

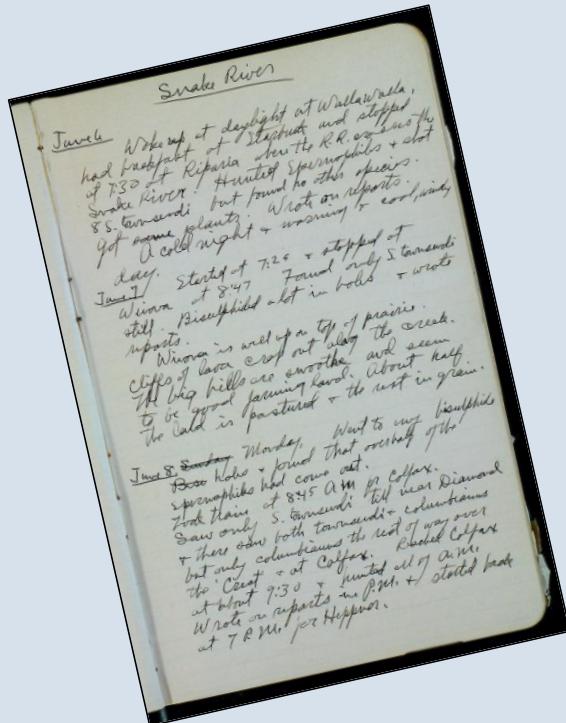
Qualitative Data Management for Interdisciplinary Research

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](#)

February 8, 2018

CONGRESSIONAL RECORD—DAILY DIGEST

D138

Subcommittee on State Department and USAID Management, International Operations, and Bilateral International Development: Senators Isakson (Chair), Risch, Rubio, Portman, Paul, Shaheen, Coons, Booker, and Udall. Senators Corker and Menendez are ex officio members of each subcommittee.

THE OPIOID CRISIS

Committee on Health, Education, Labor, and Pensions: Committee concluded hearing to examine the opioid crisis, focusing on the impact on children and families, after receiving testimony from Becky Savage, 525 Foundation, Greencastle, Indiana; Stephen W. Patrick, Vanderbilt University School of Medicine, Nashville, Tennessee; and William C. Bell, Casey Family Programs, Seattle, Washington.

BUSINESS MEETING

Committee on the Judiciary: Committee ordered favorably reported the nominations of Kurt D. Engelhardt, of Louisiana, to be United States Circuit Judge for the Fifth Circuit, Barry W. Ashe, to be United States District Judge for the Eastern District of Louisiana, Howard C. Nielson, Jr., to be United States District Judge for the District of Utah, James R. Sweeney II, to be United States District Judge for the Southern District of Indiana, and John C. Anderson, to be United States Attorney for the District of New Mexico, Brandon J. Fremin, to be United States Attorney for the Middle District of Louisiana, Joseph P. Kelly, to be United States Attorney for the District of Nebraska, Scott W. Murray, to be United States Attorney for the District of New Hampshire, David C. Weiss, to be United States Attorney for the District of Delaware, David G. Jolley, to be United States Marshal for the Eastern District of Tennessee, and Thomas M. Griffin, Jr., to be United States Marshal for the District of South Carolina, all of the Department of Justice.

INTELLIGENCE

Select Committee on Intelligence: Committee met in closed session to receive a briefing on certain intelligence matters from officials of the intelligence community.

[US Congressional Record](#)

What is qualitative data?



Author's photos

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

By MITCH SMITH FEB. 23, 2017

Crews cleaned up the Oceti Sakowin camp in North Dakota on Tuesday, the day before the deadline for protesters to leave. Rich Cai/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 campsites were taken into custody.

RELATED COVERAGE

- North Cam
- North Ruins
- South of E
- Oceti Tribe
- Arm Acre

The New York Times

NRDC @NRDC · Mar 2
Government emails show the real reason behind Trump's decision to eliminate #BearEars National Monument: oil drilling and coal mining. nytimes.com/2018/03/02/cl...

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." on.msnbc.com/2GXW9NL

106 662 625

John Schwartz @jswatz · Mar 2
The team of @EricLiptonNYT and @LFFriedman with another big scoop: internal documents show that shrinking **Bear Ears** monument was hugely about drilling and mining. nytimes.com/2018/03/02/cl...

