Qualitative data sharing and re-use for socio-environmental systems research:
A synthesis of opportunities, challenges, resources and approaches

SESYNC WHITE PAPER

Lead authors: Kristal Jones and Steven M. Alexander

Lead authors:
Kristal Jones (SESYNC) and
Steven M. Alexander (Science Advisor, Canadian Department of Fisheries and Oceans)

Contributing authors (in alphabetical order):
Nathan Bennett (University of British Columbia and Stanford University), **Libby Bishop** (UK Data Service and UK Data Archive - University of Essex), **Amber Budden** (DataONE), **Michael Cox** (Dartmouth University), **Mercè Crosas** (Harvard University), **Eddie Game** (The Nature Conservancy), **Janis Geary** (University of Alberta), **Charlie Hahn** (University of Washington), **Dean Hardy** (SESYNC), **Jay Johnson** (University of Kansas), **Sebastian Karcher** (Qualitative Data Repository), **Matt LaFevor** (University of Alabama), **Nicole Motzer** (SESYNC), **Patricia Pinto da Silva** (NOAA), **Jeremy Pittman** (University of Waterloo), **Heather Randell** (SESYNC), **Julie Silva** (University of Maryland), **Joseph Smith** (University of Maryland), **Mike Smorul** (Nava Public Benefit Corporation, formerly at SESYNC), **Carly Strasser** (Collaborative Knowledge Foundation), **Colleen Strawhacker** (National Snow & Ice Data Center), **Andrew Stuhl** (Bucknell University), **Nicholas Weber** (University of Washington), **Deborah Winslow** (National Science Foundation)

The opinions, ideas and positions presented in this paper are the authors' alone, and do not reflect the opinions, ideas or positions of the National Socio-Environmental Synthesis Center, the University of Maryland or the National Science Foundation.

EXECUTIVE SUMMARY

Researchers in many disciplines, both social and natural sciences, have a long history of collecting and analyzing qualitative data to answer questions that have many dimensions, to interpret other research findings, and to characterize processes that are not easily quantified. Qualitative data is increasingly being used in socio-environmental systems research and related interdisciplinary efforts to address complex sustainability challenges. There are many scientific, descriptive and material benefits to be gained from sharing and re-using qualitative data, some of which reflect the broader push toward open science, and some of which are unique to qualitative research traditions. However, although open data availability is increasingly becoming an expectation in many fields and methodological approaches that work on socio-environmental topics, there remain many challenges associated the sharing and re-use of qualitative data in particular.

This white paper discusses opportunities, challenges, resources and approaches for qualitative data sharing and re-use for socio-environmental research. The content and findings of the paper are a synthesis and extension of discussions that began during a workshop funded by the National Socio-Environmental Synthesis Center (SESYNC) and held at the Center Feb. 28-March 2, 2017. The structure of the paper reflects the starting point for the workshop, which focused on opportunities, challenges and resources for qualitative data sharing, and presents as well the workshop outputs focused on developing a novel approach to qualitative data sharing considerations and creating recommendations for how a variety of actors can further support and facilitate qualitative data sharing and re-use.

The white paper is organized into five sections to address the following objectives:

(1) Define qualitative data and discuss the benefits of sharing it along with its role in socio-environmental synthesis;

(2) Review the practical, epistemological, and ethical challenges regarding sharing such data;

(3) Identify the landscape of resources available for sharing qualitative data including repositories and communities of practice

(4) Develop a novel framework for identifying levels of processing and access to qualitative data; and

(5) Suggest roles and responsibilities for key actors in the research ecosystem that can improve the longevity and use of qualitative data in the future.

TABLE OF CONTENTS

List of Tables and Boxes

INTRODUCTION

Researchers in many disciplines, both social and natural sciences, have a long history of collecting and analyzing qualitative data to answer questions that have many dimensions, to interpret other research findings, and to characterize processes that are not easily quantified. Qualitative data is increasingly being used in socio-environmental systems research and related interdisciplinary efforts to address complex sustainability challenges. There are many scientific, descriptive and material benefits to be gained from sharing and re-using qualitative data, some of which reflect the broader push toward open science, and some of which are unique to qualitative research traditions. However, although open data availability is increasingly becoming an expectation in many fields and methodological approaches that work on socio-environmental topics, there remain many challenges associated the sharing and re-use of qualitative data in particular.

This white paper discusses opportunities, challenges, resources and approaches for qualitative data sharing and re-use for socio-environmental research. The paper is organized into five sections to address the following objectives:

(1) Define qualitative data and discuss the benefits of sharing it along with its role in socio-environmental synthesis;

(2) Review the practical, epistemological, and ethical challenges regarding sharing such data;

(3) Identify the landscape of resources available for sharing qualitative data including repositories and communities of practice

(4) Develop a novel framework for identifying levels of processing and access to qualitative data; and

(5) Suggest roles and responsibilities for key actors in the research ecosystem that can improve the longevity and use of qualitative data in the future.

BACKGROUND

WHAT IS QUALITATIVE DATA?

Qualitative data includes a broad range of types and forms of information, and is sometimes defined as any information or data that is unstructured. Structured data is generally defined as data that is organized based on a pre-existing schema or framework, and that is formatted in such a way as to be machine-readable and analyzable (generally speaking, this is tabular data with discrete variables, usually presented in spreadsheets or databases). Unstructured data includes all types of data that is not discrete (in that there are many possible measurements, characteristics, or dimensions present) and/or is not organized based on a predefined framework. Most raw qualitative data, including text, images, and audio and video, is considered unstructured.

In practice, qualitative data includes written responses from open-ended interview questions, transcripts of recorded interviews or focus group sessions, field notes, and written observations. Qualitative data is not just text, however, and includes audio recordings like oral history interviews, video recordings, photographs, maps and artwork, as well as policy documents, news reports, and historical archives (Goodwin and Horowitz 2002)[1]. Researchers using qualitative data might draw on primary data (gathered by the researcher doing the analysis) or secondary data (already existing as data or information that could be systematically analyzed). Qualitative data is often collected and used by social scientists and humanities researchers, but is not limited to only these disciplines. Field studies in ecology, biology and botany also often gather qualitative data in the form of written observations, sketches, and images.

WHY SHARE QUALITATIVE DATA?

There are many potential benefits to be gained from sharing qualitative data, some of which reflect the broader push toward open science, and some of which are unique to qualitative research traditions. Broadly speaking, these benefits fall into three categories: scientific, descriptive, and material. Scientific benefits include increasing transparency, supporting reliability and reproducibility, and providing an evidence base that can be used to scale up or down research findings. Descriptive benefits refer to the contribution that qualitative data sharing can make to characterizing and bearing witness to research contexts and subjects past and present, and to teaching students about research approaches and methods. Finally, material benefits of qualitative data sharing include maximizing scarce research resources (both time and funding), and minimizing the burden on research subjects and communities.

Scientific benefits

Archiving quantitative data is a well-established practice in the social sciences, but both epistemological and practical issues have limited the expansion of qualitative data sharing (Karcher et al. 2016; see also Bishop 2014). Transparency of data and methods is for many scientists from all disciplines an essential aspect of their enterprise: "What distinguishes scientific claims from others is the extent to which scholars attach to their claims publicly available information about the steps that they took to convert information from the past into conclusions about the past, present, or future" (Lupia and Elman 2014: 20). Transparency allows others to evaluate the validity and reliability of research outputs and potentially to reproduce the findings. The appropriateness of the concepts of validity and reliability are debated by qualitative social science researchers, with some epistemological approaches rejecting the possibility of

---

[1] See also this list from the Qualitative Data Repository.

external validity while others emphasize it as foundational to social scientific claims (for further discussion, see [Boorman et al. 1986](#); [Goodwin and Horowitz 2002](#); [Cho and Trent 2006](#)). The perceived scientific benefits of sharing qualitative data therefore depend in part on the orientation of the researchers sharing and potentially re-using the data. For some, data sharing allows for objective confirmation of the validity and reliability of findings, by making accessible the information and the methods used to come to conclusions. For other researchers who reject the notion of complete objectivity, data sharing can still provide systematic documentation of the research process and findings that can be used for triangulation ([Corti and Fielding 2016](#)).

An additional scientific benefit of qualitative data sharing relates to the ability to scale analyses up or down by integrating multiple data sources. One of the limitations of qualitative data, both social and ecological, is that the degree of complexity and heterogeneity requires either significant processing and winnowing for final analysis, or generating rich research outputs (like ethnographies) that are not easily comparable ([Goodwin and Horowitz 2002](#); [Osmond et al. 2004](#)). This can make it difficult to scale up findings from already completed analyses that derive from qualitative data. Sharing raw qualitative data can support further analysis that applies a common framework across many data sources in order to generate a larger sample size ([Poteete and Ostrom 2005](#)). Scaling up research findings can also be done by comparing across cases, and by sharing the range of data that constitute a case study or case study-like methodology, researchers can test both generalizability and theories thought to be invariant or universal ([Cox 2014](#); [Poteete and Ostrom 2005](#)). In contrast, sharing qualitative data can also support the scaling down of relationships and mechanisms found to exist at broad or general scales, by providing information about the contexts within which specific relationships play out. For example, agent-based modelling efforts that reflect general understandings of human decision-making can be refined and parameterized for specific settings by drawing on other, often qualitative sources of empirical information ([Lindkvist et al. 2017](#); [Janssen and Ostrom 2006](#)).

Descriptive benefits

A second set of benefits of qualitative data sharing are what we call descriptive. Qualitative data is often more expansive, inclusive and varied than quantitative data, and as a result has the potential to convey more and different information than a set of measurements organized in a spreadsheet. Anthropologists and others who utilize ethnographic research methods have long recognized the unique ability of qualitative data to document the richness of people and places, and phenomena within them. As one workshop participant noted, there is a tradition of archiving qualitative data for posterity's sake, in repositories like the [Human Relations Area Files](#). These repositories have historically focused on preserving artifacts of all types, but not necessarily in a format that makes the information easily accessible or usable for future analysis. Increasingly, however, scholars working in a variety of disciplines are highlighting the opportunities for action associated with sharing data of all types. The International Arctic Science Committee, for example, emphasizes in its ethically open access statement about data sharing the need to balance the ethics of knowledge creation with the immediate regional challenges that can be best addressed if timely and comprehensive data is freely available ([IASC 2013](#)). And as [Barbour and Barbour (2003)](#) point out, the curiosity that drives many qualitative researchers and their methods creates the possibility that qualitative data will encompass ideas, measurements, and themes that were not necessarily the focus of the original or primary study. Sharing qualitative data can therefore facilitate communication and interpretation of many additional and otherwise overlooked dimensions of a place or a problem.

Another descriptive benefit to sharing qualitative data is the opportunities it offers for teaching about research design and methods, as well as about specific topical areas. In quantitative methodologies, using shared, public or open access data has become a common practice in the classroom, and available data

sets cover a wide range of topics (e.g., King 2006; Janz 2015). Instructors in qualitative methods classes are less likely to use qualitative data (in part given its scarce availability) than those teaching quantitative methodologies. However, methods instructors who are aware of qualitative data repositories report that students learn better when using real data that reflects their own research interests (Bishop 2014). Such usage is becoming more widely recognized, recommended (Corti & Bishop 2005; Bishop 2012), and employed (Karcher 2016), and increased sharing of qualitative data can continue to broaden the topical areas in which the use of real qualitative data is feasible.

