

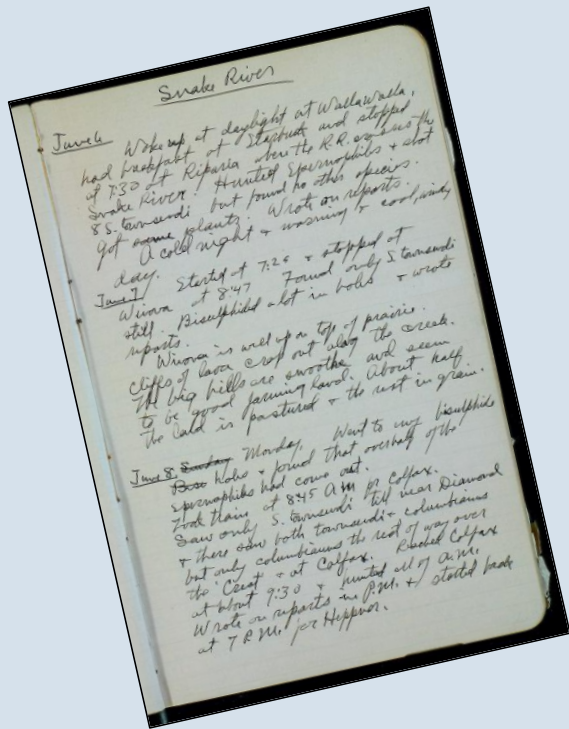
Qualitative Data Management for Interdisciplinary Research

KRISTAL JONES, RESEARCH SCIENTIST, NATIONAL SOCIO-ENVIRONMENTAL
SYNTHESIS CENTER (SESYNC), UNIVERSITY OF MARYLAND-COLLEGE PARK

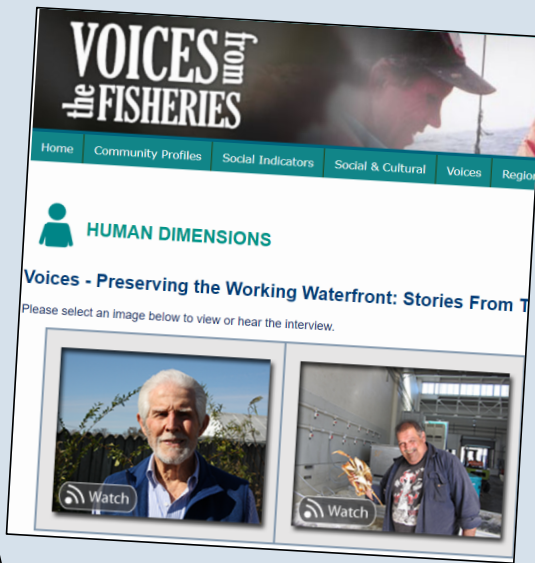
STEVEN M. ALEXANDER, MITACS SCIENCE POLICY FELLOW AND SCIENCE
ADVISOR, FISHERIES AND OCEANS CANADA

DataONE webinar, April 10, 2018

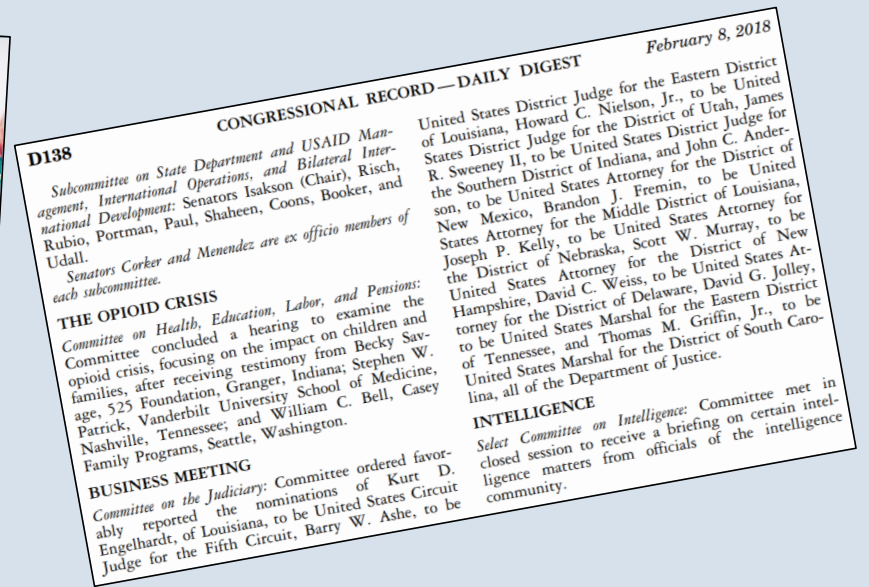
What is qualitative data?



