Should we share qualitative data? Epistemological and practical insights from conversation analysis

Joyce, Jack ; Douglass, Tom; Benwell, Bethan; Rhys, Catrin; Parry, Ruth ; Simmons, Richard; Kerrison, Adrian

DOI:
10.1080/13645579.2022.2087851

Should we share qualitative data? Epistemological and practical insights from Conversation Analysis

Abstract

Over the last thirty years, there has been substantial debate about the practical, ethical and epistemological issues uniquely associated with qualitative data sharing. In this paper we contribute to these debates by examining established data sharing practices in Conversation Analysis (CA). CA is an approach to the analysis of social interaction that relies on audio/video recordings of naturally occurring human interactions and moreover works at a level of detail that presents challenges for assumptions about participant anonymity. Nonetheless, data sharing occupies a central position in both the methodology and the wider academic culture of CA as a discipline and a community (ten Have, 2007). Despite this, CA has largely been ignored in qualitative data sharing debates and discussions. We argue that the methodological traditions of CA present a strong case for the value of qualitative data sharing (QDS) and offer open data sharing practices that might be usefully adopted in other qualitative approaches.

Keywords

Data Sharing, Conversation Analysis, Open Science, Qualitative Data

Data sharing is an important principle of the Open Science agenda but is not yet widespread within qualitative research. Debates emphasise ethical and epistemological issues uniquely associated with qualitative data that are seen as a barrier to data sharing. However, noticeably absent from these debates is the contribution of Conversation Analysis (CA), a qualitative research approach with a long history of data sharing. We address that gap through examining the formal and informal data sharing practices of CA and the underlying epistemological position that supports those practices. We argue that

the particular conception of data and context adopted in CA makes redundant the very distinction between primary and secondary data that currently frames data sharing debates and has thus significantly influenced data sharing practices amongst CA researchers. Two broad distinctions can be made with regards to data sharing: (1) *data sharing as a research practice*, contributing to the rigour of analysis within a particular project; (2) *corpus sharing*, where a complete data set is made available via a repository for other research projects. Both types of sharing are well established in the CA community with demonstrable benefits for research outcomes. Thus, we argue that the methodological traditions of CA present a strong case for the value of qualitative data sharing (QDS) and offer open data sharing practices that might be usefully adopted in other qualitative approaches. This is particularly important in a context where funders, publishers and legal requirements increasingly expect data to be made available for sharing.

Qualitative Data Sharing: Nature and Context

Interest in potential reuse/secondary analysis of qualitative data has grown since the 1990s (Heaton, 2008; Hughes et al. 2020). Arguments for sharing and reusing qualitative data include checking of findings, fostering public trust in science, and enhancing research training (DuBois et al. 2018). In addition, existing data can be analysed to produce new findings, which is time and cost-effective for researchers and avoids unnecessary burden on participants (Kuula, 2011). The first qualitative data repository, Qualidata, was established over 25 years ago (and is now part of the UK Data Service[1]). Since then, technological advances have increased capacity to store and facilitate access

[1] A resource that provides guidance on data management and includes a large archive of data and the details of other collections.

to large datasets in repositories (Chauvette et al., 2019, p. 2; Corti et al., 2016)[2]. Increasingly, research funders and publishers encourage and even mandate QDS (Antonio et al., 2019; Chauvette et al., 2019) through policies of open access, shaped by a commitment to principles of transparency and scrutiny, and to maximising the social value of publicly-funded research (UKRI, 2021). In the UK, important milestones for QDS were the adoption by all research councils in 2011 of the Common Principles on Data Policy and in 2012 of the Policy on Access to Research Outputs which required research outputs to make explicit how the research data would be made available (Bishop & Kuula-Luumi 2017). With 55% of research in UK HE funded by research councils, these policies "strongly influence research practices" (Bishop & Kuula- Luumi, 2017, p. 2). The UK Research and Innovation (UKRI) conducted a review, in the UK (published February 2022), of open access policies and the data sharing landscape which reiterated their concordat on Open Research Data (UKRI, 2021) emphasising that publicly funded research should be openly available with as few as restrictions as possible.

Nonetheless, concern about QDS persists. Resistance is often expressed on ethical and epistemological grounds (Chauvette et al., 2019; Mozersky et al., 2020a). Two large-scale surveys of scientific researchers identified a number of recurring concerns: fears relating to participant anonymity, misinterpretation in secondary analyses, invalid conclusions, data errors, being scooped, and researcher burden. Overarching and unifying guidelines, policies, and mandates are also still lacking internationally. Moreover, where they exist, data sharing approaches, policies, and repositories are mostly established with quantitative research in mind (Antonio et al., 2019; Tsai et al., 2016). For example, pre-

[2] Data repositories were initially developed to increase the transparency and sharing of data from clinical trials (Antonio et al., 2019).

registration forms required by some funders are often inadequate adaptations of forms designed for quantitative work (Humă & Joyce, frth). Recent empirical research on QDS preparedness in the US found that even repository specialists lacked experience and knowledge relevant to QDS (Mozersky et al., 2020a). Many felt unprepared to advise qualitative researchers – particularly on decisions about sensitive data. Similar limitations in knowledge and preparedness were found amongst qualitative researchers and institutional ethics committee members. This US study found little experiential knowledge of QDS; each group (researchers, ethics committees and repository staff) felt that the primary responsibilities and decisions about QDS lay elsewhere. Even if those groups were to become more experienced, there remains a lack of agreement and guidance on best practice, exacerbated by different requirements between institutions and repositories, and different countries' laws relating to cross-border data sharing. Arguably, resolving these infrastructural and practical concerns depends on also addressing the debates about QDS to which we now turn.

Debating Qualitative Data Sharing

Literature reflecting on the viability of QDS is dominated by discussion of ethical and epistemological challenges posed by the various forms of qualitative data, each with different affordances, including (inter alia) observational data (e.g. fieldnotes, audio/video recordings), participant produced data (e.g. diaries) and researcher elicited data (e.g. interviews). Across this section we review literature outlining these challenges and the impact they have on qualitative researchers' commitment to data sharing and data reuse. We discuss ethical and epistemological challenges that exist across the range of qualitative methods as well as relating to specific methods.

