Whose data is it, anyway?

Qualitative criminology, prisons, and the Open Research agenda

Ben Jarman
University of Southampton Law School
b.jarman@soton.ac.uk

2025-02-12

Abstract   Recent years have seen increasing calls from funders and publishers for researchers to share their data openly. While the underlying aims of reproducibility and transparency are laudable, they create particular challenges for qualitative researchers, whose work emerges from relationships of trust and contains sensitive personal information. This talk examines these tensions through the lens of prison research, where issues of confidentiality, consent, and institutional power are especially sharply drawn. Drawing on examples, the talk develops broader insights about the nature of qualitative data and its relationship to research openness. It describes an alternative framework for conceptualising qualitative data as Dialogic, Emergent, Abundant and Relational (DEAR, see Westbury *et al.* 2022), and suggests that this offers a more productive angle on the underlying question (transparency) than alternatives. A recent debate in US criminology is referenced to illustrate how academic disciplines might engage constructively with this underlying issue. The talk concludes by reflecting on how qualitative researchers might engage with the open research agenda while maintaining their methodological and ethical commitments.

Contents

Setting the scene

A dilemma

A few years ago, a colleague of mine was in a difficult position. They had been engaged in a fraught back-and-forth with a funder, about the publication of data from a proposed project which would involve detailed ethnographic interviews with people in prison. Of central interest to the project was how participants felt about the reasons for their imprisonment. In many cases, would involve them narrating the circumstances surrounding a serious violent offence.

The funder, as is now standard, required the full anonymised dataset to be made available in a data repository. My colleague's worry was that this requirement, imposed at the project's inception, would entail the publication of highly personal and sensitive information about participants' lives, something they might be ill-equipped to consent to. Some interviews would lead to quite deep and personal territory, and invite participants to describe deeply discrediting acts. All of this would put the researchers in a tricky position. Having secured participation on the basis that people's stories would be handled and written up with respect, they would then publish an anonymised dataset online.

My colleague had been unsuccessful in trying to push back on the funder's publication requirement, and felt that their principled objections were being ignored. They were seriously considering arguably oppositional responses./ Was it possible, for instance, in publishing the data, to specify access restrictions so stringent that the data would effectively be impossible to obtain? Or could such broad redactions be applied in de-identifying the interviews, that they would effectively become unusable by others?

I've opened with this story because it exemplifies a broader tension between what we can broadly call an Open Research agenda, and the distinctive ethical and methodological commitments involved in qualitative research. My colleague's position left me feeling conflicted. My instincts pull strongly towards openness, and towards the principles I associate with the Open Research agenda: in particular, the democratisation of knowledge, and the idea that there is a tight bond between the transparency, rigour, and credibility of knowledge.

But I could also understand my colleague's position very well. Here was someone whose commitment to research ethics I knew to be sincere, and whose judgement I saw as exemplary, considering options which would never have seemed appropriate had they not been backed into a corner. It seemed that a broad top-down agenda for openness, imposed in some sense from outside, was the cause of this anxious, frustrated, defensive response. And it got me thinking about my own plans for the interviews from my own PhD.

I had been intending to publish an anonymised version of my own interviews. Like I said, my instincts favour openness, and part of the reason I began the PhD was because of my past work for voluntary sector organisations in prisons. These are places in which there is a compelling public interest, and yet they are very difficult for any outsider to understand, particularly from

the largely silenced perspective of prisoners themselves. Unlike my colleague, I had the luxury of a choice. I was funded by an ESRC studentship, but publication is only required for grant-funded studies.

I ultimately decided *against* making my interview data openly available. I'll describe a few of the reasons for that decision later, but I was aware at the time I would not have this choice in the future, and so I've tried to keep thinking about how openness interacts with the qualitative traditions. This talk will package up some of that thinking, and situate it in some subsequent debates.

What is 'openness'?

Before delving into all of that, I want to very briefly clarify a few terms I'm going to use repeatedly today.

By Open Research, I mean what Michael Nielsen (n.d.) has called "the idea that scientific knowledge of all kinds should be openly shared as early as it is practical in the discovery process." This is the broadest notion of openness, and from it flow some other ideas.

Second, we have Open Access, which most of us will by now be familiar with. Simply put, it reflects a specific commitment that publicly-funded research should be publicly available, not hidden behind paywalls. In some sense, the research belongs to the public, and should be available to the public.

Finally, there's Open Data, which extends this principle of openness to research data itself. This is my definition, but I want to highlight the idea that it involves a set of normative claims about the relationship between research data and research quality. It is associated with the idea that we should publish not just our findings, but the underlying data that supported them, making it available for reuse and verification. And it's associated with various frameworks for thinking about data.