[National Museum of Natural History](#)



[NOAA Voices from the Fisheries](#)



[US Congressional Record](#)

What is qualitative data?



Author's photos

Standing Rock Protest Camp, Once Home to Thousands, Is Razed

By MITCH SMITH FEB. 23, 2017

RELATED COVERAGE

- North Carolina
- North Dakota
- South of the Border
- Ocean Tribes
- Army Accuses

Crews cleaned up the Oceti Sakowin camp in North Dakota on Tuesday, the day before the deadline for protesters to leave. [Click here for The New York Times](#)

MANDAN, N.D. — The final holdouts at the sprawling pipeline protest camp south of here were arrested Thursday, and the authorities began using heavy equipment to tear down the remaining structures and clear debris on the federally owned land where thousands had lived in recent months.

The arrests, of 46 people, came a day after an evacuation deadline issued by Gov. Doug Burgum. Most protesters left Wednesday of their own volition, and others departed Thursday by crossing the frozen Cannonball River to the Standing Rock Sioux Reservation. Those who remained at the main campsite were taken into custody.

[The New York Times](#)

NRDC @NRDC · Mar 2
Government emails show the real reason behind Trump's decision to eliminate #BearsEars National Monument: oil drilling and coal mining.
[nytimes.com/2018/03/02/cli...](#)
56 1.2K 1.4K

MSNBC @MSNBC · Mar 2
Interior Sec. Ryan Zinke reduced Utah national monument land by 85%, and he said it "isn't really about oil and gas at all."
Here's why a new report says oil was central in the decision:
[on.msnbc.com/2GXW9NL](#)
106 662 625

John Schwartz @jswatz · Mar 2
The team of @EricLiptonNYT and @LFFriedman with another big scoop: internal Interior documents show that shrinking Bears Ears monument was hugely about drilling and mining. [nytimes.com/2018/03/02/cli...](#)
64 862 775

Center for Bio Div @CenterForBioDiv · Mar 2
Surprise, surprise....it wasn't out of concern for "federal overreach" or "states' rights" that they decided to shrink down #BearsEars. They wanted to plunder the resources lying underneath our public lands.
[nytimes.com/2018/03/02/cli...](#)
64 1.9K 2.6K

[Twitter \(#BearsEars\)](#)

What role does qualitative data play in interdisciplinary research?

Qualitative data can...

- Provide temporal insights
- Form case comparisons
- Scale up patterns
- Scale down interpretations
- Broaden the evidence base

What are the benefits of sharing and re-using qualitative data?

Scientific

- Transparency and triangulation

Descriptive

- Expansive, inclusive, and varied aspects of a phenomena
- Opportunities for teaching & learning

Material

- Reduce the burden on individuals and communities
- Increased return on investment for funders and institutions
- Access for knowledge users outside of research institutions

Is qualitative data currently being shared?



A screenshot of the ICPSR website. The header includes the ICPSR logo and the text 'Start Sharing Data' and 'Log In/Create Account'. A navigation bar contains links for 'DATA PREPARATION GUIDE', 'CONFIDENTIALITY', and 'SUGGEST DATA TO ARCHIVE'. The main content area is titled 'Guide to Social Science Data Preparation and Archiving: Introduction' and lists several topics: 'Importance of Data Sharing and Archiving', 'Planning Ahead for Archiving and Preservation of Data', 'The Data Life Cycle', and 'Using the Guide'. A sidebar on the left lists 'Data Preparation Guide' and 'Introduction' with a numbered list of sections. A small graphic on the right shows a globe with data lines.

