



Crowdsourcing Government Information Backup: The Activism and Implications of DataRescue

Sangeeta Desai - University of North Carolina at Chapel Hill

David Pcolar - Digital Preservation Network

Michelle Padley - Department of Geography, UNC-CH

DataRescue Chapel Hill: A Grassroots Platform for Government Data Preservation and Education



datarescue | chapel hill



DataRefuge: A Grassroots Movement

...committed to identifying, assessing, prioritizing, securing, and distributing reliable copies of federal climate and environmental data so that it remains available to researchers. Data collected as part of the #DataRefuge initiative will be stored in multiple, trusted locations to help ensure continued accessibility.

Data Rescue events create trustworthy copies of federal climate and environmental data (among other things), while the Internet Archive, datarefuge.org, and a consortium of major research libraries *holds* these copies

DataRescue Chapel Hill - March 4, 2017

Funding and support from UNC-SILS and UNC-Libraries

Held at the Research Hub, Davis Library, UNC-Chapel Hill

87 volunteers from UNC, Duke, NCSU, RTP, and the local community

Included librarians, archivists, professors, graduate students, undergraduates, environmental activists, IT professionals and community members

Focused on identification, harvesting, describing and preserving government websites, documents and datasets



datarescue | chapel hill



datarescue | Chapel Hill

DATA RESCUE CHAPEL HILL

March 4, 2017 | 10:00am - 12:00pm

Research Hall

Davis Library, 208 Raleigh

Register

[DataRescue Workshop](#)

Join us for a DataRescue event, with support from the School of Environment and Natural Resources and the School of Library and Information Science at UNC-Chapel Hill. In collaboration with [DataRefuge](#), a project, DataRescue Chapel Hill aims to preserve online information about environmental programs that are identified as high risk for removal.

Roles requiring advanced technical and/or archival skills

RESEARCHERS:

Researchers review URLs the Seeders and Sorters mark as "Uncrawable". Consider this path if you have good listening and writing skills, and/or if you can make creative engaging materials.

HARVESTERS:

Harvesters figure out how to capture the uncrawable data. Consider this path if you're a "technical expert" and try to figure out how to capture it. This is a complex task that requires technical expertise and also requires different techniques for different tasks.

CHECKERS:

Checkers inspect a harvested dataset and make sure that it is complete. The main question is "will the bag make sense to a scientist?" Checkers need to have an in-depth understanding of the data and potential content variations for datasets. Previous work with scientific data is helpful.

BAGGERS:

Baggers perform a quality assurance check and package the data. Consider this path if you have archiving experience, or have strong tech skills and attention to detail. As a bagger you create a bag file (or "bag"), which includes basic technical metadata, and upload it to the DataRescue website.

DESCRIBERS:

(This may include a few people from the Baggers path) Consider this path if you have experience with scientific data, particularly climate or environmental data, or with creating metadata. This role will be very helpful on this path. Describers will create a CKAN account to help organize datasets and URLs that have been downloaded. They then link to bags of data to make the data public and accessible.

Roles requiring basic or no technical skills

DOCUMENTATION OR STORYTELLING:

Consider this path if you're on social media (Facebook, Instagram, Twitter, whatever), if you have good listening and writing skills, and/or if you can make creative engaging materials.

SEEDING AND SORTING:

This is the widest path and requires a variety of skill levels. Consider this path if you are a "technical expert" and have web experience, or just have a great attention to details.

THE LONG TRAIL:

Consider this path if you'd like to build DataRescue into the future. Future projects will also exist that exists but can't be captured in a single weekend as well as data that doesn't exist, but is needed. Experience in public engagement projects and informal STEM education settings will be helpful.

EVENT ORGANIZATION AND COORDINATION:

Consider this path if you are interested in helping to organize the DataRescue and/or the event smoothly.

