ARCHIVING & SHARING QUALITATIVE DATA: IMPLICATIONS FOR DATA MANAGEMENT PLATFORMS

A REPORT PREPARED FOR THE CADRE PROJECT

JULIE MCLEOD, KATE O'CONNOR, NICOLE DAVIS

CADRE

Archiving & Sharing Qualitative Data: Implications for Data Management Platforms

A report prepared for the CADRE Project

Julie McLeod, Kate O'Connor, Nicole Davis

Studies of Childhood, Education & Youth (SOCEY)
Faculty of Education, University of Melbourne, and School of Education, La Trobe University

Archiving & Sharing Qualitative Data: Implications for Data Management Platforms

Publisher: University of Melbourne

Version: 1.0
Date: 29 May 2024
DOI: 10.60836/pqbc-jw49
Project DOI: 10.47486/PL106

Executive Summary

Sharing and reusing research data has long been recognised to have significant potential public benefit, but this is challenged by the difficulty of locating and accessing datasets. The Coordinated Access to Data, Research, and Environments (CADRE) project, led by the Australian Research Data Commons (ARDC), aims to address this issue through the creation of a shared access management platform which operationalises the widely used 'Five Safes' model (Desai et al. 2016) of sensitive data governance (CADRE n.d.a, n.d.b, n.d.c).

This report explores the implications of the CADRE platform for qualitative data, an area which receives far less attention in data sharing research, frameworks and infrastructure development but also offers significant potential for reuse and methodological experimentation. The report unpacks the challenges and opportunities CADRE raises for archiving and sharing qualitative research data, drawing on a study design comprising: (1) engagement with prominent debates in the research literature; (2) focus groups undertaken to ascertain researcher and community views on archiving and sharing qualitative data; and (3) reflections on the authors' experiences of archiving qualitative research projects. It considers epistemological and ethical questions concerning the merits and challenges of archiving qualitative data; concerns relating to infrastructure, governance, and data ownership; and the practical and technical issues researchers confront in archiving their own projects and accessing others, pointing towards four primary findings.

First, although archiving and sharing qualitative research data is acknowledged as ethically and epistemologically complex, researchers in the qualitative social sciences are not opposed to this in principle and see value in carefully thinking through how this can be done well.

Second, the Five Safes is a useful framework for supporting the archiving and sharing of qualitative research data but on its own it situates decisions in a risk-oriented framework which leaves less scope to consider data utility and value. Indigenous data sovereignty frameworks such as CARE that foreground collective benefit, authority to control, responsibility and ethics are important complements to the Five Safes model as these support data access beyond the academy and ensure balance across different ethical dimensions. Researchers in the qualitative social sciences express strong support for the need for better research governance and infrastructure to enable responsible data sharing. But this in itself is not sufficient to resolve challenges in supporting access to research beyond the academy.

Third, resourcing is a critical issue and, despite in-principle support for the idea of qualitative data archiving and sharing, there is far less willingness from researchers working in the qualitative social sciences to engage in the amount of work and funding required to do this well. This points to the need for not only better research funding but also the importance of more extensive comprehensive protocols and practical guidelines for archiving and sharing qualitative research.

Finally, the benefits of the CADRE platform for qualitative research are restricted without the inclusion of more extensive and appropriate metadata to aid researchers trying to identify relevant datasets. While CADRE is primarily a platform to manage, access the minimal provision of metadata is particularly consequential for qualitative datasets. The relatively small size and targeted focus of many qualitative data sets means that without detailed metadata researchers will face difficulties finding and using the datasets they need, inhibiting the platform's intent to increase access and reuse. This is compounded by the relatively modest number of Australian archived qualitative datasets available and accessible for re-use. Further, decisions about access to qualitative datasets are not neatly resolved at a single point in time and may require revisiting, based on a principled-base process to support ongoing custodianship.

Table of Contents

Introduction

Sharing and reusing research data has long been recognised as having significant potential public benefit (Arthur & Hearn 2021; Arthur et al. 2021). However, interest in this has intensified in recent decades, with a rapidly changing environment influenced by the drive for open access research and data, exponential increase in the bulk digitisation of data as well as born-digital data, massive recent developments in the potential applications of Artificial Intelligence (AI), and changing ideas about how data should be governed and to whom it should be made available.

These practices are linked to principles of 'open research' or 'open scholarship' and the idea that such data should be FAIR – Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016; ARDC, n.d.a). Related principles are expressed in frameworks for the stewardship of and access to data, including the CARE (Collective Benefits, Authority to Control, Responsibility, Ethics) Principles for Indigenous Data Governance (GIDA n.d.; ARDC n.d.a; Carroll et al. 2021). As Arthur et al. (2021, 796) argue,

> Open scholarship offers a new and strategic way for universities to bridge the gap between makers and users of research – that is, the 'elite' academic world and civil society – through increased knowledge exchange and public accountability.

The difficulty of locating and accessing data in government, institutional, and other repositories has been identified as a major barrier to utilising data for public benefit and maximising the impact of research funding. At the same time, researchers, data custodians, and the people about whom this data is collected or generated, must have assurance that protocols are in place for their data to be shielded against misuse, with appropriate and robust forms of data governance in place. Further, the terms 'data sharing' might cover a wide range of data availability, including open access data, data with public or researcher access mediated by a custodian, or data to which access is heavily restricted to a certain organisation or section thereof (Arthur & Hearn 2021).

The demand for greater access to various types of data has led to increasing attention to regulatory issues, such as new legislation, platforms and frameworks. These are viewed as facilitating responsible access to data and ensuring the security of any sensitive data being shared by providing protections for the individuals and groups about whom that data is collected.

CADRE (Coordinated Access to Data, Research, and Environments) (CADRE, n.d.a), led by the Australian Research Data Commons (ARDC), is a project designed to address some of these concerns. Its purpose is to improve access for the social sciences and related disciplines to research data defined as sensitive and held in Australian research institutions and collaborating government and private sector agencies (CADRE, n.d.a). It aims to 'enable data owners and users to address the core concerns around governance, creation, management and sharing of sensitive data for research' (McEachern [et al.] 2021, 5) through a shared access management platform, the CADRE platform, which 'will coordinate necessary information to make decisions about access to sensitive data' (CADRE 2024a).

The platform operationalises an access framework building on a long-established model for sensitive data sharing that has been developed over the past two decades. Known as the 'Five Safes' (Desai et al. 2016), this model has been the basis of data access and sharing protocols for numerous organisations globally, including several Australian state and federal government legislative initiatives focused on data sharing (e.g., the Australian *Data Availability and Transparency [DAT] Act 2022* and DATA [Data Availability and Transparency Act] Scheme resulting from it (ONDC 2024); and the South Australian *Public Sector (Data Sharing) Act 2016* (McLeod et al. 2023; McEachern [et. al.] 2021, 2022, 2024); and data sharing protocols for several Federal government departments (ABS 2021, AIHW 2022).

The authors of this report lead the SOCEY project (Studies of Childhood, Education & Youth) (SOCEY n.d.), with team members based at the University of Melbourne and La Trobe University.

Overall, SOCEY promotes qualitative research archiving (McLeod et al. 2020), through academic outputs and infrastructure research and development, including a dedicated qualitative archive pilot project, located within the ADA Dataverse platform (SOCEY 2019–20).

As one of the project partners for the wider CADRE project, we have compiled this report to explore the implications of CADRE specifically for *qualitative* data and research, including the applicability of the Five Safes model. The report is informed by our participation in various CADRE work programs (see Appendix 1); it is also offered as a guidance or thinkpiece for those intending to access or share qualitative data, whether or not they are CADRE users, and canvasses considerations that might play into decision-making about this process.

The report begins with an overview of the CADRE project and the design of our study to investigate implications of the platform for qualitative data. This is then followed by discussion of, first, epistemological and ethical questions and second practical and technical issues which in combination inform our main findings.

1. The CADRE Project and Qualitative Data

The CADRE project aimed to address some of the concerns regarding data sharing through developing a CADRE Five Safes Framework, the CADRE platform, which has a linked training and engagement program. Here we consider how these project elements have accounted for qualitative data and where further work might be beneficial.

CADRE Five Safes Framework

The *CADRE Five Safes Framework* (McEachern 2021, 2022, 2024) was developed to underpin the platform's data access and release management and protocols. As noted, the CADRE platform operationalises the Five Safes Framework which is widely utilised both in Australia and internationally for data release protocols. The Five Safes are conceived of as linked and interoperable dimensions of responsible access protocols and comprise:

- Safe projects: Is the use of the data appropriate [or, does it produce public benefit]?
- Safe people: Can the researcher be trusted to use it in an appropriate manner?
- Safe data: Is there a disclosure risk in the data itself?
- Safe settings: Does the access facility limit unauthorised data?
- Safe outputs: Are the statistical results non-disclosive? (McEachern et al. 2021, 5)

As a result of discussions among CADRE work program groups, two further safes were added to the original five safes framework:

- Safe groups: does this research group (or team) all meet requirements for management of data sharing activities?
- Safe organisations: does the organisation have appropriate data management and data governance policies and practices for researcher conduct?