Material benefits

A third set of benefits associated with qualitative data sharing are material, and relate to both making the most of research investments as well as reducing the burden on research participants, communities and institutions. Research funding is perpetually in short supply, even for the most well-funded fields, and funding for qualitative research has historically been more difficult to secure. Data sharing and re-use of data in secondary analyses and synthesis research could therefore expand the potential findings from qualitative research projects, and could also support more types of qualitative analysis that are relatively low-cost and feasible for a broad range of researchers. Epistemological and practical aspects of qualitative research, including the potential impacts of researcher-participant interactions, intercoder reliability, and many other relational characteristics are often cited as reasons that qualitative data are not shared and re-used (for a summary of these debates, see Bishop 2014). As noted above, there are scientific benefits to addressing and moving forward from these concerns (in terms of scaling up and down analyses), and there are material benefits as well to maximizing the use of data gathered through resource-intensive methodologies. The ability to further learn from and interpret secondary qualitative data is especially important for early-career researchers and those not situated in academic institutions, for whom securing governmental funding is more challenging, as well as for practitioners with methodological training who sit outside of traditional research institutions but have the interest and ability to use qualitative analyses to inform their work.

An additional material benefit of sharing and reusing qualitative data is the potential to reduce the burden (i.e., research fatigue (Clark 2008)) on individuals and communities with whom data is being generated and gathered. Taking time to talk and engage with researchers places demands on research subjects that can have diminishing returns for subjects over time, a critique of qualitative and field-based research that is not new and yet has not been uniformly addressed across research communities (for a few examples of this discussion see Adams 1979; Hartter et al. 2013). For example, following a surge in corporate and governmental interest in oil and gas resources in the Mackenzie Delta of the Canadian Arctic during the 1960s and 1970s, the four primarily indigenous communities became "one of the most studied" regions in the country. Community members reported that social science data collected across the period 1950 to 1985 was duplicative, irrelevant or inapplicable to local interests, and often-incomprehensible (Brizinski 1993). Today, with increasing attention to climate change, sovereignty, and resource development in the same area, community members express continuing interest in social science research and have developed guidelines to help researchers understand local concerns, cultures, and capacities (Inuvialuit Regional Corporation n.d.). Having access to qualitative data already collected in specific places and about specific topics can provide researchers interested in new questions with background information so that they do not request redundant information. Reviewing and incorporating not only final qualitative analyses but also the rich detail that is present in primary qualitative data into new qualitative research projects can also make them more sensitive to the particular histories of places, people and problems. Utilizing existing information to reduce the burden placed on research subjects and improve the appropriateness of new research questions and projects can in turn accelerate understanding of complex problems and possible ways to address them. Efforts like those being made in the Arctic research

community to build open data systems that can increase ethical knowledge sharing, including the recent establishment of ELOKA (Exchange for Local Observations and Knowledge of the Arctic), are working to maximize the impact of already gathered data and minimize the burdens placed by repetitive research.

WHAT ROLE DOES QUALITATIVE DATA PLAY IN SOCIO-ENVIRONMENTAL SYNTHESIS?

Socio-environmental (SE) systems research as both a topical area and an approach has rapidly expanded over the past two decades, especially as broad efforts toward sustainability have increased in visibility and interest. SE systems are defined by interactions between humans and the environment. Or, per the foundational orientation of SESYNC, SE systems research is predicated on the assumption that all environmental problems are, by definition, social problems. SE systems research is therefore focused on any question, problem or phenomenon that can be understood to relate to both the non-human world and the social systems that interact with it. SE systems research is a big tent that includes many conceptual frameworks, including coupled human and natural systems (CHANS), social-ecological systems (SES), human dimensions of the environment, and a range of efforts within sustainability science. The systems approach inherent in SE systems research reinforces the assumption that humans and the environment are connected, and that researchers must explore and characterize the relationships and dynamics within and across systems in order to understand discrete outcomes or problems (for a summary of systems thinking, see Checkland 1999).

The use of qualitative data in SE systems research has steadily increased, in part due to the scientific and descriptive benefits outlined above. Qualitative data is being increasingly drawn upon in what are traditionally quantitative SE fields, including conservation biology (Parmesan and Yohe 2003; Stem et al. 2005; Pullin et al. 2013; see also the State of Alaska Salmon and People (SASAP) program) and environmental and institutional economics (Poteete and Ostrom 2005; Hicks et al. 2016; Lindkvist et al. 2017). The use of qualitative data in SE systems research has moved beyond simply contextualizing or telling the story behind quantitative analyses, and increasingly includes monitoring, assessment and impact studies. In marine conservation planning, for example, qualitative data and analysis has been integral to assessing the degree to which ocean management plans have achieved the sustainability (ecological, social, and economic) goals for which they are designed and implement (Gill et al. 2017). At a more theoretical level, many SE system frameworks, including Ostrom's (2009) multi-level approach that emphasizes the role that interactions across ecological and social scales have in determining systems dynamics, are in fact premised on the need for qualitative understanding of not only what relationships exist but how and why they evolve (see also Janssen et al. 2006).

Box 1: Voices from the Fisheries Project: Oral histories and qualitative data sharing

The National Oceanic and Atmospheric Administration's National Marine Fisheries Service is mandated, by the Magnuson's-Stevens Fisheries Conservation and Management Act (MSA) (and its amendments) to manage the commercial and recreational fisheries conducted in Federal waters. The Sustainable Fisheries Act (a 1996 amendment to the MSA) introduced 10 national standards for fisheries management. National Standard 8 states that conservation and management measures need to take into account the importance of fisheries resources to fishing communities and to provide the sustained participation of such communities while minimizing impacts to them. Since this period, NMFS has employed a growing number of anthropologists and other social scientists to collect baseline data and conduct research on fishermen and fishing communities and the relationship between these with the resources that their livelihoods depend on. They have

also conducted research on the social impacts of federal fishing regulations on these individuals, businesses and communities.

One method used to gather this type of data is oral history. Oral history involves the audio recording of first hand experiences of individuals in order to learn more about a specific past event or perspective. While oral histories were being collected in support of NOAA's mission and on-going research, there was no plan in place to ensure the long term archiving and protection of these recordings. The [Voices from the Fisheries Project](#) began in the mid 1990's to fill this void and also to try to recover audio recordings of oral histories related to US fisheries from around the country. Initially, hundreds of analog tapes were 'discovered' by calling relevant researchers and enquiring if they had any collections that were at risk. With funding from the NMFS Office of Science and Technology along with a NOAA Preserve America Grant, analog tapes were converted to .wav files. Simultaneously, a database was created at NMFS that would house these recordings in order to preserve them, but also to make them available to the public. The project aims to encourage researchers to 'create|archive|share' and the public and other researchers to 'search|listen|learn'. As such, the project provides technical support to improve the quality of oral histories produced that document this human/fisheries connection and also provides the basic sharing platform so that these oral histories can continue to inform researchers as well as anyone else interested in learning from these connections and personal experiences. To date, there are over 1,000 interviews uploaded on the site (www.voices.nmfs.noaa.gov) representing a growth rate of approximately 100 new interviews a year drawn from social scientists from NOAA and universities, historical societies, non-governmental organizations, students, and others.

In addition to an increasing emphasis on SE systems research in general, there has been an additional push over the past decade toward data-driven synthesis research to address pressing and complex sustainability challenges ([Palmer et al. 2005](#); [Young et al. 2006](#); [Hampton et al. 2013](#)). Synthesis research is broadly defined as interdisciplinary or transdisciplinary research that draws on existing data and information to ask questions and identify dynamics that span spatial, temporal, and disciplinary scales. Synthesis research engages heavily with the open science and transparency efforts that have expanded access to and re-use of a wide range of data and information. Indeed, many of the scientific, descriptive and material benefits of qualitative data re-use described above are in fact benefits of engagement with synthesis research more generally. For synthesis research to be possible, however, data of all types must be shared in formats and locations that are appropriate and accessible for synthesis researchers. Appendix B provides brief descriptions and references for synthesis projects across many topical areas that have drawn on re-used or secondary qualitative data. The use of qualitative data in SE synthesis research is increasing, but it remains difficult for many synthesis projects to identify usable forms of qualitative data to integrate into SE synthesis analyses.

Although qualitative data is at the current moment not commonly utilized in SE synthesis research, several projects at SESYNC provide examples to the contrary, and demonstrate the potential contribution of qualitative information and analyses to synthesis research. One project has utilized primary qualitative data, including interview transcripts, to characterize households and their land use and land management decisions in coastal Nicaragua. These data were coupled with primary quantitative surveys of plant diversity and secondary quantitative data on plant range and traits to not only identify relationships

between human decision-making and biodiversity, but also to interpret and explain why these relationships exist (for partial results, see Sistla et al. 2016). In a second example, a recent postdoctoral project at SESYNC looked at the development of green infrastructure in cities, and the social and ecological drivers and outcomes of different approaches to storm water management. The project drew on published storm water management plans, which exist in distributed form throughout city and county websites and which had to be identified, acquired, and coded based on characteristics hypothesized to reflect different approaches to infrastructure development[2]. An iterative process of combining analysis of the qualitative documents with geospatial data on infrastructure and environmental variables allowed the researcher to identify the drivers of infrastructure development and their outcomes. In both of these SE synthesis examples and several others, qualitative data and analysis provides both an additional source of information about the relationships across social and environmental systems, and also allows for systematic interpretation of 'why' and 'how' SE processes and dynamics unfold.

Box 2: Study of historic documents and climate change

A third domain of qualitative data re-use in SE research is the contemporary study of climate change through historic documents, a practice referred to as both climate history and historical climatology (Climate History Network 2017). The purpose, methods, and outcomes of this domain can be understood through the example of the Old Weather Project, a collaboration among climate scientists at the United Kingdom's Hadley Centre and the National Oceanic and Atmospheric Administration, archivists at the U.S. National Archives and National Maritime Museum, and other scholars (Old Weather 2017). The Project enlists the help of thousands of volunteers to crowdsource the transcription of logbooks from whaling and naval voyages to the Arctic region during 19th century, with specific attention to weather data recorded by those on board. Once digitized, climate scientists synthesize observations of pressure, wind speed, precipitation, and cloudiness to reconstruct past climates and refine climate projections (Old Weather 2017). Primary investigators of the Old Weather Project suggest their approach can be replicated to improve the global record of climate observations, whether to fill-in regions of the world with little existing data or to extend that record deeper into the past (Old Weather 2017).

---

[2] Full details about the project are available on the SESYNC website.

CHALLENGES FOR QUALITATIVE DATA SHARING AND RE-USE

Although open data availability is increasingly becoming an expectation in many fields, there remain many challenges associated with data sharing and re-use in general, and with the sharing and re-use of qualitative data in particular. In this section, we highlight a few of these challenges — practical, epistemological, and ethical — and provide background and examples of resources and approaches that have been developed to help address these challenges. (For a recent qualitative study that examines qualitative researchers' concerns about the practical and epistemological challenges of qualitative data sharing, see Broom et al. (2009)).

