



UNIVERSITY OF  
CAMBRIDGE



## Open Data and sensitive interviews: Reflecting on ethics, consent, and reproducibility

Ben Jarman

[bmhj2@cam.ac.uk](mailto:bmhj2@cam.ac.uk)

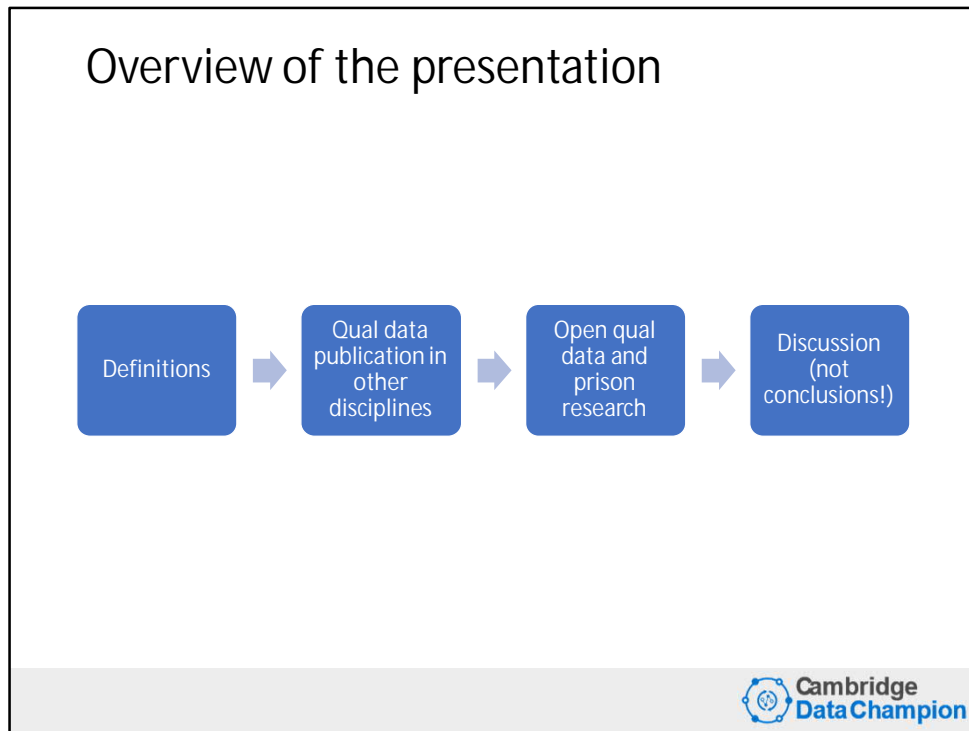
3<sup>rd</sup> December 2020

Briefly - what the data champions programme is.

Got involved because

- Was interested in Open Access and Open Research
- Briefly thought about exploring publication of the dataset from my PhD – I haven't gone ahead with this – explain why
- Was already providing some informal advice to people on things like encryption and backup – minor components in data management!
- Wanted networks, training, resources

But part of the role also involves being a resource – along with Laura K – to the department. So please get in contact if you are interested in these issues.



### Overview of the presentation

- Clarifying some definitions
- Describing some of the literature on this subject
  - Problems/sticking points
  - Some solutions
- Moving into talking about prisons and prisons research
  - Some distinctive problems
  - Some distinctive rationales
- Hopefully a discussion

My aims – not to try and draw conclusions but to think out loud about something that interests me – publication of data. I think it's ethically fraught but defensible in principle and potentially very worthwhile.

So I want to start having some reflective conversations about it.


Hence no conclusions – putting ideas forward and hoping for good chat afterwards.

## Open Research ≠ Open Access ≠ Open Data

---

<b>Open Research</b>	<u>Definition:</u> “the idea that scientific knowledge of all kinds should be openly shared as early as it is practical in the discovery process” (Nielsen n.d.)
<b>Open Access</b>	<u>Put simply:</u> make research outputs freely available (≠ paywall) <u>Implications:</u> Publicly-funded = available to public   Widest possible availability   ↑ pace of innovation   ↑ citation rates   ↓ profiteering by academic publishers
<b>Open Data</b>	<u>Various drivers:</u> Reproducibility crisis   Funder requirements   Innovation and collaboration Benefits for secondary data users  <u>Data publication:</u> established norm in some fields, increasingly common in others   Also including other outputs (e.g. analysis code, software)   DOIs, and sometimes restricted access   Data citations

---



Without getting bogged down in the detail I do want to just define a few key terms.

