nature sustainability

PERSPECTIVE

https://doi.org/10.1038/s41893-019-0434-8

Qualitative data sharing and synthesis for sustainability science

Steven M. Alexander[1,2,3*], Kristal Jones[1*], Nathan J. Bennett[4,5,6], Amber Budden[7], Michael Cox[8], Mercè Crosas[9], Edward T. Game[10], Janis Geary[11], R. Dean Hardy[1,12], Jay T. Johnson[13], Sebastian Karcher[14], Nicole Motzer[1], Jeremy Pittman[15], Heather Randell[1,16], Julie A. Silva[17], Patricia Pinto da Silva[18], Carly Strasser[19], Colleen Strawhacker[20], Andrew Stuhl[21] and Nic Weber[22]

Socio–environmental synthesis as a research approach contributes to broader sustainability policy and practice by reusing data from disparate disciplines in innovative ways. Synthesizing diverse data sources and types of evidence can help to better conceptualize, investigate and address increasingly complex socio–environmental problems. However, sharing qualitative data for re-use remains uncommon when compared to sharing quantitative data. We argue that qualitative data present untapped opportunities for sustainability science, and discuss practical pathways to facilitate and realize the benefits from sharing and reusing qualitative data. However, these opportunities and benefits are also hindered by practical, ethical and epistemological challenges. To address these challenges and accelerate qualitative data sharing, we outline enabling conditions and suggest actions for researchers, institutions, funders, data repository managers and publishers.

Contemporary socio–environmental challenges, such as biological conservation or climate change adaptation, require solutions that simultaneously account for diverse actors, institutions and environmental processes. Studying these challenges and offering sustainable solutions is further complicated by uncertain drivers of change, complex feedback loops (both social and environmental), and interactions across broad geographies and timespans. Socio–environmental synthesis as a research approach contributes to broader sustainability policy and practice goals by combining disparate disciplines to re-use data in innovative ways (see Fig. 1 for a depiction of the data use workflow in synthesis research). It is characterized by transdisciplinary inquiry that helps identify patterns across time or geographical scales[1], understand emerging socio–environmental relationships[2], and provide evidence to develop and implement policy[3,4].

Socio–environmental synthesis has accomplished much in terms of both fundamental scientific discoveries and actionable, policy-oriented results through relying primarily on existing quantitative data[5,6]. These data are readily accessed, downloaded and re-interpreted, and are often made available for re-use by the same entities that fund their collection (national government science agencies, international organizations, or grant-funded researchers). However, this growing body of research can contribute even more by expanding to include qualitative data (see Box 1 for definitions of types of data), understood as information that is not, in its initial form, depicted as discrete numerical values. Qualitative data can include text (written transcripts of interviews or focus group discussions, policy documents, journalistic articles and social media content), images (maps, photographs and artwork), video and audio artifacts (oral histories, news reports and music), and other types of unstructured information[7].

Qualitative data present significant opportunities for expanding the focus of socio–environmental synthesis research[8–10]. For example, the exhaustive study *Voices of the Poor*[11–13] drew on qualitative data sources to redefine 'ill-being' as a multidimensional concept, recognizing income poverty as only one aspect of deprivation. Similarly, qualitative data can advance multidimensional understandings of socio–environmental systems. It is particularly useful for incorporating lived experiences, needs, values and perspectives of individuals, groups and even nations, which are often left out of scientific and policy discourses. Data synthesis and reusing all available data types is especially important in regions with fewer resources to produce primary data.

Data sharing can be done directly (that is, researcher to researcher) or indirectly (that is, via a data repository), and we apply the term to the entire process by which data is managed and made available to other researchers. Data re-use refers to the secondary use of data that has been already collected or aggregated. The data

[1]National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, MD, USA. [2]Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden. [3]Environmental Change and Governance Group, Faculty of Environment, University of Waterloo, Waterloo, Ontario, Canada. [4]Institute for Resources, Environment and Sustainability, University of British Columbia, Vancouver, British Columbia, Canada. [5]Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, British Columbia, Canada. [6]Center for Ocean Solutions, Stanford University, Stanford, CA, USA. [7]DataONE, University of New Mexico, Albuquerque, NM, USA. [8]Environmental Studies Program, Dartmouth College, Hanover, NH, USA. [9]Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA. [10]The Nature Conservancy, South Brisbane, Queensland, Australia. [11]Department of Medicine, University of Alberta, Edmonton, Alberta, Canada. [12]School of Earth, Ocean, & Environment, University of South Carolina, Columbia, SC, USA. [13]Department of Geography & Atmospheric Science, University of Kansas, Lawrence, KS, USA. [14]Qualitative Data Repository, Syracuse University, Syracuse, NY, USA. [15]School of Planning, University of Waterloo, Waterloo, Ontario, Canada. [16]Department of Agricultural Economics, Sociology, and Education, Pennsylvania State University, University Park, PA, USA. [17]Department of Geographical Sciences, University of Maryland, College Park, MD, USA. [18]NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA, USA. [19]Collaborative Knowledge Foundation, San Francisco, CA, USA. [20]National Snow and Ice Data Center, University of Colorado Boulder, Boulder, CO, USA. [21]Department of Environmental Studies and Sciences, Bucknell University, Lewisburg, PA, USA. [22]Information School, University of Washington, Seattle, WA, USA. *e-mail: s22alexa@uwaterloo.ca; kjones@sesync.org

