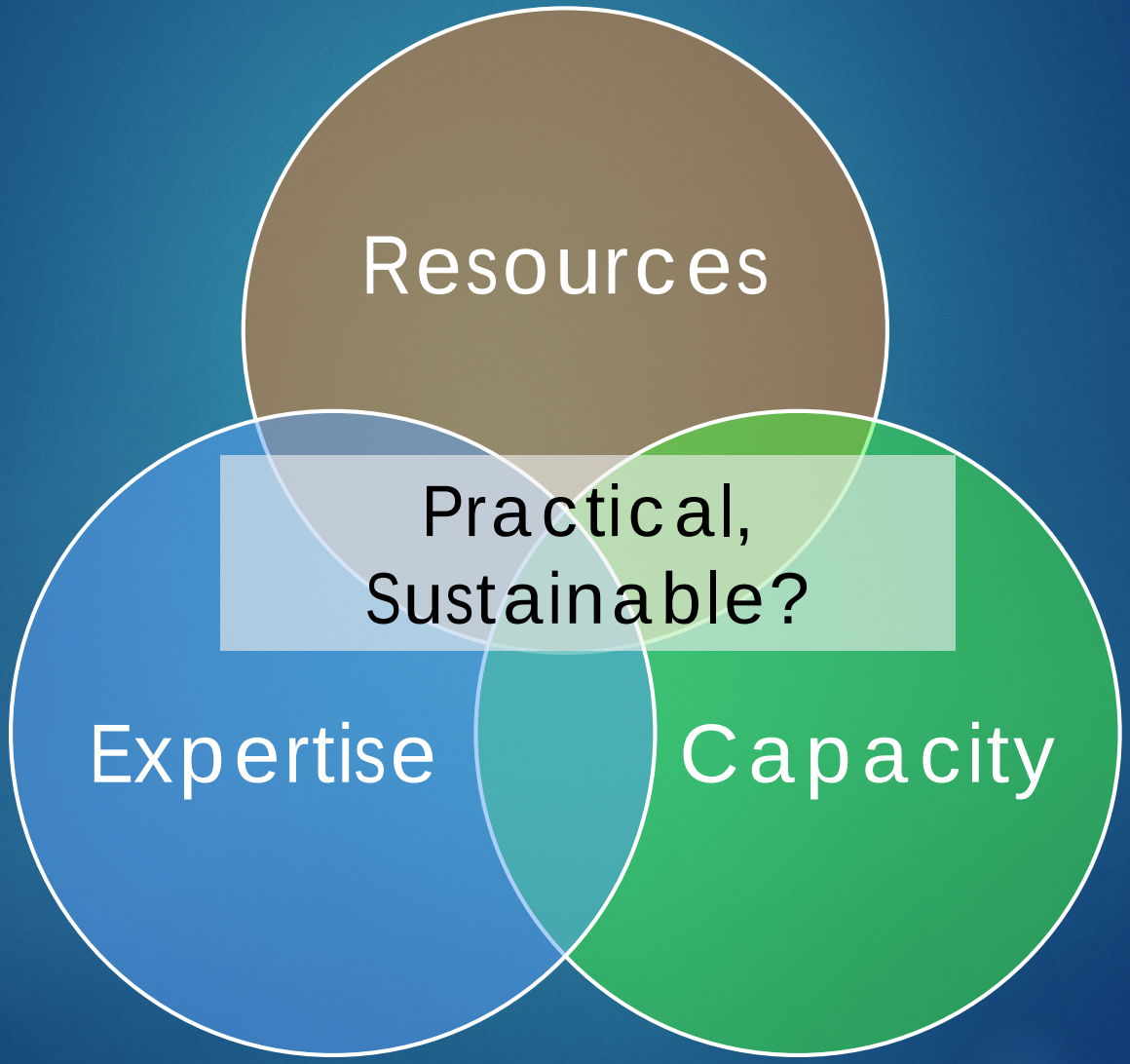
- Confirm risk level
- Collaborate with Data Centers
- Keep metadata with data
- Preserve Legacy data too
- Sync Efforts

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.