Open Research is “the idea that scientific knowledge of all kinds should be openly shared as early as it is practical in the discovery process”. Not a universal definition, but widely cited. It’s associated with other agendas besides.

Open Access is a term we are probably more familiar with in the social sciences – it’s the thing that means we put outputs in university repositories, but also the thing that is leading to new journals.

Open Data might be less familiar, but it’ll be familiar to anyone who’s had a research council grant, because it’s the requirement that the original data from a study should be published along with other outputs. I’ve had conversations with colleagues about this indicating that they feel uncomfortable with it, and I want to be clear there are good reasons for that. But – and I want to draw attention to this – although the name – and the explicit link with OA and OR – is relatively new, the practice of publishing and reusing qualitative data has been around for longer than you might think.

# Open Research ≠ Open Access ≠ Open Data

---

## Open Research

Definition: "the idea that scientific knowledge of all kinds should be openly shared as early as it is practical in the discovery process" (Nielsen n.d.)

---

## Open Access

Link to STEMM subjects/quantitative methodologies


Criminology = behind the curve on OA compared to other social sciences (Ashby 2020)

---

## Open Data

Data publication: established norm in many fields, increasingly common in others | Also including other outputs (e.g. analysis code, software, DOIs, and sometimes restricted access) | Data citations

---



Two other observations before moving on – first, that criminology is quite backwards on all these agendas, relative to the social sciences generally. And the social sciences generally are quite behind the STEMM subjects, in some of which open data publication is now the norm.

And second, that the STEMM link can mean that existing training materials are quite incomplete and offer methodological and other guidance which leaves very big gaps compared to what qualitative researchers might want to think about. The online training I've done on research data management is from the environmental sciences – and there is a massive difference between publishing spreadsheets of rainfall data and publishing interviews describing people's real lives in deep and sometimes painful detail. The ethical and methodological issues are completely distinct.

## 'Open data' / qualitative data reuse

- In (oral) history, prompted by ESRC-funded UK Data Archive
  - (e.g. Fielding and Fielding 2008; Neale and Bishop 2012; Neale 2013)
- Debate among (mostly) UK methodologists and ethicists
  - (e.g. Parry and Mauthner 2004; Bishop 2005; 2009; Hammersley 2009; 2010; Mauthner 2012; Bishop 2013; Neale 2013)
- Somewhat separately, in US anthropology, medical research and (qualitative) psychology
  - (e.g. Cliggett 2013; Tsai et al. 2016; DuBois, Strait, and Walsh 2018; Chauvette, Schick-Makaroff, and Molzahn 2019)
- Plus 'state of play' overviews, how-to's and toolkits
  - (e.g. Muhr 2000; Lin 2009; Bishop and Kuula-Luumi 2017; Strathern, Issig, and Pfeffer 2020)



Came as a surprise to me that the UK is a leader in reusing qualitative data and there's a small but interesting literature on the practice of doing this.

Quite a lot of that literature, in the UK at least, relates to the UK Data Archive, which began in 1997. Most of the papers are by historians; they don't have all the same methodological concerns as social scientists (data availability is an end in itself for them, and not a means to theory-building), but some are in common and the ethical issues that arise for them are very similar.

But it's worth saying that many of the secondary analyses described in this UK literature, now being used by historians, was originally collected by social scientists. Highlight the Fielding and Fielding – not open publication, but shows benefits of secondary analysis on Cohen and Taylor. Also some of the work by Neale and Bishop in this list relates to the publication openly of data archives assembled by Norbert Elias and others.

These papers go into some depth on ethical issues, and rather less on methodological issues. Ethical issues relate not only to the subjects of those studies – who could not possibly have given informed consent – but also to judgements made by secondary researchers about the prejudices and myopia of the primary researchers. I haven't time to go into more detail here but I want to draw attention to this small but rich and interesting literature. It deals with many of the concerns and much of the uneasiness we

might have, and makes clear that some problems are 'sticky' while others are easier to mitigate.