This adapted model directly influenced the CADRE platform design, as well as the format of compulsory training modules required for data applicants/users.
As noted in the *Five Safes Framework*, it is important to consider the specific characteristics of (sensitive[1]) qualitative data in developing access protocols and assessing the appropriateness of the Five Safes

---

[1] ARDC (n.d.b), from whence CADRE draws its definitions and approaches to 'sensitive' data notes this type of data is 'often confidential, … [and] can include identifiable personal and health/medical data, Indigenous data, ecological data, and commercial-in-confidence data'. ARDC defines sensitive data as any 'that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the

model. While the Five Safes can assist in decision-making regarding access to and the release of qualitative data, it cannot on its own make this type of data completely safe. Following Kirilova and Karcher (2017) and others, we argue that 'it is difficult to render qualitative data and its outputs' (McEachern [et al. 2021], 54) completely 'safe', when the richness of the data, including contextual information, is vital to the purpose and utility of this data.

For qualitative research, ensuring the safety of data may not necessarily require it to be fully de-identified and anonymised. Rather, data safety depends on the interoperation of all aspects of the Five Safes. This dovetails with established human ethics processes in the humanities and social sciences, which can include multilevel and multi-stage consent forms for participants to complete over the course of a project, as well as access controls for archived data (McEachern [et al. 2021], 54; McLeod & O'Connor 2021).

Platform design

In operationalising the Five Safes Framework, the platform has taken a multi-pronged approach, providing an access point for all data users to search, request, approve and access released datasets. It utilises 'the Resource Entitlement Management System (REMS) for managing access rights to resources such as research datasets and virtual environments' (CADRE n.d.b). REMS allows researchers to 'apply for access to research data, and the owners of the rights to the materials can process the received applications and manage access rights' (Elixir n.d.).

Users[2] access the platform and associated training via CILogon, an industry standard identity and access management platform widely used in research institutions (CILogon n.d.; CADRE n.d.b). This is done with either their institutional Single Sign On (SSO) (the preferred method) or via ORCiD (Open Researcher and Contributor ID), a persistent digital identifier, through which researcher outputs and profile (e.g., education, 'affiliations, grants, publications, peer review, and' research linkages can be viewed and ascertained (ORCiD n.d.).[3]

Users can then access datasets from participating repositories included on the CADRE platform.

> Access to CADRE will be coordinated from an existing source, such as the Australian Data Archive (ADA) and the necessary evaluations on whether it is the appropriate data for the project will be made by the applicant in the first instance. (CADRE 2024a, 8)

Only the basic information about datasets is contained on the CADRE platform, with minimal metadata. This is because CADRE is a platform for *accessing data and approving its release*, rather than a mechanism for supporting decisions about whether the datasets are suitable for applicants' projects. Decisions about the suitability of a dataset need to be made by the researchers before coming to CADRE, based on the protocols of whichever archival repository holds the full information (e.g.,

---

research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions' (ARDC 2022, 6). Within this realm, we acknowledge there is a continuum of types of data, both qualitative and quantitative, and some datasets or subsets can be more, or less, sensitive than others. Even so, the use of this nomenclature of 'sensitive' for data accessed through CADRE underlines our point that the CADRE framework is couched in terms of safety and risk. We acknowledge that the designation of 'sensitive' for all data accessed through the platform immediately positions this data as inherently risky, potentially foreclosing other considerations and aspects of data access and use, even though the Five Safes method, data classification plans (e.g., UNSW 2021, n.d.; University of Melbourne, n.d.) and other frameworks for approaching data governance (e.g., ARDC 2021) may allow for assessing levels of sensitivity.

[2] 'Users' here refer to anyone using the platform. See CADRE user guides (2024a, 2024b) for the formal definitions of various roles on the platform. A Data Applicant is a user that wishes to access datasets, while 'an approved data applicant is someone who has received approval to access the requested data' (CADRE 2024a).

[3] In 2023 two points of discussion among the CADRE team were that government departments do not usually use CILogon but authenticate through a similar system, VANGuard (Department of Industry, Science & Resources, n.d.); also, government and NGO or non-higher education researchers have less uptake of ORCiD. The project is currently contemplating implementing the former for authentication and access, in addition to CILogon, and encouraging (or possibly requiring) all researchers to establish an ORCiD.

metadata and descriptive information) about the dataset. Datasets will be searchable via keywords but only select datasets are visible to all users.

Once a user identifies a dataset they wish to access, they submit a Data Sharing Request (DSR) and provide relevant information about themselves and the way they will use the data in a single online form (with customisable fields to meet the needs of those making decisions about the release of data [i.e., Data Custodians or Deciders]). Using institutional credentials and/or persistent identifiers associated with a research record allows those making decisions about data access to identify parameters of the Safe People dimension of the Five Safes, establish the Data Applicant's credentials, including their membership of a 'safe organisation' and whether they have a record of relevant education and research experience.

A Data Applicant's form is then submitted to a Handler who processes the DSR. The dataset Decisionmakers can then assess the Data Applicant's suitability, and that of any Collaborators, to access and use the data. The information enables them to:

- establish Data Applicants' credentials to partly determine a Safe People, Safe Organisations, Safe Groups; and
- gather information on the Project for which the data will be used; including, project summaries, funding sources, ethical clearances, data management plans, personnel, and projected outputs, in order to assist in establishing Safe Projects, Safe Settings, Safe Outputs, Safe Groups.

This assists those reviewing and approving the DSR to ensure requests fit the dimension of Safe Data in light of the other Safes, after which they can then request further information or release the data for use.

For the purposes of this Report, it is important to note that specific access requirements for *qualitative* data are not part of the access protocols themselves. However, reviewing a user's ORCiD profile can provide important information about a Data Applicant's capabilities, such as training and publication record in qualitative research and relevant methods, enabling the platform Decisionmakers to determine the suitability of the data applicant/user for release of data.

Training and engagement

A final component of the work has been the development of one-off training modules that Data Applicants and other users must complete prior to utilising CADRE and accessing the datasets available through this platform (CADRE n.d.d). These include:

- Module 1: Five Safes for Sensitive Data Management
- Module 2: Roles and Responsibilities

The development of these modules was informed by a series of training and module-testing workshops, which engaged CADRE partners as well as external participants from academic institutions, government organisations, and community stakeholders.

The training addresses a central CADRE objective – to fill gaps related to CADRE in research ethics training and data management in universities (and other research institutions or agencies). This includes 'training in the Five Safes, to augment … [ethics and data management training] researchers already get through academic supervision and on-the-job training' (Jones 2022).

Module 1 is broad-ranging and includes training in each of the Five Safes (+ 2). It explains the Safes, outlining details of each dimension and utilising case studies and a variety of assessment waypoints throughout to cement the learners' understanding. This includes explanation of how various dimensions might intersect and the assessment process for potential release of dataset release. Other key areas are also addressed, namely:

- related legislation such as the DATA Scheme which aims to increase the availability and use of Australian government data;
- benefits of and necessity for retaining (archiving), sharing and reusing data;
- ethical uses of data, including that accessed via CADRE; and
- benefits of and obstacles to data archiving, sharing, and reuse.

The second module is briefer, focusing on the roles and responsibilities of data users, including:

- data management and data governance;
- details of relevant legislation;
- how CADRE and the DATA Scheme interoperate; and
- summaries of user roles and responsibilities.

As with the first module, case studies are used to convey information via scenarios and users complete assessment waypoints throughout.

International best practice recommends that access to and release of qualitative data requires data users (as well as the original researchers) to undertake training in data management and ethics so that the requested material is as safe as possible (QDR n.d.c, n.d.d) According to Kirilova and Karcher (2017, 5), it is necessary to educate 'researchers how to be "safe people" and how to plan for "safe projects" – when accessing such data and using them for secondary analysis – and providing long-term "safe settings" for the data, including via de-identification and appropriate access controls and protocols.

There are several advantages to the CADRE system for both data applicants/users and data custodians which can potentially assist in streamlining access requests and release. While initially the requirement for module training and application protocols for the data user may seem time consuming, training is a one-off requirement and stored for any future requests. Data Applicants and Collaborators can also update their details as appropriate for specific datasets Information about the Data Applicant, their current application, and past data requests/usage can immediately be seen by those making decisions about the requests, facilitating the process of determining access. Additionally, the system will:

- show progress updates in real time;
- allow data custodians through one platform to see applicant information, request more details, and approve or deny access;
- allow data applicant/user to respond to/see the above stages; and
- access and download approved data.

Qualitative study design

This report unpacks the challenges and opportunities CADRE raises for archiving and sharing qualitative research data. It draws on four strands of work we undertook, which focused on (1) researcher and community views on archiving and sharing qualitative data and the work of the CADRE project; (2) literature reviews of prominent debates; and (3) reflections on the authors' experiences of archiving qualitative research projects; and (4) engaging with the technical affordances and limitations of the CADRE platform.