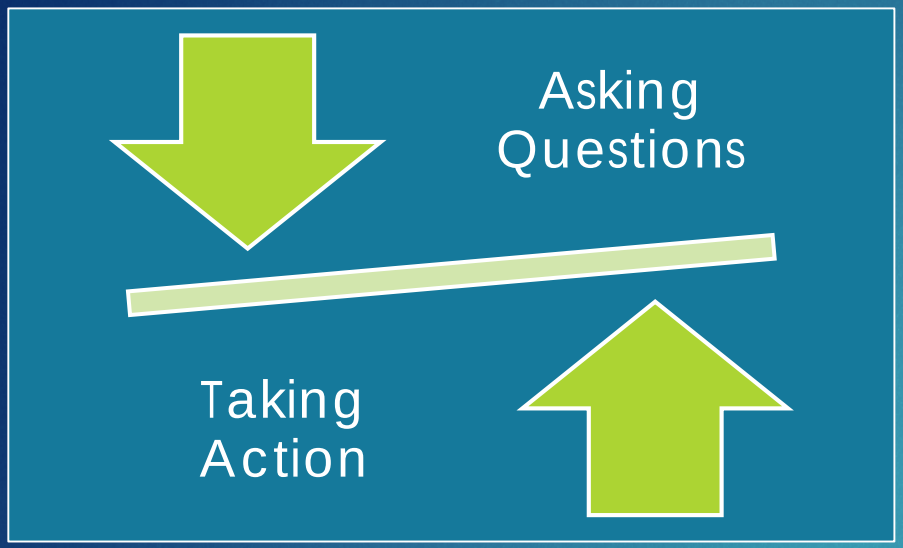
4. Time and money are scarce.

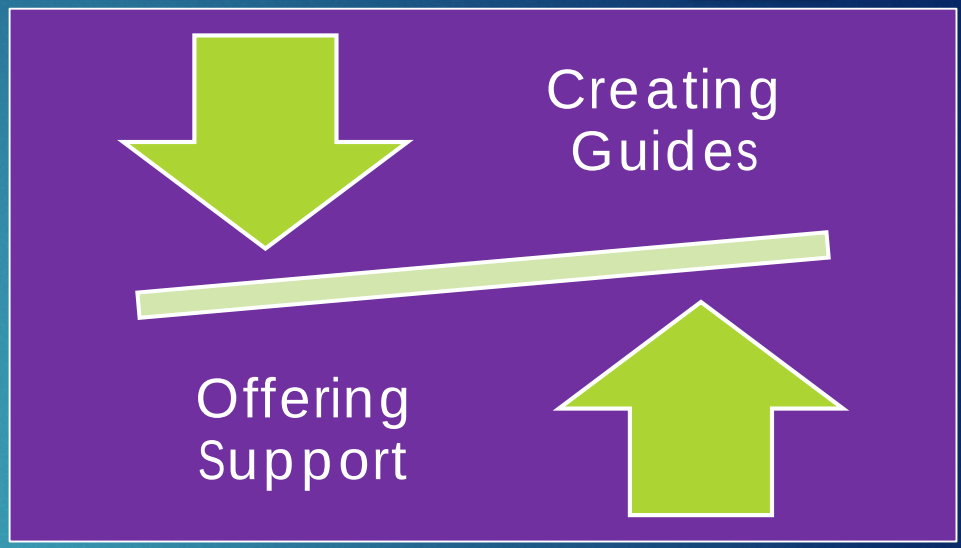
- Volunteer Based
- Limited Infrastructure
- Distributed responsibilities?
- Large File-sharing solutions?