Increasing frequency in US literature, including (to my surprise) from anthropologists who have published field observations and notes, and from qualitative psychs. The anth literature draws attention to sharing of data – anths passing on archives to other anths. Argues that this should be less about passing on privilege and more about publication.

Not that many – but some – practical publications, e.g. the most recent one of these, a technical report on a tool for anonymising qualitative transcripts. Also some retrospectives, overviews and practical stuff – making clear that Finland and Germany are also key leaders in qual data publication. But practical tools few and far between.



There is a fair amount of consistency about the risks and sticking points. I’ve borrowed one particular typology of these, from a review by US qualitative psychologists.

That’s quite a list and the objections on it ‘feel’ quite close to some of the conversations I’ve had with various people around the department on this issue.

But the existing literature covers these issues very well. I don’t want to go into too much depth about that, because you can find the papers yourself, but there’s a broad consensus about how to mitigate most of these problems, as well as about which ones are more intractable. For example, a set of practices has arisen relating to the partial publication of what we might call ‘curated’ or ‘prepared’ (meaning anonymised and sometimes partially redacted) datasets, and relating to regulating access to these in some cases where the data are especially sensitive. These strategies essentially vary the degree of openness and the nature of the data.

That’s one example, but I want now to move on and bracket the issues on this slide. I’m happy to return to them a bit later, and to say more about how researchers in other fields have mitigated these issues. But I now want to focus quite tightly on the middle two parts of this typology, and apply them to my field, prison research.

## What about prisons research?

(framework from DuBois, Strait, and Walsh 2018)

### Scientific quality

- Quality standards – reproducibility (or ‘warrant’)?
- Epistemology
- (Importance of context and relationships → which data get shared?)

### Ownership

- Informed consent
- Whose work, whose data?

### Confidentiality and harms

- Protecting participants from...
- Deductive identification & invasion of their privacy
- Other harms (e.g. consequences of disclosure)

### Researcher resistance

- “It’s my work”
- Difficulty of preparing work for archiving (metadata, adequate deidentification)



The prison setting is distinctive, and I think it raises some distinctive questions about qualitative data publication in relation to the second and third points in this typology: informed consent, and to protecting participants

I’ve picked them out in red. But to be clear, I’m not trying to say that the other issues don’t apply in this field – they do. Points 1 and 4 apply to prisons, but the issues that arise are not distinctive to prisons, and they are quite well dealt with elsewhere. So I think if you’re interested in them, you can draw food for thought and good strategies by using the existing literature I’ve mentioned.

But for the next few minutes I’m going to concentrate on what’s distinctive about prison research.



## Prisons as distinctive social/research settings

(e.g. Cohen and Taylor 1977; Liebling 1999; 2001; Crewe 2014; Ugelvik 2014; Beyens et al. 2015)

Closed environments

Secretive

Public interest (prurience?) vs. public ignorance

Difficult to gain access

Prison life is governed both formally and informally

Highly prescriptive regulation of research(ers)

Yet also much interstitial 'space'

'Informed consent' (vs. 'interpersonal trust'?)



I won't spend very long here describing my PhD, save to say that it's based on long semi-structured qualitative interviews with men serving life sentences for murder in two prisons.

The prison setting is distinctive, and that needs to be taken into account when thinking about data publication. I want to spell out why I think that's so, but also point out that some of the factors I identify cut more than one way; I don't have strongly formed opinions one way or the others.

Closed environments –

- They are secretive places. Secretiveness and public ignorance point to greater openness and potential benefits from data being more widely available – particularly to non-academic users
- But there are risks of prurience and interest in prisons and prisoners not always benign – (this is also true within criminology)
- Difficulty of access suggests that secondary analyses are also really worthwhile – it takes a LOT to collect the data

So here I think there's at least an arguable case for opening these institutions to more scrutiny, and one way of doing that is to make data available for uses that it hasn't been considered for.

Social life is formally and informally governed –

- Has a number of consequences – God only knows how the NRC would approach data publication
  - Strong reason to suppose that prison security departments would be against it
- But access is formal and about gatekeeping, and prison researchers have often written about comparative freedom once they are inside the prison
- Thus informal regulation – with one consequence being that informed consent is often difficult to secure – ‘Gerald’ vs. ‘Mark’

I don’t have a very clear conclusion to offer here about what this means for publication. I think its implications are hard to unpack; but I also think there are other areas of social life where research data are published and they are similarly subject to different regimes of governance.