Ethical Debates

Some qualitative researchers argue that it is impossible to ensure that research participants know what they are consenting to when it comes to data sharing – how data might be used in future projects and by other researchers (Parry & Mauthner, 2004; see also Chauvette et al., 2019). Consent for data sharing can only ever be secured in a general manner for and about the process itself, rather than for specific research (Irwin, 2013, p. 297). The issue, then, is whether it can be considered ethical to share data when *fully informed* consent for the myriad ways data might be used can never be achieved.

The counterargument, however, is that it is impossible to be fully 'informed' about all aspects of research (even research questions, for example, may not be formed prior to data collection in some qualitative approaches) (Bishop, 2009). Hence, this is not a reason to dismiss QDS. Additionally, qualitative research participants, who invest time but also emotionally in the research, generally seem to support data sharing and to assume it occurs to a greater extent than it does (Kuula, 2011). Indeed, participants in sensitive qualitative studies, interviewed by Mozersky and colleagues (2020b), reported broad support for QDS where data are anonymised, although, when pressed, expressed concern about confidentiality and potential misuse/misunderstanding in future research. This research also suggests that qualitative research participants trusted research institutions and their researchers to be sufficiently transparent with data collection and sharing plans. These findings echo similar work on the trust and value of research, such as Parry et al. (2016) who surveyed research participants and reported that most regard qualitative video-based research as acceptable, and Williams et al. (2010) who reported an overwhelming majority of participants believing that recording was worthwhile.

There are, however, other potential ethical issues. Qualitative data can be highly sensitive, confessional and intimate. Ensuring confidentiality and anonymity, and

protecting participants from unintended identification is vital, but rigorous anonymisation is time and labour intensive and challenging. In some forms of qualitative research, even if all possible measures of anonymisation are taken, there is still the possibility for identification as total anonymisation is impossible (Hopkins, 1993). For example, in longitudinal data collection or in other forms of data collection that link different sets of data together, the accumulation of information and associations potentially present a greater disclosive risk (Law, 2005). Similarly, qualitative research that takes place in small communities or on phenomena that are rare may be particularly hard to fully anonymise (Chauvette et al., 2019; Hardy et al., 2016).

This poses a challenge to QDS on two-fronts: the integrity and quality of certain forms of data is compromised when it is digitally altered to ensure participant confidentiality (such as video data) (Bishop, 2009: 262), and the reuse of original (unedited) data runs the risk that researchers outside of the original project may not know what should be anonymised and how it should be anonymised. These are challenges whose answers may have profound consequences for the scientific analysis (Corti et al., 2000).

It is true that there are no straightforward or 'one size fits all' answers to these issues (see Humă & Joyce, frth). However, it has also been argued[3] that "too often the critics of reusing qualitative data have narrowly construed the debate to focus solely on participants – to the exclusion of other agents – and rights – to exclusions of duties. Such arguments do not do justice to the depth of moral debate required" (Bishop, 2009, p. 258). In this sense, advocates of QDS argue for a broader range of ethical considerations. From

[3] This is influenced by deontological ethics. See Bishop (2009, p. 257-260) for an overview.

this perspective, the benefits to knowledge, policy and society from data sharing (weighed against, in the majority of cases, minimal risks to participants) need to be given greater emphasis (Bishop, 2009). For example, data sharing can mean avoiding the unnecessary intrusion and burden on participants that result from collecting data that already exist.

Epistemological Debates: The Problem of Context

The central epistemological debate in QDS literature concerns the contextual nature of qualitative data. One sense of this contextuality relates to the relationship between the researcher and the overall research process. In this sense, qualitative data is argued to be contextual in that it reflects the original researcher's positionality, beliefs, judgments, disciplinary assumptions and boundaries, as well as their theoretical and methodological inclinations and intentions within those disciplinary boundaries (Irwin, 2013). These aspects are embedded in the data, uniquely shaping its constitution and analysis. Reflexivity is thus a central practice of the qualitative paradigm, requiring the researcher to explicitly examine how these underpinning beliefs and practices have shaped the data and its analysis. However, for many qualitative approaches this practice is not available in secondary data analysis (Mauthner et al., 1998). As a consequence, many qualitative researchers maintain that only they (and their team) can analyse their data in a contextually fitted and adequately reflexive manner.

A second aspect of the contextuality of qualitative data focuses more narrowly on different conceptualisations of context, particularly with respect to how it relates to talk/text. van den Berg (2008, p. 186-188) describes three ways of conceptualising context: as (broad) *extra-discursive template,* where the relation between text and context is predefined (e.g. Critical Discourse Analysis); as (narrow) *intra-discursive product*, where context is only relevant when demonstrably made relevant in a participant's talk (e.g. Conversation Analysis); and as (intermediate) *conditions of discursive production,*

where the necessary contextual information depends on the focus of the research and the data being used (e.g. ethnography of communication). The broad and intermediate conceptions of context view qualitative data as generated in a specific time and setting of which the primary researcher necessarily has first-hand, intimate experience. In this sense, ethnographic fieldnotes, for example, may be difficult or impossible to meaningfully interpret by a researcher who did not participate in the original research (Chuavette et al., 2019). As Hammersley (2010, p. 3) notes, "in the process of data collection researchers generate not only what are written down as data but also implicit understandings and memories of what they have seen, heard, and felt, during the data collection process.". According to this perspective, the extent of contextual understanding in secondary analysis will necessarily be more limited and interpretive (Hammersley, 2010). This understanding of context, however, is not in harmony with CA's perspective which we will consider later.

Other contributors to context as an intra-discursive product posit that data are constructed and not independent of the research process – that, in other words, 'context' is not ontologically separate from data (Mauthner & Parry, 2009). Ethnomethodologists (Garfinkel et al., 1981; Lynch, 1982) have questioned how "data" is even first granted that status by researchers, and how a discipline's technical language and concepts must be deployed to alert others to the presence of data (Maynard & Clayman, 1991). From this perspective, researchers do not 're-use' data, because data are constituted for the first time in a particular research project (Moore, 2007; see also Bishop, 2007).

The past two sub-sections have explored the existing literature concerned with data sharing in qualitative research. This has shown gaps in researchers' agreement on the value of – and commitment to the practice of – QDS; significant interpretive and practical difficulties associated with this; and a series of ethical and philosophical

192 questions regarding the sharing and reuse of data. The paper now turns to understandings
193 and approaches to data sharing in CA and considers their potential to address these gaps,
194 difficulties and questions.