I'll come back to those frameworks later, but for now I want to note that these principles are not operationalised identically across different academic disciplines. In STEMM subjects and some quantitative social sciences, this normative agenda is particularly strongly articulated, for good reasons associated with reproducibility and the cost/benefit implications of secondary data reuse.

The social sciences more broadly have moved more slowly, and criminology especially so (Ashby 2020). There is some evidence of unease among qualitative researchers about Open Data (Bucerius & Copes 2024; e.g. Hanchard & San Roman Pineda 2023). As my colleague's example illustrates, this isn't a coincidence. Qualitative data give rise to particular ethical and practical questions.

Even so, data publication is increasingly becoming a default practice, driven from above by funders. Currently, UKRI expects data arising from its funding "to be made as open as possible and as restricted as necessary".[1] There have been similar developments with publicly funded research in the United States (White House Office of Science and Technology Policy (OSTP) 2022) and elsewhere.

---

[1] This policy is under review ('UKRI developing new research data policy framework' 2024), and it appears that UKRI might follow the recommendations of an ESRC-commissioned review, which found that current policies underrecognised both the diversity of data used in the social sciences, and the fact that some forms of data *cannot* be made fully open (Allanson *et al.* 2024).

So this is the wider context for my talk: something we might call an 'Open Data agenda', favouring data publication by default, and making the case in terms derived from the STEMM subjects and quantitative social sciences. How can we as qualitative researchers engage with this agenda?

Prison research as an analytical lens

I want now to say a little more about my field, which is prison research. I claimed before that prisons are places of great public interest but very little public understanding.

This is because, sociologically, they are highly distinctive, in several important ways. First, they are closed institutions, deliberately separated from public view. Given the costs and contested moral status of imprisonment, we should want an informed debate about what happens in them. Yet access by researchers and other outsiders is difficult, and thus so is holding prisons accountable for what they do. Those of us who do gain access occupy a privileged position, and we should try and use it to promote understanding. When I said before that my instincts favoured openness, I meant this, more than I meant a claim about research quality.

A further complication arises from the governance of access. On paper, research is *highly* regulated. You write careful ethics applications saying how you will handle data, and you go through a highly demanding gatekeeping process which rejects many research applications, on grounds which include security, but also prisoners' vulnerability. As a result, there are clear boundaries around matters such as confidentiality.

But researchers have often noted that complex and contested organisational cultures influence prison life, to a far greater degree than the rules. Your relationships with prison officers and prisoners are crucial in conducting the research. Those relationships can be tricky, and sometimes ethically complicated. There are intense and murky subcultures among both staff and prisoners which impose demands that rules and policies do not always foresee. Consent is negotiated; trust is contingent and often temporary; motivations aren't always clear; strong social undercurrents affect who talks to you and what they're prepared to say.

Prisoners, in particular, often have very different reasons for participating in research. Many are cynical about the potential for research to improve things, but some badly want their opinions to be heard; others hope there will be some intangible benefit to participation; some simply want to be let out of their cell for a couple of hours, or welcome the chance to talk with someone from the outside world. Prisoners engage with formal requirements like consent forms in varied ways, so that you sometimes have to really work to make the claim of informed consent a substantive, not a tenuous one.

Paradoxically, these complicated, hidden dynamics are often *exactly* what insiders—both prisoners and staff—say they wish the outside world knew more about.

Let me illustrate this through a specific example. One of my participants, whom I'll call Martin, was in prison for a murder committed in the grip of psychosis and involving extreme forms of violence. When I asked if Martin had a pseudonym he would prefer to be known by, he insisted he'd like to be known by his real name. He was proud of the changes he had made in prison, and the insights he had gained into the way he used and thought about violence. He was proud, too, that he could now walk away from the endless provocations of prison life, without ruminating on them endlessly or seeking vengeance. Martin wanted it known, including by those he'd harmed in the past, that he was no longer that person. But I couldn't

grant this request, because it's a condition of access that no participant be identifiable, on the basis that this might cause further harm to their victims.

I'd have liked to use Martin's real name, since on a human level I think his record of desisting from extreme and habitual violence is something worthy of respect, acknowledgment, and perhaps celebration. But the wider rules of the research game mean I can't do that. It's a worthwhile illustration of my theme: who does Martin's story belong to?

My key claim here is that three key tensions that run through all qualitative research, but prison research brings them into especially sharp relief.

First, there is a tension between our transparency and credibility as researchers, and our duties to safeguard our participants and respect their confidentiality. This tension is particularly evident when those participants have inflicted serious harm on others, and when we require them to dwell in places of seclusion and separation.