PRACTICAL CHALLENGES FOR QUALITATIVE DATA SHARING AND RE-USE

Two major challenges for sharing and re-using all types of data are the identification of appropriate infrastructure for depositing and accessing data, and the creation and standardization of metadata that can provide adequate information for the re-use of data in new analyses. Quantitative data communities and fields have much to offer in addressing both of these challenges, and at the same time, there remain challenges that are unique to qualitative and unstructured data that will require specific efforts to overcome.

Data sharing can take many forms, from deposits in well-known and managed repositories to notes at the end of articles suggesting that data is available "on request," "on the author's personal webpage," or as supplemental material to a journal article or book. This wide of range of approaches to sharing reflects the histories of specific disciplines, research organizations and technological developments, and all types of sharing have the potential to bring about the benefits outlined above. However, one goal of making data available for re-use is to accelerate understanding and discovery, and to draw on as much information as possible to address complex questions. When data are shared in ways that require extreme effort by the receiving researcher to access and format the data for re-use, or when raw or disaggregated data are not shared at all, it becomes much harder to include such information in analyses and ultimately, to answer some difficult questions. Especially as digital technologies have taken a dominant role in all aspects of data management and analysis, some of the challenges associated with idiosyncratic data sharing have begun to be addressed through guidelines and processes meant to improve the sharing–to–re-use pipeline.

In an already seminal paper, Wilkinson et al. (2016) outline four core principle to guide the infrastructure on which data should be shared. Such data should be FAIR, for Findable, Accessible, Interoperable, and Re-usable. The FAIR guidelines focus in particular on the ability of both humans and machines to access data and metadata. They contain, therefore, a strong emphasis on standards and metadata for how to display and document data. In effect, many data repositories are already "FAIR" according to most of the 11 specific principles listed in the foundational article. From an author or project leader's perspective, FAIR principles and infrastructure designed with them in mind can greatly facilitate the logistics of sharing data: simply by identifying and depositing in a suitable data repository, they can assure the FAIRness of their data. From the perspective of re-use, FAIRness is particularly salient for synthesis research. By emphasizing standards and interoperability, data can be identified and retrieved systematically, facilitating research across different disciplines as well as research using multiple types of data. In addition, some of

the "repositories of repositories," including Dataverse and DataONE, have built upon the FAIR guidelines by creating networks of allied data repositories and collections that share common organizational systems across which users can query and search for data sources.

Table 1: FAIR principles and definitions

Principle	Definition
Findable	• Data are described with rich metadata • Data or metadata have unique and persistent identified • Metadata includes the identifier of the data it describes • Data or metadata are registered/indexed in searchable resource
Accessible	• Data or metadata are retrievable by identifier using a standardized protocol • Protocol is open, free and universally implementable • Protocol allows for authentication and/or authorization when needed • Metadata remain accessible even if data are no longer available
Interoperable	• Data and metadata use a formal, accessible, shared and applicable language to represent content and knowledge • Data and metadata use vocabularies that follow FAIR principles • Data and metadata include and describe references to other data/metadata sources
Re-usable	• Data and metadata have rich description and plurality of accurate and relevant attributes • Data and metadata have clear and accessible usage license • Data and metadata include detailed provenance information • Data and metadata meet domain-relevant community standards

Adapted from Wilkinson et al. (2016)

One of the challenges identified by FAIR standards reflects not so much the infrastructure and process for depositing and housing data, but rather the content of those data entries and the associated metadata that makes the data re-usable. Metadata standards are well established within and across many quantitative fields of study (for two examples, see the Federal Geographic Data Committee and the International Barcode of Life), and are generally built into digital data repositories in the form of fields that researchers must fill out to characterize their data. For data from the social, economic, behavioral and health sciences, the Data Documentation Initiative (DDI) provides extensive metadata guidance for many forms of quantitative human subjects research. A DDI working group has developed a data model for qualitative data, which includes a wide range of characteristics of the data and objects related to the data that must ideally be documented in order to provide adequate information to make qualitative data re-useable (see Hoyle et al. 2013 for details). The UK Data Archive has extended the DDI model to create the Qualitative Data Exchange Schema (QuDEx), with a primary goal of addressing the need to heavily 'mark up' digital files to create comprehensive metadata for qualitative data. As noted by the DDI working group, a major challenge of generating comprehensive metadata for a qualitative collection, or set of primary data artifacts that relate to one another, is how to account for the particularities of specific portions (or segments) of qualitative data.

It is important to note that FAIR is not a maximalist position for data sharing. For example, the guidelines do not make specific reference to citing data or otherwise crediting their creators (Katz 2017). In addition, the guidelines specifically do not mention freely accessible data (often referred to as "open data"), and Mons et al. (2017) highlight the fact that accessible is qualified with "under well-defined conditions."

How repositories approach safeguarding data is of particular concern for many qualitative researchers and the data they might be interested in sharing, which often contains sensitive information about human subjects. Social science data repositories like ICPSR and QDR have clear protocols for restricting access to certain types of data that might require, for example, proof of human subjects research ethics training. We discuss these levels of access and levels of data processing later in this white paper (see Section 4). In addition, there are important questions about what constitutes data in qualitative research and how to incorporate the practical steps of making available the information upon which claims are made into the epistemological orientations from which some qualitative data is generated (Asher and Jahnke 2013).

EPISTEMOLOGICAL AND ETHICAL CHALLENGES FOR QUALITATIVE DATA SHARING

The wide range of methodologies and approaches for qualitative data gathering and analysis derive from the diversity of epistemological orientations and ethical commitments of researchers that generate and use qualitative data (for discussion of social science concepts related to these orientations, see Moon and Blackman 2014). Epistemology (how we know what we know) focuses on the process through which knowledge is generated, and the "relationship between the knower and the known" (Maxwell 2011: 10). By definition, qualitative data is a type of information that is unstructured, and so can be generated and used within many different research and knowledge-generating processes. In other words, qualitative data as artifact does not demand a particular epistemological orientation. The epistemological approach that guides qualitative gathering, however, can greatly influence the likelihood that a researcher will feel it appropriate to share that data or re-use it for synthesis purposes, and epistemology also impacts the form, content and extent of data shared. A related but distinct dimension of qualitative data methods and outputs relates the ethical (or axiological) commitments made by the researcher throughout the research process. Many scholars from across the spectrum are beginning to make the distinction between epistemological principles that preclude qualitative (or any other) data sharing, and ethical challenges that can potentially be addressed through careful process and management approaches (for further discussion, see for example Haraway 2001; Bishop 2009; Biddle and Schafft 2015).

Epistemological challenges

There are two broad epistemological approaches to qualitative research within modern scientific inquiry: constructivism/subjectivism, and positivism/objectivism. In addition, there is an increasing body of literature that articulates an indigenous or traditional knowledge epistemological frame that falls outside of the scientific paradigm. Positivism is, in brief, the familiar frame of the scientific method, with an emphasis on an underlying and immutable true ('objective') nature of reality, which researchers work to systematically uncover and characterize with increasing precision and completeness. Positivism is often called empiricism, because knowledge of reality is taken to be gained through observations and measurements that are material, tangible, and/or discrete. In contrast, constructivism is an epistemological orientation that starts from the premise that reality, and therefore knowledge, is relational, generated through perception, experience and position within contingent systems that are themselves constructed. From the constructivist frame, which is often assumed to be the only legitimate one from which to conduct qualitative research, knowledge of reality is gained by understanding the ('subjective') particularities of perspective and interpretation. Reality, in other words, is different for each individual, and cannot be reduced to common underlying patterns, mechanisms or laws. For a detailed summary of the history and subdivisions of these epistemological categories, see for example Maxwell (2011).

Abstract exploration of epistemology might seem out of place in concrete discussions about data sharing and re-use, but it is important to understand how the most fundamental orientation of a researcher and research project impacts what counts as data, and how that data can and should be analyzed and

interpreted (for further discussion, see [Hammersley 1997](#)). It is also important to note, as many qualitative researchers have, that there are many nuances within each epistemological tradition that influence if and how qualitative data might be shared and re-used to generate new knowledge (for just two examples of this discussion, see Denzin and Lincoln (2008), and [Bryman (1984)](#)). There are some epistemological orientations that will always and completely reject the notion that qualitative data could be re-used by anyone other than the original researcher (or anyone at all, including the same researcher in the future). And there are other epistemological frames that guide many qualitative researchers that can and do include the possibility of data sharing and re-use to generate further or new knowledge.

A purely positivist approach to research should have no epistemological problem sharing data, as the data are seen to be discrete and defined representations of the objective world, and so there is no concern that subsequent re-use of those data could result in a different interpretation and therefore understanding of a phenomenon. While few qualitative researchers subscribe to such an ideal understanding of data and knowledge, there are many qualitative researchers who do take a broad objectivist view, and who see qualitative data as representing complex and partial but still empirical realities (see for example Becker 1996). From this epistemological stance, it is possible, at least theoretically, to provide enough context and metadata alongside shared qualitative data to allow for appropriate and accurate re-use in subsequent analyses. In contrast, constructivist epistemology is likely in pure form to reject the notion that qualitative data could be re-used in appropriate fashion, since the data themselves are generated through the relational process of an individual researcher engaging with research subjects (this is the 'researcher as instrument' construct). From this point of view, the information needed to contextualize and accurately re-interpret the data is as large as the research process itself, and could not be transferred or documented in a structured way to allow someone else to use the data. Again, although some qualitative researchers work from a constructivist frame that precludes any re-interpretation or re-use of data, many others hold epistemological orientations that view knowledge as constructed and yet reflective of either empirical or relational patterns that can be documented, categorized and therefore appropriately re-interpreted with adequate background understanding (for empirical discussion, see [Broom et al. 2009](#)).

What do these abstract observations mean for the actual mechanics of qualitative data sharing? Throughout the workshop discussions that were the precursor to this white paper, there was a general consensus that the inclusion of adequate and appropriate metadata alongside qualitative data can address many of the epistemological concerns potentially raised by the prospect of data sharing. Metadata can provide detailed information about the context of the research process, everything from methodological and practical considerations that influence the content captured in the data to reflections from the researcher on possible meaning associated with specific content. In short, metadata can be considered as a downstream form of research proposal and field notes. The upside of the development of metadata standards for qualitative data and an orientation toward FAIR principles (as outlined in Table 1) is that much of this metadata can be organized and presented in ways that are then commensurate across data sources, which in turn increases the likelihood of re-use for synthesis. The downside of relying on metadata to address epistemological challenges is the amount of time it can take for a researcher to generate 'adequate' content and provide enough detail to assure both researcher and research subjects' perspectives are thoroughly documented. Continued discussion is needed about the types and structures of metadata that can provide necessary, sufficient and appropriate details that reflect various epistemological orientations. Some qualitative researchers will continue to reject any form of data sharing, while others will need to provide more metadata than they might otherwise be inclined to do. In the middle are many researchers who gather qualitative data and who will likely find it increasingly possible to articulate and operationalize their own epistemological understandings if metadata standards and expectations for data sharing are transparent and well-articulated.