A screenshot of the SESAMAD website. The header includes the Dartmouth | SESMAD logo and navigation links for 'Home', 'Manual', 'Cases', 'Components', and 'Theories'. The main heading is 'Social-Ecological Systems Meta-Analysis Database: Cases'. Below the heading is a paragraph of text describing the database and a large illustration of a bird in flight.

A screenshot of the Harvard Dataverse website. The header includes the Harvard logo and the text 'Dataverse'. Below the header is a 'Metrics' section showing '3,155,384'. A search bar contains the text 'qualitative'. Below the search bar is a list of results: 'Dataverses (8)', 'Datasets (513)', and 'Files (182)'. Each result has a checkmark icon.

What challenges exist to accelerating qualitative data sharing and re-use?

Practical challenges

- Resources: Time, expertise, financial support
- Infrastructure: Where to deposit?

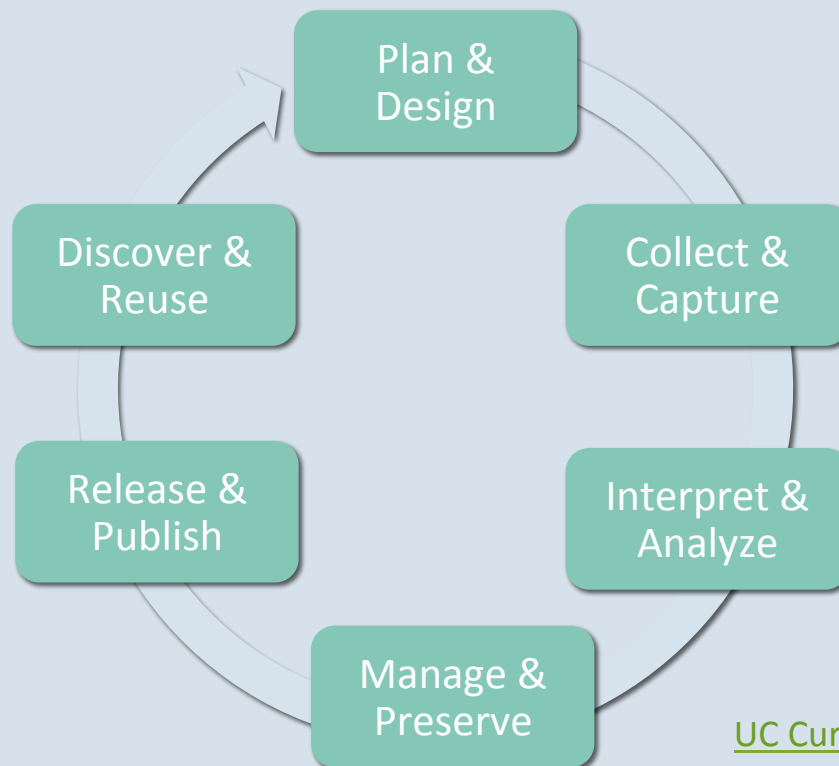
Ethical

- Confidentiality, representation, consent

Epistemological

- Spectrum from pure positivism to pure constructionism, with lots of pragmatic space in the middle

Research Data Lifecycle



Who plays a role in addressing these challenges?