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 #BearEars. They wanted to plunder the resources lying underneath our public lands. nytimes.com/2018/03/02/cl...

64 1.9K 2.6K

Twitter (#BearEars)

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?



Dartmouth | SESMAD

Social-Ecological Systems Meta-Analysis Database: Cases

This page contains all of the social-ecological cases that have been entered into the SESMAD database. Each case consists of the following three types of components: (1) environmental commons, which are managed by (2) actor groups, via (3) governance systems. The relationships between a set of components are described by a set of interactions that are created once the components have been coded. Each case in the database can be considered to be a very basic case study unto itself, as well as being directly comparable to other cases in the database. Each case is viewable in its own summary page via the table below. This page lists all of the components and interactions that have been attached to the case, as well as a visual rendering of these elements.



ICPSR Start Sharing Data

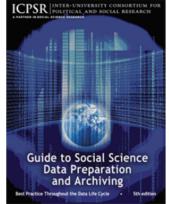
Log In/Create Account

Data Preparation Guide CONFIDENTIALITY SUGGEST DATA TO ARCHIVE

Guide to Social Science Data Preparation and Archiving: Introduction

- Importance of Data Sharing and Archiving
- Planning Ahead for Archiving and Preservation of Data
- The Data Life Cycle
- Using the Guide

Importance of Data Sharing and Archiving



HARVARD Dataverse

Metrics 3,155,384 D

qualitative

[Dataverses \(8\)](#)

[Datasets \(513\)](#)

[Files \(182\)](#)

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

Practical challenges

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

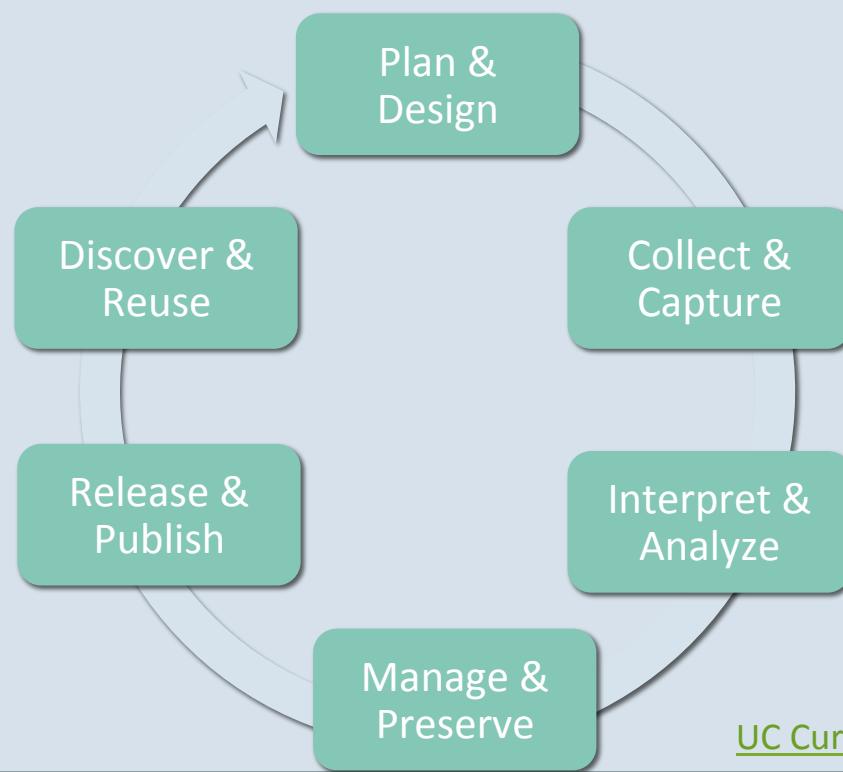
Ethical

- 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?