Ethical challenges

In addition to abstract epistemological concerns associated with data sharing and re-use, there are ethical challenges that must be fully considered throughout the data life cycle. As Bishop (2009) notes, ethical research decision-making includes very important concerns for protecting participants' rights, and extends as well to responsibilities that researchers have to the scholarly community and the public at large. Institutional review boards (IRBs) tend to focus on the narrower (though by no means simple) ethical issues associated with human subjects research. Broader ethical considerations are not governed by any specific institution, but are often present in the design and execution of many qualitative research projects and how researchers choose to construct them (Biddle and Schafft 2015). Ethical commitments have the potential to both limit and motivate qualitative data sharing, and efforts to engage ethical discussions and positions will be necessary to further support both sharing and re-use of qualitative data.

The ethical challenges associated with informed consent, confidentiality and anonymity in human subjects research are well documented and are largely policed by IRBs and scientific integrity bodies. The Federal government recently updated the Federal Policy for the Protection of Human Subjects (often called the Common Rule), which sets guidelines for ethical decision-making to maintain privacy and insure research participant consent (GPO 2017). For data sharing, both federal and institutional requirements of privacy and confidentiality require that researchers remove all identifying information from any data or analysis that will be shared beyond the approved research team. In practice, this can mean removing names, exact locations, and other details about people and places that could be combined to identify specific individuals. This work takes time, and when considered in conjunction with the creation of adequate metadata (as discussed above), seems to dissuade many qualitative researchers from considering depositing their data in open repositories. Similarly, informed consent at the outset of research projects, which is mandated by research ethics bodies and protocols, often does not encompass the full scope of possible future uses to which the data could subsequently be put - as Bishop (2009: 263) notes, "all consent is partial." There are ethical questions about representation and trust for researchers who consider archiving their data when they are unable to request consent after the research project is complete (the inability to return to research subjects can be practical, a problem of time and funding, due to the sensitive nature of the research topic, or even due to the loss of contact through death or displacement of the research subjects) (Bishop 2009; Hartter et al. 2013).

The actions required by IRBs, designed to protect human subjects, can limit the usability of shared data for future research, especially research that seeks to understand the interactions between people and their environment. If the phenomena of interest occur at fairly fine spatial scales (for example, if a researcher is interested the relationship between proximity to a polluting power plant or a natural disaster, and opinions about climate change), there is often a mismatch between the granularity of non-human data (about the natural and built environment) and the need to aggregate or scramble the spatially identified aspects of human subjects data (Hartter et al. 2013). In addition, limiting the archiving and sharing of human subjects data because of the risk of re-identification or because of a lack of consent deprives not only researchers but end users of research with rich sources of information that could contribute to public goods and useful knowledge. Bishop (2009) and others (Lupia and Elman 2014; DuBois et al. 2017) highlight these additional ethical challenges faced by all researchers who gather data about humans and social systems, and raise questions about how to balance the ethical obligations researchers have to individual research subjects with the ethic of research transparency and contributing to the public good.

It is important to note that the ethical challenges faced by researchers considering sharing data do not apply only to those with primary qualitative data. Often, systematic gathering and analysis of secondary qualitative data, including photos, maps or policy documents, can highlight inequalities, incongruences

and unintended consequences of policies and practices. By archiving these data sources and pointing to them (via analysis) as evidence of a problem, researchers must grapple with questions of transparency and completeness. In other words, the ethic of openness (in this case, submitting all evidence considered) is challenged by the desire to discover something new (for further discussion, see Kapiszewski and Kirilova 2014; Lupia and Elman 2014). There are also ethical challenges associated with data re-use, in terms of representation and a lack of engagement in the re-use/synthesis process by original research participants (Bishop 2009). As Turner (2016) points out, it is overly simplistic to assume that research subjects would not want data that they shared to be re-used, especially if they care about the research topic at hand (and avoiding research fatigue in their community). And at the same time, it is incumbent on researchers engaged in secondary analysis to be transparent themselves about the methods they use to gather, analyze and interpret qualitative data, and how their research process relates to their findings. This is part of the 'ethical openness' described by the International Arctic Science Committee (IASC 2013) and other science networks interested in leveraging the evidence base to address challenging questions in ways that are accurate, appropriate and respectful of the original research process and subjects.

LANDSCAPE OF RESOURCES FOR QUALITATIVE DATA SHARING AND RE-USE

Although sharing and re-use continues to be much less common with qualitative than with quantitative data, there are many resources and options for actors across the research ecosystem to build their own capacity for engaging with qualitative data. The most comprehensive source of information on all parts of the qualitative data life cycle is a public Zotero library of articles, policy documents, guidelines and other resources that is managed by the Qualitative Data Repository (QDR).[3] As of the end of 2017, this resource library has over 1200 entries covering everything from guidance on data management and curation to copyright and citation issues to pedagogical resources for teaching about qualitative data depositing and re-use. We highly recommend that readers of this white paper turn to this resource as they narrow in on the specific types of resources needed for their own projects and interests. In this section, we highlight a few of the most comprehensive and common resources that can facilitate and build capacity for qualitative data sharing and re-use. These resources include repositories in which researchers can both deposit (share) and discover (for re-use) qualitative data, technical resources to support many different actors in the qualitative data ecosystem, and examples of networks and communities of practice that are defining their own approaches to managing, sharing and re-using qualitative data.

REPOSITORIES

While archiving data in some form has a long history in certain disciplines, including for survey-based social sciences (see Corti 2012 for a summary), there has been a push over the past 20 years to greatly expand the expectations and opportunities for depositing research data (National Research Council 1995; Van den Eynden and Corti 2017; Bishop and Kuula-Lummi 2017). This push has come from journals (Fairbairn 2011), funders (Northwestern University has an extensive a list of Federal agency funder requirements and histories), and many in the academic research community (McNutt 2016). As a result of the move toward open science and open data, myriad data repositories have been established and expanded. These repositories include those hosted and maintained by libraries and research institutions, government agencies, disciplinary or topical communities of practice, and increasingly, by data management and curation organizations.

There are some clearinghouse lists of data repositories, including the Directory of Open Access Repositories (OpenDOAR) and re3data (now a service of DataCite), that allow users to search both lists of repositories and a limited amount of information about repository holdings. Some data management organizations and projects, including DataONE and Dataverse, also provide users the ability to search across data repositories to identify data collections and entries of interest. As part of the preparation for the workshop that generated this white paper, we created a list of the major data repositories, and identified whether their holdings include qualitative data, and whether they have standards and guidelines for depositing qualitative data and/or human subjects data. A static version of the list has been archived on the SESYNC website.

Many data repositories do have qualitative data entries, especially those that are federated or support data deposits by a diversity of teams, projects and researchers, and many have guidelines for depositing human subjects data, be they qualitative or quantitative. However, very few repositories have specific guidance and procedures for archiving qualitative data, and to our knowledge there are only a very few

---

repositories that focus on qualitative data. In the United States, the Qualitative Data Repository ([QDR](#)) has led the push to create standards, protocols and tools for archiving qualitative research data. This repository currently focuses mostly on data from political science (because of its original funding source and an interest from that disciplinary community), and it is continuing to expand its holdings to engage a broader set of primarily social science disciplines. In the United Kingdom, the Qualidata archive, which was established in the late 1990s, has led the way in qualitative data archiving and has more recently merged with the [UK Data Archive](#) of the UK Data Service ([Bishop and Kuula-Lummi 2017](#))[4]. In addition to these leaders, several repositories and services provide specific support for including qualitative data in data deposits and data management systems. The Inter-university Consortium for Political and Social Research ([ICPSR](#)), for example, includes [guidance for managing and preserving qualitative data](#) within its overall tools for researchers wishing to deposit their data, including extensive documentation and protocols for restricting access to and use of human subjects data. The [Dataverse](#) software, which is used by both the Qualitative Data Repository and Harvard Dataverse repository, accepts any qualitative file formats and allows the user to extend metadata to support any type of qualitative data across research fields. The Dataverse software provides options to restrict the data files and add any type of terms of use or license associated with the data, while the descriptive metadata is public to make the dataset findable.[5]

The scope of the qualitative data available for re-use varies widely, but by any metric pales in comparison to the volume of archived quantitative data. QDR, which has a fairly small budget and focused topical area, and has functioned more as a pilot project that has generated extensive guidance on qualitative data depositing, has about 30 data project entries as of 2017. The UK Data Archive reports that as of 2016 it held about 1000 entries that include qualitative data ([Bishop and Kuula-Lummi 2017](#)), while a coarse scan of ICPSR in late 2017 suggested that about 10% of its 4300 entries include qualitative data. The Harvard Dataverse repository supports mostly quantitative data but it also accepts qualitative data. Currently 1,000 of its 25,000 deposited datasets include qualitative data[6]. Of course, it is likely that there is qualitative file data associated with many more project and data entries across many data repositories, but it remains the case that there is a lack of clear and consistent identification of qualitative data entries in many repositories and in their data deposit guidelines.

TECHNICAL RESOURCES

Technical resources tend to focus on data management plans that address the entire data life cycle, which tend to lead a researcher through the process of data gathering all the way to data sharing. Many researchers are familiar with the request for a data management plan from funding agencies, which is common and standard across most public funders in the United States and the European Union.[7] [Guidelines from the US National Science Foundation](#), for example, reflect both broad government mandates for transparency in taxpayer-funded research, as well as the specific needs and approaches relevant to a given directorate or division. While these guidelines highlight broad categories of data management and preservation that researchers must consider, they do not provide much detail in terms of the options that researchers have and what actions they can take throughout the research process. Many researchers have begun to use templates and online (free) services like the Data Management

---

[4] See a [recent blog post](#) from Louise Corti for discussion of European qualitative data collections.

[5] At the time of this writing, the Dataverse software is being integrated with the [DataTags](#) projects to support sensitive data in version 5.x.

[6] The [Murray Research Archive](#) hosted at Harvard Dataverse is an example of one of the collections that contains both quantitative and qualitative data in psychology and sociology.

[7] For a summary of definitions, guidelines and expectations for data management plans across several US and European funding agencies and repositories, see [Hodson and Molloy (2014)](#).

Planning Tool (DMPTool), which walk researchers through the decisions that must be made at all stages of the data life cycle to ensure appropriate access. The DMPTool includes considerations of qualitative data formats as well as human subjects research issues, including levels of access and processing (as will be discussed below). The DMPTool also has example data management plans for qualitative research projects and primary qualitative data.[8] However, it does not offer specific guidance on repositories that are most appropriate for qualitative (or quantitative) data, nor does it provide explicit discussion about how researchers can achieve the balance of openness and accountability with the ethical and epistemological commitments that underpin the specific research process.