Connect

[Tweet #DataRescueCH](#)

[Find us on Facebook](#)

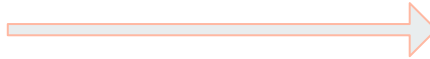


Teams and Roles for Participants

* Primer Writing Group

Seeding and Sorting (Web Archiving)

DataRescue Team



The Long Trail

*Event Organization and Coordination

Researchers

Harvesters

Checkers

Baggers

Describers



Why DataRescue Chapel Hill was a Success

Preservation:

WebArchiving - Worked through the entire Department of Water primer
Datasets - Harvested and Bagged at least 10 datasets

Education about:

Endangered Data and its Causes
Government Data Public Rights
Digital Preservation Basics (Always Back it Up!)

WebArchiving and Data Archiving Experience

Opportunity/Outlet to Act

Community



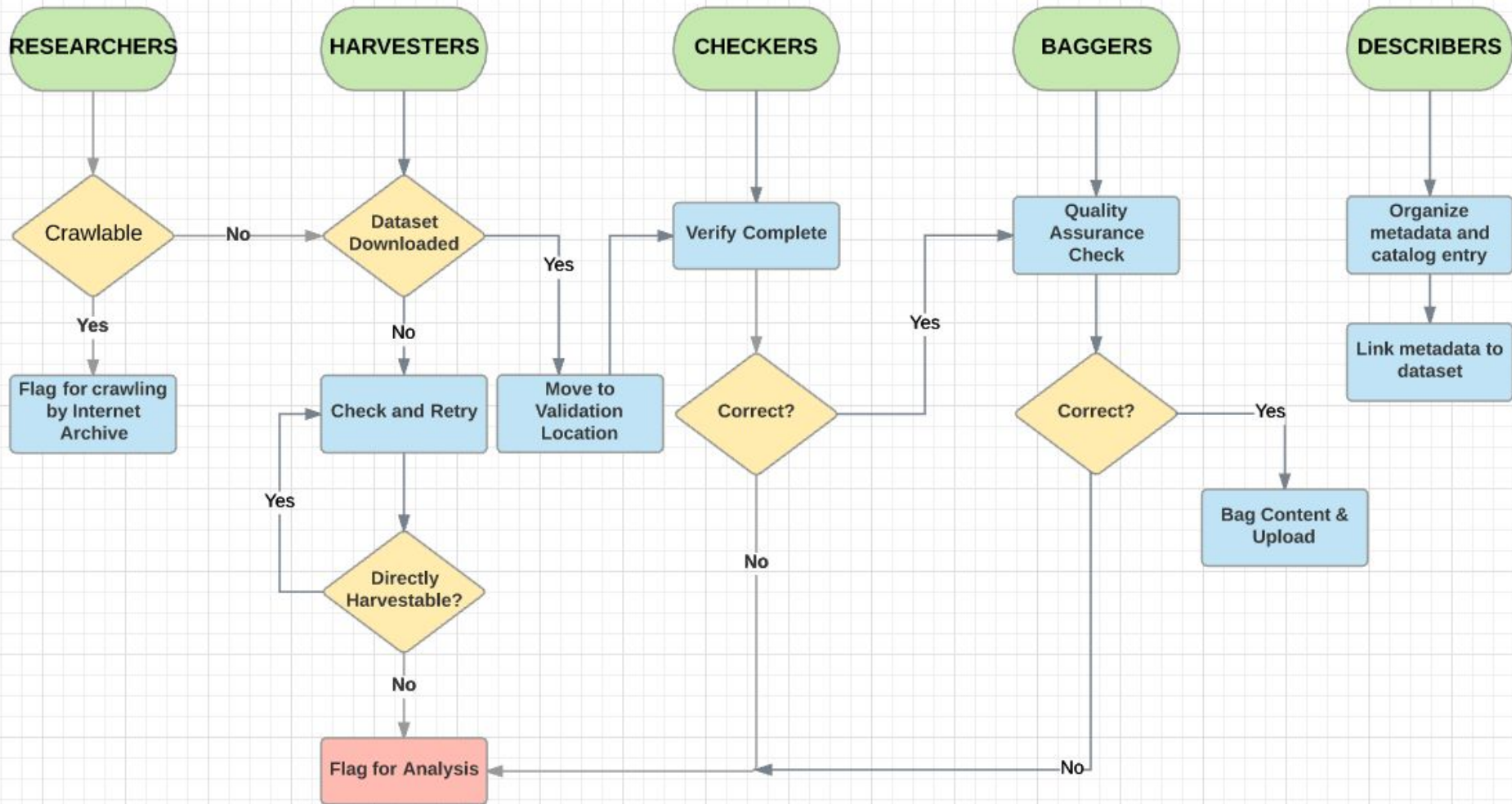
DataRescue: Popularity, Trust, and Provenance