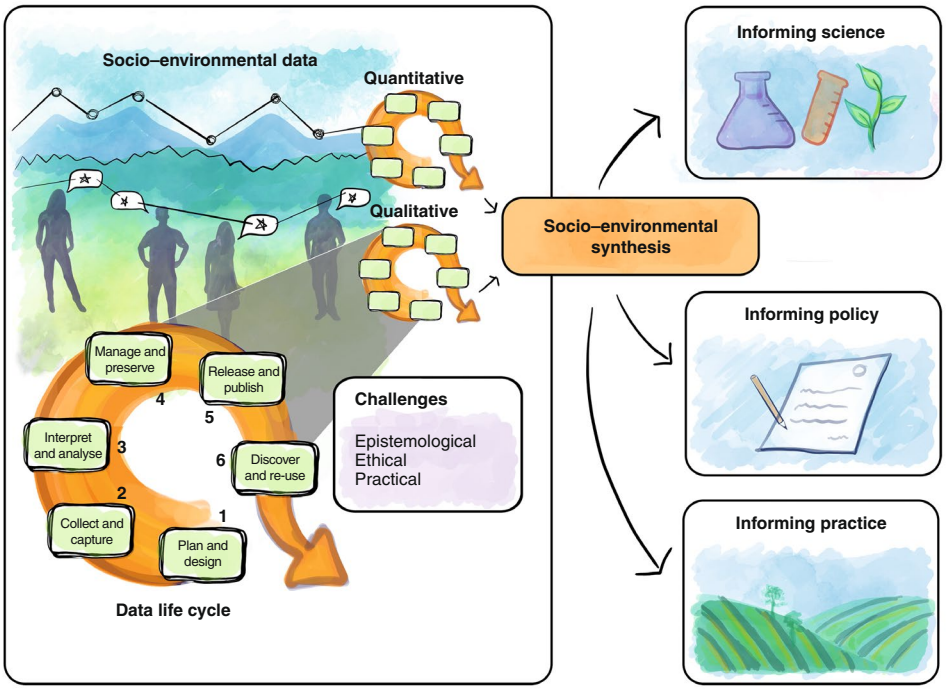

Fig. 1 | Qualitative data sharing for re-use. The data and research life cycle (depicted in orange) is connected to socio-environmental synthesis via the last stage (data re-use), illustrated by the small black arrows. Data-driven synthesis can take many forms including: (1) qualitative data only; (2) quantitative data only; or (3) a combination of qualitative and quantitative data. Socio-environmental synthesis as a research approach enhances and informs science, policy and practice that is oriented toward addressing sustainability challenges. Credit: image courtesy of Sofia Jain-Schlaepfer, www.wiseart.net.

life cycle includes all of the stages and processes necessary for the successful collection, management and preservation of data for use and re-use. Figure 1 depicts the data life cycle for qualitative data included in socio–environmental synthesis.

Despite its great potential, this rich and diverse source of data remains largely absent from socio–environmental synthesis research[5,6]. In this Perspective, we discuss the benefits to sustainability science, policy and practice from adding qualitative data to the socio–environmental synthesis evidence base, and the challenges to data sharing and re-use. Three main types of barriers limit this incorporation of qualitative data: epistemological differences surrounding data collection, sharing and re-use[14,15], ethical commitments[16], and practical issues associated with building and maintaining knowledge commons[17]. To address the challenges associated with sharing qualitative data for re-use, we conclude by articulating a framework for qualitative data management that adds nuance to the possible forms of data that could be shared. Finally, we outline actionable steps for researchers, institutions, funders, data repository managers and publishers to maximize corresponding benefits for sustainability.

Benefits of qualitative data sharing for re-use

Reusing qualitative data facilitates the inclusion of important concepts (for example, governance and aspirations), relationships (for example, power structures and place attachment), and indicators (for example, well-being and trust) that are not easily quantifiable. Without these elements, socio–environmental research is limited in its ability to account for human perceptions, values or motivations[8,9,14]. Qualitative data can add essential information about a given setting or system, and also social, cultural and historical context[10]. Such information can lead to more relevant and actionable solutions or facilitate monitoring outcomes. In this section, we highlight the benefits of incorporating qualitative data to address sustainability challenges in three main domains (domains that are not intended to be mutually exclusive).

Informing science. In socio–environmental synthesis, qualitative data about natural and social systems and their interactions can enhance understanding spatiotemporal variability[18]. While quantitative analyses can identify statistical patterns and relationships between drivers and outcomes, qualitative data informs the underlying reasons for those relationships, using rich descriptions and analysis of patterns in context[19]. In addition, qualitative data are particularly well suited to measure multidimensional concepts like equity and efficacy, which are increasingly important for sustainability. Sharing data about these concepts can also facilitate new types of triangulation and estimations of internal validity and accuracy[20,21]. For example, interviews conducted repeatedly with a population affected by extreme environmental events can illuminate shifting attitudes or perceptions of vulnerability associated with displacement[22]. Digitized historical photographs are also a valuable qualitative data source that can help track environmental change in non-numerical ways; such as to document the loss of large recreationally caught trophy fish over decades[23].

When researchers have access to qualitative data derived from many cases they can: (1) conduct novel cross-case and multi-level comparisons of patterns and contexts; (2) explore the reasons for covariance observed in socio–environmental systems when quantitative measures suggest no causation between drivers and outcomes; and (3) increase the likelihood of generalizability beyond what is often feasible for a single researcher or research team[24]. Comparing cases allows identifying consistent patterns in relationships between human decision-making and observed environmental change[25] and can both confirm and challenge supposedly universal theories[24]. This approach can help build theory[26], as well as highlight how context (for example, social, historical or political) affects socio–environmental outcomes[10].

Informing policy. Repeatedly, scholars have called for increasing the evidence base of effective management and intervention

Box 1 | What is qualitative data?

Qualitative data can be defined in terms of how the data are organized, formatted and managed[69]. Defining data as either structured or unstructured helps to understand the diversity of qualitative data types. Structured data are organized based on an a priori schema or framework. They are formatted to be machine readable (often in tabular form with discrete variables and observations of those variables). In contrast, unstructured data include any type of information that is not organized into singular and discrete categories. In this sense, all qualitative data may be considered unstructured, because a single piece of information—a word, a photograph or a statement—can measure many different concepts or phenomena depending on the theoretical and analytical approaches used to interpret the information.