## Prisons as distinctive social/research settings

(e.g. Cohen and Taylor 1977; Liebling 1999; 2001; Crewe 2014; Ugelvik 2014; Beyens et al. 2015)

Intense (and murky) culture/ economies	Low levels of trust, front-stage vs. back-stage
	Intra-institutional knowledge = 'thick' & consequential
	∴ participants highly recognisable even in deidentified data?
Shame and stigma	Qualitative data often (though not exclusively) highly sensitive
	Risk of identification (e.g. through media) high
	Risk of harm from identification high (but uneven?)



### Intense and murky cultures and economies

- Despite what I said about personal trust, trust is contingent and temporary, and it's hard to be sure you've got a read on people's motives
- 'Ray' or 'Michael' among my participants – difficulty of having conversations in public places, different motivations disclosed in more private setting
- Strong undercurrents including things you won't know about
- Risky disclosures – and potentially highly identifiable data, if data in the 'wrong' hands

And yet these more complicated, less rule-bound, more 'embarrassing' sides to prison life are the very things that some prisoners (and staff) say they wish the outside world knew about – so it cuts both ways

### Shame and stigma

- Specifically in relation to my PhD – everyone interviewed is serving a life sentence for murder. In nearly all cases it's easy to find coverage of trials online; some are very high-profile
- They are all easy to Google – and thus the risk to them of identification based on details of their offence would be high
- The risk of harm also high – but think of Martin and his wish to be identified under his real name – shame and pride cut both ways – the benefits of a counter-narrative

denied to many prisoners?

In the end, with my own data, I've decided against publication: technical reason of the PIS (lack of planning), practical reason that it's hard to build this stuff in along the way, and ethical reason that the research is very focused on individual lives and not institutional forces, and therefore too difficult to untangle.

But it's an issue I'm going to have to grapple with if I do publicly-funded research in future, and I think there's enough of a case in favour of publication that the question of how to do it needs taking seriously.

And here are some key principles I think are important.

- Respecting the participant's autonomy
- But also having honest conversations with them

That's as far as I'm prepared to go, prescriptively speaking, and I think the rest requires more thought. There are some guidelines in the literature on these – but as I've said, I'm not aware of anyone in criminology or prisons research who's asking these questions.

## Points for discussion?

Open research a newish area – particularly for criminology

- (Anecdotally) there's not much excitement about it among senior academics
- Thus good for ECRs – good topic for methods/ethics writing and development of resources?

I've outlined some risks and benefits in prisons research – what am I missing?

- Anything I said that you want to take issue with?

Has anyone else been thinking about whether/how to publish their PhD data?  
Why?

- What boundaries or 'red lines' exist for you, ethically, personally, or otherwise?

How do the issues raised by Open Data vary in other criminological sub-fields?

I said I was interested in a discussion – so here are some prompts.