Second, there is a tension between the formal governance structures we are subject to as researchers, and the underlying reality—especially evident in qualitative research—in which we carry out our investigations. This tension is pronounced when (as in prisons) social life is not entirely, or even mostly, governed formally.

The third tension is between voice and control. On one hand, we have the idea that one reason to use qualitative methods is to centre neglected perspectives, or (more vaguely) to in some sense give participants a 'voice', to respect their autonomous understanding of the world. On the other hand, we face the reality that sometimes the right to a 'voice' is contested, or controversial, and that sometimes, people may find the exposure that comes with having a voice uncomfortable, or regrettable.

I'm claiming that anyone who uses qualitative methods has to think about these dynamics, but also that they are particularly evident in relation to the prison setting.

I now want to step back from that setting, and examine how this kind of insight might cause us to think about managing research data.

Understanding 'openness' and qualitative data

Theoretical frameworks and their limitations

FAIR

The dominant framework, as many of us know, bears the acronym FAIR (Wilkinson *et al.* 2016). It posits that data should be prepared and published to be Findable, Accessible, Interoperable, and Reusable.[2]

These principles, while valuable, emerge from assumptions that don't necessarily hold for the kind of research I've been talking about. One such assumption: the idea that data exists independently of research relationships, that it can be separated from context and reused without a loss of meaning, and that removing identifiers renders it safe to share.

---

[2]Other frameworks are available: one that describes humanities data (CORE—Collected, Organised, Recontextualised and Explained, see Gilby *et al.* n.d.) and one relating to indigenous knowledge (CARE—Collective benefit, Authority to control, Responsibility and Ethics, see Carroll *et al.* 2020).

Returning to Martin's case illustrates some of these problems. The details of Martin's offence, and those of all but one of my other interviewees—can easily be found by searching online for contemporary media coverage of their court cases. It's hard to eliminate the risk that a de-identified transcript of his interview might identify him to anyone with half an hour and the use of a search engine. Carelessness here could have have serious consequences. Some of Martin's counterparts, for instance, were serving time for the murders of children, or offences against other vulnerable victims. Although it is rare, such people have been murdered by other prisoners. We cannot remove everything from which Martin's identity could be deduced, without rendering the data less reusable and less accessible.

Moreover, even were details of the index offence to be entirely redacted, Martin's references to prison social life, in a setting where everyone knows everyone else's business, could make him identifiable to insiders. More generally, participants might express attitudes about violence risk assessments or rehabilitative interventions, which might complicate their progression through the sentence, or otherwise delay their release from prison. What Martin might say to me, he might not choose to say to a parole board.

Once again, this exemplifies broader challenges with qualitative data sharing. I have dwelled mostly on risks associated with him becoming identifiable. But I could have mentioned others, including informed consent. Martin went to prison before the iPhone was invented; some of the men I interviewed had been in prison since I was a toddler, in the early 1980s. In this context, it's really hard to see how I could adequately explain what it meant for anonymised data to be deposited in a research repository. How could I say these individuals had given their informed consent, even if they had signed a carefully worded consent form?

So the very details that make Martin's story valuable for understanding the phenomena we're studying, are the same details that make him identifiable, and it's far from clear that these are data which can be interoperable, or reusable.

So what are we to do? One answer is to wish away the standardised governance, and simply refuse data publication. This is the choice I made with my PhD interviews, but I no longer have the same luxury. Another answer is to comply with data publication mandates, but to do so reluctantly or even with a certain covert resistance, depositing the data while also redacting it so thoroughly or restricting access so stringently that it becomes less useful. Another is to comply fully, and to side-line worries about protecting participants. We can pay lip service to open data, while at the same time rendering the data findable, inaccessible, un-interoperable, and hard to reuse.

These are imperfect choices. But that's true of many choices in research. What I find helpful in such situations is knowing that other researchers with similar standpoints are also grappling with these issues, so that I can look to them for ideas. And that's what I want to shift to now.

I'm going to suggest two things: first, that we need different ways of thinking about what qualitative data is, and how we might approach questions of openness; and second, that we need to do that not from the standpoint of a single, top-down agenda for Open Research, but rather by engaging with that agenda from the distinctive positions we hold as researchers who use qualitative methods.

DEAR

In 2021–22, I worked with an interdisciplinary working group at Cambridge, where I did my PhD, to develop an alternative framework for open research discourse. We were interested in how to respect qualitative researchers' methodological integrity and ethical commitments, while also recognising the reality of funder requirements.