Several data repositories and data management organizations provide more specific guidance on how to include qualitative data in a data management plan. The broadest set of resources is maintained by QDR and includes both general background on how to conceptualize qualitative data management, as well as specific templates, checklists and examples to guide researchers through the planning process.[9] QDR also provides detailed guidance for specific aspects of the qualitative data management process, including writing informed consent documents that confirm the possibility of future data sharing. Similarly, ICPSR provides extensive guidance on human subjects data management plans, and is a useful resource for researchers starting to think about how to operationalize ethical concerns around confidentiality and risk. For more general guidance, DataONE offers a series of educational modules, including presentation slides and hands-on exercises that teach users about best practices in data management across the data life cycle. Both of these latter resources, and more like them, include only brief discussion of text-based data types, and do not focus on the specific ethical and epistemological dimensions of qualitative research and data sharing. In addition, by looking across all of these technical resources, we have identified a need for more guidance and possibly new tools to manage the qualitative data sharing and re-use workflow.

Box 3: Digital Humanities Bootcamp

The evolving scholarly community of the Digital Humanities provides a variety of technical resources useful for qualitative data sharing and re-use. Digital Humanities refers to the application of digital and computational methods to humanistic inquiry (Burdick et al. 2012). The "bootcamp" is a popular collaborative environment in which experienced digital humanists and interested, inexperienced scholars create, test, master, and apply research tools. Attesting to their accessibility and potential for productive interactions, bootcamps can feature graduate students, librarians, archivists, museum professionals, administrators, managers, and funders (THATCamp 2017). Bootcamp sessions often include sessions relevant to qualitative data sharing and re-use. For instance, the 2017 iteration of the Digital Humanities Summer Institute, an annual two-week workshop at the University of Victoria that began in 2001, offered courses on "Open Access and Open Social Scholarship," "Ethical Collaboration in the Digital Humanities," "RDF and Linked Open Data," "Wrangling Big Data for DH," and "Beyond TEI: Metadata for Digital Humanities" (Digital Humanities Summer Institute 2017). Since the early 2000s, digital humanities bootcamps have grown in number and distribution. THATCamps (The Humanities and Technology Camps) are now regularly held at university humanities departments, research laboratories, and conferences around North America and Europe (THATCamp Directory 2017).

---

[8] For one example relevant to socio-environmental research, see the DMP for "A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy".
[9] For example, QDR has webinar on secure management of qualitative data can also be found here.

Many of the resources overviewed above are oriented toward researchers looking to learn about the tools and skills needed to deposit their data (of any type) in FAIR fashion - that is, in a location and format that facilitates discovery and the potential for further re-use. However, many other actors in the research data ecosystem have a role to play to support and encourage qualitative data sharing, and there are a few sources of resources for them as well. QDR offers some publications and resources for managers of data repositories, as well as for IRBs and ethics officers. For librarians and other research data archiving and discovery professionals, the International Association for Social Science Information Services and Technology (IASSIST) offers an extensive set of resources, including presentations, trainings and best practice guidelines that have been contributed by IASSIST members. Most of these resources are freely available, and include guidance on specific topics like qualitative data visualization, specific data curation needs of qualitative researchers, and information about qualitative data management and analysis software.

NETWORKS AND COMMUNITIES OF PRACTICE

In addition to the specific resources overviewed above, there are many networks and communities of practice that are explicitly engaging with various aspects of qualitative data sharing and re-use. In this section, we highlight just a few examples of networks that offer a wide range of resources and that are engaging in innovative approaches to supporting qualitative data sharing and re-use. The Research Data Alliance, an international network of professionals working across the research data ecosystem, includes several interest groups (meant for discussion) and working groups (meant to generate concrete tools and outputs) related to qualitative data management, the ethics of data sharing, and the creation of metadata standards for FAIR data sharing. Similarly, IASSIST provides members (membership costs $50 annually) with access to networks and discussions with research data professionals via a listserv, discussion boards and interest groups (for example, as of 2017 there is currently a Qualitative Social Science and Humanities Data Interest Group). The Association for Computers and the Humanities is another professional organization that provides access to publications and resources via paid membership, and hosts as well an open discussion board for all topics related to digital humanities approaches and tools.

In addition to organized networks of research data professionals that have interest in qualitative data sharing and re-use, there are also many communities of practice, comprised of researchers and other practitioners, that share topics or geographies in common and that are developing innovative approaches to qualitative data sharing and re-use. For example, the Arctic research community, which includes a wide diversity of disciplines and actors, and is supported by many governments and international efforts, has invested heavily in both data management and sharing infrastructure, as well as guidelines for ethical data access, for qualitative data and different forms of Indigenous Knowledge. The Exchange for Local Observations and Knowledge of the Arctic (ELOKA), supported by the National Science Foundation (NSF), has developed guidelines, communities of practice, and technical solutions for ethical data sharing and supports several open data initiatives to advance Arctic understanding. DataARC, another NSF-funded effort, has built prototype web-based data discovery tools and a conceptual ontology to link qualitative data (largely historical records) to quantitative measurements of environmental characteristics. Related networks of Indigenous data professionals have developed the US Indigenous Data Sovereignty Network, which engages a broad community of professionals in discussions about how to ensure ethical and appropriate data gathering, use, sharing and re-use with indigenous and native communities, and provides links to data repositories.

LEVELS OF PROCESSING AND ACCESS FOR QUALITATIVE DATA SHARING AND RE-USE

As has been discussed and alluded to throughout this white paper, there are both many benefits and many challenges to sharing and re-using qualitative data, depending on the types of data gathered and the approaches used to gather them. In this section, we overview two aspects of qualitative data sharing that can be adapted to reflect specific characteristics of the original data gathering process and that in turn impact how data can be re-used. We introduce the idea of levels of data processing, borrowing the heuristic used by NASA for earth science data, and describe possible levels of processing for different types of qualitative data. We also overview the levels of access to data and the types of access restrictions that are often offered by repositories that include human subjects data and other sensitive, often qualitative, data. Finally, we combine these two aspects of qualitative data sharing, and provide examples of different combinations of data formats and access options. We highlight as well the re-use possibilities for different levels of data, and note the benefits and trade-offs of data with different levels of processing and access.

LEVELS OF PROCESSING

Data processing levels have been well defined in the earth science community, led by NASA's Earth Observation System Data and Information System (EOSDIS) data products. EOSDIS provides data products at five levels of processing (on a 0 to 4 scale[10]), ranging from unprocessed, totally raw data from individual sensors (Level 0) to model outputs or analytical results like derived variables (Level 4). There are two additional dimensions to the level of processing, beyond simple aggregation, that are important to note as well. The first is the inclusion of metadata (time, location, etc.) and associated geophysical variables as the level of processing increases. The second additional dimension is the standardization of spatial and temporal scales as the level of processing increases, which culminates in analytical outputs at level 4. Similar data processing scales and dimensions are defined by other physical science data gathering efforts, including the US Department of Energy's AmeriFlux project (which characterizes data processing using a two-dimensional framework of spatio-temporal representativeness and quality) and the National Ecological Observatory Network.

Data processing is also discussed for social science data, and often focuses on similar issues of data cleaning, aggregating, and standardizing measurements and variables. The Finnish Social Science Data Archive provides the most extensive guidance on qualitative data processing (as well as separate resources for quantitative data processing), including discussion of transcription, organization and naming of data files, and the generation of metadata. Separate guidance is provided for anonymization and protection of confidentiality, with a specific section on these issues with qualitative data. Other social science repositories provide more general guidelines for data processing and anonymization of human subjects data. ICPSR, for example, includes data 'enhancement' in the data ingest process, focusing on adding appropriate metadata and creating consistent fields and descriptors. QDR discusses data processing in terms of the types of entries to the repository, as well as approaches for redaction and partial reduction of identifying details as a means of processing for anonymity (Kirilova and Karcher 2017).

Because the data generated in the course of research involving human participants often involves promises of confidentiality and/or anonymity to participants, most discussion of social science data

---

[10] NASA provides a full description of the scale for earth observation data.

processing focuses on these challenges. This is particularly true for qualitative data: due to the richness of the information included in, for example, transcripts of qualitative interviews, they may contain contextual information (often referred to as "indirect identifiers") that allow knowledgeable readers to infer the interviewee even if basic identifying information is removed. Characterizing levels of processing for qualitative data will certainly include discussion of how the researcher deals with confidentiality. However, given the breadth of types of qualitative data, and taking a cue from the additional dimensions of processing that have been articulated for biophysical data, workshop discussions concluded that there are many additional possible processing considerations for qualitative data. We explore these in detail as we develop a framework for qualitative data sharing at the end of this section.

LEVELS OF ACCESS

The concept of "open data" dominates the discourse on data sharing. Open data is "made available without restriction, on a non-discriminatory basis, for no more than the cost of reproduction and distribution" (National Research Council 1995). Many social science data repositories offer access controls for such situations. Sometimes misunderstood as impediments to open sharing of data, such controls actually make sharing of sensitive data possible in the first place. Such data should be, using the EU's Horizon 2020 principles, "as open as possible, as closed as necessary" (H2020 Programme 2016). These principles reflect as well the "ethical open access" statement of the International Arctic Science Committee (IASC 2013), which emphasizes the ethical impetus to balance free access (for transparency and to increase knowledge generation and dissemination) and individual rights (including confidentiality and local knowledge)[11].

Levels of access and the methods used to limit that access vary by repository and service. Sweeney et al. (2015), for example, suggest six levels of access that are characterized by encryption and authentication practices. ICPSR, which handles a wealth of sensitive social science data, both qualitative and quantitative, offers four levels of access[12]: public-use (cleaned for confidentiality but otherwise fully available), restricted-use (researchers commit to use restrictions and are files delivered through secure channels); virtual data enclave, which allows remote access to data that remains on ICPSR servers; and physical enclave, which requires researchers to access data at ICPSR facilities. Other repositories like QDR offer multiple types of restricted access that are increasingly stringent in requirements that must be met by the researcher interested in accessing the data. Providing multiple levels of access for different types of data and for specific artifacts within a single research project highlights the distinction between FAIR principles for data sharing and a purist orientation toward open data. Providing discussion about how and why different levels of access are appropriate for different types of qualitative data can help to engage some of the epistemological and ethical concerns initially raised by qualitative researchers.

A FRAMEWORK FOR LEVELS OF PROCESSING AND ACCESS FOR QUALITATIVE DATA

Based on the existing resources that define levels of processing and levels of access for various types and sources of data, we propose a corollary framework that is specific to qualitative data in its many forms. This framework allows researchers to consider both aspects of data sharing at the same time, and highlights the relationship between data type, and appropriate levels of processing, and access.

---

[11] For a presentation on ethical data use in the Arctic research community a recent YouTube video.
[12] For full description of these levels of access, see the ICPSR guide for data depositing.

Table 2: Definitions of levels of processing for qualitative data

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

The levels of processing proposed in Table 2 reflect the fact that processing both increases the confidentiality and protection of sensitive information within data artifacts, as well as provides increasingly standardized and summarized description of the research context and methodology as well as the data themselves. Highlighting both aspects of processing underscores the point that 'raw' data is not always more useful analytically or for future re-use, and also reflects the common observation in qualitative research that all data is a representation of reality (and therefore already processed in the sense of being partial or somewhat summative (for further discussion see Temple and Young (2004))). An entire data deposit can have a single, consistent level of processing, or different artifacts or types of qualitative data within a deposit can be provided at different levels of processing, depending on their characteristics.