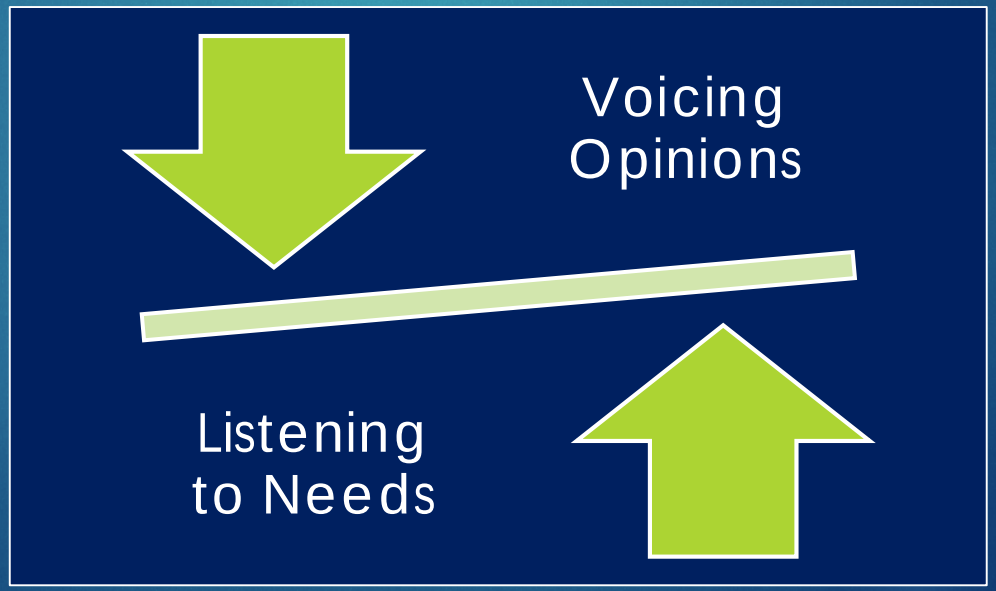
5. Politically charged = Difficult to discuss.

- 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...