In the process of preparing our report (Westbury *et al.* 2022), a central challenge was methodological heterogeneity. There isn't just one kind of qualitative data, and so approaches to openness need to vary by field. As I've tried to show, each research tradition might consider greater transparency in relation to distinctive concerns, worries, or advantages. In my case, the desire to see more informed debates about imprisonment is one such example.

Members of the working group were therefore not always in accord; but rather than trying to resolve these differences prescriptively, we asked whether a broader, better conceptualisation of qualitative data itself might be a starting point for more generative conversations about 'openness', *within* disciplines and research communities. The idea was to find a more productive terminology with respect to qualitative data than FAIR, and hopefully to facilitate discussion of the underlying issues in a way that did not feel quite so much like being dictated to.

The result was what we called the DEAR framework—standing for the idea that qualitative data is Dialogic, Emergent, Abundant, and Relational. Let me explain what we meant by each of these terms, returning to the prison setting to illustrate:

- **Dialogic**: The data from my conversations with Martin and others didn't simply exist, waiting to be collected. It emerged through dialogue, in a context of trust, which developed during a series of interpersonal conversations about what an interview would involve and why they might talk to me about their lives. A dialogic understanding of data has implications for matters like the accessibility and reusability of the data.
- **Emergent**: The findings from my PhD weren't linear, standardised, or predictable in their relationship with the data. Each conversation led to subtly different insights and modified my understanding of life in the two prisons where I conducted the research. Had I spoken with different participants, or interviewed the same people in a different order, a different view might have emerged. This, again, has implications for ideas like 'interoperability' or 'reusability'.
- **Abundant**: In total, I interviewed 66 men in three prisons about their lives, and made copious fieldnotes and notes from prison documents. Though a small N by some standards, this generated over a million words of interview transcripts, and many observational notes and notes from prison documents. Truly anonymising this body of data would involve a non-trivial quantity of labour, again with implications for what it takes to make it findable and accessible.
- **Relational**: Most importantly, the knowledge the PhD produces exists within, and depended upon, a web of relationships—among me, them, and a set of institutions. I can document and describe those relationships, but the meaning of the data are hard to separate, and there are implications for reuse.

Who exactly 'owns' the data, and the knowledge that comes from it? My funder? Me? My university? My participants? The Prison Service? What I'm trying to say is that I think this isn't a very good question. Instead of asking 'whose data is it?', we might ask, 'what responsibilities

7

arise from how this knowledge came into being?' This shifts us from questions of ownership to questions of stewardship, from property to relationship - and this, I would suggest, offers us more productive ways to think about openness.

I want to shift now to briefly describe a recent debate among American criminologists, which I think covers some of the same ground. I want to comment on how it arose, and how I think it exemplifies the kind of intradisciplinary conversation I'm calling for.

A current debate

In January 2024, the US journal *Criminology* announced that all authors it published must, barring truly exceptional circumstances, deposit their data and code in a repository. This policy mirrored a change in federal research funding rules (White House Office of Science and Technology Policy (OSTP) 2022), and was presented with reference to terms like "reproducibility" and "generalizability". It would require qualitative researchers to upload interview transcripts, fieldnotes, and coding trees, with minimal restrictions.

Two members of the journal's editorial advisory board, both qualitative researchers, published an article (Bucerius & Copes 2024) in the American Society of Criminology's newsletter *The Criminologist*, outlining their dissent from this policy. It was signed by a number of other international editors from the journal's board, and argued that the policy change made a series of unacceptable trade-offs between its stated goals (transparency and quality), and several undesirable outcomes, some of which overlapped with those I've described today.

What's particularly interesting is how quickly the debate (which is ongoing— see La Vigne (2025)) moved beyond data publication, to address wider questions of about the relationship between transparency and quality, and whether Open Data, as conventionally understood, was an appropriate frame for this debate. Some contributions (e.g. Greene-Colozzi & Freilich 2024; La Vigne 2025; Young 2024)addressed how the idea of 'transparency' should be operationalised by qualitative researchers. Some reaffirmed the value of qualitative and mixed methods, but also offered robust criticisms of how difficult it can be to evaluate qualitative research, because of poor documentation of how it was done.

Encouragingly, some responses have also engaged constructively with the practical challenges associated with data publication. Jacques & Wheeler (2024) and Dickinson (2024) have explored technical solutions for anonymisation, while Young (2024) and Greene-Colozzi & Freilich (2024) have advocated principles for the documentation of research workflows, and frameworks to balance participant protection with academic accountability.

What's crucial is that this debate represents some reflection on the kinds of stewardship responsibility which I've suggested are incumbent on qualitative researchers. These changes— the shift to digital data, increasing expectations of transparency—are established realities that require constructive engagement. I want to close with a few suggestions about some practical ways forward.

Ways forward?

