



Show Me the Data: research reproducibility in qualitative research

🕒 September 18, 2018 (<http://blog.ukdataservice.ac.uk/show-me-the-data/>)

👤 Neil Dymond-Green (<http://blog.ukdataservice.ac.uk/author/neil-dymond-greenmimas-ac-uk/>)

💬 Leave a comment (<http://blog.ukdataservice.ac.uk/show-me-the-data/#respond>)

Louise Corti, Director of Collections Development and Data Publishing for the UK Data Service explores research reproducibility in qualitative research.

In quantitative methods, reproducibility is held as the gold standard for demonstrating research integrity.

However, threats to scientific integrity, such as fabrication of data and results, have led to some journals now requiring data, syntax and prior registration of hypotheses to be made available as part of the peer-review. While qualitative research reproducibility has been questioned in the past, it has been protected from the recent transparency agenda.

But for how long? What if journals mandated the sharing of data and analysis for qualitative research?

These issues were addressed at a session in this year's NCRM Research Methods Festival, held in Bath, where a panel of speakers debated whether there was indeed a 'crisis' and what 'reproducibility' approaches and standards might look for qualitative research.

I position myself as a neutral research methodologist, and from this position, in this session I wanted us to explore contrasting views. I understand the concerns of qualitative colleagues, yet – like many others in the social sciences – I value the principles of open research and open science and the concepts of transparency and ground-truthing.

Terms used in this debate space can seem threatening, such as:

- Verification
- Replication
- Reproducibility
- Transparency
- Integrity
- Restudy

The speakers in our session took various positions:



Louise Corti:

How qualitative researchers might respond creatively to a reproducibility crisis

Slides (https://www.ukdataservice.ac.uk/media/622138/corti__rmf_qualtransperency_bath2018.pdf) (PDF)



Sarah Nettleton:

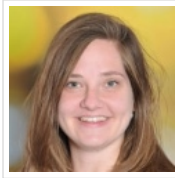
How various 'crises' surrounding 'transparency in qualitative research have emerged and how data sharing might help mitigate this

Slides (https://www.ukdataservice.ac.uk/media/622141/nettleton__rmf_qualtransperency_bath2018.pdf) (PDF)

Nicole Janz:

Practical strategies for teaching replication in the quantitative tradition in political science

Slides (https://www.ukdataservice.ac.uk/media/622140/janz__rmf_qualtransperency_bath2018.pdf) (PDF)



Maureen Haaker:

Practical examples of what reproducibility might look like, based on existing archived data collections

Slides

(https://www.ukdataservice.ac.uk/media/622139/haaker__rmf_qualtransperency_bath2018.pdf) (PDF)

Starting with my data sharing/archivist hat on, my professional role is to promote the sharing of data, context and methods to support the re-use of public research resources. In this vein, I have spent over thirty years evangelising on the sharing of qualitative data leading this strand of work for the UK Data Service, and Qualidata before that, dating back to 1994.

The UK is seen as a leader in archiving qualitative data, with methods which are mature in practice and culture. I do believe that **every researcher can share something** from their methods and data, and that we do need to be ready to come up with positive arguments and solutions to calls for replication.

Is there a crisis?

We cannot have failed to note the increasing drive for openness and sharing, value and transparency in our daily lives. GDPR and open access, as well as concerns about how data has been used in referendums and elections have all raised public awareness of this drive.

Funders, Professional Societies and Journals are also driving open research mandates. The 2007 OECD Principles and Guidelines for Access to Research Data from Public Funding (<https://www.oecd.org/sti/sci-tech/38500813.pdf>) (PDF), declared data as a public good, prompting a swathe of action.

This data sharing ideology promotes value for money through:

- more and subsequent analysis of data;
- opportunities for collaboration and comparison;
- richer exposure of methods; and
- research validation.

Indeed, the concept of replication has also become ever more prominent in the research narrative. The idea is appealing: that sharing of the data which underpins research publications can help counter mistrust in research findings,

This mistrust has been increased by problems that have occurred in peer-reviewed publications from unverifiable results and faking of data (for instance, in the field of psychology, there are examples from the Netherlands

(<http://www.telegraph.co.uk/news/worldnews/europe/netherlands/8868337/Dutch-social-psychologist-found-to-have-faked-data.html>) and Belgium (<https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html>)). Janz reminded us that a 2015 study (<https://doi.org/10.1126/science.aac4716>) which attempted to replicate the results of 98 psychology papers found that more than half couldn't be reproduced,

More recently, in the UK, there has been a UK Parliamentary Science and Technology Committee inquiry on Research Integrity (<https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2017/research-integrity-17-19/>), looking at the so-called 'crisis in reproducibility' of research.