195 **Centrality of Data Reuse in Conversation Analysis**
196
197 Before detailing CA's contributions to data sharing debates, it is necessary to provide a
198 brief overview of CA's history and its arguably unique relationship with data sharing[4].
199 CA draws focus on practices and social actions rather than people or experiences which
200 means it is commonplace to ask very different questions of reused data. In the early days
201 of CA, data collections tended to be limited to audio only. Initially, Harvey Sacks, while
202 a researcher at The Suicide Prevention Centre, analysed recorded phone calls for his PhD
203 thesis in 1966[5].

204 From its inception, data has been central to CA's concerns – with much of the
205 early development of CA by Sacks drawing on two collections of audio recordings: calls
206 to the suicide hotline, and group therapy sessions (Sacks, 1992). Over the following
207 decades, a number of phone call corpora were created, notably recordings taken from
208 around Santa Barbara in California which the CA community refer to as the Newport
209 beach corpus or more commonly: "Classic data", and Elizabeth Holt's "Holt corpus" of
210 phone calls recorded by a British family over three years. These are data which are widely
211 available and widely reused and on which much of the ground-breaking work in CA is

[4] The cumulative relationship between CA and QDS is unique but the specific practices and procedures are not unique to the approach.

[5] For a fuller picture of the founding of CA see: Psathas (1994) ten Have (2007) Sidnell (2011) Silverman (1998).

based[6]. The (re)use of this data speaks to a core ideal in CA – that data used in research should be made available to check findings (often in the form of transcripts, but also with visual representations (see Walker, 2017) or in the sharing of audio/video data). Sacks illustrates this point:

> "It was not from any large interest in language or from some theoretical formulation of what should be studied that I started with tape-recorded conversations, but simply because I could get my hands on it and I could study it again and again, and *also, consequentially, because others could look at what I had studied and make of it what they could, if, for example, they wanted to be able to disagree with me"*
>
> (Sacks, 1984, p. 26, emphasis added).

The technology of the time influenced the practices of the discipline: recordings could be replayed, scrutinised by others and made available for future studies by other researchers. Sacks goes on to explain how he chooses the data he works with:

> "People often ask me why I choose the particular data I choose. [..] And I am insistent that I just happened to have it, it became fascinating, and I spent some time at it" (Sacks, 1984, p. 27).

Data which CA researchers *just happen to have* has been used and reused in a number of studies addressing a wide range of interactional phenomena. CA with its grounding in ethnomethodology examines the observable practical common-sense reasoning as revealed in the data itself to make sense of how the social world is constituted in local environments. This means the distinction between primary and secondary analyses

[6] Most modern CA research no longer draws on classic data with that collection being normally reserved for teaching.

disappears because the source of evidence is always constituted for the first time. This fits with Moore's (2007) understanding that in data reuse, analysis is always primary but of a *different order of data*, Hughes et al. (2020) extend this position by articulating "the range of approaches and practices involved in producing different orders of data" (p. 568). One example is Gibson's (2019) primary (rhetorical) analysis of Milgram's classic obedience experiment data which offers a reinterpretation of the core insights around obedience and persuasion. Similarly, Hughes et al. (2020) show how interview data might be reused to examine features of relational dynamics. In this way QDS can open up novel avenues of research and lead to further scrutiny of prior findings[7].

Opening up novel avenues of research is one benefit of QDS, other benefits are explored by Jepsen and colleagues who reflect on reasons for creating their primary care consultations archive: a corpus of recordings of GP consultations, linked survey responses and patient records named the 'One in a Million' corpus:

> "Data sharing provides considerable added value in terms of minimising data collection costs, reduced environmental impact, and patient and practice burden. This will support low-cost studies including doctoral-level research, thus building research capacity in primary care." (Jepson et al., 2017, p. 350).

Their argument expands Sacks' point – that a chief reason for sharing data is so that other researchers (particularly those who are at an early stage of their careers) can just *happen to have* that data. This is a point the paper will return to later, but first it is necessary to describe the kinds of data that CA usually works with and the essential characteristics of that data.

---

[7] See Humă & Joyce (frth) for a discussion on the relationship between the culture of data sharing and the culture of continuous refinement and replication in CA.

CA draws on recordings of naturally occuring[8] social interaction. This can include any site where interaction occurs between participants, including (but not limited to): Online chat logs (e.g. Meredith & Stokoe, 2014), Institutional encounters (e.g. Drew & Heritage, 1992), Phone calls (Holt, 1996), AI (e.g. Suchman, 2007; Mair et al., 2020), Video recordings (e.g Mondada, 2018) and so on. In this vein, research interviews can be viewed as a site of interaction (Potter & Hepburn, 2012). CA research can be broadly placed within one of two camps: 'pure CA' and 'Applied CA'. Antaki (2011) explains that 'pure CA' focuses on interactional practices and procedures detached from *any* type of context (e.g. Jefferson, 1988) and takes an endogenous orientation to the conversation itself rather than drawing on analytic insights about the institutional context. Compared to 'applied CA' which focuses on interactional practices *within* a certain setting (e.g. Drew & Heritage, 1992; ten Have, 2007) and provides an evidence-base for interventions (e.g. Stokoe, 2014; Wilkinson, 2015). A discussion of the debate regarding the two terms can be found in Antaki (2011).

The procedures for both camps are largely the same – recordings of social interaction are gathered and analysis proceeds with 'unmotivated looking'; that is, as Psathas (1990) explains, the researcher discovering what is happening in the recordings and not searching for predetermined phenomenon. Data collection of interaction recordings do not normatively involve the researcher which thus enhances the usefulness of the data for reuse and reanalysis. To outsiders the unmotivated looking and efforts to

[8] "Naturally occuring" is a slogan in the CA enterprise and usually contrasts with researcher elicited data or scripted talk, but see the debate in Discourse Studies which problematises the 'natural' and 'non-natural' data distinction (Speer, 2002a; 2002b; ten Have, 2002; Lynch, 2002; Potter, 2002).

remain exogenous to the data collection process may seem unstructured and haphazard, but the methodological technology imposes a high degree of rigour to account for and evidence unmotivated discoveries (see Liddicoat, 2007, p. 9 and Schegloff, 1996a, p. 172-173 on accounting for phenomena). In short, data analysed in this way aims to avoid mediation by the subjective perspective of the researcher.