In relation to the first strand, while much has been written about the possibilities for and barriers to researchers (or data creators) archiving and sharing their own data, less attention has been given to barriers for potential end-users of such data. This is an issue acknowledged in data sharing initiatives, including CADRE. Yet questions regarding how, at a granular and practical level, researchers perceive, search for, request and experience the process of accessing and reusing qualitative datasets, particularly datasets that may have mediated access, remains under-researched.

To address this, we held interviews and focus groups with qualitative researchers to explore their interest in and capacity to access archived research data for reuse in their work as well as gauge their interest in archiving their own qualitative data sets. Insights from these discussions provide valuable perspectives on the lived experience and perceptions of researchers and end-users, as well as providing a contribution to building a more substantial evidence base for understanding researcher experience of data archiving and reuse.

In total four group and individual interviews were held with 16 qualitative researchers, with sessions ranging from one to seven participants. Most participants (14) were from the university sector, with two participants from community organisations that work with children and young people. Several of the academic participants regularly work with government, NGOs and community groups in research projects and in this respect their responses are also relevant to work in such sectors. Participants included Higher Degree Research Students, academics from early career to senior levels, and some had (or previously had) careers in government and community organisations.

While this is a relatively small number of participants, many are at high levels in their organisations and have long-term academic careers, with sound understanding of qualitative data research methods and innovation. Further, our aim was to shed light on current views among a mix of researchers rather than undertake a large-scale comprehensive study. Additionally, the focus of these sessions was informed by findings from previous workshops held by SOCEY members in 2018 and 2019, which included academic (2018), non-government, community, and local, state, and federal government (2019) participants interested in or practitioners of qualitative research methods and data.

Participant recruitment was targeted, with most being invited due to their use of qualitative data methods in their research and their engagement with or interest in the potential for data archiving and sharing. Most participants had some knowledge of or experience with archiving qualitative data sets or reusing archived qualitative data in their own research. Others were considering archival options for future research projects. Some participants (or their organisations) had previously been involved in the 2018 and 2019 SOCEY events.

We adopted a semi-structured form of interviewing, with participants encouraged to provide their own input and speak freely to the issues raised, guided by a set of discussion prompts developed and led by the researchers. These concerned:

- participants' work, projects and experiences with archiving, sharing, accessing, re-using qualitative data;
- their views on the benefits, barriers and technical requirements for archiving and re-using qualitative data;
- their views on the relevance and implications of the Five Safes Framework; and
- their views on the CADRE platform and its potential implications.

Prior to the discussion, participants were provided with an introduction to the CADRE project and the earlier work of the SOCEY team as well as given background reading material explaining this in further detail.

Before commencing our work on the CADRE project, we had previously established the SOCEY Repository in 2019-2020, a dedicated archive for qualitative studies on childhood, education and youth housed at the ADA, and archived six qualitative data sets and projects within this repository supported by funding from the ARDC. This experience is discussed in detail in our discussion paper, *Doing*

Research Differently (McLeod et al. 2020). As part of our study design, we revisited these findings in the context of the CADRE project and platform.

We also reflected on our attempts to recruit further projects for the SOCEY Repository as part of the current CADRE project (2020–2024), considering the challenges of doing this in light of what they suggest regarding the nature of researcher interest and reluctance in data archiving. Additionally, we conducted detailed literature reviews focused on community archiving, qualitative research practices and data governance debates and undertook testing and training within the CADRE platform.

2. Epistemological and Ethical Barriers to Archiving and Sharing Qualitative Research Data

There is a growing body of research exploring how to make qualitative data more open that has gathered pace in the past 15 years, including in Australia. Over the last two decades, researchers have examined the possibilities and drawbacks, barriers to and complexities of qualitative data sharing and reuse, addressing ethical, epistemological and methodological challenges and affordances (Hammersley 1997; Mauthner et al. 1998; Corti et al. 2000; Parry & Mauthner 2004; Bishop 2005; Cheshire et al. 2009; Bishop 2009; Elman & Kapiszewski 2013; Childs et al. 2014; Thompson & McLeod 2015; Borgerud & Borglund 2020; McLeod & O'Connor 2021; Arthur & Hearn 2021; Arthur et al 2021; Moore et al. 2021 to name just a selection. Further discussion and details can be found in McLeod et al. 2020).

Qualitative data is diverse in type and rich in scope, with understanding the context of its creation essential to how it is interpreted. The very nature of qualitative data and methodologies, and their capacity to elicit contextual detail and illuminate the specificity of experiences and subjective perspective is crucial to its epistemological and knowledge making contributions. The benefits of making such data available, whether through fully open or mediated access, include providing greater transparency in research processes and methods and generating opportunities for methodological innovation and experimentation (McLeod & O'Connor 2021; Moore et al. 2021).

Making research data available can also provide a rich ground for future research that might reuse past datasets to gain new insight into that past or the methodologies employed. Such data might also be re-examined from new contexts, bringing fresh questions and perspectives that might help reveal novel and unexpected insights into the data, the study that collected it, or other aspects of life at the time it was conducted (McLeod et al. 2020; Savage 2010; Andrews 2011; Childs et al. 2014; Borgerud & Borglund 2020). There are also potential social benefits in making data available, particularly to the communities or groups that were part of the original data collection. As Moore et al. (2021) have argued, opening up qualitative research data to broader publics can assist in collaborative endeavours that can bring new insights and open up further possibilities for community engagement through use of that data.

Documented ethical challenges in sharing qualitative research data include: ensuring the security of data (and the personal information of individuals and groups); managing potential data misuse; resistance to sharing data by researchers; the ease with which potential users can locate and access datasets; disconnects between the actual experience of trying to archive and re-use data and policy precepts; uncertainty of funding research and infrastructure to maintain data security; and the time and resource intensive nature of digital archiving (Mauthner & Parry 2013; Barbera 2013; McLeod et al. 2020; Borgerud & Borglund 2020; Hostler 2023. Arthur et al. 2021, Table 1, summarise some of the many barriers). These challenges might be heightened by the type of data being shared; for example, when there are cultural or community responsibilities or highly personal information and sensitivities at risk.

The very richness and complexity of qualitative data renders it both highly sensitive and difficult to fully de-identify in such a way that guarantees participants or communities are not traceable through any contextual material, even when names of individuals, businesses, suburbs etc have been deidentified. The potential sensitivity of data and the need to protect participants has been seen as a major barrier in the past to sharing such data (Cheshire et al. 2009; Moore et al. 2021; Joyce et al. 2022). Yet, one of the valuable aspects of qualitative data is precisely its rich contextual detail; stripping detail from this data so that it renders individuals, communities and groups and their private information fully unidentifiable risks reducing the usefulness of the data (Hammersley 1997; Mauthner et al. 1998; Broom et al. 2009; Andrews 2011; Tsai et al. 2016; McLeod & O'Connor 2021; Joyce et al. 2023).

Finally, the limited socialisation of the principles and practicalities of data sharing, and documentation of its benefits for the social sciences are inhibiting factors for data archiving and sharing and share research material. Older paradigms of data re-use have been dominated by risk management and risk averse approaches to protection of data and participants. Other factors influencing researcher reluctance can include well-intentioned advice from mentors to either not share or proforma destroy interview and other qualitative materials after a set period of time, and lack of understanding of how data might be ethically shared (Tsai et al. 2016; Borgerud & Borglund 2020; McLeod et al. 2020; Davis & Neish 2022; Joyce et al. 2023).

Researcher perceptions of the ethics of data sharing

The discussions held with our focus group participants addressed these larger epistemological and ethical questions, with participants raising the following concerns:

- The importance of documenting purpose and context in the creation of qualitative research data and as part of laying the foundations for future archiving.
- The forms of training needed to understand particular datasets such as interview transcripts.
- Questions of data ownership, including the manner in which archives can and should be managed by repositories and data stewards; and who has the authority to decide on access and re-use.
- Anonymising data and the potential for data to be made re-identifiable, including through future uses of Artificial Intelligence (AI).
- The reification of producing the 'new' in research and how reanalysis of existing data is therefore positioned.
- The personal, emotive and affective dynamics of qualitative data collection and whether that can ever be captured in an archive.
- Community engagement with archives and whether data archiving and sharing facilitates that engagement or allows researchers to bypass such engagement.
- Whether some participant groups might be more likely to be put off by the prospect of their data being archived and the implications of this for research practice and qualitative datasets.
- Risks associated with opening data up for public scrutiny and misinterpretation and the potential for data to be used in ways that might be harmful to the participants or in terms of the project itself, as well as unintended harms and risks.

Many of these risks were seen as particularly heightened for research with marginalised and vulnerable communities. One participant spoke about their experience working on a large ARC-funded research project, which focused on the experiences of underage refugees and migrants. For that project, the team made the decision to archive and share project survey data that could be easily anonymised but not the project interviews, due to concerns about potential use of the data for political purposes in the context of concerns about anti-refugee sentiment in the wider community.