A wide diversity of gathering approaches for qualitative data influence data management approaches at the downstream end of the research process. Like quantitative data, qualitative data may be collected directly by the person who will interpret it (primary data gathering) or it might exist prior to the research process, like data collected from historical archives of news stories, personal communications or visual materials. Qualitative data are most often collected and utilized by researchers trained in social science and humanities, but are not limited to these disciplines. Ecology, biology and botany also gather qualitative data in the course of field studies, as written observations, sketches and images. These are just some examples of the many types of qualitative data.

strategies to address complex socio–environmental problems[27,28]. However, calls for evidence-based policy currently default to quantitative methods and data and would benefit from greater engagement with a plurality of methods to "provide a more complete picture on which to base management decisions"[29]. Incorporating qualitative data broadens the evidence base and contributes to contextualizing and translating results for use in sustainability policy elsewhere. Cross-case comparisons using qualitative data can also increase the reliability of patterns and relationships found to inform sustainability policies and governance[24,28]. Similarly, contextualizing drivers of change within socio–environmental systems at scales appropriate to both governance and ecosystem function is important to translate science into actionable policy[30].

As sustainability policy is increasingly entering the purview of local to global governments and international bodies, learning from on-the-ground cases of success and failure is indispensable[26,31]. For example, in a recent study of large marine protected areas, qualitative data on management proved to be more relevant than quantitative metrics in uncovering how and why certain sustainability outcomes were realized in some locations but not others[32]. Qualitative data can also inform further research to understand human behaviour and decision-making within specific contexts or under different scenarios, to predict the impacts of prospective policies[26,33]. Understanding why some drivers affect outcomes differently across cases can help adapt sustainability policies better to specific socio–environmental contexts. For example, Gill and colleagues[34] drew on qualitative data to identify characteristics of the political, management and social context of marine protected areas that conditioned the success of protection policies, and to highlight the linkages between policy context and implementation outcomes.

Informing practice. Sustainability practitioners and applied researchers increasingly incorporate qualitative data gathering into their approaches to increase the legitimacy and efficacy of their activities. For example, there has been a move toward supporting place-based responses and solutions to regional or global environmental change. Planning processes and assessments that draw on qualitative data, such as focus group discussions about local environmental history or minutes from public meetings[35], can represent the experiences and perspectives of local stakeholders better than standardized methods limited to discrete quantitative measures[29]. Sharing and reusing all types of data, when appropriate, can also reduce the burden placed on communities underrepresented in science and decision-making where their perceptions and experiences are repeatedly solicited[36]. Making qualitative data available for future re-use also increases returns on investment for funders and researchers by maximizing the use of existing information and focusing expenditures of time and money on gathering new data.

Many sustainability practitioners also recognize the value of qualitative data to improve equity and inclusion. Hicks and colleagues[8] drew attention to the value of blending qualitative and quantitative data in socio–environmental assessments of multidimensional human well-being that reflect local needs and values. World views are often best represented with qualitative data and are important contributions to the evidence base, particularly in data-poor nations, communities and environments[35,37]. For example, Tengö and colleagues[38] make a call beyond including qualitative data in evidence bases, to widening them by incorporating diverse knowledge systems (understood as systems of agents, practices and institutions that organize the production, transfer and use of knowledge). These knowledge systems include those rooted in Western science and those based on Indigenous or intersectional identities.

Challenges of qualitative data sharing for re-use

Open science and data accessibility (via data repositories and metadata standards) are increasingly expected in many fields. However, several challenges must be addressed for data sharing and associated sub-processes (for example, archiving and accessing), particularly with qualitative data. These challenges are mainly in three areas: epistemological, ethical and practical (Table 1).

Epistemological challenges. Epistemology focuses on the process that generates knowledge and the "relationship between the knower and the known"[39]. The epistemology a researcher brings to a project influences everything from research design to the collection, analysis and interpretation of data (for further discussion, see refs. [10,40]). Accordingly, we acknowledge that data originating from certain epistemological perspectives might never be deemed appropriate for sharing and re-use (see Box 2 for description of epistemological perspectives). For some qualitative researchers, efforts to analyse and re-interpret qualitative data for synthesis purposes could be seen as invalid because a new individual (the synthesis researcher) could not bring experience with the original research process to bear on a new analysis[41]. However, acknowledging the different origins of diverse data sources does not preclude the possibility to use such data in tandem during analysis—what Nightingale[42] refers to as "productive tensions". For example, data sharing for re-use could lead to qualitative data being integrated with quantitative data or even other qualitative data into some synthetic or comprehensive picture of a social–ecological system[43]. Qualitative data could also be used in synthesis research to parameterize, interpret or validate findings from analysis of other data sources.

Not all epistemological orientations preclude data sharing for re-use (see Box 2). Researchers with a generally positivist epistemological orientation might gather and analyse qualitative data for the purposes of triangulation or to increase explanatory power, and could feel comfortable sharing that qualitative data for others to do the same. Conversely, researchers who largely work from a constructivist epistemological starting point might feel comfortable sharing parts of their qualitative data that describe empirical phenomena[44]. For example, field notes describing what happened