Despite the growing range of data that CA researchers draw upon, core data sharing principles remain unchanged since its foundation – that others ought to have access to the data, including, ideally, the original video/audio recordings, so to scrutinise the analysis of the researcher and that data is usually made available for pre-publication data sharing sessions (referred to as "data sessions") as an integral part of the method. Recordings which are particularly sensitive may be subject to greater sharing restrictions which may be mitigated by heavily anonymising the recordings (e.g. voice altering, video manipulation etc.), or by asking data session participants to sign a non-disclosure agreement and return all materials after a meeting. These more extreme measures are often the result of ethics committee requirements, and not of the science itself, with many sensitive anonymised data sets shared without such restrictions in place.

To summarise, over the course of CA's history the core principle that data should be shared and reused has established formal and informal practices for handling data. The remainder of the paper describes CA's understanding of and approach to ethics and epistemology, and explores CA's established procedures and practices for sharing data with the intent of widening ongoing debates and allaying some of the persistent concerns in qualitative research about data sharing.

Ethics, Context and Conversation Analysis

Ethics

Conversation Analysts deal with many of the same ethical dilemmas experienced by

other forms of qualitative research. For CA studies, which collect recordings of social interaction, ensuring anonymity for participants in shared data can be technically complex. It minimally requires the deletion of names, dates and locations. However, other components that make participants identifiable require more complex anonymisation decisions, for instance, anonymising voices and faces, or whether to remove specific details such as references to a participant's medical condition. Moreover, a participant may, in the course of a recording, indicate (in the recording) that some part should be anonymised through either explicit mention (e.g. Speer & Hutchby, 2003), or by blocking the recording equipment (e.g. Mondada, 2014). There is a profound understanding in CA that simply removing names, dates and locations might not always be sufficient for anonymisation.

It is not possible to predict every ethical dilemma which may arise in the course of research hence ethical solutions cannot be prescribed *a priori*. These ethical questions will persist as (hopefully all) researchers endeavour to protect their participants from harm. However, when the possibility of data sharing is built into research procedures ethical safeguards become even more central to the research design (see Albert & Hofstetter, frth for a discussion). This includes providing information to participants about data reuse (and its associated risks) along with consent forms which allow participants to decide whether, and in what forms/contexts, their data may be shared for future research[9]. This has, for example, been the approach taken in the

[9] Participants are rarely in a position to fully understand the research process and a discussion of this and how ethics panels are not geared to handle qualitative data sharing warrants a future paper (but see Hammersly, 2014; ten Have, 2007: 79-81).

National Institute for Health Research (NIHR) funded CA project known as 'Real Complaints' (Real Complaints, 2021).

Context

CA can uniquely contribute to debates about the problem of context in QDS. Precisely what is meant by 'context' is arguably "fuzzy" (van Dijk, 2007, p. 285) across the social sciences, as it can be a shorthand to denote a specific situation, or the historical/geographical/cultural environment, of the object of investigation. However, CA's particular way of dealing with 'context' does not entertain contextual explanations of phenomena. Handling context in CA has been debated at length[10] (see ten Have, 2007, p. 58-59 and Wooffitt, 2005, p. 168-179 for reviews). In short, CA does not assume that aspects of context such as social categories (race, gender, power, class etc.) are relevant *a priori*. Rather, context is dealt with analytically if, and only if, it is *procedurally relevant* and *demonstrably attended* to by the interlocutors themselves[11] (see Schegloff, 1992). Hence Irwin's (2013) concerns about the contextual qualities of qualitative data are not normally relevant in CA.

This should not be read as necessarily advocating for this way of handling context in qualitative research generally. The point being made is that when sharing data, it cannot be foreseen how it may be (re)used. The endogenous understanding of context espoused

[10] This was discussed and responded to at length between Emmanuel Schegloff (1997; 1998; 1999b; 1999c), and Margaret Wetherell (1998) and Michael Billig (1999a; 1999b) who took issue with Schegloff's original 1997 paper.

[11] It is this additional step that distinguishes CA from other inductive approaches such as ethnography and grounded theory.

by CA means that it does not carry any 'burden' of externally imposed context to delimit what it can be used to demonstrate. It cannot be expected that the data which researchers share will only be used by those within the same discipline or even those who share similar interests – rather, researchers ought to anticipate that the shared data may be used beyond the scope of the original research, (see the previous discussion on the practices of producing different orders of data (Hughes et al. 2020)).

Returning to discussion of the extent to which data may be meaningfully interpreted by researchers outside of the original research, new and fruitful avenues of investigation can be found in the reuse of data collected for alternate purposes (e.g. Gibson, 2019; Hughes et al. 2020). For certain qualitative approaches the arguments by Irwin (2013) and Chuavette et al. (2019) are salient, but different approaches with different epistemologies may make use of data in ways unforeseen by the original researchers. Data collected for one particular purpose may be meaningfully reinterpreted with CA because of its focus on phenomena demonstrably enacted and treated as relevant by participants in the discourse.

Data Sharing in Conversation Analysis: Practical Aspects

As a fundamentally collaboratively discipline CA has fostered a culture[12] and tradition of data sharing out of which has emerged a community of practice:

> "CA is a *community*, although with various degrees of intensity. As it has become established as a quite solidly and specifically defined approach in the human sciences, you can, by working in the CA tradition, become 'a member' of that community" (ten Have, 2007, p. 11)

[12] We refer to "culture" in the sense of disciplinary culture rather than epistemology.

Typically, research which makes its data available does so following completion of the project – whether defined as the publication of an (final) article, or the overall conclusion of the funding period. Data collections may be described on websites such as the Open Science Framework during the research process but are not commonly available until after project completion. We refer to this widespread form of QDS as 'corpus sharing' distinguished from the practices and solutions employed by the CA community to add levels of transparency and rigour to the analysis, which we refer to as 'data sharing as a research practice'. This section returns to the fears outlined previously and discusses the practical aspects of the CA approach to data sharing both as corpus sharing and as established research practice.