Another participant talked about their research with marginalised girls and problems of how these participants would feel about sharing this research with governments, given recent actions such as the application of computer-automated debt recovery to Centrelink social security clients in the 2010s (the 'Robodebt' scandal) (Victoria Legal Aid, 2023). Other participants had similar concerns and spoke about how strong their positionality and their personal relationships with participants and communities has influenced their own research. One participant commented that they felt uncomfortable with others using their data in ways that might be contrary to the purposes for which it was collected, or where secondary analyses did not share the ethical commitments and community relationships they had worked hard to cultivate.

However, despite these concerns, in general participants were not opposed to archiving and sharing qualitative research data. They recognised that universities and funding bodies were moving towards requiring this and that many of the issues raised by archiving and sharing qualitative research are also present in the broader practice of qualitative research. In one focus group discussing the potential for others to misinterpret shared data, participants noted that this is an issue that extends beyond the use of archived data; for example, disagreements with how project partners have used data in projects that have been worked on collaboratively. In another focus group, one academic talked about the challenges they are currently facing in a very large multi-disciplinary and multi-campus project, seeing this as having some parallels with the issues raised in terms of the challenges of navigating different institutional infrastructures and sharing data with colleagues with very different theoretical and methodological approaches to understanding that data.

Another participant, a historian who regularly uses archives, reflected on the philosophical dimensions of interpreting archived data. They suggested that instead of seeing this as a matter of misrepresenting data that it could be more productively viewed as providing an alternative account, a different reading; these debates are very much part of the practice of history where archival materials are always being reinterpreted. As another participant put it, the issues raised don’t ‘mean that we can’t share qualitative research, it means we need to be careful with the trade-offs”.

Several participants also spoke about the value they saw in continued use of datasets, both in terms of getting the most out of the data collected and ensuring the particular communities, particularly marginalised communities, are not asked to retell their stories over and over again. These comments came up in all sessions, and included the following:

> There’s also a general need to or a desire to reach out and look for research that exists, rather than having to go back to communities that have often already provided this, this information or these stories already. (Interview 1)
>
> One of the things that I’ve been extremely aware of with every ARC project I’ve done is just how much how much more you could do with the material, because you’re lucky if you can get … the book or … a significant lot of analysis of a project down, but you just know how much more there is in it. (Interview 3)
>
> There are obviously some benefits around and like we do a lot of research and … we don't always write about all the things you could write about, and so when you … get young people to spend the time, it also makes sense that you really sort of made the most of that of the data. (Interview 4)
>
> I think it’s a great idea. And I think that, you know, I do a lot of work in schools, and with young people, and it’s really hard to recruit. So … anything that can … make the most of the data that is collected, I think, is a great idea. (Interview 4)

In the same vein, one participant from a community organisation spoke about a large project they have custodianship of, which was not set up for data sharing:

> If I were to do it again, I would definitely get consent from the participants about further use of the research data, because it’s incredibly rich … [and] there’s only so much that I can do with that beautiful dataset … it’s almost, I feel like it’s disrespectful in a way that data wasn’t used as much

> as it really could have done to change policy and programs in the way that we really wanted to, we just didn't have the resources and, as a Not-For-Profit, we never [will]. (Interview 3)

There was also significant support expressed for community ownership of research data and for returning research data to communities, and some of the participants were engaged in this work.

The direction and tenor of these focus group discussions were quite different in tone to a series of earlier discussion held with Australian qualitative social sciences researchers in the late 2000s (e.g., Broom et al., 2009; Cheshire et al., 2009). In commentary on these earlier discussions, attitudes towards data archiving and sharing were reported as typically more negative and focused on the protection of participant and researcher rights (although with important differences between disciplinary traditions noted) (Broom et al., 2009). While many of the same big ethical and epistemological issues were canvassed, in our focus groups these were seen as important conversations to open up, not reasons to foreclose qualitative archives altogether.

Finding 1: Although archiving and sharing qualitative research data is acknowledged as ethically and epistemologically complex, researchers in the qualitative social sciences are not opposed to this in principle and see value in carefully thinking through how this can be done well.

3. Infrastructure, Governance and Data Ownership

Data sharing research, frameworks and infrastructure development have frequently concentrated on quantitative data, often with a focus on big datasets (Borgerud & Borglund 2020; McLeod & O'Connor 2021). Likewise, when data sharing infrastructure (platforms, frameworks, legislation) is developed, 'data' is often considered a homogenous whole, without reference to or consideration of differences between data types (McLeod et al. 2023). Moreover, qualitative data and associated research methodologies have tended to receive less attention. This is particularly so for qualitative data in the humanities, arts and social sciences and humanities (HASS), despite the remit of open research principles to address the breadth of disciplines.

In Australia, there are no dedicated qualitative or mixed-methods data archives, despite attempts to establish a central qualitative data archive (AQuA) in the late 2000s to early 2010s (Cheshire et al. 2009); however, the Australian Data Archive (ADA) does hold a significant amount of qualitative or mixed-methods data. The ADA includes extensive guidance for researchers on how to prepare their data with ethical considerations in mind (ADA 2021) but this guidance does not fully attend to the needs and ethical concerns of qualitative researchers as it tends to, approach qualitative and quantitative data as a homogenous whole (although researchers can contact the archive for advice. ADA n.d.). This is a gap that the SOCEY team SOCEY Repository has partly attended to with the development of the SOCEY Repository, specifically developed for qualitative data, and the accompanying guidelines for deposit.

Access to much qualitative data is and must arguably remain mediated through custodians to ensure both the data and the people at the centre of that data are treated in an ethical and respectful manner (McLeod et al. 2020; McLeod & O'Connor 2021). The development of infrastructure to enable researchers to share data more confidently, with protocols in place to allow for differing levels of access to data and to vet those that request access to that data, is essential to promoting the sharing of all data. However, as argued, given the character of qualitative data, additional different protocols for access and release of archived materials are required. The relative lack of facilities dedicated to the archiving and sharing of qualitative data has been seen as a significant barrier to researcher confidence in more open (though still mediated) scholarship (Mauthner et al 1998; Corti & Thompson 1997; Broom et al. 2009;

Wolski 2011; Childs et al. 2014; QDR n.d.a). This situation has been somewhat alleviated through the creation of dedicated qualitative data databases or repositories (or broader archives in which the archiving of qualitative data has been attended to in the frameworks and protocols that govern them), including Qualidata in the UK (now part of the UK Data Archive), over a dozen similar repositories in Europe (UK Data Service n.d.), and the Qualitative Data Repository (QDR n.d.a) in the US.

Despite these initiatives, infrastructure for qualitative data still lags substantially behind those designed for quantitative data or for repositories that do not distinguish between different data types at all. Such generic repositories, where all types of data are held, may be a way forward for archiving of qualitative data but qualitative researchers will likely not be confident or comfortable to commit to such practices without better guidance and support for preparing data for deposit.

In response to the ethical and epistemological issues canvassed above, researchers have also argued for innovative ethical approaches to qualitative data sharing that might render it safer to release. This includes research data management processes, long utilised in social sciences, such as designing appropriate anonymisation and deidentification protocols, building in greater clarity for informed consent from participants, and fostering greater co-creation of research projects and data by researchers, participants and broader publics (McLeod & O'Connor 202; Mozersky 2020; Reanimating Data n.d.; Moore et al. 2021; Pascoe Leahy 2022; Joyce et al. 2023).

As Moore et al. (2021) note, 'the archiving and reuse of qualitative data appears to threaten established ethical practice in qualitative research' and can present an insurmountable barrier for many researchers. However, they argue, methodological innovation may be necessary to further the uptake of qualitative data archiving and sharing and thus harness its potential to the extent that is being seen with quantitative data. Informed by the practices of community archiving, Moore et al.'s work unsettles traditional dichotomies of 'researcher' and 'participant'. Those that would have in the past been the 'subject' of research are deemed as co-creators of such research and have greater decision-making over both the care of their data as well as the production of outputs that report on that research (Cook 2012; Bastian & Flinn 2020; Joyce et al. 2023; Walter et al 2020). As Joyce et al. (2023) note, in many of these cases, the data sharing itself can become an integral part of the research practice and methodology as archival materials are shared and discussed with co-creators/participants (also, Moore et al. 2021; Reanimating Data n.d.).

These approaches to sharing and archiving qualitative data have been influenced by larger debates about data ownership and stewardship and the rights of people and communities to access and/or control their own data. In particular, this work has critiqued the extractive nature of academic and government research on marginalised people and communities and affirmed the need for ongoing care, curation and stewardship of research data once collected.