Google searches:

mirroring "government data"... About 263,000 results

datarefuge... About 70,700 results

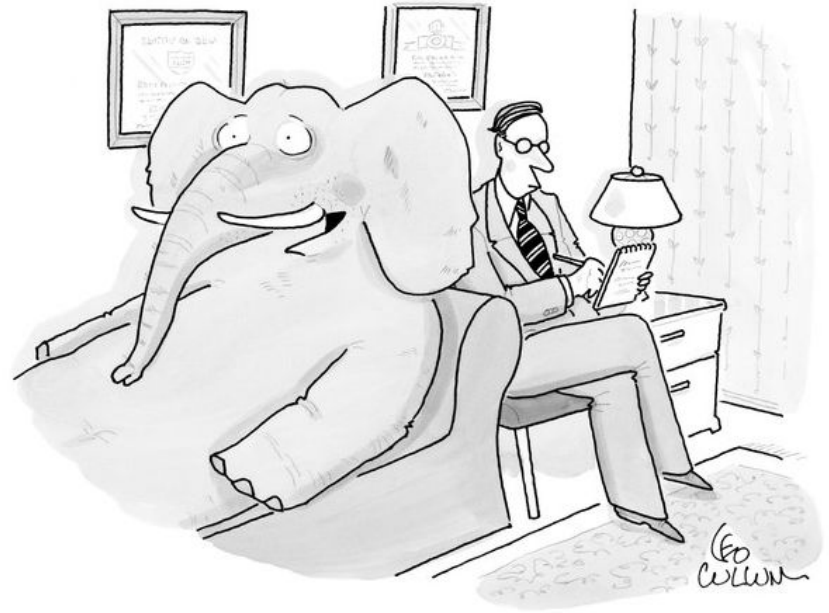




Trust, Threats, & Reuse


Is the harvested data trustworthy?

- Establishing provenance
- Certifying chain of custody
- Verifying authenticity



"I'm right there in the room, and no one even acknowledges me."

October 27, 2014

A decorative horizontal line with a teal segment on the left and an orange segment on the right.

DataRefuge is a public, collaborative project designed to address the following concerns about federal climate and environmental data:

- What are the best ways to safeguard data?
- How do federal agencies play crucial roles in data collection, management, and distribution?
- How do government priorities impact data's accessibility?
- Which projects and research fields depend on federal data?
- Which data sets are of value to research and local communities, and why?



Cooperation and Collaboration

- There are no guarantees of perfectly safe information
- We can create safe and trustworthy copies
- Working with data providers will assure trust
- Making the data reliably discoverable is key to reuse



**SAFETY IS
THE PRESENCE
OF POSITIVES**
not the absence
of negatives



Working with Data Users

- Subject matter experts are not necessarily data experts
- The crucial role of librarians



Librarians and archivists are quite literally in the business of preserving and organizing data.

<https://pennwic.wordpress.com/tag/data-refuge/>

Barriers to Working with Data Users

- Data users and subject matter experts may take data availability for granted
- Lack of technical training
- Hierarchical challenges and communication issues
- Disciplinary baggage when talking to graduate students and faculty





Questions?

Sangeeta Desai - sdesai11@live.unc.edu

School of Information & Library Science
University of North Carolina at Chapel Hill

David Pcolar - dave@dpn.org @pcolar

Michelle Padley - mpadley@live.unc.edu

Department of Geography
University of North Carolina at Chapel Hill