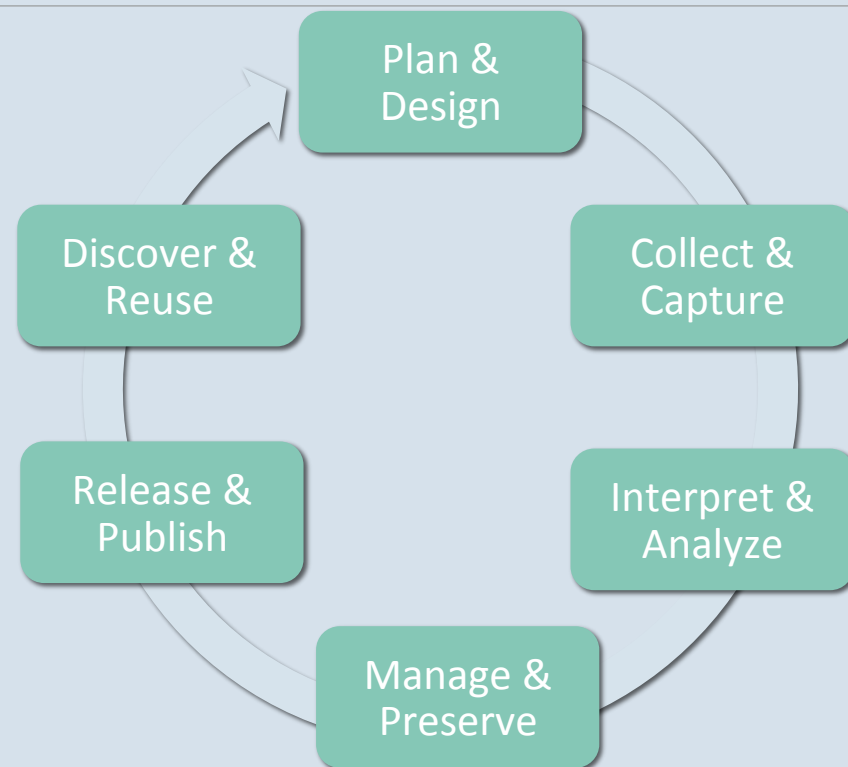
Researchers

Research institutions

Journals and publishers

Funders

Data repositories



What resources exist to address qualitative data management challenges?

Primary data lifecycle:

Plan and Design

- Data management planning:
<https://qdr.syr.edu/guidance/managing/planning-data-management>
- IRB : <https://qdr.syr.edu/node/20260/>
- Develop shared protocols:
<https://www.atkinson.cornell.edu/collaborations/oxfam-cu.php>
- Teaching qualitative data management webinar from IASSIST:
<https://www.youtube.com/watch?v=aATIKsF96Ro&feature=youtu.be>

What resources exist to address qualitative data management challenges?

Primary data lifecycle:

Collect and Capture, Interpret and Analyze

- Myriad resources from environmental anthropology and sociology, human geography
- Communicate methodologies, approaches and assumptions:
Cox, M. (2015). A basic guide for empirical environmental social science. *Ecology and Society* 20(1): 63.
<http://dx.doi.org/10.5751/ES-07400-200163>
- Sharing code from qualitative data software:
RQDA <https://cran.r-project.org/web/packages/RQDA/RQDA.pdf>
Code books from Atlas.ti, NVivo, MAXQDA

What resources exist to address qualitative data management challenges?

Primary data lifecycle:

Release and Publish, Discover and Re-use

What is unique about publishing qualitative data for re-use in interdisciplinary research?

What resources exist to address qualitative data management challenges?

Secondary data lifecycle:

Plan and Design, Collect and Capture

What is unique about interdisciplinary research using secondary qualitative data?

What resources exist to address qualitative data management challenges?

Secondary data lifecycle:

Interpret and Analyze

- Overview of synthesis methods: Dixon-Woods et al. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research and Policy* 10(1): 45-53. <http://dx.doi.org/10.1258/1355819052801804>
- Meta-analysis: Cox, M. (2014). Understanding large social-ecological systems: introducing the SESMAD project. *International Journal of the Commons* 8(2): 265-276. <http://doi.org/10.18352/ijc.406>
- Text mining: <http://tm.r-forge.r-project.org/>
- Regular expressions: <https://sesync-ci.github.io/text-mining-lesson/2016/09/14/>

What resources exist to address qualitative data management challenges?