Data Sharing as a research practice

CA is a community of practice with a particularly democratic impulse – that both the analysis and the research process builds from the ground up with students, practitioners, and experienced CA researchers able to contribute insights through data sharing sessions. Data sharing is thus "baked into" the research process. During the research process, there are options for qualitative researchers to share data and findings in progress at conferences, seminars and research meetings but CA is distinctive in that focused data sharing meetings (referred to as 'data sessions') are an integral part of the scientific process and disciplinary culture – researchers regularly share their data at data sessions.

Data sessions are structured research meetings where direct access to the recordings is made available to other researchers to scrutinise[13]. Although data is often

[13] Examples of groups include the Conversation Analysis Reading and Data Sessions (CARDS) at Ulster University, and the long-standing Discourse and Rhetoric Group (DARG) at Loughborough University. A list of groups is maintained here: https://rolsi.net/data-sessions/

386 analysed by multiple researchers during the research process across the gamut of
387 qualitative (and quantitative) research approaches, data sessions are distinctive in the
388 sense that data is subject to scrutiny and analysis by others outside of immediate research
389 teams and institution. Findings can be independently checked, and ideas collaboratively
390 explored (ten Have, 2007).

391 The procedures of a data sharing session can vary amongst research groups, but
392 usually the data presenter shares recordings and transcripts with the group. The transcripts
393 will contain more detail than is perhaps necessary for the data owner's interest "because
394 even if, say, pauses or overlaps are not germane to the current analysis, some other
395 researcher might want to use the same materials for checking findings or for novel
396 analytic purposes" (Jordan & Henderson, 1995, p. 48). Participants see/hear the recording
397 several times as transcripts are recognised to be a static and partial representation of
398 interaction. After a moment or two of quiet 'thinking' time, the group will propose
399 observations. There is usually some rule that group members are initially limited to a
400 single observation to encourage a collaborative, democratic ethos with equal access to
401 contribute.

402 Crucially, sharing data allows for a more transparent analytic process for the data
403 owner and for learners of the method. The practices and procedures of the data session
404 might be usefully adopted by other qualitative approaches to fit with the Open Science
405 movement (see Humă & Joyce, frth) by not only corroborating findings, but also making
406 explicit the discussions and analysis of data often done behind closed doors. We argue
407 that collaborative analysis of data adds another level of rigour in the analytic process
408 where banal observations may be retold as composed and refined analytic points and
409 flawed analysis, or invalid conclusions recognised and corrected. The value of the data
410 session cannot be overstated and highlights possibilities for data sharing beyond making

data sets available in a repository post-project. Worries about being scooped by sharing one's data prior to publication are greatly outweighed by the benefits of the data session, and indeed, new projects may be launched, and collaborations proposed following such sessions. Sharing is thus baked into the research process through the tradition of the data session through which both new and seasoned researchers and, where relevant, participants involved in the data, are invited to witness and scrutinise data on their own terms.

Corpus sharing

'Data sharing' typically refers to post-project data sharing in a repository – where a final corpus of data is made available to other researchers. This paper treats a corpus of data as data (in whatever form) associated with a single project, whereas a repository of data is a resource where multiple corpora are stored and made accessible to others. The gold standard of data sharing is typically regarded as unrestricted access to data, which is shared with accessibility in mind, and is fully described and indexed so that other researchers can easily search and understand the data (e.g. Mass Observation Archive, British Library Sound Archive), and importantly, check analysis. CA is not unlike other qualitative disciplines in that corpora are held in various places (some, such as TalkBank, are specific to communication data), and although the ideal is unrestricted sharing, in practice there may be gatekeepers or restrictions on accessing data. Moreover, data is often shared through informal networks rather than through a formal repository.

It is impossible to predict how data may be reused which carries with it benefits and drawbacks. Within CA the units of analysis are discursively realised practices and social actions, rather than people or experiences. The approach allows for the study of phenomena whose context is endogenously constituted within the talk itself. In this way, subsequent researchers can ask very different questions about the data, focusing on what

is *being done* in the interaction and not the original purpose of the data collection (which for CA researchers is a context that is not relevant to the analysis). For example, data from Heritage and colleagues' 2007 study investigating how doctors encourage patients to voice concerns in consultations was used by Heritage (2012) in a study focusing on the relationship between epistemic status and stance. Similarly, interview data from a study by Hepburn and Brown (2001) asking how secondary school teachers use 'stress' to manage their accountability and make sense of their institutional role was used by Potter and Hepburn (2005) to critique the (over)use of interviewing in qualitative psychology. This illustrates that *all* data irrespective of method of collection might be repurposed to generate novel findings potentially in novel ways.

The generation of findings is, however, only one argument for post-project data sharing. Conversation Analysts have long advocated for direct access to the data[14] presented in empirical articles. Providing corroborative evidence, which is prepared effectively (see Walker, 2017), for research claims allows others to independently check those claims. To repeat Sacks's observation, "others could look at what I had studied and make of it what they could" (Sacks, 1984, p. 26). Much of the foundational CA work which reused recordings (e.g., from the Newport Beach corpus or the Holt corpus) had such an impact in the community because fellow researchers were familiar with the data and able to independently check findings.

Conclusion

This paper contributes to discussions around the viability and usefulness of QDS, adding insights from the established traditions of CA to widen those discussions, and to advocate

[14] 'Direct access' may be confused with access to the in-the-moment encounter, but here we refer to the original recording of the encounter.

for a more open and flexible approach to QDS. Ongoing debates emphasise the ethical and epistemological barriers to QDS and are framed by a distinction between primary and secondary data. We argue that CA's conception of data and context makes this distinction redundant. Currently, the demands of funders, Open Science and legal restrictions influence decisions about what gets shared and how, but inexperience and lack of consensus on best practice for QDS persist. Our aim has been to reflect on the long history of sharing data in CA, the impetus for sharing within the CA community, and how these procedures might be drawn on by other qualitative approaches.

We discuss two types of data sharing that are baked into the design of CA studies: sharing as research practice, and corpus sharing. We show how the 'data session', while not unique to CA (see also, grounded theory), enables the research process to build from the ground up and that this collaborative analysis adds rigour to the analytic process. Corpus sharing is a more traditional understanding of data sharing – and while matters of context and ethics present as barriers to reusing data, for conversation analysts, having access to the original recordings of analysed encounters is considered gold-standard. The expectations, tools and procedures of CA facilitate more transparent QDS within the community but as with most other approaches they rely on the researcher(s) having sufficient means to engage in QDS.