Table 3: Definitions of levels of access for qualitative data

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 obtaining IRB or accessing data through virtual environment)
C - Controlled	Data is available for use when the user is approved by the original researcher (access could depend on research questions and intended analysis, 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)

The levels of access defined in Table 3 integrate the language and approaches used by several of the leading data repositories for social and/or qualitative data (including specifically the policies of QDR, ICPSR and the UK Data Archive). The specific mechanisms and technologies used to facilitate each level of access are variable, and in this framework, we focus not on how but on why the data are subject to certain level of access restrictions. Open data are those that because of their origin or level of processing do not require any access restriction beyond general commitment to use and cite them in accordance with standard practices. Restricted data are those deemed sensitive from a research ethics point of view, and so must be accessed after meeting certain criteria that ensure that the original research subjects' rights are respected. Access to controlled data is restricted not only because of ethical concerns but also because of epistemological or interpretive concerns. Requests to use controlled data will be decided on by the original researcher, and access could be granted to all or part of the data at various levels of processing. Finally, closed data are those deposits that are made for archival purposes but are entirely non-accessible to any other user. These data might be embargoed until research results are published, or they might remain closed until a certain amount of time or a sensitive issues has passed.

By explicitly articulating levels of processing and levels of access as two distinct and related aspects of qualitative data sharing for re-use, we hope to open the conversation for more qualitative researchers to consider sharing their data in some way. To further clarify the data sharing opportunities that can exist, in Table 4 below we provide examples of types of qualitative data that might reasonably be shared with specific combinations of levels of processing and access. Table 4 depicts types of qualitative data and analysis for four example projects as defined by the type of qualitative data (each in a different color), and situates possible levels of processing and associated levels of access for each type of data. The project shown in red text represents research using public policy documents as secondary qualitative data. The project in green text represents research using images of a political rally or protest as primary and/or secondary qualitative data. The project in blue represents research using interview transcripts and field notes as primary qualitative data. Finally, the project in pink text represents research using photographs and ethnographic field notes of sensitive locations as primary qualitative data.

Table 4: Levels of processing and access for four example types of qualitative data

Level of processing	Level of access			
	A [open]	B [restricted]	C [controlled]	D [closed]
0 [Raw data]	Public policy documents	Images of a political event	Raw interview transcripts or field notes	Photographs of sensitive sites or individuals
1	Public policy documents with search terms as metadata	Images of a political event with faces blurred	Interview transcripts with names and locations redacted	Ethnographic field notes of sensitive sites or events
2	Public policy documents with code for web scraping	Images of a political event with faces blurred and metadata about context of event Interview transcripts with names and locations redacted and metadata about setting of interviews	Interview transcripts with names and locations redacted and metadata about setting of interviews Excerpts of ethnographic field notes and metadata about research sites or events	*Empty boxes depict levels of processing less commonly associated with closed (Level 4) access*
3	Public policy documents organized by theme and with code for thematic analysis	Images from political event with faces blurred, metadata and initial analysis of meaning Interview excerpts with names and locations redacted and metadata including thematic codes	Excerpts of photographs and field notes that represent site or event characteristics and metadata including thematic codes	
4 [Research findings/ output]	Descriptive summary of themes within policies with methodology explained Selected images from a political event with faces blurred representative of analytical categories Summary of thematic analysis of interview transcripts with methodology explained	Summary of thematic analysis of interview transcripts with methodology explained Written summary of sensitive sites or events with reference to specific photographs and notes	Written summary of sensitive sites or events with reference to specific photographs and notes	

These four examples are only a small sample of possible qualitative content. However, this matrix and set of examples highlights the diversity of data types, research settings, and ethical and epistemological commitments that must be accounted for when making decisions about the level of access and level of processing at which to share qualitative data. As shown in Table 4 below, for example, increased processing does not necessarily mean that more open access will be immediately appropriate, if the research context is sensitive or if the researcher has not provided adequate metadata for interpretation of the data.

ROLES AND RECOMMENDATIONS FOR ACTORS IN QUALITATIVE DATA SHARING AND RE-USE

Based on the benefits, challenges and frameworks for sharing and re-using qualitative data that have been identified and synthesized throughout this white paper, in this section we highlight the roles and recommendations for key actors in the data ecosystem. Individual researchers, research institutions, data repositories and cyberinfrastructure organizations, journals and publishers, funders, and knowledge users all have distinct and interconnected roles to play in supporting and expanding the likelihood that qualitative data will be shared and re-used in future research projects. Based on workshop discussion and the breadth of professional diversity that is represented by the authors of this paper, we identify these roles and make recommendations of specific actions that each type of actor can take to lower the barriers that prevent qualitative data sharing and re-use.

RESEARCHERS

Researchers from many disciplinary and institutional backgrounds are foundational to the process of generating, sharing and re-using qualitative data, and they play a key role as well in building networks and capacity for data sharing and data re-use. In the research design and data gathering process, researchers who utilize qualitative methods often spend significant time articulating the epistemological and ethical positions and safeguards that are reflected in their approach. By identifying these dimensions early in the research process, researchers create a foundation from which to make decisions about if and how to share qualitative data. Researchers must also take into account practical constraints like time and knowledge of appropriate data processing and sharing mechanisms. Finally, researchers who choose to re-use qualitative data in secondary or synthesis research projects are often looking for new sources of information to scale up or scale down their analyses, and to leverage scarce resources by drawing on existing data. Appropriate citation of others' data products builds trust within the research community and ensures reciprocal benefits of data sharing and data re-use over time[13,14].

RECOMMENDATIONS FOR RESEARCHERS

When generating new data

When qualitative and mixed methods researchers initiate new projects that will generate qualitative data, it is imperative that they plan at the outset to systematically document the data generation process and to manage the data in uniform and standardized ways. This planning and data management process should include generating metadata as quickly as possible following data gathering activities (be they in the field, in the archives or on the Internet). Researchers are encouraged to utilize data management planning and data curation tools at the beginning of a research project and to include budget lines for data management and processing in grant proposals and funding requests.[15]

When depositing data

Researchers should follow the FAIR principles[16] when making decisions about where, how and what data to deposit. This includes selecting an appropriate repository, making decisions about the level of access that is ethically and epistemologically appropriate, and depositing data that has been processed and includes documentation to a degree that will allow other researchers to potentially integrate the data into new analyses. Researchers are encouraged to provide data that has been processed only enough to be used appropriately by subsequent users, and to set access restrictions that are as open as is ethically possible.

When accessing and re-using data

When researchers access and re-use qualitative data for new analyses, they should provide full data citations and identify all associated content that contributes to the data re-use. Researchers are encouraged to read and engage all provided metadata to ensure appropriate and accurate interpretation of qualitative data.

---

[13] For discussion of trust in open data, see Corti and Fielding (2016) and Lin and Strasser (2014).

[14] For example, data sharing has been shown to be associated with article citation rates (see Piwowar et al. (2007)).

[15] See the Data Management Planning Tool and Purdue University's Data Curation Profiles as two example resources.

[16] For discussion of FAIR principles, see Wilkinson et al. (2016).

RESEARCH INSTITUTIONS

Research institutions, both academic and government agencies, and their staff, including administrators, librarians, and institutional review boards (IRBs), have several distinct roles to play in supporting and facilitating both sharing and re-use of qualitative data. Administrators at research institutions can make decisions about what kinds of human capacity and technological infrastructure in which to invest that could support data sharing, as well as how researchers receive credit for sharing data and conducting synthesis research. Librarians are often on the front lines of data sharing and data discovery for re-use, working with researchers to manage the data sharing and discovery process (Bracke 2011). Finally, IRBs have responsibility for ensuring that researchers gain appropriate and informed consent from research participants about how and by whom data gathered during a research project will be used for analysis. Sharing human subjects data, especially qualitative data, in repositories for future re-use is a fairly new possibility for researchers and many research institutions do not have the policies or resources in place for dealing with this possibility in either new or legacy research projects.

Recommendations for research institutions

Administrators

Administrators should prioritize investments in human capacity, including in research data librarian positions, and technological infrastructure, that can support and encourage researchers to deposit all types of data, including qualitative data. Administrators should also work to shift the tenure and promotion guidelines at their institutions to give credit to researchers for sharing data products, as well as for conducting synthesis research[17]. Administrators are encouraged to identify gaps in the resources and capacities for data sharing at their institutions, and to adjust investments and policies to encourage qualitative data sharing and re-use.

Librarians

Research librarians should become familiar with the resources available to support researchers interested in depositing qualitative data, as well as those interested in discovering qualitative data for re-use[18]. Librarians are encouraged to gain an understanding of the roles and responsibilities of different actors within the qualitative data sharing and re-use ecosystem, and to apply their skills to data curation activities[19].

Institutional Review Boards

IRBs should review their policies and procedures to ensure that they can provide support to researchers interested in sharing qualitative data for possible re-use[20]. IRBs are encouraged to develop guidelines for qualitative data sharing from legacy (already completed) research projects, and to provide researchers with guidance on the levels of access and levels of processing that can help meet ethical human subjects research standards.

---

[17] Examples of such guidelines include those from the health sciences, where data papers are common (see Breeze et al. (2012) and Chavan and Penev (2011)) and from the digital humanities (see statements on the creation and re-use of digital products from the Modern Languages Association and the American Historical Association).
[18] A few examples of these resources include Purdue University's Data Curation Profiles and the Qualitative Data Repository's extensive guidance on qualitative data sharing for re-use.
[19] For further discussion and examples the role of research librarians, see MacMillan (2014) and Johnston (2014).
[20] For an example informed consent template that includes language about data sharing, see Cornell IRB forms.

DATA REPOSITORIES AND OPEN SCIENCE ORGANIZATIONS

As the open science approach to sharing research data has been increasingly adopted by individual researchers and research communities of practice, public and private investments have supported the creation and maintenance of both data repositories and open science organizations that develop and pilot related cyberinfrastructure. Data repositories and open science organizations support the sharing of qualitative data by developing standards, including those for metadata and levels of access[21], and the re-use of qualitative data by developing software for data management and curation[22]. Data repositories and open science organizations also often act as 'brokers' to facilitate data discovery and synthesis research, through federation and aggregation of existing data resources[23]. Finally, specific data repositories that focus on human subjects and/or qualitative data have extensive capacity building materials and efforts to increase qualitative data sharing and re-use[24].

RECOMMENDATIONS FOR DATA REPOSITORIES AND CYBERINFRASTRUCTURE ORGANIZATIONS

Develop standards and software for qualitative data sharing

Although there have been some efforts to establish standards for both metadata and levels of access for qualitative data, more can be done to test and refine these guidelines, and to link them into FAIR principles[25] and standards for quantitative data to facilitate interdisciplinary research[26]. Data repositories and open science organizations are encouraged to invest in efforts to develop standards and tools for qualitative data sharing that can be integrated with other open science initiatives, including open-source software.