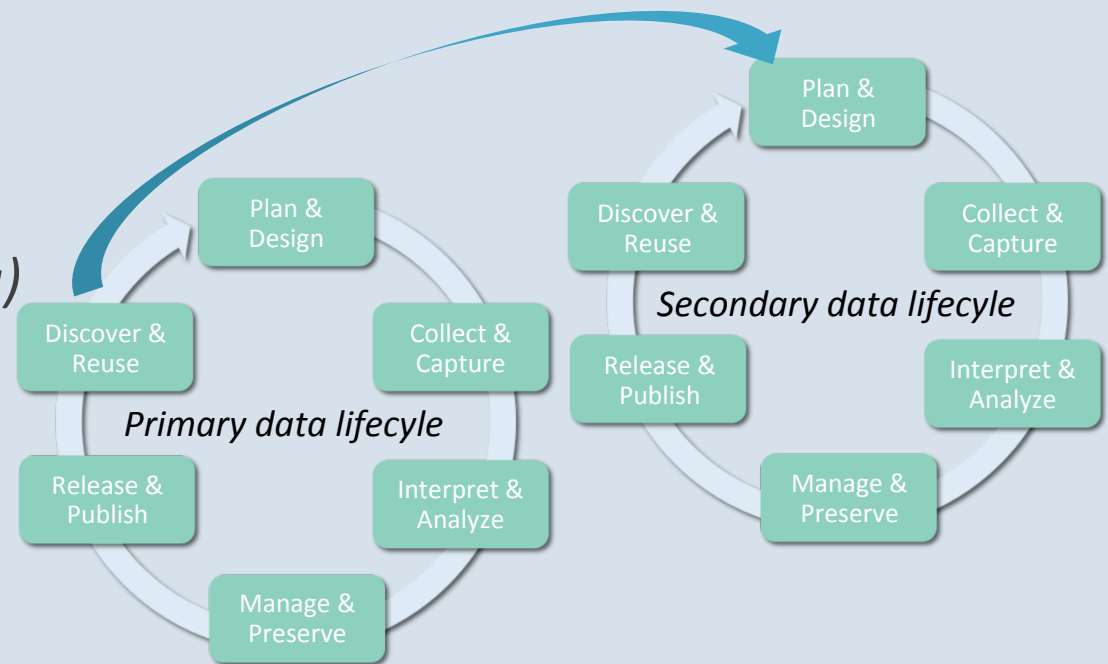
Secondary data lifecycle:

Manage and Preserve, Release and Publish

What is unique about publishing qualitative data for re-use in interdisciplinary research?

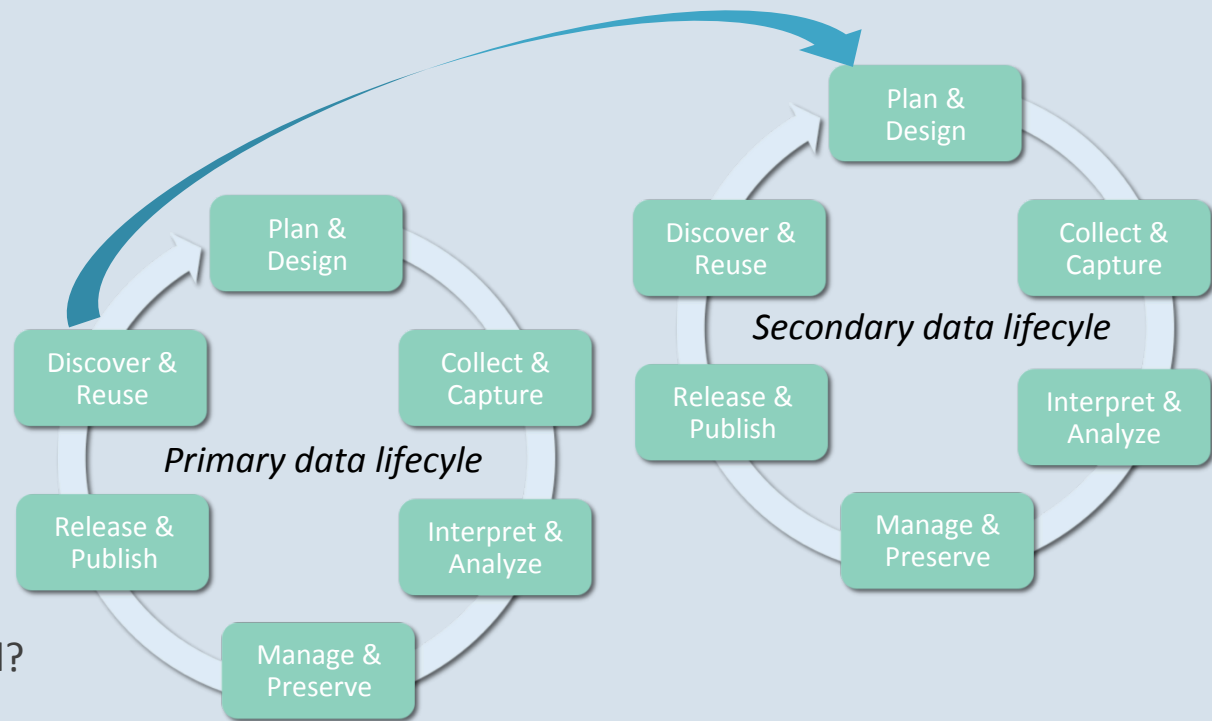
What resources exist to address qualitative data management challenges?