Beyond the practical barriers to engaging in sharing as a research practice or establishing and sharing a data corpus, many authors point to significant ethical and epistemological barriers for QDS. For ethics, data reuse presents challenges for informed consent and the high level of anonymisation potentially required might make the data difficult to work with. While we accept that attempting to solve all ethical issues relating to participants consent or prescribing ethical solutions *a priori* is a fool's errand, building in the possibility of data sharing into the research design, as CA does, foregrounds ethical

safeguards which can alleviate potential dilemmas. For epistemology, many qualitative approaches consider reflexivity a central practice making secondary analyses impossible and indeed, different qualitative approaches conceive of 'context' very differently which again, makes the data difficult to work with. CA, as an illustration, does not face these concerns. The distinct way that CA conceives of 'context' dissolves the distinction between 'primary' and 'secondary' data meaning that data is always constituted for the first time. Unlike a number of qualitative approaches, CA is not 'burdened' by externally imposed context to delimit what it can be used to demonstrate. To be clear, we are not advocating that all qualitative approaches follow CA's conception of context but instead are arguing that future use of any data can never be predicted and that all data, irrespective of method of collection, might be repurposed to generate novel findings in potentially novel ways.

We have demonstrated that sharing insights from CA can allay fears and barriers to QDS and that the long-established and refined tools of CA make QDS much more achievable. CA is, however, not a panacea for all QDS challenges and for many qualitative researchers, particularly early-career researchers, or those in marginalised areas, barriers to QDS – whether ethical, epistemological, or economic – can prove difficult to overcome without sufficient support and funding and thus they may be reluctant to share their data (Pownall et al., 2021) which can adversely impact career outcomes (Siegel & LaMarre, 2019). This is a crucially important topic which we have not discussed at length and so encourage further scholarship on this issue.

We conclude by reiterating Sacks' (1984, p. 26) explanation of how he came to study the data that he did: "simply because I could get my hands on it and I could study it again and again, and also, consequentially, because others could look at what I had studied and make of it what they could". CA was built on the ideal that data *should* be

shared for the benefit of the primary investigator and the research community. The overall intention of this paper was to spark further discussion of what QDS could look like across the range of qualitative approaches.

References

Albert, S. & Hofstetter, E. (frth). Data Management 1: Privacy, Security and Access.

Antaki C. (2011). Six Kinds of Applied Conversation Analysis. In C. Antaki (Ed.), *Applied Conversation Analysis* (pp. 1-14). Palgrave Advances in Linguistics. Palgrave Macmillan, London.

Antonio, M. G., Schick-Makaroff, K. Doron, L.S., White, L. & Molzahn, A. (2019). Qualitative Data Management and Analysis within a Data Repository. *Western Journal of Nursing Research,* 42(8), 640-648. https://doi.org/10.1177/0193945919881706

Berg, H. v. d. (2008). Reanalyzing qualitative interviews from different angles: the risk of decontextualization and other problems of sharing qualitative data. *Historical Social Research,* 33(3), 179-192. https://doi.org/10.17169/fqs-6.1.499

Billig, M. (1999a). Whose terms? Whose ordinariness? Rhetoric and ideology in conversation analysis. *Discourse & Society,* 10, 543–58. https://www.jstor.org/stable/42888872

Billig, M. (1999b). Conversation analysis and the claims of naivety. *Discourse & Society,* 10, 572–6. https://doi.org/10.1177/0957926599010004007

Bishop, L. (2007). A reflexive account of reusing qualitative data: Beyond primary/secondary dualism. *Sociological Research Online, 12*(3), 43-56. https://doi.org/10.5153/sro.1553

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

Bishop, L. & Kuula-Luumi, A. (2017). Revisiting Qualitative Data Reuse: A Decade On. *SAGE Open*, 1-15. https://doi.org/10.1177/2158244016685136

Chauvette, A. Schick-Makaroff, K. & Molzahn, A.E. (2019). Open Data in Qualitative Research. *International Journal of Qualitative Method,* 18, 1-6. https://doi.org/10.1177/1609406918823863

Corti, L., Day, A. & Backhouse, G. (2000). Confidentiality and informed consent: Issues for consideration in the preservation of and preservation of access to qualitative data archives. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Researc*h http://www.qualitative-research.net/fqs-texte/3-00/3-00cortietal-e.htm.

Corti, L. Fielding N. & Bishop, L. (2016). Editorial for Special Edition, Digital Representations: Re-Using and Publishing Digital Qualitative Data. *Sage Open*, 1-3. DOI: 10.1177/2158244016678911

Drew, P. & Heritage, J. (1992). *Analyzing Talk at Work. Interaction in Institutional Settings*. Cambridge: Cambridge University Press.

DuBois, J. M., Strait, M. & Walsh, H. (2018). Is It Time to Share Qualitative Research Data? *Qualitative Psychology,* 5(3), 380-393. doi: 10.1037/qup0000076

Garfinkel, H., Lynch, M. & Livingstone, E. (1981). The work of a discovering science construed with materials from the optically discovered pulsar. *Philosophy of the Social Sciences,* 11(2), 131–158.

Gibson, S. (2019). *Arguing, Obeying and Defying: A Rhetorical Perspective on Stanley Milgram's Obedience Experiments*. Cambridge University Press.

Hammersley, M. (2010). Can We Reuse Qualitative Data Via Secondary Analysis. Notes on Terminological and Substantive Issues. *Sociological Research Online,* 15(5), 1-7. https://doi.org/10.5153/sro.2076

Hardy, L. J., Hughes, A., Hulen, E. & Schwartz, A. L. (2016). Implementing qualitative data management plans to ensure ethical standards in multi-partner centers. *Journal of Empirical Research on Human Research Ethics,* 11(2), 191 198. https://doi.org/10.1177/1556264616636233

Harris, J., Theobald, M. A., Danby, S. J., Reynolds, E., & Rintel, S. (2012). "What's going on here?" The pedagogy of a data analysis session. In A. Lee & S. J. Danby (Eds.), *Reshaping doctoral education: International Approaches and Pedagogies* (pp. 83–96). London: Routledge.

Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research/Historische Sozialforschung,* 33(3), 33-45.