Table 1 | Challenges for qualitative data sharing

Challenge	References
Epistemological	
Epistemological traditions influence whether and how qualitative data might be shared and re-used.	67
Potentially incommensurate 'nature' of data derived through and from different epistemological frames.	42
Critical epistemologies, which start from an understanding of knowledge and evidence that is partial and unique to a given individual, context and interpretation, may not be comfortable sharing qualitative data for re-use that seeks generalizability.	47
Ethical	
Informed consent, confidentiality and anonymity associated with data to be shared is difficult to guarantee without losing value of data.	50
Sometimes data from past projects are of renewed interest for sharing, but were gathered before such an option was common in informed consent statements.	62
Lack of representation of and engagement with original research participants in synthesis work.	50
Ethics of openness can be made vulnerable to desires to leverage an existing evidence base and discover something new, regardless of whether such a discovery is transparent, complete or appropriate.	48
Practical	
Fewer options than quantitative researchers for repositories that can support the diversity of data types, access restrictions and metadata needs of qualitative and multi-modal data.	53,68
Long-term financial resources necessary to maintain the infrastructure remain difficult to secure.	64
Lack of adequate metadata standards to ensure the appropriate and accurate re-use of qualitative data in future synthesis research.	54
Lack of incentives to encourage use of the resources, additions of refinements to existing data and metadata, contributions of new data, and contribution of value-added data that ultimately enhance the overall value of the commons.	56

or who was at an event might be shared, but not explanations from individuals about the meaning of the event.

There are important questions to address in order to re-use data outside of its original epistemological frame. While from a constructivist standpoint data interpretation and meaning is contingent on context, it can also show empirical patterns that can be documented and categorized independent of the original researcher[45]. Provided there is sufficient background and metadata (descriptive information about data gathering and measurement categories), from a constructivist standpoint it is possible to share and re-use qualitative data without compromising the complex empirical realities of human subjects[46].

The richness of qualitative data is beneficial for sustainability science. However, this descriptive complexity also implies challenges to include qualitative data in synthesis research aimed at generalizable or transferable findings[7]. Here, lessons learnt from similar limitations of field observations in ecology[18] are useful. To share and synthesize ecological data, researchers have attempted to address the limitations of small *n* studies by utilizing less processed ('raw') data from many individual studies[2].

Reusing qualitative data has additional epistemological complexity, since many researchers who gather primary qualitative data do not see generalizability as a goal or a possibility. Critical epistemologies like those that underpin the concept of situated knowledge[42], for example, hold that evidence and knowledge are partial and unique to a given individual, context and interpretation. Researchers working from this starting point are thus unlikely to be comfortable sharing qualitative data with an eye toward synthesis and generalized analysis[47].

Ethical challenges. While not exclusive to qualitative data, the extent of ethical challenges, ranging from protecting participants' rights to a primary researcher's responsibilities to the scholarly community and the public good[48,49], set qualitative data apart. Ethical concerns associated with informed consent, confidentiality and anonymity are well documented and largely overseen by institutional review boards (IRBs) and scientific integrity bodies[50]. For anonymization, researchers often need to remove all identifying information from any data or analysis shared beyond approved research teams. In practice, this redactive work requires time and judgement about what is considered identifying, personal or otherwise sensitive information. Furthermore, anonymizing can have a disproportionately limiting impact for projects exploring interactions between people and their environments, and for studies conducted at fine spatial scales[36]. By potentially removing spatially explicit information about the human dimensions of socio–environmental dynamics, generic privacy requirements that can be contextually dependent may limit the contributions that such rich sources of information can make to the public good.

Informed consent processes, in which researchers make clear to research subjects how the information gathered will be used, stored and shared, can also prevent researchers from sharing qualitative data after a project is complete. Although many IRBs now offer language and guidance about how to include information about data sharing in informed consent statements, researchers are often not aware of this possibility at the study outset. Often, data from past projects are of renewed interest for sharing, but were gathered before such an option was common in informed consent statements. These are often called legacy projects and an increasing number of IRBs have policies to guide data sharing in these cases.

Even when consent is granted and the sharing and re-use of human-subjects data is possible, further ethical concerns arise in the re-use process. These include concerns about a lack of engagement with original research participants in subsequent synthesis work, or that new analysis will not represent their original meaning[50]. It is overly simplistic to assume that research subjects would not want their data to be re-used, especially if they care about the research topic and to avoid research fatigue[51]. But research participants and the research community must trust that individuals engaged in secondary data use and analysis will be transparent and respectful. To build such trust, commitments to 'ethical openness' can be made, as described by the International Arctic Science Committee[52]. A leading example is that of ELOKA (the Exchange for Local Observations and Knowledge of the Arctic), a data and knowledge-sharing platform for climate change stakeholders, practitioners, Indigenous

Box 2 | Epistemological approaches in qualitative research

We summarize below three broad epistemological approaches to qualitative research that are most common within modern scientific inquiry: positivism (or objectivism), constructivism, and subjectivism. An indigenous or traditional knowledge epistemological frame that falls outside of the scientific paradigm is also increasingly highlighted by many scholars working in these settings. The following definitions are derived from refs. [39,70].

Positivism. Data arise from systematic inquiry (classical scientific process) into the underlying and immutable true (objective) nature of reality. Positivism largely embraces notions of any data sharing, focuses on reliability and accuracy of discrete measurements, and uses qualitative data alongside other data to identify generalizable patterns and principles.

Constructivism. Data arise from human interpretation (constructing meaning and understanding) of the empirical (material) world. Constructivism largely sees data sharing as requiring extensive documentation of the context within which data were gathered in order to systematically characterize how and why knowledge was constructed in a certain way.

Subjectivism. Data arise from unique, relational processes that emerge only when an individual researcher engages with research subjects and interpretation of information. Subjectivism largely rejects notions of qualitative data sharing for re-use by anyone other than the original researcher (including the same researcher in the future).

Arctic residents, and researchers to learn from and build upon previous work and observations in the region.

Practical challenges. Practical challenges are especially acute for qualitative data. Qualitative researchers and practitioners have fewer options than quantitative researchers for repositories that support the diversity of data types, access restrictions and metadata needs of qualitative and multi-modal data[53]. Data sharing of all types takes many forms, including depositing in well-managed repositories, to supplemental materials in a journal article, or statements of availability upon request. This wide range of approaches has some potential to deliver the benefits outlined above. However, sharing qualitative data in opaque, disjointed or overly burdensome ways undermines the goals of data sharing and ultimately may dissuade researchers from using synthesis approaches.