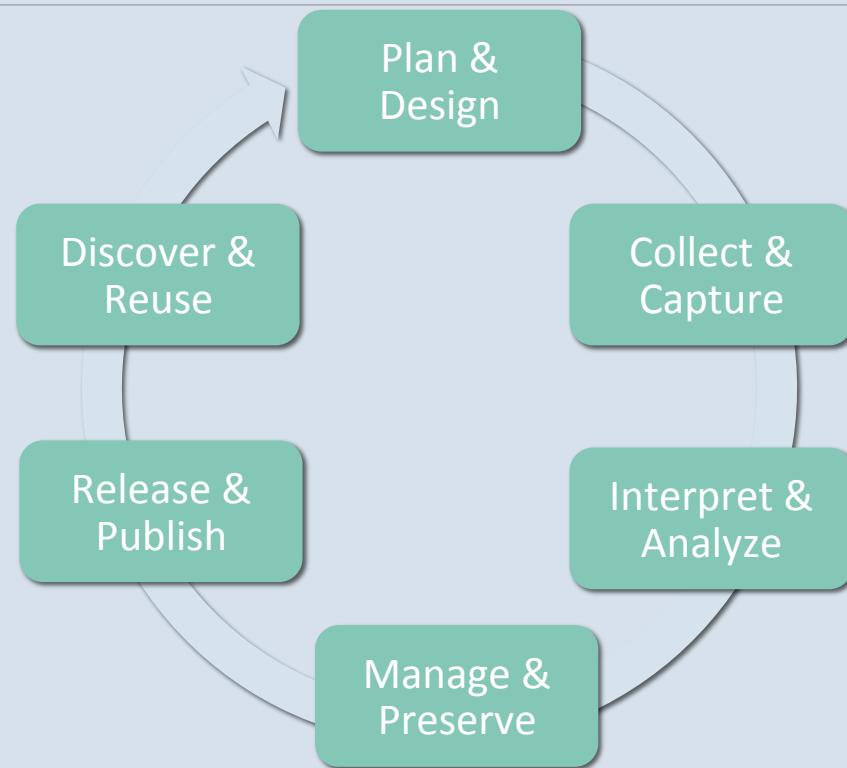
Researchers

Research institutions

Journals and publishers

Funders

Data repositories



Want a summary of the benefits, challenges, resources and recommendations?



SESYNC WHITE PAPER

Lead authors: Kristal Jones and Steven M. Alexander

SESYNC

UNIVERSITY OF MARYLAND

The cover of the SESYNC White Paper is shown at an angle. The title is "Qualitative data sharing and re-use for socio-environmental systems research: A synthesis of opportunities, challenges, resources and approaches". Below the title, it says "SESYNC WHITE PAPER" and "Lead authors: Kristal Jones and Steven M. Alexander". The SESYNC logo is at the bottom right, and the University of Maryland logo is at the bottom left.

[SESYNC Qualitative Data Initiative](#)

SUPPORTING QUALITATIVE DATA SHARING AND RE-USE:
RECOMMENDATIONS FOR RESEARCH INSTITUTIONS

BACKGROUND

Qualitative data is increasingly being used in socio-environmental systems research and related interdisciplinary efforts to address complex sustainability challenges. Qualitative data includes transcripts of recorded interviews or focus group sessions, field notes, audio and video recordings like oral history interviews, photography, and maps as well as policy documents, news reports, and historical archives. There are many benefits to be gained from sharing open data, some of which reflect the broader move toward open science. However, although open data is increasingly becoming an expectation in many fields and methodological approaches that work on interdisciplinary topics, there remain many challenges associated with the sharing and re-use of qualitative data in particular.

BENEFITS AND CHALLENGES TO QUALITATIVE DATA SHARING AND RE-USE

Broadly speaking, the benefits of qualitative data sharing and re-use fall into three categories. Scientific benefits can be used to scale up or down research findings. Descriptive benefits refer to the contribution that qualitative data sharing can make to bettering our understanding of research contexts and subjects past and present. Finally, material benefits of qualitative data sharing include maximizing scarce research resources (both time and funding), and minimizing the burden on researchers, subjects and communities.

PRACTICAL CHALLENGES TO QUALITATIVE DATA SHARING AND RE-USE

Practical challenges to sharing and re-using qualitative data include maximizing data gathering, for synthesis and depositing and accessing data, and the creation of standardized metadata that can provide adequate information for data reuse. In addition, epistemological approaches that guide qualitative data gathering can greatly influence the likelihood that a researcher will feel appropriate to share their information about context of the research process, and the methodological and ethical considerations that influenced data gathering, can address many of the concerns potentially raised by the prospect of qualitative data sharing.