Moving from principles to practice

One thing that's important is to understand the landscape better. As I said before, qualitative research is heterogeneous. What do we, and our colleagues, think about open data?

A recent report originating from the University of Sheffield offers some preliminary, non-representative clues. It found that while many qualitative researchers support openness in principle, they struggle with implementation—particularly around anonymisation and how to negotiate openness while respecting the relationality which goes with qualitative inquiry.

The Sheffield team proposed 're-renderability' as an alternative term, to use in place of the more familiar notion of 'reproducibility'. Like some of the US criminologists, they suggested that transparent analytical processes, better documented interpretive approaches, and rich contextual documentation, could render findings more transparent. They considered these measures to be *compatible* with data publication, where that is appropriate, but they do not require the publication of data.

The Sheffield team's work also highlighted how we need institutional support to exercise responsible stewardship of qualitative data. This means:

- Technical support for managing sensitive data
- Resources to prepare data appropriately
- Recognition for the time this work requires
- Space for disciplinary communities to develop appropriate approaches

All of these ideas have implications for how we behave as researchers, and hence there's a basis for constructive engagement there.

Key messages

I'd like to offer a few key messages, in closing.

First, the underlying shifts driving the open data agenda are already underway. Data is normally digital first, and the idea that there is some relationship between transparency and quality isn't going away.

Second, then, our task is to engage constructively. This might still include declining to publish data.

Third, instead of asking 'whose data is it?', or behaving as if the answer to that question is "ours", we should ask what responsibilities arise from how our research knowledge came into being. As stewards of our research relationships and the knowledge they generate, we can find ways to honour both our obligations to participants and our commitment to advance knowledge.

And finally, those are conversations which we need to have both within the wider community of those using qualitative, interpretive methods, but also within our disciplines and and faculties.

[Offer a final note on the UKRI change…]

Allanson, P., Daly, A., Elahi, A., … Tetley-Brown, L. (2024). *Doing ESRC data better: a study for the Economic and Social Research Council (ESRC),* University of Dundee. doi:10.20933/100001312

Ashby, M. P. J. (2020). The Open-Access Availability of Criminological Research to Practitioners and Policy Makers. *Journal of Criminal Justice Education*, **0**(0), 1–21.

Bucerius, S., & Copes, H. (2024). Transparency trade-off: the risks of criminology's new data sharing policy. *The Criminologist*, **50**(2), 6–9.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., … Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, **19**(1). doi:10.5334/dsj-2020-043

Dickinson, T. (2024, December 18). Review of 'A plea for open access to qualitative criminology', Review article, crimRxiv. doi:10.21428/cb6ab371.e42226cb

Gilby, E., Ammon, M., Leow, R., & Moore, S. (n.d.). Open research and the arts and humanities: opportunities and challenges, Cambridge University Working Group on Open Research in the Humanities.

Greene-Colozzi, E., & Freilich, J. D. (2024). The importance of transparency in open-source criminological research. *The Criminologist*, **50**(6), 1–5.

Hanchard, M., & San Roman Pineda, I. (2023). *Fostering open qualitative research - final project report* (report), The University of Sheffield. Retrieved from https://orda.shef.ac.uk/articles/report/Fostering_Open_Qualitative_Research_-_Final_Project_Report/24807987/1

Jacques, S., & Wheeler, A. (2024, October 31). A plea for open access to qualitative criminology: with a python script for anonymizing data and illustrative analysis of error rates. doi:10.21428/cb6ab371.15d7c59e

La Vigne, N. (2025). Redefining rigor: embracing mixed methods research in criminology. *The Criminologist*, **51**(1), 1–6.

Nielsen, M. (n.d.). Michael Nielsen: open science now! | TED talk. Retrieved 3 February 2025, from https://www.ted.com/talks/michael_nielsen_open_science_now

UKRI developing new research data policy framework. (2024, December 13). Retrieved 3 February 2025, from https://www.ukri.org/news/ukri-developing-new-research-data-policy-framework/

Westbury, M., Candea, M., Gabrys, J., … Sharma, C. (2022). *Voice, Representation, Relationships: Report of the Open Qualitative Research Working Group*, Apollo - University of Cambridge Repository. Retrieved from https://www.repository.cam.ac.uk/handle/1810/344820

White House Office of Science and Technology Policy (OSTP). (2022, August 25). Ensuring Free, Immediate, and Equitable Access to Federally Funded Research ('The Nelson Memo'), Executive Office of the President of the United States. Retrieved from https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, **3**(1), 160018.

Young, J. T. N. (2024). Expanding the scope of transparency in criminology. *The Criminologist*, **50**(6), 7–9.