Though repositories, open data communities and support for research data archiving are growing across institutions, long-term financial resources necessary to maintain the infrastructure remain difficult to secure. In addition, there is a dearth of standards for metadata and documentation to facilitate qualitative data re-use[54], and many data repositories that support open access do not have adequate standards to ensure the appropriate and accurate re-use of qualitative data in future research[54]. New guidelines and processes address idiosyncratic data sharing in an effort to improve the sharing-to-re-use pipeline. For example, the FAIR principles[55]—that data must be Findable, Accessible, Interoperable and Re-usable—are often taken as a starting point to address the challenges associated with any type of data access and re-use. Similarly, the UK Data Archive has created the Qualitative Data Exchange Schema to support the standardization of comprehensive metadata for qualitative data.

Even if all of the above-mentioned challenges can be adequately addressed (and many are currently receiving much attention), increasing qualitative data sharing is in the end a collective action problem. The commons and public goods are vulnerable to social dilemmas that produce surmountable barriers to collective action. However, in the case of the knowledge commons (for example, sharing and archiving qualitative data and data sets), collective action challenges are different. Actors must work together to manage the resource effectively, and also to create it[17]. The lack of defined community boundaries makes it difficult to exclude users, which increases 'free-riding' (using but never contributing to the knowledge commons) and decreases incentives to those who might otherwise contribute their data. As a result, the main governance goal for the knowledge commons is to encourage contributions of new data and their use (for example, new analyses), and improvements to existing data and metadata, which ultimately enhance the overall value of the knowledge evidence[56].

Enabling qualitative data sharing

Realizing the benefits of sharing qualitative data for re-use requires commitment, support and coordination from an array of actors and institutions. Below, we offer a framework for qualitative data sharing that addresses some of the challenges outlined by considering both data access and data processing levels. We then present an agenda for progress, drawing attention to clear and tangible actions for: (1) researchers; (2) research institutions; (3) funders; (4) data repositories; and (5) journals and publishers to accelerate qualitative data sharing for re-use. Intentionally broad, the agenda is relevant for all disciplines, fields and topics that can benefit from and contribute to socio–environmental synthesis and the qualitative evidence base. While we discuss the five actors and institutions as discrete entities, they are part of a system, and reforms in any one agent will have implications for others. For example, funder mandates for data archiving may be in tension with IRB insistence on the protection of human subjects.

Qualitative data access and processing levels. Open data, which are "made available without restriction, on a non-discriminatory basis, for no more than the cost of reproduction and distribution"[57], represent a major component of contemporary data sharing. Data from human-subjects research, however, often involve confidentiality agreements with participants, and many qualitative researchers take an epistemological approach that precludes non-contextual re-analysis or interpretation of data. Table 2 presents a framework (adapted from ref. [58]) for addressing some of the practical, ethical and epistemological challenges associated with qualitative data sharing for re-use.

Our framework combines levels of data access with data processing levels. Many data repositories offer access level controls to facilitate the ethical re-use of sensitive data, like requirements that anyone who wishes to access and re-use those data must have IRB clearance from their own institutions. In addition to placing limits on how qualitative data can be accessed, the framework presented in Table 2 builds on the tradition from Earth systems science of data processing levels[59]. For qualitative data, these levels move from totally raw data to partially redacted to completely summarized research findings. In many cases, both secondary and primary qualitative data can be processed (redacted or summarized to eliminate sensitive or personal information) and shared with fewer restrictions than raw data, while still maintaining most of the content of interest for future research. For some researchers with a constructive epistemological orientation, providing data at a 'higher' level of processing might mean including extensive metadata and thematic coding of data, rather than the raw interview transcripts. Combining access levels with clear guidelines about the levels of data processing that are acceptable could facilitate increased sharing of certain types of information while alleviating epistemological and ethical concerns about misinterpretation.

Actions for researchers. For researchers generating and sharing qualitative data, there are clear steps that will increase potential re-use by others. Before research begins, researchers must consider the

Table 2 | Framework for qualitative data access and processing levels

Data framework	Definition or example
Level of processing	*Definition*
0 (raw data)	Full data (for example, text, images or audio), with no redaction or aggregation.
1	Full data with redaction for direct identifiers (for example, name and address).
2	Full data with redaction for direct and indirect identifiers (for example, name, geographic information, gender, age and dates of events).
3	Excerpted or partial data with redaction for direct and indirect identifiers. Aggregated or summarized information (for example, initial analysis).
4 (final research product)	Summarized data with no sensitive information—final analysis.
Level of access	*Definition*
Closed	Data deposit exists only for archival purposes (data are not available for any further use).
Controlled	Data are available for further use when a user is approved by the original researcher (case-by-case access of data depending on intent and qualifications).
Restricted	Data are available for further use when a user meets standard criteria (obtaining IRB and accessing data through a restricted environment).
Open	Data are freely available for use with general use agreement and attribution.
Processing–Access	*Example*
0, closed	Photographs that show individuals' faces.
0, controlled	Interview transcripts with names and locations.
1, restricted	Social media data with user name removed but location included.
2, controlled	Internal government documents.
2, open	Ethnographic field notes.
3, restricted	Thematic coding of focus group transcripts.
4, closed	Analytical results embargoed until publication.
4, open	Evaluation reports.