Hepburn, A. & Brown, S. D. (2001). Teacher stress and the management of accountability. *Human Relations,* 54(6), 691-715.

Heritage, J., Robinson, J. D., Elliott, M. N., Beckett, M. & Wilkes, M. (2007). Reducing patients' unment concerns in primary care: the difference one word can make. *Journal of Gen Intern med,* 22(10), 1429-33.

Heritage, J. (2012). Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language & Social Interaction,* 45(1), 1-29. https://doi.org/10.1080/08351813.2012.646684

Holt, E. (1996). Reporting on Talk: The use of Direct Reported Speech in Conversation. *Research on Language and Social Interaction,* 29(3), 219-245.

Hopkins, M. (1993). Is anonymity possible? Writing about refugees in the United States. In C. B. Brettell (Ed.), *When they read what we write: The politics of ethnography* (pp. 121-129). Westport, CT: Bergin & Garvey.

Hughes, K., Hughes, J. & Tarrant, A. (2020). Re-approaching interview data through qualitative secondary analysis: interviews with internet gamblers. *International Journal of Social Research Methodology*, 23(5), 565-579. https://doi.org/10.1080/13645579.2020.1766759

Humă, B. & Joyce, J. B. (frth). 'One Size Doesn't Fit All'. Lessons from Interaction Analysis on Tailoring Open Science Practices to Qualitative Research.

Irwin, S. (2013). Qualitative Secondary Data Analysis: Ethics, Epistemology and Context. *Progress in Development Studies,* 13(4), 295-306. https://doi.org/10.1177%2F1464993413490479

Jefferson, G. (1988). On the sequential organization of troubles talk in ordinary conversation. *Social Problems,* 35(4), 418-442.

Jepson, M., Salisbury, C., Ridd, M. J., Metcalfe, C., Garside, L. & Barnes, R. (2017). The 'One in a Million' study: creating a database of UK primary consultations. *British Journal of General Practice*, 67(658), e345-e351.

Jordan, B. & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences,* 4(1), 39-103.

Kuula, A. (2011). Methodological and Ethical Dilemmas of Archiving Qualitative Data. *IASSIST Quarterly*, 34(3-4), 12-17.

Law, M. (2005) Reduce, reuse, recycle: Issues in the secondary use of research data. *IASSIST Quarterly,* 29, 5–10.

Liddicoat, A. (2007). An Introduction to Conversation Analysis. London: Continuum.

Lynch, M. (2002). From naturally occuring data to naturally organized ordinary activities: comment on Speer. *Discourse Studies*, 4(4), 531-537.

Lynch, M. (1993). Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science. New York, Cambridge University Press.

618 Mair, M., Brooker, P., Dutton, W. & Sormani, P. (2020). Just what are we doing when
619 we're describing AI? Harvey Sacks, the commentator machine, and the
620 descriptive politics of the new artificial intelligence**.** *Qualitative Research,*
621 21(3), 341-359. https://doi.org/10.1177%2F1468794120975988

622 Mauthner, N.S., Parry, O. & Backett-Milburn, K. (1998). The data are out there, or are
623 they? Implications for archiving and revisiting qualitative data. *Sociology, 32*(4),
624 733-745.

625 Mauthner, N.S. & Parry, O. (2009). Qualitative data preservation and sharing in the
626 social sciences: On whose philosophical terms?. *Australian Journal of Social*
627 *Issues,* 44(3), 291-307. https://doi.org/10.1002/j.1839-4655.2009.tb00147.x

628 Maynard, D. & Clayman, S. E. (1991). The diversity of ethnomethodology. *Annual*
629 *Review of Sociology,* 17, 385–418

630 Meredith, J. & Stokoe, E. (2014). Repair: Compaing Facebook 'chat' with spoken
631 interaction. *Discourse & Communication,* 8(2), 181-207.
632 https://doi.org/10.1177%2F1750481313510815
633
634 Milgram, S. (1974). *Obedience to Authority: An Experimental View*. Harper & Row.
635
636 Mondada, L. (2014). Ethics in Action: Anonymisation as a Participant's Concern and a
637 Participant's Practice. *Human Studies,* 37(2), 179-209.
638
639 Mondada, L. (2018). The multimodal interactional organization of tasing: Practices of
640 tasting cheese in gourmet shops. *Discourse Studies,* 20(6), 743-769.
641 https://doi.org/10.1177%2F1461445618793439
642
643 Moore, N. (2007). (Re) using qualitative data?. *Sociological Research Online,* 12(3), 1-
644 13.

645 Mozersky, J. Walsh, H. Parsons, M. McIntosh, T. Baldwin, K. and DuBois, J.M.
646 (2020a). Are we ready to share qualitative research data? Knowledge and
647 preparedness among qualitative researchers, IRB Members, and data repository
648 curators. *The International Association for Social Science Information Service*
649 *and Technology Quarterly,* 8(43), 1-26. doi:10.29173/iq952.

650 Mozersky, J., Parsons, M., Walsh, H. Baldwin, K. McIntosh, T. & DuBois, J.M.
651 (2020b). Research Participant Views Regarding Qualitative Data Sharing. *Ethics*
652 *and Human Research,* 42(2), 13-25. https://doi.org/10.1002/eahr.500044

653 Parry, O. & Mauthner, N.S. (2004). Whose Data Are They Anyway? Practical Legal
654 and Ethical Issues in Archiving Qualitative Research Data. *Sociology,* 38(1),
655 139-152. https://doi.org/10.1177%2F0038038504039366
656
657 Parry, R., Pino, M., Faull, C. & Feathers, L. (2016). Acceptability and design of video-

based research on healthcare communication: evidence and recommendations. *Patient education and counselling*, 99(8), 1271-1284.

Potter, J. (1996). Discourse Analysis and Constructionist Approaches: Theoretical Background. In J.T.E. Richardson (ed.) *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*. Leicester: BPS. Books: 125 -140.

Potter, J. (2002). Two kinds of natural. *Discourse Studies*, 4(4), 539-542.