Broker qualitative data discovery

Data repositories and open science organizations have the opportunity to facilitate qualitative data discovery and re-use, as well as interdisciplinary synthesis research, by expanding the data resources that are included in federation and aggregation efforts. Data repositories and open science organizations are encouraged to develop search options to discover the qualitative data resources that can be accessed through existing federation and aggregation, and to broker new relationships across disciplinary and research community boundaries.

Expand training and capacity building

Open science organizations have developed extensive training materials for biophysical and quantitative data management[27], and as qualitative research and data analysis becomes increasingly digitized, there is a need for comparable resources. Open science organizations are encouraged to develop modules and training materials for qualitative data management and analysis that draw on open-source tools and software.

---

[21] See for example Dataverse software and ICPSR guidelines for qualitative data levels of access.
[22] For a discussion of open-source R packages for use with qualitative data, see Estrada (2017).
[23] Examples of aggregation efforts include DataCite and SHARE, while DataONE is an example of a federation effort.
[24] Qualitative data management resources include those from the Qualitative Data Repository and ICPSR.
[25] For discussion of FAIR principles, see Wilkinson et al. (2016).
[26] For one example of these efforts, see the UK Data Archive's Qualitative Data Exchange Schema.
[27] For example, Data Carpentry and Software Carpentry both provide online and in-person training modules to improve research data management and computing skills.

JOURNALS AND PUBLISHERS

A recent collaborative review of the role that journals and publishers can play in promoting data sharing and open science in general concluded there is a need for a "social contract" among all actors within the research ecosystem to support and ensure data sharing for both transparency and potential future re-use (Lin and Strasser 2014). Much like data repositories often act as brokers for data discovery, journals and publishers can act as both catalyst and enforcer to encourage data sharing and appropriate data citation during re-use[28]. For example, journals and publishers are increasingly making publication of accepted articles contingent upon data deposition[29], with some nuance about different levels of access and processing for different types of data. Journals are also increasingly clear in their instructions to authors about the need for consistent data citations[30], which both provides research transparency and gives credit to the original data author, thereby incentivizing further data deposition in the future.

Recommendations to journals and publishers

Establish data deposition and availability requirements

Although many journals do have explicit data deposit and availability requirements, many do not, especially in the social sciences[31], and those journals that do are often not entirely clear about if and when requirements will be enforced. Journals and publishers are encouraged to be explicit about the levels of access and levels of processing that are mandated or acceptable for different types of data and research contexts.

Develop specific and appropriate requirements for qualitative data deposition and availability

While journals and publishers uniformly include ethical considerations in their requirements for archiving and making data available for re-use, qualitative data requires direct engagement with specific practical and epistemological considerations as well. We encourage journals and publishers to develop qualitative data deposition and availability requirements that reflect the time and resource intensity of creating metadata, as well as the epistemological positions held by the original researchers, and to offer alternative approaches to transparency when necessary[32].

Develop standards and training to ensure data citation

As researchers take the time to deposit data and research institutions begin to give credit for data products as professional outputs, it is incumbent on journals to ensure that secondary data is appropriately and fully cited. This is especially important for social science data, both quantitative and qualitative, as both data sharing and data re-use via repositories is newer than the same in biophysical sciences[33]. We encourage journals and publishers to train their editors to assess and ensure appropriate data citation practices in all accepted articles.

---

[28] For discussion of the impact of publishing requirements on data deposition, see Vines et al. (2013).

[29] See for example PLOS ONE's Data Availability Policy and the Data Deposition requirements from *Science*.

[30] For guidance to journals on how to implement data citation the Joint Declaration on Data Citation Principles, see Cousijn et al. (2017).

[31] For extensive resources overviewing journal data sharing requirements, see Gary King's research page at Harvard.

[32] For discussion of the impact on qualitative data sharing of transition from 'open access' to 'transparency', see Corti and Fielding (2016).

[33] See Mooney (2011) for review of data citations from one well-established social science repository (ICPSR).

RESEARCH FUNDERS

Much like journals, publishers and repositories, public and private research funders sit at a critical nexus in the research and data ecosystem to facilitate and encourage qualitative data sharing and re-use. Public and foundation funds have supported the establishment of many common research data repositories, but the resources for long-term maintenance of these repositories vary considerably and very few focus on qualitative data[34]. For individual researchers, funders (especially public agencies) are increasingly setting expectations for data management and data sharing that put an emphasis on transparency, appropriate access[35] and accountability of funds[36]. Funders are also increasingly highlighting the ability of researchers to include the costs of data curation or data access into their funding proposals[37]. The structure and estimation of these budget items, including allowing for costs to be charged before the end of the grant cycle but prior to all data curation tasks are complete, can greatly affect the likelihood that data are made available in a way that adheres to FAIR principles.

Recommendations for research funders

Long-term funding for data repositories

There are many possible models for securing long-term financial stability for research data repositories, funders (both public and private) investing in research for the public good and holding their grantees accountable to open science principles have a vested interest in the creation and maintenance of repositories that reflect FAIR principles. We encourage research funders to commit to long-term funding of data repositories, especially those with infrastructure for qualitative data, and to explore investment partnerships with research institutions, journals, and publishers.

More accountability and detail in data management plans

Many funders require a data management plan as part of their grant proposal process, but ongoing assessment or accountability for these plans is highly variable, in part because of the high degree of flexibility needed across diverse funding portfolios[38]. We encourage funders to set clear guidelines for what constitutes data, especially qualitative data, and access within broad categories of research, and to encourage grantees to provide detail and reporting about levels of processing and levels of access as part of their data management plans and funded research.

Encourage appropriate allocation of resources for data curation and data access

Because data curation, especially for qualitative data, requires significant investments of time and possibly financial resources (as one way to maintain data repositories), it is important for researchers to include these costs in proposal budgets. Accessing secondary qualitative data also often requires significant time investment, especially when the level of access is somewhat restricted. Funders are encouraged to support budget lines for researcher time for data curation, data deposits (if relevant as a future repository funding strategy) and data access.

---

[34] For a recent review of possible funding models for data repositories, see Erway and Rinehart (2016).

[35] For example, the NSF Data Management Plan guidelines direct researchers to consider the "lowest level of aggregated data" that is appropriate to share for a given research community and topic.

[36] A list of recent US federal agency data management policies is maintained by Northwestern University's library.

[37] For example, see the NSF's guidelines on post-end date costs, and the NSF Sociology Program.

[38] For a recent assessment of the history, application and effects of data management plans, see Metcalf (2017).

REFERENCES

Adams, A. (1979). An open letter to a young researcher. *African Affairs, 78*(313), 451-479. http://www.jstor.org/stable/721752.

Asher, A.D. & Jahnke, L.M. (2013). Curating the ethnographic moment. *Archive Journal, 3*, http://www.archivejournal.net/essays/curating-the-ethnographic-moment/.

Barbour, R.S., & Barbour, M. (2003). Evaluating and synthesizing qualitative research: The need to develop a distinctive approach. *Journal of Evaluation in Clinical Practice, 9*(2), 179-186. https://doi.org/10.1046/j.1365-2753.2003.00371.x.

Becker, H.S. (1996). The epistemology of qualitative research. In R. Jessor, A. Colby, & R.A. Shweder (eds.), *Ethnography and Human Development* (pp. 55-71). Chicago, IL: University of Chicago Press.

Biddle, C. & Schafft, K.A. (2015). Axiology and anomaly in the practice of mixed methods work: Pragmatism, valuation, and the transformative paradigm. *Journal of Mixed Methods Research, 9*(4), 320-334. https://doi.org/10.1177/1558689814533157.

Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues, 44*(3), 255-272. https://doi.org/10.1002/j.1839-4655.2009.tb00145.x.

Bishop, L. (2012). Using archived qualitative data for teaching: practical and ethical considerations. *International Journal of Social Research Methodology, 15*(4), 341–350. https://doi.org/10.1080/13645579.2012.688335.

Bishop, L. (2014). Re-using qualitative data: A little evidence, on-going issues and modest reflections. *Studia Socjologiczne, 3*(214), 167-176. http://www.data-archive.ac.uk/media/492811/bishop_reusingqualdata_stsoc_2014.pdf.

Bishop, L., & Kuula-Lummi, A. (2017). Revisiting qualitative data reuse: A decade on. *SAGE Open, 7*(1). https://doi.org/10.1177/2158244016685136.

Boorman, K.M., LeCompte, M.D., & Goetz, J.P. (1986). Ethnographic and qualitative research design and why it doesn't work. *American Behavioral Scientist, 30*(1), 42-57. https://doi.org/10.1177/000276486030001006.

Bracke, M.S. (2011). Emerging data curation roles for librarians: A case study of agricultural data. *Journal of Agricultural & Food Information, 12*(1), 65-74. https://doi.org/10.1080/10496505.2011.539158.

Breeze, J.L., Poline, J.B., & Kennedy, D.N. (2012). Data sharing and publishing in the field of neuroimaging. *Gigascience, 1*(1), 1-3. https://doi.org/10.1186/2047-217X-1-9.

Brizinski, P.M. (1993). The Summer Meddler: The Image of the Anthropologist as Tool for Indigenous Formulations of Culture. In *Anthropology, Public Policy, and Native Peoples in Canada*, edited by Noel Dyck and James B. Waldram, 146-165. Kingston, Ontario: McGill-Queen's University Press.

Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology, 43*(6), 1163-1180. https://doi.org/10.1177/0038038509345704.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *The British Journal of Sociology, 35*(1), 75-92. https://doi.org/10.2307/590553.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital Humanities*. Cambridge, MA: MIT Press. https://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf.

Chavan, V., & Penev, L. (2011). The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC Informatics, 12*(Supp 15), S2. https://doi.org/10.1186/1471-2105-12-S15-S2.

Checkland, Peter. 1999. Systems thinking. In *Rethinking Management Information Systems: An Interdisciplinary Perspective*, edited by Wendy L Currie and Bob Galliers, 45–56. New York: Oxford University Press.

Cho, J., & Trent, A. (2006). Validity in qualitative research revisited. *Qualitative Research, 6*(3), 319-340. https://doi.org/10.1177/1468794106065006.

Clark, T. (2008). 'We're over-researched here!': Exploring accounts of research fatigue within qualitative research engagements. *Sociology, 42*(5), 953–70. https://doi.org/10.1177/0038038508094573.

Corti, L. (2012). Recent development in archiving social research. *International Journal of Social Research Methodology, 15*(4), 281-290. https://doi.org/10.1080/13645579.2012.688310.

Corti, L., & Bishop, L. (2005). Strategies in teaching secondary analysis of qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 6*(1). http://www.qualitative-research.net/index.php/fqs/article/view/509.

Corti, L. & Fielding, N. (2016). Opportunities from the digital revolution: Implications for research, publishing, and consuming qualitative research. *SAGE Open, 6*(4), 1-13. https://doi.org/10.1177/2158244016678912.

Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Murphy, F., Polischuk, P., Martone, M., & Clark, T. (2017). A data citation roadmap for scientific publishers. *BioRxiv*. https://doi.org/10.1101/100784.