In Australia and globally, many of these debates have been led by Indigenous scholars and communities calling for Indigenous Data Sovereignty and Indigenous Data Governance (Kukatai & Taylor, 2016; Walter et al., 2020). One important development has been the drafting of the aforementioned CARE principles which affirm the need for ethical data governance and 'the crucial role of data in advancing Indigenous innovation and self-determination' (GIDA n.d.). These principles include:

- **C**ollective Benefit: 'Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data' (RDAIIDSIG n.d., 2).
- **A**uthority to Control: 'Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered' (RDAIIDSIG n.d., 3).
- **R**esponsibility: 'Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit' (RDAIIDSIG n.d., 4).
- **E**thics: Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem. (RDAIIDSIG n.d., 5).

These were developed to complement the 2016 FAIR principles (data that is Findable, Accessible, Interoperable and Reusable). As Walter et al. (2020) discuss, frameworks like FAIR and Five Safes make no mention of the concerns of Indigenous people regarding their data and are therefore insufficient on their own to address all ethical dimensions related to sharing Indigenous Data (and, by extension, other forms of data). These need to be complemented by frameworks that foreground questions of ethics and collective benefit such as CARE. A similar argument is also made by Rose et al. (2023) in their development of the Indigenous Data Ownership, Custodianship, and Stewardship (InDatOCS) governance model, which they propose in order to clearly delineate between the functions of owner, custodian and steward in the data lifecycle and to critique models of data governance which conflate all roles under the purview of national governments.

While this work specifically addresses the governance of Indigenous Data, its valuable insights have potential application to broader ethical data governance practices for qualitative research. Alongside greater informed consent from research participants, ceding full researcher control over data and co-creating research with those previously considered subjects/objects, these principles and paradigms can arguably open up greater possibilities for the sharing of data and open scholarship. They invite an approach to data sharing not only in terms of safety and risk but also in utility and value to others; this reframing is vital to supporting access beyond the academy and ensuring balance across different ethical dimensions.

Researcher perceptions of the governance arrangements embedded in the CADRE platform and Five Safes Framework

In relation to the CADRE platform and the *Five Safes Framework*, participants were generally supportive of the need for work in this area and for better governance arrangements. Most saw the CADRE platform as quite 'bureaucratic' but also felt that some forms of oversight were necessary.

As part of these discussions, one participant spoke about becoming the custodian of a longitudinal study conducted by a community service organisation they had joined, seeing this as offering as a parallel perspective to researchers handing their data over to others to work on. This participant experienced a significant workload in taking on this project which had been conducted by others and had a significant qualitative component but had not been set up or managed by qualitative researchers. One part of this was the need to establish retrospectively a lot of metadata about why data was being collected. Later in the focus group, the same researcher also commented in relation to concerns the CADRE platform was too bureaucratic:

> I think that people that I work with, … from other organizations would find it a hindrance. But for me as a researcher, … I think it safeguards and protects people's data a lot better than the haphazard way that … was going on before I stepped in. So, I think that it's sort of, yes, I think it would be onerous, and people would, you know, they keep on me, but … the thing is that, like, I am very concerned about the ethics of data sharing. And if we don't have a framework, it makes it very unsafe. So, I think, yeah, I'm all for it. (Interview 3)

This same participant also noted that their organisation regularly shares and reuses data with like-minded organisations, but this requires a lot of back-and-forth discussion, which a framework like the Five Safes might help structure. The participant commented:

> We share data quite a bit, like with consent from the participants, of course, and with all these different approvals. It is quite fiddly. … So, lots of conversations about the stakeholders, lots of conversations about how it's going to be used. And we learned very early on that if anybody is writing anything about using from other people's data, that we check for that the person who collected that data first to see that with it, it's a spirit, whether it's secure, as well. So, we do a lot of checks and balances. But … it's all very informal. And it would be really good to use the 5 Safes Framework. (Interview 3)

Others also made similar comments, acknowledging that the CADRE platform and the 5 Safes Framework did feel overly bureaucratic and similar to how universities handle ethics processes, but also that some kind of governance is required. One participant commented,

> I'm not sure what the answer is, because you obviously need some of it. You need it. But in a lot of ways, whether some of that actually gets at where the problems are? I'm not sure. (Interview 3)

Another agreed with this, stating:

> I do like the idea of the framework. I completely agree with [another participant] about the bureaucracy of it, but I think … to have a kind of structure around it so that everyone's not reinventing it all the time is important. (Interview 3)

However, they also compared the framework and platform with the introduction of ethics committees, noting that while this did result in some losses about the kind of research which could be done it added important protections for people so there are always trade-offs and complexity in terms of how new structures work and their implications.

Related to this, participants in another focus group criticised the language of the Five Safes Framework and raised questions about its efficacy in opening up research. One participant commented that she found the language of safety to be blocking rather than enabling. She suggested that the terminology 'could put people off' because it felt 'a bit gate-keepy', noting 'that language of safety feels a bit like you're trying to access a crime database'. Another stated that they felt that the CADRE platform and the Five Safes Framework did not address 'institutional barriers for communities to access information about themselves' but were simply about 'replicating the researcher's own vision of what data is for' and that other non-digitised archival repositories she has accessed 'have less barriers than … this'. This was followed up by a comment from a third participant, this time a representative from a community organisation, that they felt that concern was valid and that the CADRE platform 'wouldn't actually reduce any of those barriers to accessing information for communities and using this information to support communities to do the sorts of things that they might use this for' (Interview 1).

As these comments highlight, the question of who oversees the archive and monitors requests for access is not straightforward and, while some of this work might be supported by the CADRE platform, wider epistemological and ethical questions about data ownership are not. As one participant put it, 'should you or I be the people who decide, in terms of openness of knowledge, what we like can be done' (Interview 3). In the discussions, the researchers we spoke with both supported community-led archiving *and* saw interpreting qualitative data as a skilled practice not readily achieved without appropriate training, further complicating this question. The question of the intended audience for the archive is critically important, and not easily resolved or necessarily opened up by the new platform or the Five Safes Framework.

Finding 2: The Five Safes is a useful framework for supporting the archiving and sharing of qualitative research data but on its own it situates decisions in a risk-oriented framework which leaves less scope to consider data utility and value. Indigenous data sovereignty frameworks like CARE that foreground collective benefit, authority to control, responsibility and ethics are important complements to the Five Safes model as these support data access beyond the academy and ensure balance across different ethical dimensions. Researchers in the qualitative social sciences express strong support for the need for better research governance and infrastructure to enable responsible data sharing. But this in itself is not sufficient to resolve challenges in supporting access to research beyond the academy.

4. Practical and Technical Issues

Resourcing constraints

Research on archiving qualitative research projects, including our own, has drawn attention to the significant resourcing constraints researchers face in preparing datasets for deposit (O'Connor et al., 2024; McLeod & O'Connor 2021; McLeod et al. 2020).

In our focus groups, those who had prior experiences of archiving their own projects were very aware of resourcing constraints, and the time and personnel requirements for archiving qualitative research data. One participant had led a large longitudinal research project comprising both quantitative and qualitative data collection, which is archived with the ADA, and noted the importance of having a dedicated staff member to manage this and the issues arising when that is not supported. Other participants also spoke about workload intensification and the challenge of finding time and resourcing to archive projects adequately. They noted limitations with current technical solutions and the problems evident in finding a software platform that meets ethics security requirements, supports collaborative work across institutions, and is also easily accessible when it comes to archiving at the completion of a project.

In relation to CADRE in particular, one participant commented that they do see the CADRE platform as providing significant benefits in reducing the risks associated with sharing data but not necessarily with the workload required for overseeing this:

> I think it really gives a bit of a partial relief in terms of who is requesting access to the data, and under what conditions. So that's definitely, I think, … a plus. But then I think the issues of actual management of the request for access might be another thing … that generates a lot of additional work … [so] what are the incentives for me as a researcher if I collected data to go through this extra set of steps to allow someone else under these conditions to access the data? (Interview 3)

These focus groups underlined that there is currently a lack of specific practical instructions, guidelines and protocols to support qualitative researchers in their archiving practice. There is also limited research that discusses the intertwined aspects of qualitative research practice and infrastructure in relation to data sharing, ethics practice and training, although this has been the subject of recent publications and initiatives in both academic institutions and private organisations (O'Connor et al. 2024; QDR n.d.a, n.d.b; QDS n.d.; Lee 2023). Much more work needs to be done to better understand how to support researchers in the practical processes of preparing their data for archiving and future sharing, as well as promoting the benefits of such practice. Safely sharing and reusing data must also involve providing social scientists with the infrastructure, methods and tools to undertake this practice.

Our original intention for the CADRE during 2020 to 2023 was to archive one or two new datasets, alongside those already loaded to the SOCEY Repository, and our experiences attempting this further support these concerns. We anticipated that this would extend findings from the earlier pilot study, as well as allow us to gather further examples of datasets for CADRE platform testing. However, despite extensive promotion of the Repository to both broad and targeted audiences and the ready availability of funding (and personnel) to complete this work, we were not able to secure another project for archiving.