From a researcher's perspective, how do we operationalize sharing (preserving and publishing) oriented toward re-use (discover and design)?



What resources exist to address qualitative data management challenges?

- What kind of metadata is necessary?
- Has the data been cleaned and made anonymous?
- Will the data be discoverable?
 - What counts as data?
 - How does epistemology shape what can be shared and re-used?



Linking qualitative data sharing and re-use: Levels of access

Level of access	Definition
A – Open	Data is freely available for use in accordance with general use agreement of repository and standard citation practices
B - Restricted	Data is available for use when user meets standard criteria set by data repository to ensure ethical use of data (could include use agreement, obtaining IRB or accessing data through virtual environment)
C - Controlled	Data is available for use by trained users in a controlled environment (access could depend on secondary research questions and intended analysis, controls on access method and amount of data shared is decided by original researcher)
D - Closed	Data deposit and citation exist for archival purposes but no data are currently available (could be embargoed until publication of results, change in sensitive situation, death of a participant, or certain duration of time from collection)

Linking qualitative data sharing and re-use: Levels of processing

Level of processing	Definition	
0 – Raw data	Full text, image or audio No aggregation or analysis	No redaction - all identifiers included No additional information about context and methodology
1	Full text, image or audio No aggregation or analysis	Redaction for direct identifiers Idiosyncratic information about context and methodology
2	Full text, image or audio No aggregation or analysis	Redaction for direct and indirect identifiers Standardized information about context and methodology
3	Excerpted text, image or audio Thematic or topical aggregation	Redaction for direct and indirect identifiers Standardized information about context and methodology
4 – Research output	Summarized text, image or audio Thematic or topical analysis	Redaction for direct and indirect identifiers Summarized information about context and methodology

Level of processing	Level of access			
	A [open]	B [restricted]	C [controlled]	D [closed]
0 [raw]	Public policy documents			Raw interview transcripts or field notes
1	Public policy documents with search terms as metadata		Interview transcripts with names and locations redacted	
2	Public policy documents with code for web scraping	<i>Interview transcripts with names and locations redacted and metadata about setting of interviews</i>	<i>Interview transcripts with names and locations redacted and metadata about setting of interviews</i>	
3	Public policy documents organized by theme and with code for thematic analysis	Interview excerpts with names and locations redacted and metadata including thematic codes		
4 [Research outputs]	Descriptive summary of themes within policies with methodology explained <i>Summary of thematic analysis of interview transcripts with methodology explained</i>	<i>Summary of thematic analysis of interview transcripts with methodology explained</i>		

Takeaways

Researchers should consider data sharing and re-use across all stages of the lifecycle

- This often requires more planning at the outset (IRB, metadata documentation, etc.) for qualitative data than quantitative data

Interdisciplinary research requires working across systems, vocabularies, tools, etc.

- There are many resources out there, but they often aren't used in an integrated workflow

Support for this work comes from the National Socio-Environmental Synthesis Center (SESYNC), which is supported under funding received from the National Science Foundation DBI-1052875.

SESYNC Qualitative Data Initiative:

<https://www.sesync.org/for-you/cyberinfrastructure/research-and-tools/qualitative-data-initiative>

Contact information:

Kristal Jones kjones@sesync.org

Steven M. Alexander s22alexa@uwaterloo.ca