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.

Denzin, N.K., & Lincoln, Y.S. (2008). *The Landscape of Qualitative Research*. Thousand Oaks, CA: Sage.

DuBois, J.M., Strais, M., & Walsh, H. (2017). Is it time to share qualitative research data? *Qualitative Psychology*, Advance online publication. https://doi.org/10.1037/qup0000076.

Estrada, S. (2017). Qualitative analysis using R: A free analytic tool. *The Qualitative Report, 22*(4), 956-968. http://nsuworks.nova.edu/tqr/vol22/iss4/2.

Erway, R. & Rinehart, A. (2016). If you build it, will they fund? Making research data management sustainable. Dublin, Ohio: Online Computer Library Center. https://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016.pdf.

Fairbairn, D.J. (2011). The advent of mandatory data archiving. *Evolution, 65*(1), 1-2. https://doi.org/10.1111/j.1558-5646.2010.01182.x.

Gill, D.A., Mascia, M.B., Ahmadia, G.N., Glew, L., Lester, S.E., Barnes, M. et al. (2017). Capacity shortfalls hinder the performance of marine protected areas globally. *Nature, 543*(7647), 665-669. https://doi.org/10.1038/nature21708.

Goodwin, J. & Horowitz, R. (2002). Introduction: The methodological strengths and dilemmas of qualitative sociology. *Qualitative Sociology, 25*(1), 33-47. https://doi.org/10.1023/A:1014300123105.

GPO. (2017). *Federal Policy for the Protection of Human Subjects*. Government Publishing Office, Federal Register, Vol. 82, No. 12. https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/pdf/2017-01058.pdf.

H2020 Programme. (2016). *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. Horizon 2020, European Research Council. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Hammersley, M. (1997). Qualitative data archiving: some reflections on its prospects and problems. *Sociology, 31*(1), 131-142. https://doi.org/10.1177/0038038597031001010.

Hampton, S.E, Strasser, C.E., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., & Porter, J.H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment 11*(3): 156-162. https://doi.org/10.1890/120103.

Haraway, D. (2001). Situated knowledges: The science question in feminism and the privilege of partial perspective. In M. Lederman & I. Bartsch (eds.), *The Gender and Science Reader* (pp. 169-188). New York, NY: Routledge.

Hartter, J., Ryan, S.J., MacKenzie, C.A., Parker, J.N., & Strasser, C.A. (2013). Spatially explicit data: Stewardship and ethical challenges in science. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.1001634.

Hicks, C.C., Levine, A., Agrawal, A., et al. 2016. Engage key social concepts for sustainability: Social indicators, both mature and emerging, are underused. *Science 352*(6281), 38-40. https://doi.org/10.1126/science.aad4977.

Hodson, S., & Molloy, L. (2014). *Current Best Practice for Research Data Management Policies*. Memo for Danish e-Infrastructure Cooperation and the Danish Digital Library. https://doi.org/10.5281/zenodo.27872.

Hoyle, L., Corti, L., Gregory, A., Martinez, A., Wackerow, J., Alvar, E., et al. (2013). *A qualitative data model for DDI*. Data Documentation Initiative Working Paper No. 5. https://www.ddialliance.org/sites/default/files/AQualitativeDataModelForDDI.pdf.

IASC. (2013). Statement on principles and practices for Arctic data manage. *International Arctic Science Committee*. https://iasc.info/images/data/IASC_data_statement.pdf.

Inuvialuit Regional Corporation. Guidelines for Research in the Inuvialuit Settlement Region. https://nwtresearch.com/sites/default/files/inuvialuit-regional-corporation.pdf.

Janssen, M.A., Bodin, O., Anderies, J.M., Elmqvist, T., Ernstson, H., McAllister, R.R.J., Olsson, P., & Ryan, P. (2006). Toward a network perspective of the study of resilience in social-ecological systems. *Ecology and Society, 11*(1), art15. http://www.ecologyandsociety.org/vol11/iss1/art15/.

Janssen, M. A., & Ostrom, E. (2006). Empirically based, agent-based models. *Ecology and Society, 11*(2), art37. http://www.ecologyandsociety.org/vol11/iss2/art37/.

Janz, N. (2015). Bringing the gold standard into the classroom: Replication in university teaching. *International Studies Perspectives, 17*(4), 392-407. https://doi.org/10.1111/insp.12104.

Johnston, L. (2014). *A workflow model for curating research data in the University of Minnesota Libraries: Report from the 2013 data curation pilot*. University of Minnesota Digital Conservancy. http://hdl.handle.net/11299/162338.

Kapiszewski, D. & Kirilova, D. (2014). Transparency in qualitative security studies research: Standards, benefits, and challenges. *Security Studies, 23*(4), 699-707. https://doi.org/10.1080/09636412.2014.970408.

Karcher, S. (2016). Teaching with Qualitative Data: An Example. 28 September. https://qdr.syr.edu/qdr-blog/teaching-qualitative-data-example.

Karcher, S., Kirilova, D., & Weber, N. (2016). Beyond the matrix: Repository services for qualitative data. *International Federation of Library Associations and Institutions, 42*(4), 292-302. https://doi.org/10.1080/09636412.2014.970408.

Katz, D. S. (2017, June 22). FAIR is not fair enough. https://danielskatzblog.wordpress.com/2017/06/22/fair-is-not-fair-enough/.

King, G. (2006). Publication, publication. *PS: Political Science & Politics, 39*(01), 119–125. http://j.mp/2owxpXM.

Kirilova, D. & Karcher, S. (2017). Rethinking data sharing and human participant protection in social science research: Applications from the qualitative realm." *Data Science Journal, 16*(September). https://doi.org/10.5334/dsj-2017-043.

Lin, J. & Strasser, C. (2014). Recommendations for the role of publishers in access to data. *PLoS Biology, 12*(10), e1001975. https://doi.org/10.1371/journal.pbio.1001975.

Lindkvist, E., Basurto, X., & Schlüter, M. (2017). Micro-level explanations for emergent patterns of self-governance arrangements in small-scale fisheries—A modeling approach. *PloS One, 12*(4), p.e0175532. https://doi.org/10.1371/journal.pone.0175532.

Lupia, A., & Elman, C. (2014). Openness in political science: data access and research transparency. *PS: Political Science & Politics, 47*(01), 19–42. https://doi.org/10.1017/S1049096513001716.

MacMillan, D. (2014). Data sharing and discovery: What librarians need to know. *The Journal of Academic Librarianship, 40*(5), 541-549. https://doi.org/10.1016/j.acalib.2014.06.011.

Maxwell, J.A. (2011). Epistemological heuristics for qualitative research. In H. Soini, E.L. Kronqvist, & G.L. Huber, *Epistemologies for Qualitative Research*, (pp. 10-27). Tubingen, Germany: Center for Qualitative Psychology.

McNutt, M. (2016). #IAmAResearchParasite. *Science, 351*(6277), 1005. https://doi.org/10.1126/science.aaf4701.

Metcalf, J. (2017). *Data management plan: A background report*. Council for Big Data, Ethics and Society. http://bdes.datasociety.net/council-output/data-management-plan-a-background-report/.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Santos, da S., Bonino, L. O., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use, 37*(1), 49–56. https://doi.org/10.3233/ISU-170824.

Moon, K., & Blackman, D. (2014). A guide to understanding social science research for natural scientists. *Conservation Biology, 28*(5),1167–77. https://doi.org/10.1111/cobi.12326.

Mooney, H. (2011). Citing data sources in the social sciences: Do authors do it? *Learned Publishing, 24*(2), 99-108. https://doi.org/10.1087/20110204.

National Research Council. (1995). *On the Full and Open Exchange of Scientific Data*. Washington, D.C.: The National Academies Press. https://www.nap.edu/catalog/18769/on-the-full-and-open-exchange-of-scientific-data.

Osmond, B., Ananyev, G., Berry, J. et al. (2004). Changing the way we think about global change research: Scaling up in experimental ecosystem science. *Global Change Biology, 10*(4), 393-407. https://doi.org/10.1111/j.1529-8817.2003.00747.x.

Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. *Science, 325*(5939), 419-422. https://doi.org/10.1126/science.1172133.

Palmer, M.A., Bernhardt, E.S., Chorensky, E.A., et al. 2005. Ecological Science and sustainability for the 21st century. *Frontiers in Ecology and the Environment, 3*(1): 4-11. https://doi.org/10.1890/1540-9295(2005)003[0004:ESASFT]2.0.CO.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature, 421*(6918), 37-42. https://doi.org/10.1038/nature01286.

Piwowar, H.A., Day, R.S., & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE, 2*(3), e308. https://doi.org/10.1371/journal.pone.0000308.

Poteete, A., & Ostrom, E. (2005). Bridging the qualitative-quantitative divide: Strategies for building large-N databases based on qualitative research. Paper presented at the American Political Science Association Annual Meeting, 1-4 Sept. Washington, DC. http://hdl.handle.net/10535/5890.

Pullin AS, Sutherland W, Gardner T, Kapos V, Fa JE. 2013. Conservation priorities: identifying need, taking action and evaluating success. *Key Topics in Conservation Biology 2*, 3–22. https://doi.org/10.1002/9781118520178.ch1.

Sistla, S.A., Roddy, A.B., Williams, N.E., Kramer, D.B., Stevens, K., & Allison, S.D. (2016). Agroforestry practices promote biodiversity and natural resource diversity in Atlantic Nicaragua. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0162529.

Stem, C., Margoluis, R., Salafsky, N., & Brown, M. (2005). Monitoring and evaluation in conservation: A review of trends and approaches. *Conservation Biology, 19*(2), 295-309. https://doi.org/10.1111/j.1523-1739.2005.00594.x.

Sweeney, L., Crosas, M., & Bar-Sinai, M. (2015). Sharing sensitive data with confidence: The datatags system. *Technology Science* 16 October. https://techscience.org/a/2015101601.

Temple, B., & Young, A. (2004). Qualitative research and translation dilemmas. *Qualitative Research, 4*(2), 161-178. https://doi.org/10.1177/1468794104044430.

Turner, D. (2016). Archiving qualitative data: Will secondary analysis become the norm? *Quirkos Blog*, 24 Nov. https://www.quirkos.com/blog/post/qualitative-archives-secondary-analysis-software.

Van den Eynden, V. & Corti, L. (2017). Advancing research data publishing practices for the social sciences: From archive activity to empowering researchers. *International Journal on Digital Libraries, 18*(2), 113-121. https://doi.org/10.1007/s00799-016-0177-3.

Vines, T.H., Andrew, R.L., Block, D.G., Franklin, M.T., Gilbert, K.J., Kane, N.C., Moore, J.C., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal, 27*(4), 1304-1308. https://doi.org/10.1096/fj.12-218164.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, sdata201618. https://doi.org/10.1038/sdata.2016.18.

Young, O.R., Lambin, E.F., Alcock, F., et al. 2006. A portfolio approach to analyzing complex human-environment interactions: Institutions and land change. *Ecology and Society 11*(2): 31. http://www.ecologyandsociety.org/vol11/iss2/art31/.