Examples of the types of qualitative data that might be most appropriately deposited with different combinations of processing and access are listed. Data type, research setting, epistemological orientation and ethical requirements all influence how processed data should be in order to be shared, and how narrow the access to those data should be. The examples demonstrate that the relationship between the level of processing and level of access is not linear, in that a higher level of processing does not always correspond to a more open level of access. Note that the framework could also be envisioned as a matrix with one axis concerning levels of processing and the other with levels of access.

potential ethical implications of re-use, and include such considerations in the design and informed consent process, as well as in the development of a data management plan. A critical component of any data-sharing plan is metadata and documentation[60]. Early on, consideration should be given to the levels of processing for data that will be shared and archived (for example, raw versus aggregated or transformed data). Researchers will also need to allocate the necessary resources (for example, time and financial) and build this into their funding proposals and operating budgets to support data management and long-term archiving. Furthermore, they will need to identify options for data storage repositories (for example, university or offsite repository) that are appropriate for qualitative data. Developing and articulating expectations for what constitutes appropriate data-sharing behaviour will encourage participation in the research commons, and will make it easier to hold people accountable for negative behaviours. Sustainability science and qualitative researchers can learn from the many disciplines and communities of practice that have successfully established governance structures for research commons that adhere to agreed-upon ethical standards[61].

Actions for research institutions. Research institutions can facilitate sharing and re-use of qualitative data in three main ways: (1) adapting research ethics policies (for example, those governed by IRBs or science integrity bodies); (2) increasing resources for libraries and data librarians; and (3) establishing appropriate incentives for researchers to share data. IRBs can adopt and establish clear guidelines and policies for informed consent, qualitative data processing and retention in perpetuity, and access options associated with data sharing. Such adjustments should include consideration of data sharing associated with legacy projects, which in some cases may not be feasible due to conditions of prior consent. In this data-intensive research ecosystem, libraries and librarians are quickly becoming the de facto clearinghouse for data management on university campuses. Despite this new role and responsibility, they often lack the necessary resources to support the sharing and re-use of qualitative data. Dedicated positions focused on data management and archiving help to address this gap, as is additional training on information handling and storage for other library personnel[62]. Finally, research institutions can offer appropriate incentives so that researchers receive the necessary recognition for the creation and sharing of data products. These should include considering a data product as equivalent to a research publication for tenure purposes, and supporting graduate research assistants to produce and disseminate data products derived from their research[63].

Actions for funders. Funders of all types can encourage data sharing for re-use by allowing budget lines for data preparation and by recognizing the increased cost that comes with preparing qualitative data and generating adequate metadata for sharing. For example, preparing qualitative data for sharing can be more costly because de-identifying large amounts of interview transcripts requires significant manual labour. In addition, funders could require a review of the existing accessible evidence base to justify research proposing new data collection. This would increase all types of data re-use and ward against research participant fatigue, which is especially common for qualitative methodologies that demand significant time from participants[36].

Finally, funders will be critical for securing long-term financial stability for data repositories[64]. It will be important to consider whether there may be differential impacts of any new requirements, such as on researchers from low- and middle-income countries, and ensuring there are solutions or necessary modifications so all can participate in both sharing and reusing qualitative data.

Actions for data repositories. Data repositories will be essential to accelerating the sharing and re-use of qualitative data. Repositories serve as data brokers, providing a catalogue of their available data, as well as discovery and indexing services to facilitate re-use. Repositories should at a minimum follow FAIR data principles[55], and they can help establish the necessary standards associated with key aspects of qualitative data sharing such as metadata and access levels[65]. Data repositories should continue to provide and expand their training and capacity building within the research community, especially among disciplines where qualitative data sharing is new. In addition, data repositories with costs to deposit could have waivers for researchers from low- and middle-income countries to ensure this is not a barrier to participation in data sharing and re-use. Infrastructure managed by data repositories can also help address some of the challenges outlined above, including options for embargo on data until the release of publications, options for different access levels depending on level of data processing, and the assigning of digital object identifiers (DOIs) to deposited data sets to allow primary researchers to receive credit for those deposits.

Actions for journals and publishers. While data sharing guidelines and policies are becoming commonplace in journals, there are few policies that specifically address some unique aspects of qualitative data. Researchers interested in sharing qualitative data will likely be more comfortable publishing in journals that have explicit policies allowing multiple access levels and types of data processing. In addition, journals should require a data accessibility statement prior to publication, and allow researchers to explain the rationales for access restrictions, processing and storage location for qualitative data (for one example, see the *PLOS ONE* data availability guidelines). Journals that publish qualitative or multi-method research should ensure that their policies cover qualitative data by including both clear guidance on the information expected to be shared, and robust exceptions where ethical concerns or cultural considerations preclude data sharing. Finally, journals can help incentivize and document authors' commitment to open data through the use of 'badges' or other forms of certification based on author actions (like making data available, useable and so on)[66].

When appropriate safeguards against epistemological, ethical and practical concerns are taken into account, qualitative data sharing can be beneficial for socio–environmental synthesis. Still, sharing and re-use are currently under-emphasized and under-incentivized. Accelerating socio–environmental synthesis through qualitative data sharing will require considering ethics and data sovereignty, allocating resources for data management and long-term archiving, tailoring policies and guidelines to the attributes of qualitative data, and providing appropriate incentives. At the core, qualitative data sharing and re-use demands commitment, support and coordination from the entire research community.