LEVELS OF PROCESSING AND LEVELS OF ACCESS FOR QUALITATIVE DATA

To address some of the challenges and maximize the benefits associated with qualitative data sharing and re-use, all actors across the research data life-cycle should be aware that there are many levels of access and levels of processing of which qualitative data can be shared for future reuse. Different combinations of processing and access will appropriate for different types of data and research contexts.

Level of processing and definition

- Full text, image, etc.
- No reduction
- Text, image, etc.
- Identifiers redacted
- Image, etc.
- Text, image, etc.
- Analysis
- Standardized metadata
- Summarized metadata

Level of access and definition

- A Data freely available for use in accordance with general use agreement or repository and standard citation practices
- B Data originally available for use when user meets criteria (e.g. by data repository to ensure ethical data use)
- C Data available for use when user is approved by intended researcher (access could depend on intended analysis)
- D Data dependent on further discussion of opportunities, obligations, and responsibilities for use (e.g. embargoed until change in research situation)

For further discussion of opportunities, obligations, and responsibilities for use of qualitative data sharing and re-use see the [SESYNC Guide](#).

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:

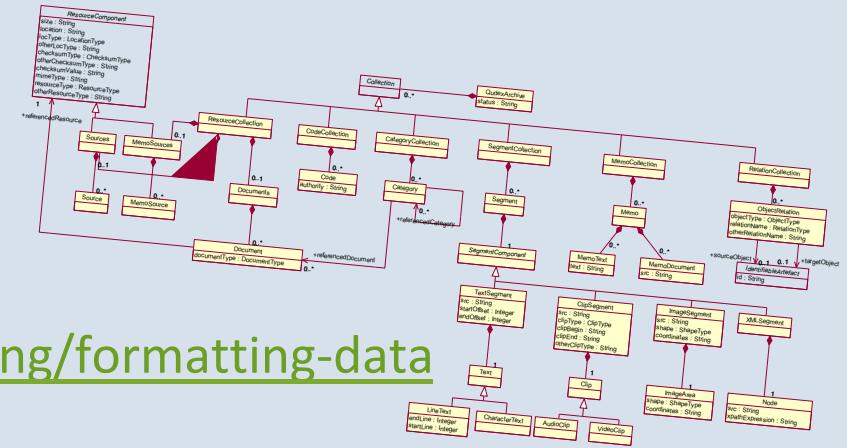
Collect and Capture, Interpret and Analyze

- 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: Manage and Preserve

- IRB and de-identification:
<https://qdr.syr.edu/node/20260/>
- File formats:
<https://qdr.syr.edu/guidance/managing/formatting-data>
- DDI metadata standards:
<https://www.ddialliance.org/sites/default/files/AQualitativeDataModelForDDI.pdf>



[UK Data Archive](#)

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?

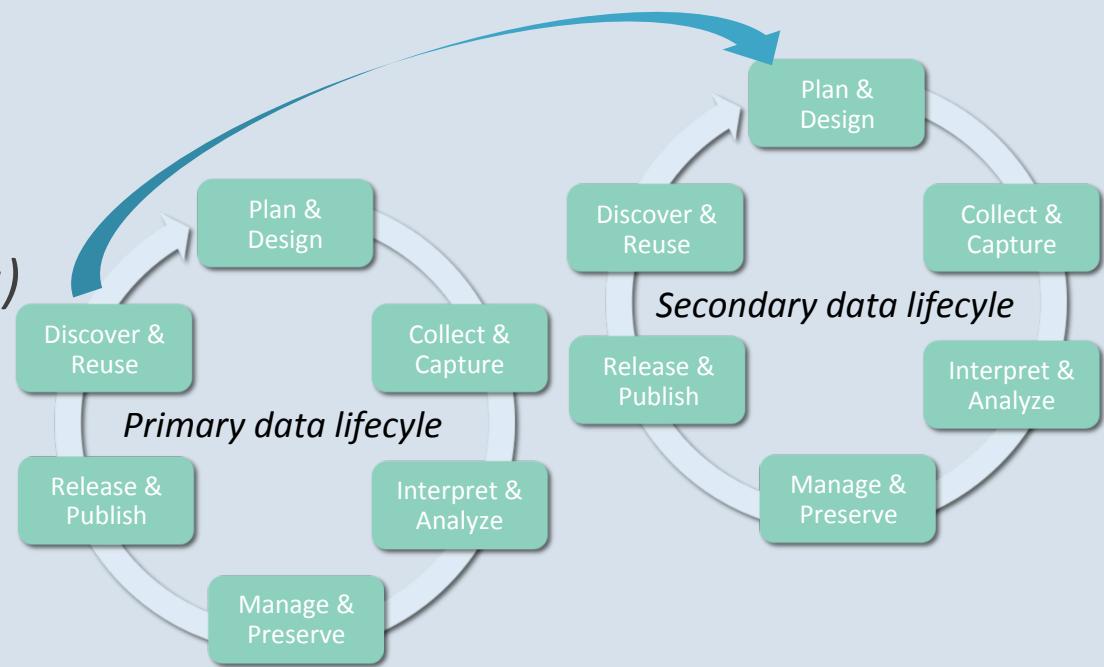
Secondary data lifecycle:

Manage and Preserve, Release and Publish

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

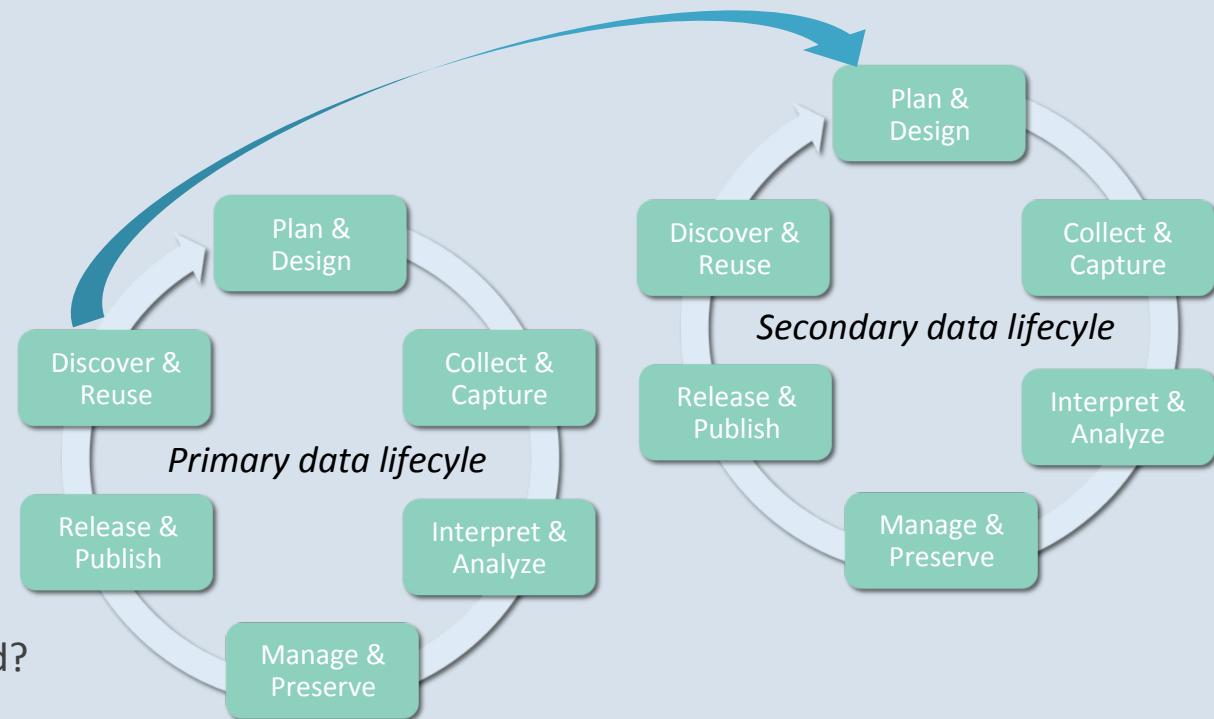
What resources exist to address qualitative data management challenges?

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