Potter, J. & Hepburn, A. (2005). Qualitative interviews in psychology: problems and possibilities. *Qualitative Research in Psychology,* 2, 281-307. https://doi.org/10.1191/1478088705qp045oa

Pownall, M., Talbot, C. V., Henschel, A., Lautarescu, A., Lloyd, K. E., Hartmann, H., Darda, K. M., Tang, K. T. Y., Carmichael-Murphy, P. & Siegel, J. A. (2021). Navigating Open Science as Early Career Feminist Researchers. Psychology of Women Quarterly, 45(4), 526-539. https://doi.org/10.1177%2F03616843211029255

Psathas, G. (1990). Introduction: Methodological issues and recent developments in the study of naturally occurring interaction. In G. Psathas (Ed.), *Interaction Competence* (pp. 1-30). Washington DC: International Institute for Ethnomethodology and Conversation Analysis.

Psathas, G. (1994). *Conversation Analysis: The Study of Talk-in-Interaction* (Qualitative Research Methods). SAGE Publications.

Real Complaints (2021). *Information for Participants*. Accessed 18.04.21. Available at: https://www.realcomplaints.org/research.

Sacks, H. (1984). Notes on Methodology. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 21-27). Cambridge: Cambridge University Press.

Sacks, H. (1992). *Lectures on Conversation*, vols. 1 and 2, edited by G. Jefferson. Oxford: Basil Blackwell.

Schegloff, E. A. (1992). In Another Context. In A. Duranti & C. Goodwin (Eds.), *Rethinking Context: Language as an Interactive Phenomenon* (pp. 191-227). Cambridge: Cambridge University Press.

Schegloff, E.A. (1997). Whose text? Whose context? *Discourse & Society,* 8, 165–87. https://doi.org/10.1177%2F0957926597008002002

Schegloff, E.A. (1998). Reply to Wetherell. *Discourse & Society,* 9, 413–16. https://doi.org/10.1177%2F0957926598009003006

Schegloff, E. A. (1996a). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology,* 104, 161-216. https://www.jstor.org/stable/2782190

Schegloff, E.A. (1999b). "Schegloff's Texts" as "Billig's Data": a critical reply. *Discourse & Society,* 10, 558–72. https://doi.org/10.1177%2F0957926599010004006

Schegloff, E.A. (1999c). Naivety vs. sophistication or discipline vs. self-indulgence: a rejoinder to Billig. *Discourse & Society,* 10, 577–82

Sidnell, J. (2011). *Conversation Analysis: an introduction.* London: Wiley-Blackwell.

Siegel, J. A. & LaMarre, A. (2019). Navigating "publish or perish" as qualitative researchers. *Nature behavioural and social sciences.* https://socialsciences.nature.com/posts/54648-navigating-publish-or-perish-as-qualitative-researchers

Silverman, D. (1998). *Harvey Sacks and Conversation Analysis.* Cambridge: Polity Press.

Speer, S. A. (2002a). 'Natural' and 'contrived' data: a sustainable distinction? *Discourse Studies*, 4(4), 511-525.

Speer, S. A. (2002b). Transcending the 'natural'/'contrived' distinction: a rejoinder to ten Have, Lynch and Potter. *Discourse Studies*, 5(5), 543-548.

Speer, S. A. & Hutchby, I. (2003). From Ethics to Analytics. Aspects of Participants' Orientations to the Presence and Relevance of Recording Devices. *Sociology,* 37(2), 315-337. https://doi.org/10.1177%2F0038038503037002006

Stevanovic, M. & Weiste, E. (2017). Conversation-analytic data session as a pedagogical institution. *Learning Culture and Social Interaction,* 15, 1-17.

Stokoe, E. (2014). The Conversation Analytic Role-play Method (CARM): A method for training communication skills as an alternative to simulated role-play. *Research on Language and Social Interaction,* 47(3), 255-265.

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge: Cambridge University Press.

ten Have, P. (2002). Ontology or methodology? Comments on Speer's 'natural' and 'contrived' data: a sustainable distinction? *Discourse Studies*, 4(4), 527-530.

ten Have, P. (2007). *Doing Conversation Analysis*. SAGE Publications.

Tsai, A.C., Kohrt, B.A., Matthews, L.T., Betancourt, T.S., Lee, J.K., Papachristos, A.V., Weiser, S.D. & Dworkin, S.L. (2016). Promises and pitfalls of data sharing in

qualitative research. *Social Science & Medicine,* 169*,* 191-198. https://doi.org/10.1016/j.socscimed.2016.08.004

UK Research and Innovation (UKRI) (2021). *Open research.* Available from: https://www.ukri.org/about-us/policies-standards-and-data/good-research-resource-hub/open-research/ [Accessed 15th April 2021]

Van Dijk, T. A. (2007). Comments on Context and Conversation. In N. Fairclough, G. Cortese, & P. Ardizzone (Eds.), *Discourse and Contemporary Social Change* (pp. 290-295). Bern: Peter Lang.

Walker, G. (2017). Visual Representations of Acoustic Data: A Survey and Suggestions. *Research on Language and Social Interaction,* 50(4), 363-387. https://doi.org/10.1080/08351813.2017.1375802

Wetherell, M. (1998). Positioning and interpretive repertoires: conversation analysis and poststructuralism in dialogue. *Discourse & Society,* 9, 387-412. https://doi.org/10.1177%2F0957926598009003005

Williams, B., Dowell, J., Humphris, G., Themessl-Huber, M., Rushmer, R., Ricketts, I., Boyle, P. & Sullivan, F. (2010). Developing a longitudinal databse of routinely recorded primary care consultations linked to service use and outcome data. *Social Science & Medicine,* 70(3), 473-478. DOI: 10.1016/j.socscimed.2009.10.025

Wilkinson, R. (2015). Conversation and aphasia: advances in analysis and intervention. *Aphasiology,* 29(3), 257–268. https://doi.org/10.1080/02687038.2014.974138

Wooffitt, R. (2005). *Conversation analysis and discourse analysis: a comparative and critical introduction*. London: Sage.

Acknowledgements

We would like to thank Emily Hofstetter, Saul Albert and Helen Baron who commented on an early draft of the paper. We would also like to thank the reviewers and editors for their constructive suggestions.

Contributions

TD conceived the initial idea for the paper. The manuscript was drafted by JJ and TD. BB, CR and JJ contributed to the development of key ideas, the structure, the rewriting of several sections of the manuscript, and the editing of the manuscript. RP and RS helped develop the initial ideas on which the paper was written, and, along with AK offered feedback and guidance on drafts of the paper.