In 2012 the US political science community introduced a practical approach to assessing replicability. The DA-RT (<https://www.dartstatement.org/>) (Data Access and Research Transparency) initiative was aimed at journal editors, and sought to instate a replication threshold for quantitative paper submissions. The DA-RT helpfully separates out data, analytic and production transparency.

In support of **analytic transparency**, for journals signed up to the DA-RT statement, authors are required to submit analysis code along with their article, that must be fully replicable. Indeed some journals rerun code to check it.

But what is the equivalent of this exercise in qualitative research?

Nettleton reminds us that typically data are co-produced, co-constructed, embedded in and by contexts and the conditions of production are inextricably interlinked with process of analysis and interpretation. Further, interrogation of data, coding and reduction are highly iterative processes, where small decisions are likely hard to capture.

Software like NVivo can capture coding frames and annotations and memos can help support the documentation of decisions surrounding how codes were assigned, changed and so on; and this 'evidence' could be saved and viewed by future researchers.

In the US the political science community has set up a Qualitative Transparency Deliberations (<https://www.qualtd.net/>) (QTD) platform to encourage discussion of the meaning, costs, benefits, and practicalities of research transparency for a broad range of qualitative empirical approaches in Political Science.

Following from this, Elman and colleagues in the US set up an Annotation for Transparent Inquiry (<https://qdr.syr.edu/ati>) (ATI) initiative which encourages scholars to annotate specific passages in an article that can amplify the text, by adding links to and notes about data sources underlying a claim. While there is value in this approach, especially in encouraging source data to be citable and revisited, we should not veer towards mandating the evidencing of claims.

Focus on production transparency

Given that much fieldwork is impossible to fully replicate, the idea of **production transparency** (the methods used to collect data) is likely to be more appealing to the qualitative researcher.

Aside from any ethical issues that can arise in sharing data, how useful is it to upload a 'replication dataset' – probably a subset – alongside a paper? Part of a larger data collection from a study presents a far less useful device for supporting transparency than having the opportunity to explore the 'whole' collection of data. In qualitative research, how might one even select the 'subset'?

It is useful to consider the spectrum of immersivity in qualitative research, for example from passive observation through to participatory research or ethnography that will likely require different layers of description.

Evidence to demonstrate production transparency will vary from a description of research design, sampling, fieldwork, fieldwork materials such as topic guide and thumbnail sketches of interviews can provide a good deal of context. These could be sufficient for a future researcher to undertake a restudy or to revisit the raw data.

For ethnography, this will be harder, but various fieldwork diaries, audio-visual sources, might help put the reader back in the original fieldwork scenario.

We have lots to learn from historians on methods for assessing provenance and assessing the value of a data collection by scrutinising available context; and the National Centre for Research Method's 'archaeology' (<http://bigqlr.ncrm.ac.uk/2016/11/08/research-team-blog-4/>) metaphor for reusing large collections of qualitative data is appealing.

We also need to consider what kinds of documentation and materials might help us demonstrate **process transparency**. The methods section in an article can often be restricted and unlikely to be helpful in our quest for transparency. When data are shared, supporting 'grey' materials are archived alongside data, providing context.

We are fortunate that the ESRC has supported the sharing of over 1000 data collections we can explore to see what UK qualitative researchers have already archived

The UK Data Service promotes the sharing of context alongside data and has developed systems and approaches to encourage the future-proofing of data, including the handbook **Managing and Sharing Data – A Guide to Good Practice** (<https://uk.sagepub.com/en-gb/eur/managing-and-sharing-research-data/book240297>), published by Sage in 2014.

Haaker (https://www.ukdataservice.ac.uk/media/622139/haaker__rmf_qualtransperency_bath2018.pdf) (PDF) shows us examples from the UK Data Service data catalogue of supporting contextual materials from archived research datasets that help researchers understand the data and the research process:

- Quali Election Study of Great Britain, 2015 (<https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6861>)
- Anti-politics: Characterising and Accounting for Political Disaffection, 2011-2012 (<https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7855>)
- Conservation, markets and justice – Part 2: Ethnographic participatory video data (<https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=852476>)

Data journals such as **The Journal of Open Health Data** (<https://openhealthdata.metajnl.com/>) further provide a further valuable outlet for describing the rationale and methods that created a published dataset.

A researcher's experience of transparency

Nettleton recounts her experiences of archiving her data from her post millennial study of how the day-to-day experiences and perspectives of UK doctors were influenced by cultural, political and socio-economic shifts in their working environment. (Becoming a Doctor: a Sociological Analysis, 2005-6 (<https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6124>)).

She is surprised and delighted that these data have been used for teaching medical students as well as research. She had agonized over the level of anonymity at the time of depositing, but on reflection believes that it was a helpful experience for her, and that she has contributed to being transparent.

She notes that a researcher's immediate gut instinct will be to not share data; it's easier, and feels safer; yet she also notes that the great value in future-proofing rich primary data for future research.