This experience has itself revealed some of the challenges in the quest to archive and share qualitative data, which occur even when some of the more obvious barriers (e.g., time, money, infrastructure) have seemingly been addressed. Some projects did not proceed with archiving at the time due to reasons such as researchers being interested but the project not yet at the stage to archive; others had obligations to archive data elsewhere; some did not have the appropriate consent from participants to archive material; while others did not have the time or resources to devote to organising the material in readiness for archiving.

Several researchers who consulted us did use the Repository guidelines as a model at the beginning of their project when preparing their DMPs and ethics applications and may archive with us in the future. We also currently (March 2024) have one project at the EOI stage which has to go through approvals for archiving with another institution and the NGO that commissioned the work, but which has consent to archive: a process and case that bodes well for future archiving.

Difficulty in recruiting projects points to some of the practical barriers to archiving, including time restrictions or lack of consent, but also suggests a hesitancy among researchers to deposit and share work, despite the removal of some barriers (money and some time in this case). This speaks to a deeper reluctance undertake this work – for example, perhaps due to an informed decision not to engage or indeed opposition to the practice on epistemological or other grounds, ethical concerns, or a sense of protectiveness of their data, among other possibilities.

For several datasets considered for archiving during the CADRE project, there were insufficient underpinnings in place to archive: often this was lack of consent to do so from participants or other stakeholders at the project design stage. This was the case with an oral history project conducted for the centenary of Canberra in 2014, where educators from the University of Canberra were interviewed about the history of learning in the ACT. These were utilised in an online exhibition, with interview recordings of key figures in ACT education, excerpts of which were public access until the time of the archiving of the site in the mid 2010s, and still accessible via Wayback Machine (University of Canberra, n.d). While we were keen to archive the full recordings (originally planned for the Australian National Museum of Education), that no consent was obtained at the time presented a major obstacle.

Finding 3: Resourcing is a critical issue and, despite in-principle support for the idea of qualitative data archiving and sharing, there is far less willingness from researchers working in the qualitative social sciences to engage in the amount of work and funding required to do this well. This points to the need for not only better research funding but also the importance of more comprehensive protocols and practical guidelines for archiving and sharing qualitative research.

Accessing datasets

One significant limitation in accessing shared qualitative research data in Australia is the small numbers of research projects that have created accessible datasets (owing to the challenges canvassed above) and the lack of infrastructure to search across multiple repositories. This is something the CADRE platform has the potential to address fruitfully. Where datasets that are not already on an easily findable or searchable platform, they would benefit from being included in a repository that is affiliated with CADRE and their release would be streamlined (particularly if the original researcher or data owner cannot be located).

However, as CADRE is a platform for *accessing data and approving its release*, rather than making decisions about whether the datasets are suitable for applicants' projects, only limited metadata is available. For qualitative research, this is a potentially significant limitation. Given the often relatively small size and targeted focus of many qualitative datasets, detailed metadata is particularly important. Without this, researchers are unlikely to find and use the datasets they need, inhibiting the platform's intent to increase access and reuse. If a researcher sees something of interest while searching in the platform, they might not be able to decide if this database is suitable for them due to lack of description and detailed metadata.

Qualitative research projects in the social sciences, however, might be archived in diverse repositories, including less accessible institutional repositories or indeed on university share drives. We archived our datasets in the ADA, which is more known for its quantitative or mixed methods datasets, rendering them more likely to be located by interested researchers. Over the past three years, we have had a number of people seek permission to use datasets from the archive, but many similar datasets prove difficult to locate.

These experiences prompted us as data custodians to consider the questions we needed to ask of researchers who wished to use the data, including ascertaining a detailed understanding of their credentials and training, how they intended to use the data and the ethical safeguards they had in place for such usage, often with significant back and forth discussion with those requesting access mediated by the repository. In retrospect, our questions could have been better guided by the Five Safes framework, supported by the CADRE platform.

However, issues have arisen in mediating these requests which are less addressed by the infrastructure offered through CADRE. In some cases, although requests came from researchers with university affiliations, the usage of the data for the purposes they proposed was not in alignment with the release and uses of that data at that time. This indicates that automated data release or access for researchers, streamlined through the CADRE system (although it is not currently a feature of the present iteration of the platform), may not regularly be a possibility and would likely often require scrutiny by a decisionmaker.

This also points to some of the complexity in how access decisions are made and raises questions as to who should be considered authorised to make such decisions, both in the short and long term. These are questions not neatly resolved by the CADRE system or the Five Safes Framework but require thought and deliberation, and this may include consideration of expertise in qualitative research as a specific research skill. It should be noted that protocols will also need to be put in place in the future to pass data custodianship on to another responsible party to ensure that data access requests can continue to be managed appropriately and in line with the intent of the project participant consent. While the Five Safes framework itself can be a useful tool for data custodians in deciding whether to release sensitive data, providing an interoperative 'checklist' that considers different aspects of the data application and what release of this data would mean for the various stakeholders might be needed as an additional support.

The CADRE training materials authorising access to the platform also focus predominantly on issues of data safety and do not address any of the concerns about community access and control noted in the previous section. While these training materials provide important insights for those interested in archiving, sharing and reusing qualitative data, their focus is very much on technical and practical concerns and the larger epistemological and methodological questions discussed in the previous section are largely unaddressed.

Finding 4: The benefits of the CADRE platform for qualitative research are restricted without the inclusion of more extensive and appropriate metadata to aid researchers trying to identify relevant datasets. While CADRE is primarily a platform to manage access, the minimal provision of metadata is particularly consequential for qualitative datasets. The relatively small size and targeted focus of many qualitative data sets means that without detailed metadata researchers will face difficulties finding and using the datasets they need, inhibiting the platform's intent to increase access and reuse. This is compounded by the relatively modest number of Australian archived qualitative datasets available and accessible for re-use. Further, decisions about access to qualitative datasets are not neatly resolved at a single point in time and may require revisiting, based on a principled-base process to support ongoing custodianship.

Conclusion

This report has explored the implications of the CADRE platform for qualitative data, a much-neglected area in data sharing research, frameworks and infrastructure development but also a space of considerable importance and potential social and researcher benefit. It has considered the challenges and opportunities CADRE raises for archiving and sharing qualitative research data over three sections focusing on (1) epistemological and ethical questions concerning the merits and challenges of archiving qualitative data; (2) concerns relating to infrastructure, governance, and data ownership; and (3) the practical and technical issues researchers confront in archiving their own projects and accessing others, pointing towards four primary findings as noted throughout.

First, although archiving and sharing qualitative research data is acknowledged as ethically and epistemologically complex, researchers in the qualitative social sciences are not opposed to this in principle and see value in carefully thinking through how this can be done well.

Second, the Five Safes is a useful framework for supporting the archiving and sharing of qualitative research data but on its own it situates decisions in a risk-oriented framework which leaves less scope to consider data utility and value. Indigenous data sovereignty frameworks such as CARE that foreground collective benefit, authority to control, responsibility and ethics are important complements to the Five Safes model as these support data access beyond the academy and ensure balance across different ethical dimensions. Researchers in the qualitative social sciences express strong support for the need for better research governance and infrastructure to enable responsible data sharing. But this in itself is not sufficient to resolve challenges in supporting access to research beyond the academy.

Third, resourcing is a critical issue and, despite in-principle support for the idea of qualitative data archiving and sharing, there is far less willingness from researchers working in the qualitative social sciences to engage in the amount of work and funding required to do this well. This points to the need for not only better research funding but also the importance of more extensive comprehensive protocols and practical guidelines for archiving and sharing qualitative research.

Finally, the benefits of the CADRE platform for qualitative research are restricted without the inclusion of more extensive and appropriate metadata to aid researchers trying to identify relevant datasets. While CADRE is primarily a platform to manage access, the minimal provision of metadata is particularly consequential for qualitative datasets. The relatively small size and targeted focus of many qualitative data sets means that without detailed metadata researchers will face difficulties finding and using the datasets they need, inhibiting the platform's intent to increase access and reuse. This is compounded by the relatively modest number of Australian archived qualitative datasets available and accessible for re-use. Further, decisions about access to qualitative datasets are not neatly resolved at a single point in time and may require revisiting, based on a principled-base process to support ongoing custodianship.