Received: 18 November 2018; Accepted: 23 October 2019;
Published online: 25 November 2019

References

1. Palmer, M. A. et al. Ecological science and sustainability for the 21st century. *Front. Ecol. Environ.* **3**, 4–11 (2005).
2. Hampton, S. E. et al. Big data and the future of ecology. *Front. Ecol. Environ.* **11**, 156–162 (2013).
3. Romulo, C. L. et al. Global state and potential scope of investments in watershed services for large cities. *Nat. Commun.* **9**, 4375 (2018).
4. Siegel, K. J. et al. Sovereign states in the Caribbean have lower social-ecological vulnerability to coral bleaching than overseas territories. *Proc. R. Soc. B* **286**, 20182365 (2019).
5. Rodrigo, A. et al. Science incubators: synthesis centers and their role in the research ecosystem. *PLoS Biol.* **11**, e1001468 (2013).
6. Palmer, M. A., Kramer, J. G., Boyd, J. & Hawthorne, D. Practices for facilitating interdisciplinary synthetic research: the National Socio-Environmental Synthesis Center (SESYNC). *Curr. Opin. Environ. Sustain.* **19**, 111–122 (2016).
7. Goodwin, J. & Horowitz, R. Introduction: the methodological strengths and dilemmas of qualitative sociology. *Qual. Sociol.* **25**, 33–47 (2002).
8. Hicks, C. C. et al. Engage key social concepts for sustainability: social indicators, both mature and emerging, are underused. *Science* **352**, 38–40 (2016).
9. Bennett, N. J. et al. Conservation social science: understanding and integrating human dimensions to improve conservation. *Biol. Conserv.* **205**, 93–108 (2017).
10. Moon, K. et al. Expanding the role of social science in conservation through an engagement with philosophy, methodology, and methods. *Methods Ecol. Evol.* **10**, 294–302 (2019).
11. Narayan, D. *Voices of the Poor: Can Anyone Hear Us?* (Oxford Univ. Press, World Bank, 2000).
12. Narayan, D., Chambers, R., Shah, M. K. & Petesch, P. *Voices of the Poor: Crying out for Change* (Oxford Univ. Press, World Bank, 2000).
13. Narayan, D. & Petesch, P. (eds) *Voices of the Poor: From Many Lands* (Oxford Univ. Press, World Bank, 2002).
14. Moon, K., Brewer, T. D., Januchowski-Hartley, S. R., Adams, V. M. & Blackman, D. A. A guideline to improve qualitative social science publishing in ecology and conservation journals. *Ecol. Soc.* **21**, 17 (2016).
15. Cox, M. A basic guide for empirical environmental social science. *Ecol. Soc.* **20**, 63 (2015).
16. Kukutai, T. & Taylor, J. (eds) *Indigenous Data Sovereignty: Toward an Agenda* Vol. 38 (ANU Press, 2016).
17. Hess, C. & Ostrom, E. in *Understanding Knowledge as a Commons: From Theory to Practice* (eds Hess, C. & Ostrom, E.) 3–26 (MIT Press, 2011).
18. Osmond, B. et al. Changing the way we think about global change research: scaling up in experimental ecosystem science. *Glob. Change Biol.* **10**, 393–407 (2004).
19. Riedlinger, D. & Berkes, F. Contributions of traditional knowledge to understanding climate change in the Canadian Arctic. *Polar Rec.* **37**, 315–328 (2001).
20. Cook, C. N. & Hockings, M. Opportunities for improving the rigor of management effectiveness evaluations in protected areas. *Conserv. Lett.* **4**, 372–382 (2011).
21. Fielding, N. G. Triangulation and mixed methods designs: data integration with new research technologies. *J. Mixed Methods Res.* **6**, 124–136 (2012).
22. Peek, L., Fothergill, A., Pardee, J. W. & Weber, L. Studying displacement: new networks, lessons learned. *Sociol. Inq.* **84**, 354–359 (2014).
23. McClenachan, L. Documenting loss of large trophy fish from the Florida Keys with historical photographs. *Conserv. Biol.* **23**, 636–643 (2009).
24. Poteete, A. & Ostrom, E. Bridging the qualitative-quantitative divide: strategies for building large-N databases based on qualitative research. In *101st Annual Meeting of the American Political Science Association* W05I-27, 1–53 (American Political Science Association, 2005); http://hdl.handle.net/10535/5890
25. Magliocca, N. R. et al. Closing global knowledge gaps: producing generalized knowledge from case studies of social-ecological systems. *Glob. Environ. Change* **50**, 1–14 (2018).
26. Janssen, M. A. & Ostrom, E. Empirically based, agent-based models. *Ecol. Soc.* **11**, 37 (2006).
27. Charnley, S. et al. Evaluating the best available social science for natural resource management decision-making. *Environ. Sci. Policy* **73**, 80–88 (2017).
28. Game, E. T. et al. Cross-discipline evidence principles for sustainability policy. *Nat. Sustain.* **1**, 452–454 (2018).
29. Bennett, N. J. Using perceptions as evidence to improve conservation and environmental management. *Conserv. Biol.* **30**, 582–592 (2016).
30. Wyborn, C. & Bixler, R. P. Collaboration and nested environmental governance: scale dependence, scale framing, and cross-scale interactions in collaborative conservation. *J. Environ. Manag.* **123**, 58–67 (2013).
31. Cash, D. W. et al. Scale and cross-scale dynamics: governance and information in a multilevel world. *Ecol. Soc.* **11**, 8 (2006).
32. Ban, N. C. et al. Social and ecological effectiveness of large marine protected areas. *Glob. Environ. Change* **43**, 82–91 (2017).
33. Lindkvist, E., Basurto, X. & Schlüter, M. Micro-level explanations for emergent patterns of self-governance arrangements in small-scale fisheries—a modeling approach. *PloS ONE* **12**, e0175532 (2017).
34. Gill, D. A. et al. Capacity shortfalls hinder the performance of marine protected areas globally. *Nature* **543**, 665–671 (2017).
35. Biedenweg, K., Harguth, H. & Stiles, K. The science and politics of human well-being: a case study in cocreating indicators for Puget Sound restoration. *Ecol. Soc.* **22**, 11 (2018).