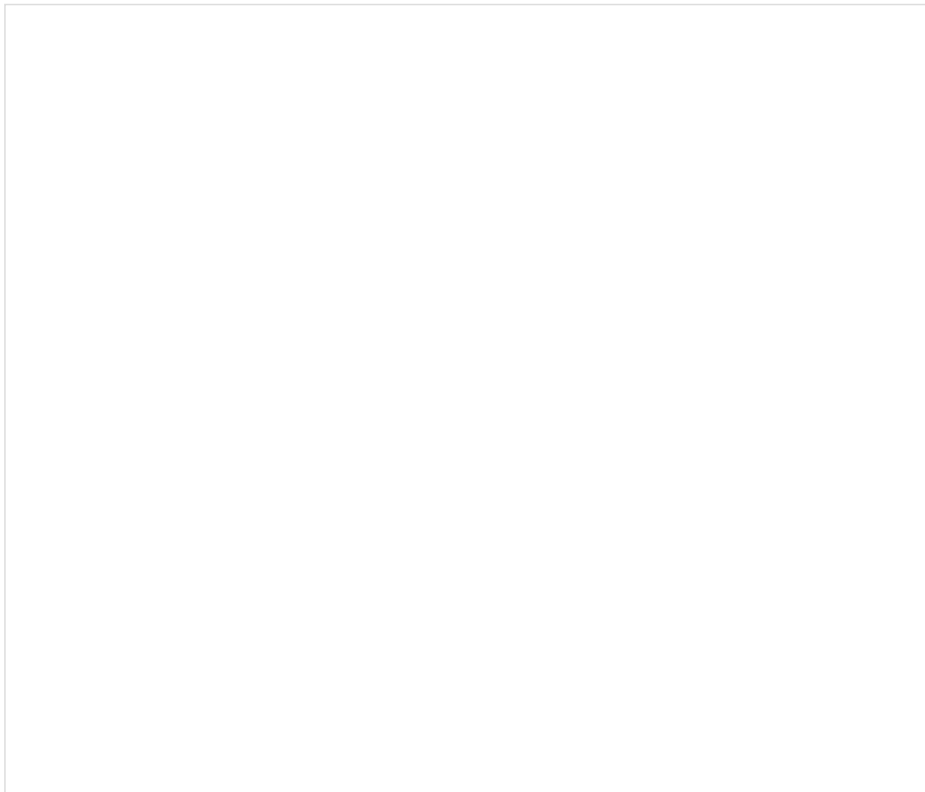
Thinking back to previous research on a sensitive topic and – having had time to mull over 'archiving' and its place in the knowledge economy, she now believes that she may have been able to share some of the data.

She does note the resource burden of preparing data and considers that we may need to come up with a better way to support this at the researcher and institutional level, anticipating how much time it will take and building it into research plans. Data management planning may help grant applicants think about this upfront, but it rarely solves the issue of allowing adequate resourcing for archiving at the end of a project. Publishing becomes the ultimate goal. However, having data available to support papers can help in the quest for transparency.

In agreement with Nettleton, and from her grounded experience of sharing data, at the UK Data Service we agree that the processes of archiving qualitative data needs to be rigorous yet sufficiently nuanced to allow for its flexibility and messiness.

Supporting transparency in research publication

As part of the RMF session, we collectively came up with a very simply summary table for our audience that classifies activities supporting various kinds of transparency, based on the DA-RT classification. There are some subtle differences between good practice for quantitative and qualitative approaches, but these are not huge.



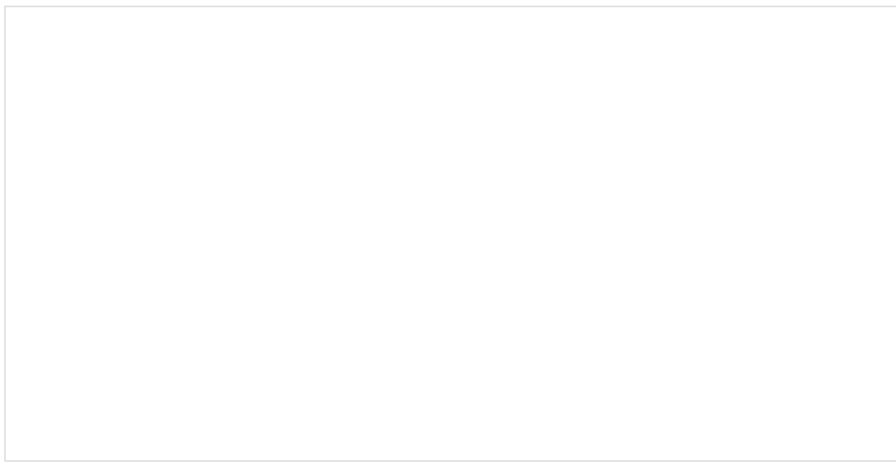
Getting in early: transparency in undergraduate dissertations

Given QAA's 2017 report on contract cheating, third party services and essay mills (https://www.qaa.ac.uk/docs/qaa/quality-code/contracting-to-cheat-in-higher-education.pdf?sfvrsn=f66af681_8) (PDF), providing early guidance for students on the importance of academic rigor and integrity is vital.