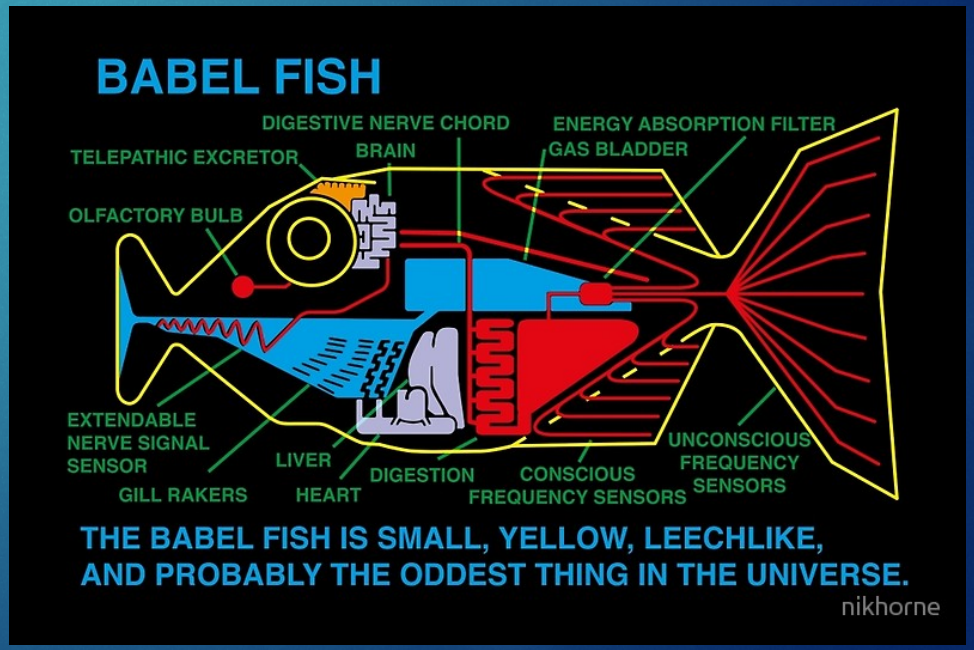




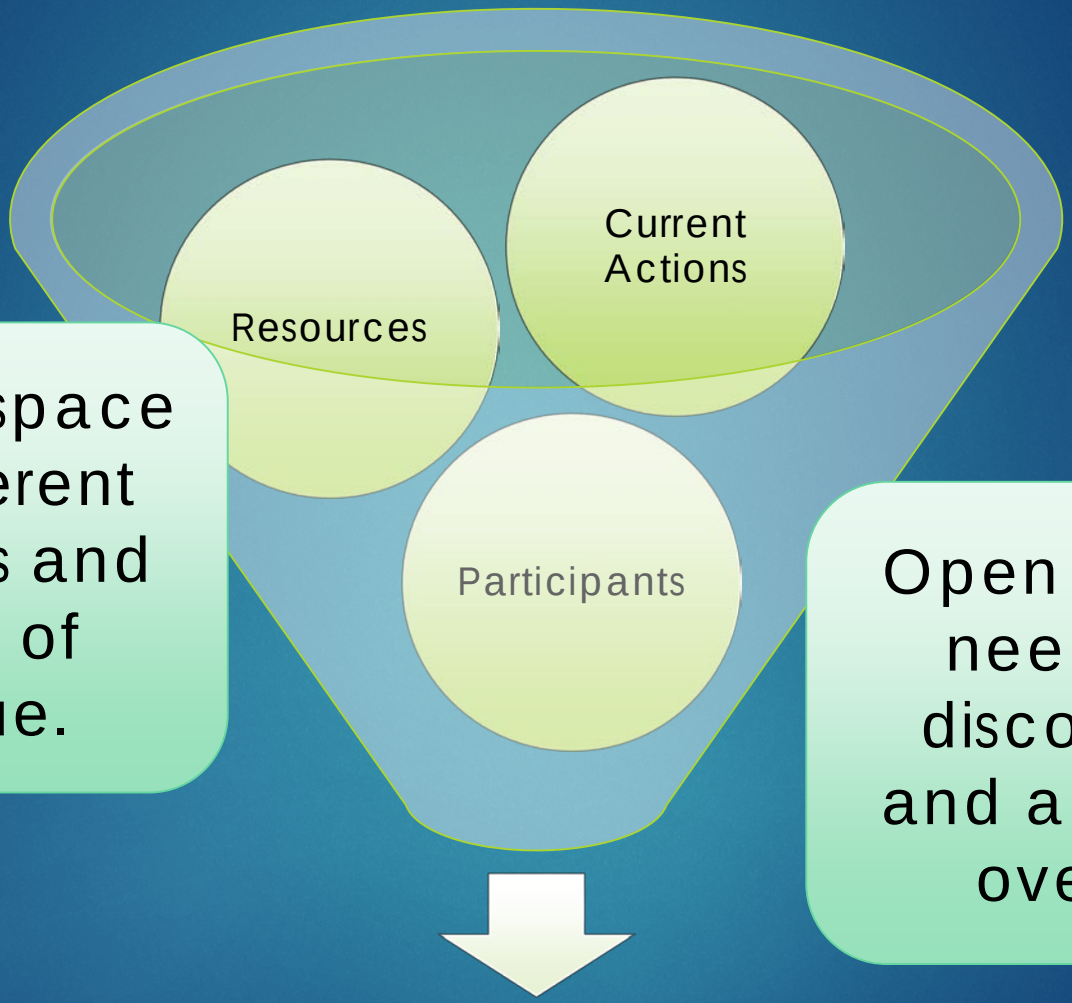
What's next?

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.

Open materials need to be discoverable and adaptable overtime.

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

Michener, W. K., Allard, S., Budden, A., Cook, R. B., Douglass, K., Frame, M., … Vieglais, D. A. (2012). Participatory design of DataONE—Enabling cyberinfrastructure for the biological and environmental sciences. *Data Platforms in Integrative Biodiversity Research*, *11*, 5–15. https://doi.org/10.1016/j.ecoinf.2011.08.007

https://www.washingtonpost.com/news/energy-environment/wp/2018/04/24/pruitt-to-unveil-controversial-transparency-rule-limiting-what-research-epa-can-use/?noredirect=on&utm_term=.7654d75b7482

Mayernik, M., Downs R., Duerr, R., Hou, S., Meyers, N.; Ritchey, N., Thomer, A., Yarmey, L. (2017) Stronger together: the case for cross-sector collaboration in identifying and preserving at-risk data. ESIP. Paper. https://doi.org/10.6084/m9.figshare.4816474.v1


Questions?