# References

- Ashby, Matthew P. J. 2020. 'The Open-Access Availability of Criminological Research to Practitioners and Policy Makers'. *Journal of Criminal Justice Education* 0 (0): 1-21. <https://doi.org/10/gbh7n7>.
- Beyens, Kristel, Philippe Kenes, Sonja Snacken, and Hanne Toumei. 2015. 'The Craft of Doing Qualitative Research in Prisons'. *International Journal for Crime, Justice and Social Democracy* 4 (1). <https://doi.org/10/9f6jrn>.
- Bishop, Libby, and Arja Kuula-Luumi. 2017. 'Revisiting Qualitative Data Reuse: A Decade On'. *SAGE Open* 7 (1): 2158244016685136. <https://doi.org/10/gf5z7w>.
- Bishop, Libby. 2005. 'Protecting Respondents and Enabling Data Sharing: Reply to Parry and Mauthner'. *Sociology*, April. <https://doi.org/10/bx6t58>.
- Bishop, Libby. 2009. 'Ethical Sharing and Reuse of Qualitative Data'. *Australian Journal of Social Issues* 44 (3): 255-72. <https://doi.org/10/qz4tbc>.
- Bishop, Libby. 2013. 'The Value of Moral Theory for Addressing Ethical Questions When Reusing Qualitative Data'. *Methodological Innovations Online*, August. <https://doi.org/10/gcd422>.
- Chauvette, Amel, Kara Schick-Makortoff, and Anita E. Molzahn. 2019. 'Open Data in Qualitative Research'. *International Journal of Qualitative Methods*, January. <https://doi.org/10/gbqzbf>.
- Cohen, Stanley, and Laurie Taylor. 1977. 'Talking About Prison Blues'. In *Doing Sociological Research*, edited by Colin Bell and Howard Newby. London: Allen and Unwin.
- Crewe, Ben. 2014. 'Not Looking Hard Enough: Masculinity, Emotion, and Prison Research'. *Qualitative Inquiry* 20 (4): 392-403. <https://doi.org/10/gf6jng>.
- DuBois, James M., Michelle Strait, and Heidi Walsh. 2018. 'Is It Time to Share Qualitative Research Data?' *Qualitative Psychology, Sharing Data*, 5 (3): 380-93. <https://doi.org/10/gbc3li>.
- Fielding, Noel G., and Jane L. Fielding. 2008. 'Resistance and Adaptation to Criminal Identity: Using Secondary Analysis to Evaluate Classic Studies of Crime and Deviance'. *Historical Social Research / Historische Sozialforschung* 34 (4 (7-9)). <https://www.hsr.ch/doi/10.1017/hsr.2008.004>.
- Hammersley, Martyn. 2009. 'Against the Ethicists: On the Evils of Ethical Regulation'. *International Journal of Social Research Methodology* 12 (3): 211-25. <https://doi.org/10/bp374n>.
- Hammersley, Martyn. 2010. 'Can We Re-Use Qualitative Data via Secondary Analysis? Notes on Some Terminological and Substantive Issues'. *Sociological Research Online* 15 (1): 47-53. <https://doi.org/10/c3khhb>.
- Liebling, Alison. 1999. 'Doing Research in Prison: Breaking the Silence?' *Theoretical Criminology* 3 (2): 147-73. <https://doi.org/10/fn7tz7>.
- Liebling, Alison. 2001. 'Whose Side Are We on? Theory, Practice and Allegiances in Prisons Research'. *The British Journal of Criminology* 41 (3): 472-84. <https://doi.org/10/dh5shz>.
- Lin, Li-Chen. 2009. 'Data Management and Security in Qualitative Research': *Dimensions of Critical Care Nursing* 38 (3): 132-37. <https://doi.org/10/crcf69>.
- Mauthner, Natascha S. 2017. 'Accounting for Our Part of the Entangled Web We Weave': Ethical and Moral Issues in Digital Data Sharing. In *Ethics in Qualitative Research*, by Tina Miller, Maxine Birch, Melanie Mauthner, and Julie Jessop. London: Sage. <https://doi.org/10.1080/10439862.2017.1311111>.
- Neale, Bren, and Libby Bishop. 2012. 'The Timescapes Archive: A Stakeholder Approach to Archiving Qualitative Longitudinal Data'. *Qualitative Research* 12 (1): 53-65. <https://doi.org/10/gbh594>.
- Neale, Bren. 2013. 'Adding Time into the Mix: Stakeholder Ethics in Qualitative Longitudinal Research'. *Methodological Innovations Online* 8 (2): 6-20. <https://doi.org/10/gcdt6s>.
- Nielsen, Michael. n.d. Open Science Now! Accessed 24 November 2020. [https://www.ted.com/talks/michael\\_nielsen\\_open\\_science\\_now](https://www.ted.com/talks/michael_nielsen_open_science_now).
- Parry, Odette, and Natascha S. Mauthner. 2004. 'Whose Data Are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data'. *Sociology*, February. <https://doi.org/10/c9mip7>.
- Parry, Odette, and Natascha S. Mauthner. 2004. 'Whose Data Are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data'. *Sociology*, February. <https://doi.org/10/c9mip7>.
- Strathern, Winko, Moritz Isia, and Jürgen Pfeffer. 2020. 'QualAnon - a Tool for Anonymizing Interview Data'. Technical Report. Munich: Technical University of Munich. <https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63883-p0011-7>.
- Ugelvik, Thomas. 2014. 'Prison Ethnography as Lived Experience: Notes From the Diaries of a Beginner Let Loose in Oslo Prison'. *Qualitative Inquiry* 20 (4): 471-80. <https://doi.org/10/gf6jnd>.