At RMF 18, we launched a new Resource Pack, entitled **Dissertations and their Data: Promoting Research Integrity** aimed at lecturers or tutors responsible for running undergraduate dissertation support classes.

The pack was developed by Haaker and colleagues for a session created for an undergraduate sociology module at the University of Essex, called **Thinking Ahead**, helping third years prepare for their dissertations. The teaching builds on the programme of capacity building work done by the UK Data Service, seeking to apply core principles of excellence in data management, description and sharing to the classroom. The dissertation may be the student's first encounter with data collection or data management, so encouraging students to think about some of the practicalities of their data collection, and how they can demonstrate rigorous and transparent research, at this early stage are useful.

Haaker has written more about **Thinking Ahead** in this blog post (<http://blog.ukdataservice.ac.uk/thinking-ahead-data-management/>).



Janz has been running inspiring teaching in her discipline for some years, holding ‘master classes’ in duplication and replication.

She distinguishes between these as findings from a true replication could be different from – by adding to or challenging – the original paper. She also reminds her students to be as professional and diplomatic when communicating failed replications (Janz, 2016 (<https://doi.org/10.1111/insp.12104>))!

Summary

If journals seek to extend their reproducibility agenda to qualitative research, they could usefully start by encouraging **process transparency** while not mandating replicability. Indeed, the potential to critically examine and re-use data is unlikely to include reproduction.

Sharing data is one step to help us to build an evidence base on which to build knowledge, but Nettleton reminds us that archiving data cannot and should not be done in response to addressing the transparency crisis.

Contextualisation of data to support results can usefully be incorporated into everyday research practice, so that ‘packaging’ it and submitting it with an article for publication might be easier.

I certainly do not advocate trying to find an equivalent of ‘replication code’. Hence ‘methods materials’ should not be cast aside as a project finishes but kept for future users to make sense of research findings and archived data.

The ESRC’s NRCM can train researchers in the excellence and innovation in methods, while the UK Data Service can show them how to document the methods to support you generated.

For us, the RMF debate session was a safe space to debate issues of transparency that can, at times, get fairly heated. Participants really enjoyed the session, and we have been invited to run a similar event again.

We hope that readers will take away a positive key message back to their colleagues and students, inviting them to embrace the idea of trust in data, and to seek to maximise transparency of their research findings, in creative ways.

The UK Data Service will be running another session on **Thinking Ahead: How to be Reproducible in Research** in the New Year, so watch our news and events page (<https://www.ukdataservice.ac.uk/news-and-events>) to find out about our future training events.

Follow

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment

Name *

Email *

Website

Post Comment

« Previous Post (<http://blog.ukdataservice.ac.uk/working-long-term-illness/>)

Next Post » (<http://blog.ukdataservice.ac.uk/brexit-uncertainty-uk-exports/>)

Categories

Select Category ▼

Posts by month

Select Month ▼

Search the Data Impact blog

- [UK Data Service \(https://www.ukdataservice.ac.uk\)](https://www.ukdataservice.ac.uk)
- [Our impact \(https://impact.ukdataservice.ac.uk/\)](https://impact.ukdataservice.ac.uk/)
- [Impact Case Studies \(https://beta.ukdataservice.ac.uk/impact/case-studies\)](https://beta.ukdataservice.ac.uk/impact/case-studies)
- [Data Impact Fellows \(http://blog.ukdataservice.ac.uk/category/data-impact-fellows/\)](http://blog.ukdataservice.ac.uk/category/data-impact-fellows/)
- [Impact and Innovation Lab \(http://blog.ukdataservice.ac.uk/category/impact-and-innovation-lab/\)](http://blog.ukdataservice.ac.uk/category/impact-and-innovation-lab/)
- [Accessibility \(http://blog.ukdataservice.ac.uk/accessibility/\)](http://blog.ukdataservice.ac.uk/accessibility/)

UK Data Service

Funded by the Economic and Social Research Council (ESRC), the UK Data Service provides access to international social, economic and population data.

The Data Impact Blog

The Data Impact blog is run by the UK Data Service and is a hub for anyone interested in maximising the impact of social, population and economic data in research and policy.

Latest Post

Using data for effective health service evaluation: why we need data stewards (<http://blog.mimas.ac.uk/ukdataservice/?p=5238>)