References

Andrews, M. 2011. Never the last word: revisiting data 1. In M. Andrews, C. Squire & M. Tamboukou (eds), *Doing Narrative Research*. Sage. DOI: 10.4135/9780857024992

Arthur, P.L. & L. Hearn. 2021. Toward open research: a narrative review of the challenges and opportunities for open humanities. *Journal of Communication* 71(5): 827–85**3**. DOI: 10.1093/joc/jqab028

Arthur, P.L., L. Hearn, L. Montgomery, H. Craig, A. Arbuckle & R. Siemens. 2021. Open scholarship in Australia: a review of needs, barriers, and opportunities, *Digital Scholarship in the Humanities* 36(4): 795–812. DOI: 10.1093/llc/fqaa063

Australian Bureau of Statistics (ABS). 2021. Five safes framework: data confidentiality guide. https://www.abs.gov.au/about/data-services/data-confidentiality-guide/five-safes-framework

Australian Data Archive (ADA). 2020. Common disclosure risk factors. https://docs.ada.edu.au/index.php/Common_Disclosure_Risk_Factors

Australian Data Archive (ADA). n.d. Why deposit with ADA … . https://ada.edu.au/depositing-data/

Australian Institute of Health & Welfare (AIHW). 2022. *Data Governance Framework 2021 (Public Edition)*. https://www.aihw.gov.au/getmedia/3117bc9c-46ee-4891-a423-2ec0d66e12e7/aihw-data-governance-framework-2022.pdf.aspx

Australian Institute of Health & Welfare (AIHW). 2023. The Five Safes framework. https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework

Australian Research Data Commons (ARDC). 2021. Publishing and sharing sensitive data – flowchart. DOI: 10.5281/zenodo.7259737

Australian Research Data Commons (ARDC). 2022. Publishing sensitive data. DOI: 10.5281/zenodo.7259742

Australian Research Data Commons (ARDC). n.d.a. FAIR data. https://ardc.edu.au/resource/fair-data/

Australian Research Data Commons (ARDC). n.d.b. Sensitive data. https://ardc.edu.au/resource/sensitive-data/

Barbera, M. 2013. Linked (open) data at web scale: research, social and engineering challenges in the digital humanities. *Italian Journal of Library, Archives and Information Science* 4(1): 91. DOI: 10.4403/jlis.it-6333

Bastian, J.A. & A. Flinn (eds). 2020. *Community Archives, Community Spaces: Heritage, Memory and Identity*. Facet Publishing.

Bishop, L. 2005. Protecting respondents and enabling data sharing: reply to Parry and Mauthner, *Sociology* 39(2): 333–336.

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

Borgerud, C. & E. Borglund. 2020. Open research data, an archival challenge? *Archival Science* 20, 279–302. DOI: 10.1007/s10502-020-09330-3

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

Carroll, S.R., E. Herczog, M. Hudson, K. Russell & S. Stall. 2021. Operationalizing the CARE and FAIR principles for Indigenous data futures. *Scientific Data* 8. DOI: 10.1038/s41597-021-00892-0

Cheshire, L., A. Broom & M. Ellison. 2009. Archiving qualitative data in Australia: an introduction. *Australian Journal of Social Issues* 3, 239–254.

Childs, S., J. McLeod, E. Lomas & G. Cook. 2014. Opening research data: issues and opportunities. *Records Management Journal* 24(2): 142–162. DOI: 10.1108/RMJ-01-2014-0005

CILogon. n.d. CILogon: an integrated identity and access management platform for science. https://www.cilogon.org/

Cook, T. 2012. Evidence, memory, identity, and community: four shifting archival paradigms. *Archival Science* 13, 95–12. DOI: 10.1007/s10502-012-9180-7

Coordinated Access for Data, Research and Environments (CADRE). n.d.a. About: Five Safes. https://cadre5safes.org.au/about/five-safes/

Coordinated Access for Data, Research and Environments (CADRE). n.d.b. Access request management. https://cadre5safes.org.au/2022/10/03/access-request-management/

Coordinated Access for Data, Research and Environments (CADRE). n.d.c. https://cadre5safes.org.au. Project_DOI: 10.47486/PL106

Coordinated Access for Data, Research and Environments (CADRE). n.d.d. *Learning*. https://learning.cadre5safes.org.au/

Coordinated Access for Data, Research and Environments (CADRE). 2024. *CADRE User Guide*. Australian Research Data Commons (ARDC). https://cadre5safes.org.au/help

Coordinated Access for Data, Research and Environments (CADRE). 2024. *CADRE User Guide – Decisionmakers* [V2.0]. Australian Research Data Commons. https://cadre5safes.org.au/help

Corti, L. & P. Thomson. 1997. Research/resources note: latest news from the ESRC Qualitative Data Archival Resource Centre (QUALIDATA). *Journal of Social History* 30(4): 963–966. https://www.jstor.org/stable/3789790

Corti, L., A. Day & G. Backhouse. 2000. Confidentiality and informed consent: issues for consideration in the preservation of and provision of access to qualitative data archives. *FORUM: Qualitative Social Research Sozialforschung* 1(3).

Davis, N. & Neish, P. 2022. *Institutional Underpinnings – Ethics and Data Discussion Paper*. University of Melbourne. DOI: 10.26188/20500101

Department of Industry, Science & Resources (Commonwealth). n.d. VANguard. https://www.industry.gov.au/government-government/vanguard

Desai, T., F. Ritchie & R. Welpton. 2016. *Five Safes: Designing Data Access for Research*. Economics Working Papers in Economics, No. 1601. University of the West of England, Bristol. DOI: 10.13140/RG.2.1.3661.1604; https://www2.uwe.ac.uk/faculties/bbs/Documents/1601.pdf

Elixir. n.d. AAI REMS. https://www.elixir-finland.org/en/aai-rems-2/

Elman, C. & D. Kapiszewski. 2013. *A Guide to Sharing Qualitative Data: Version 1.3 for Pilot Projects*. Syracuse: Qualitative Data Repository (QDR), Center for Qualitative and Multi Method Inquiry (CQMI).

Fitzgerald, E. 2020. Appendix1: Schooling Memories: Educating the Adolescent, 1930s–1970s. In McLeod, J., O'Connor, K. & Davis, N. *Doing Research Differently: Archiving & Sharing Qualitative Data in Studies of Childhood, Education and Youth*. University of Melbourne, 30–36. DOI: 10.25916/5e9e28eec21a1

Global Indigenous Data Alliance (GIDA). n.d. CARE principles for Indigenous data governance. https://www.gida-global.org/care

Hammersley, M. 1997. Qualitative data archiving: some reflections on its prospects and principles. *Sociology* 31(1): 131–141.

Hostler, T.J. 2023. The invisible workload of open research. *Journal of Trial & Error*. https://doi.org/10.36850/mr5

Jones, Y. 2022. Are there gaps in researcher training? *CADRE 5 Safes*. July 18 [blog post]. https://cadre5safes.org.au/2022/07/18/are-there-gaps-in-researcher-training/

Joyce, J.B., T. Douglass, B. Benwell, C.S. Rhys, R. Parry, R. Simmons & A. Kerrison. 2023. Should we share qualitative data? Epistemological and practical insights from conversation analysis. I*nternational Journal of Social Research Methodology* 26(6): 645–659. DOI: 10.1080/13645579.2022.2087851

Kirilova, D. & S. Karcher. 2017. Rethinking data sharing and human participant protection in social science research: applications from the qualitative realm. *Data Science Journal*, 16(43): 1–7. DOI: 10.5334/dsj-2017-043

Kukutai, T., & J. Taylor. 2016. Data sovereignty for Indigenous peoples: current practice and future needs. In T. Kukutai, & J. Taylor. *Indigenous Data Sovereignty: Toward an Agenda* 38, 1–22. ANU Press. http://www.jstor.org/stable/j.ctt1q1crgf

Lee, R.H. 2023. Introducing the Qualitative Data Sharing (QDS) Toolkit. *ELSIhub*. 26 July [blog post]. https://elsihub.org/news/introducing-qualitative-data-sharing-qds-toolkit

Melbourne Graduate School of Education (MGSE). 2017. Curriculum Policies Archive. *eScholarship Research Centre*. https://scpp.esrc.unimelb.edu.au/

Melbourne Graduate School of Education (MGSE). 2008. Curriculum Policies Project. https://web.education.unimelb.edu.au/curriculumpoliciesproject/

McEachern, S. 2022. Access request management. *CADRE Five Safes*. https://cadre5safes.org.au/2022/10/03/access-request-management/

McEachern, S. [et al.]/CADRE. 2021. *CADRE Five Safes Framework v.1.0*. DOI: 10.5281/zenodo.5748611

McEachern, S. [et al.]/CADRE. 2022. *CADRE Five Safes Framework v.1.1*. DOI: 10.5281/zenodo.5748610

McEachern, S. [et al.]/CADRE. 2024. *CADRE Five Safes Framework v.1.2*. DOI: 10.5281/zenodo.5748610

McLeod, J. 2016. Making Futures working paper. *Making Futures*. https://makingfutures.net/wp-content/uploads/2016/03/Making-Futures-Working-Paper-1.pdf

McLeod, J. & N. Davis. n.d. Collingwood College. *Making Futures*. https://makingfutures.net/schools-and-communities/collingwood/collingwood-college/

McLeod, J. & K. O'Connor. 2020. Ethics, archives and data sharing in qualitative research, *Educational Philosophy & Theory* 53(5): 523–535. DOI: 10.1080/00131857.2020.1805310

McLeod, J., K. O'Connor & N. Davis. 2020. *Doing Research Differently: Archiving & Sharing Qualitative Data in Studies of Childhood, Education and Youth*. University of Melbourne. DOI: 10.25916/5e9e28eec21a1