36. Hartter, J., Ryan, S. J., MacKenzie, C. A., Parker, J. N. & Strasser, C. A. Spatially explicit data: stewardship and ethical challenges in science. *PLoS Biol.* **11**, e1001634 (2013).
37. Fernández-Llamazares, Á. & Cabeza, M. Rediscovering the potential of indigenous storytelling for conservation practice. *Conserv. Lett.* **11**, e12398 (2017).
38. Tengö, M., Brondizio, E. S., Elmqvist, T., Malmer, P. & Spierenburg, M. Connecting diverse knowledge systems for enhanced ecosystem governance: the multiple evidence base approach. *Ambio* **43**, 579–591 (2014).
39. Maxwell, J. A. in *Epistemologies for Qualitative Research* (eds Soini, H., Kronqvist, E.-L. & Huber, G. L.) 10–27 (Center for Qualitative Psychology, 2011).
40. Hammersley, M. Qualitative data archiving: some reflections on its prospects and problems. *Sociology* **31**, 131–142 (1997).
41. Denzin, N. K. & Lincoln, Y. S. *The Landscape of Qualitative Research* (Sage, 2008).
42. Nightingale, A. J. Adaptive scholarship and situated knowledges? Hybrid methodologies and plural epistemologies in climate change adaptation research. *Area* **48**, 41–47 (2016).
43. Mahajan, S. L. et al. Systems thinking for planning and evaluating conservation interventions. *Conserv. Sci. Pract.* **1**, e44 (2019).
44. Reyes, V. Three models of transparency in ethnographic research: naming places, naming people, and sharing data. *Ethnography* **19**, 204–226 (2018).
45. Broom, A., Cheshire, L. & Emmison, M. Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology* **43**, 1163–1180 (2009).
46. Barnett-Page, E. & Thomas, J. Methods for the synthesis of qualitative research: a critical review. *BMC Med. Res. Methodol.* **9**, 59 (2009).
47. Hartsock, N. in *The Socialist Feminist Project: A Contemporary Reader in Theory and Politics* (ed. Holmstrom, N.) 350–359 (Monthly Review Press, 2002).
48. Lupia, A. & Elman, C. Openness in political science: data access and research transparency. *PS Political Sci. Politics* **47**, 19–42 (2014).
49. DuBois, J. M., Strais, M. & Walsh, H. Is it time to share qualitative research data? *Qual. Psych.* **5**, 380–393 (2018).
50. Bishop, L. Ethical sharing and re-use of qualitative data. *Austral. J. Soc. Issues* **44**, 255–272 (2009).
51. Turner, D. Archiving qualitative data: will secondary analysis become the norm? *Quirkos* (24 November 2016); https://go.nature.com/32U7hH7
52. *Statement of Principles and Practices for Arctic Data Management* (International Arctic Science Committee, 2013); https://go.nature.com/32KBpEG
53. Bishop, L. & Kuula-Luumi, A. Revisiting qualitative data re-Use: a decade on. *SAGE Open* **7**, 1–15 (2017).
54. Hoyle, L. et al. *A qualitative data model for DDI* Data Documentation Initiative Working Paper No. 5 (Data Documentation Alliance, 2013); https://go.nature.com/2MJGFmw
55. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
56. Schofield, P. N. et al. Post-publication sharing of data and tools. *Nature* **461**, 171–173 (2009).
57. National Research Council *On the Full and Open Exchange of Scientific Data* (The National Academies Press, 1995); https://go.nature.com/32MBY0I
58. Jones, K. et al. *Qualitative Data Sharing and Re-use for Socio-environmental Systems Research: A Synthesis of Opportunities, Challenges, Resources and Approaches* SESYNC White Paper (SESYNC, 2018); https://doi.org/10.13016/M2WH2DG59
59. Savtchenko, A. et al. Terra and Aqua MODIS products available from NASA GES DAAC. *Adv. Space Res.* **34**, 710–714 (2004).
60. Elman, C. & Kapiszewski, D. Data access and research transparency in the qualitative tradition. *PS Political Sci. Politics* **47**, 43–47 (2014).
61. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286–290 (2003).
62. Mannheimer, S., Pienta, A., Kirilova, D., Elman, C. & Wutich, A. Qualitative data sharing: data repositories and academic libraries as key partners in addressing challenges. *Am. Behav. Sci.* **63**, 643–664 (2019).
63. Miguel, E. et al. Promoting transparency in social science research. *Science* **343**, 30–31 (2014).
64. Lynch, C. How do your data grow? *Nature* **455**, 28–29 (2008).
65. Karcher, S., Kirilova, D. & Weber, N. Beyond the matrix: repository services for qualitative data. *IFLA J.* **42**, 292–302 (2016).
66. Kidwell, M. C. et al. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Biol.* **14**, e1002456 (2016).
67. Denzin, N. K. & Lincoln, Y. S. *The Landscape of Qualitative Research* (SAGE, 2008).
68. Corti, L. Recent development in archiving social research. *Int. J. Soc. Res. Methodol.* **15**, 281–290 (2012).
69. Boulton, D. & Hammersley, M. in *Data collection and Analysis* 2nd edn (eds Sapsford, R. & Jupp, V.) Ch. 10 (SAGE Research Methods, 2006).
70. Moon, K. & Blackman, D. A guide to understanding social science research for natural scientists. *Conserv. Biol.* **28**, 1167–1177 (2014).

Acknowledgements

S.M.A., K.J., R.D.H., N.M. and H.R. were supported by the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1052875. S.M.A. was also supported by the Social Sciences and Humanities Research Council of Canada. S.K. and N.W. were supported, in part, by an Alfred P. Sloan Grant (2018-11217). J.A.S. was supported by funding from the National Science Foundation (BCS-1042888; BCS-0746528; BCS-1413999) and the National Aeronautics and Space Administration (NNX13AB72G).

Author contributions

S.M.A. and K.J. designed and performed the research, led the collaborative work and the writing of the manuscript. All other authors contributed to the research, analysis and writing based upon their specific expertise, and they are listed alphabetically in the author list following N.J.B.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to S.M.A. or K.J.