McLeod, J., K. O'Connor & N. Davis. 2023. *Governance of Qualitative Data Across Sectors: Access, Sharing and Re-using Qualitative Research Data*. CADRE Report, Deliverable 4 [V2.0] [Interim report]. DOI: 10.5281/zenodo.7814759

McLeod, J. & K. Wright. 2020. Schooling Memories: Educating the Adolescent, 1930s –1970s. ADA Dataverse, V1. DOI: 10.26193/R8ACQQ

Mauthner, N., O. Parry & K. Backett-Milburn. 1998. The data are out there, or are they? implications for archiving and revisiting qualitative data. *Sociology* 32(4): 733–745. DOI: 10.1177/0038038598032004006

Mauthner, N.S., & O. Parry. 2013. Open access digital data sharing: principles, policies and practices. *Social Epistemology* 27(1): 47–67. DOI: 10.1080/02691728.2012.760663

Moore, N., N. Dunne, M. Karels & M. Hanlon. 2021. towards an inventive ethics of carefull risk: unsettling research through DIY academic archiving. *Australian Feminist Studies* 36(108): 180–199. DOI: 10.1080/08164649.2021.2018991

Mozersky, J., M. Parsons, H. Walsh, K. Baldwin, T. McIntosh & J.M. DuBois, J.M. 2020. Research participant views regarding qualitative data sharing. *Ethics & Human Research* 42, 13–27. DOI: 10.1002/eahr.500044

O'Connor, K., J. McLeod. & N.J. Davis. 2024. The ethics of archiving and sharing qualitative data. In B.M. Smyth, M.A. Martin, & M. Downing (eds), *The Routledge Handbook of Human Research and Integrity in Australia*. Routledge. Forthcoming, October.

Office of the National Data Commissioner (ONDC). 2024. Introducing the DATA Scheme. *Office of the National Data Commissioner*. https://www.datacommissioner.gov.au/the-data-scheme

ORCiD. n.d. https://orcid.org/

Pascoe Leahy, C. 2022. The afterlife of interviews: explicit ethics and subtle ethics in sensitive or distressing qualitative research. *Qualitative Research* 22(5): 777–794. DOI: 10.1177/14687941211012924

Parry, O. & N. Mauthner. 2004. Whose data are they anyway? practical, legal and ethical issues in archiving qualitative research data. *Sociology* 38(1): 139–152.

Piccini, J. 2023. 'Thinking in Papua New Guinean terms': the sensitive files case of 1972 and Australia's migrated archive. *History Workshop Journal* 96, 115–133. DOI: 10.1093/hwj/dbad018

Popovski, M. 2020. Appendix 5: Making Futures. In McLeod, Julie, O'Connor, Kate & Davis, Nicole, *Doing Research Differently: Archiving & Sharing Qualitative Data in Studies of Childhood, Education and Youth.* University of Melbourne, 57–60. DOI: 10.25916/5e9e28eec21a1

Qualitative Data Repository (QDR). n.d.a. About the Qualitative Data Repository. https://qdr.syr.edu/about

Qualitative Data Repository (QDR). n.d.b. Deposit process. https://qdr.syr.edu/deposit/process

Qualitative Data Repository (QDR). n.d.c. Human participants – general guidance. https://qdr.syr.edu/guidance/human-participants

Qualitative Data Repository (QDR). n.d.d. Managing data. https://qdr.syr.edu/guidance/managing

QDS Qualitative Data Sharing Toolkit. n.d. About the QDS project. https://qdstoolkit.org/about/

Reanimating Data. n.d. *Reanimating Data: Experiments with People, Places & Archives.* https://reanimatingdata.co.uk/

Reese, H. 2020. Appendix 4: The Curriculum Policies Project. In McLeod, Julie, O'Connor, Kate & Davis, Nicole, *Doing Research Differently: Archiving & Sharing Qualitative Data in Studies of Childhood, Education and Youth.* University of Melbourne, 47–56. DOI: 10.25916/5e9e28eec21a1

Research Data Alliance International Indigenous Data Sovereignty Interest Group (RDAIIDSIG). September 2019. *CARE Principles for Indigenous Data Governance.* The Global Indigenous Data Alliance. https://static1.squarespace.com/static/5d3799de845604000199cd24/t/6397b363b502ff481fce6baf/1670886246948/CARE%2BPrinciples_One%2BPagers%2BFINAL_Oct_17_2019.pdf

Rose, J., M. Langton, K. Smith & D. Clinch. 2023. Indigenous data governance in Australia: towards a national framework. *International Indigenous Policy Journal* 14(1). DOI: 10.18584/iipj.2023.14.1.10987

Savage, M. 2010. *Identities and Social Change in Britain since 1940.* Oxford University Press.

Studies of Childhood, Education and Youth (SOCEY). n.d. https://www.socey.net

Studies of Childhood, Education and Youth (SOCEY). 2019–20. Repository. https://www.socey.net/repository/

Thomson, R. & J. McLeod. 2015. New frontiers in qualitative longitudinal research: an agenda for research. *International Journal of Social Research Methodology* 18(3), 243–250. DOI: 10.1080/13645579.2015.1017900

Tsai, A.C, B.A. Kohrt, L.T. Matthews, T.S. Betancourt, J.K. Lee, A.V. Papachristos, S.D. Weiser & S.L. Dworkin. 2016. Promises and pitfalls of data sharing in qualitative research. *Social Science & Medicine* 169, 191–198. DOI: 10.1016/j.socscimed.2016.08.004

UK Data Service. n.d. Other Data Providers: European qualitative data archives. *UK Data Service.* https://ukdataservice.ac.uk/help/other-data-providers/other-data-providers-qualitative-data/european-archives/

University of Canberra [via *Wayback Machine*]. n.d. 100 years of learning: celebrating lessons learnt from Canberra's education history. https://web.archive.org/web/20140303053019/http://centenary.estem-uc.edu.au/

University of Melbourne. n.d. Research Data Classification Tool. https://dataservices.research.unimelb.edu.au/services/80/

University of New South Wales (UNSW). 2021. Data classification standard. https://www.unsw.edu.au/content/dam/pdfs/governance/policy/2022-01-policies/datastandard.pdf

University of New South Wales (UNSW). n.d. Classifying your research data. https://research.unsw.edu.au/classifying-your-research-data

Vaarzon-Morel, P., L. Barwick, & J. Green. 2021. Sharing and storing digital cultural records in Central Australian Indigenous communities. *New Media & Society* 23(4): 692–714. DOI: 10.1177/1461444820954201

Victoria Legal Aid. 2023. Learning from the failures of Robodebt. https://www.legalaid.vic.gov.au/learning-from-the-failures-of-robodebt

Walter M., R. Lovett, B. Maher, B. Williamson, J. Prehn, G. Bodkin-Andrews & V. Lee. 2020. Indigenous Data Sovereignty in the era of Big Data and Open Data. *Australian Journal of Social Issues.* 1–14. https://onlinelibrary.wiley.com/doi/10.1002/ajs4.141

Wilkinson, M., M. Dumontier & I. Aalbersberg et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018. DOI: 10.1038/sdata.2016.18

Wolski, M., J. Richardson & R. Rebollo. 2011. Shared benefits from exposing research data. *32nd Annual IATUL Conference Proceedings*. https://research-repository.griffith.edu.au/server/api/core/bitstreams/d1c5ed0f-32d1-5d6b-9ed7-1f2ac2074bf2/content

Yates, Lyn. 2020. The Curriculum Policies Project. ADA Dataverse, V1. DOI: 10.26193/6JG0DM

Appendix 1: Work Packages Undertaken

Content Working Group: This group was responsible for the development of the *CADRE Framework*, through which the platform is operationalised. The team provided a significant section on the significance of qualitative data to the project. This work was informed by associated research outputs completed as part of the SOCEY project, since its foundation in 2018.

Work Package 2 (WP2): Protocols. Responsible for 'designing a metadata (information) exchange for the CADRE Platform. The SOCEY team provided feedback for various protocols and standards that were developed during this Work Package.

Work Package 3 (WP3) and Technical Working Group. Responsible for developing the platform infrastructure and access protocols. The SOCEY team contributed to the initial design phase through the contribution of user stories, [and the other things], and later provided feedback on the platform design.

Work Package 5 (WP5): Engagement & Training. This allowed CADRE to socialise the framework, conduct a gap analysis and examine opportunities that the project can take up. The SOCEY team contributed significantly to this aspect of the project through developing material focused on qualitative data for the training modules and assisting in the running of workshops in Sydney and Melbourne that provided a testing space for the CADRE training and socialisation.

See the CADRE website for further information on the Project Phases and a description of all Work Packages for the project.

FRONT COVER:SCHOOL BAND PERFORMING AT EDUCATION WEEK, QUEEN STREET MALL, BRISBANE, 1990. QUEENSLAND DEPARTMENT OF EDUCATION, DEPT NO.BW90-5488 D. QUEENSLAND STATE ARCHIVES, ITM436117.