NOV 28, 2023

OPEN ACCESS

Data sharing barriers in a viral pandemic: semi-structured interview protocol

Yo Yehudi[1,2], Lukas Hughes-Noehrer[2], Carole Goble[2], Caroline Jay[2]

[1]OLS; [2]University of Manchester

Yo Yehudi
OLS

DOI:
dx.doi.org/10.17504/protocols.io.n2bvj3j7xlk5/v1

External link:
https://covid19-data-sharing-study.github.io/

Protocol Citation: Yo Yehudi, Lukas Hughes-Noehrer, Carole Goble, Caroline Jay 2023. Data sharing barriers in a viral pandemic: semi-structured interview protocol. **protocols.io** https://dx.doi.org/10.17504/protocols.io.n2bvj3j7xlk5/v1

MANUSCRIPT CITATION:
Preliminary results: https://doi.org/10.5281/zenodo.5234416 (extended abstract at the Data for Policy conference).
Preprint (full results): https://doi.org/10.48550/arXiv.2205.12098

ABSTRACT

In 2020, the COVID-19 pandemic resulted in a rapid response from governments and researchers worldwide. As of May 2022, over 6 million people died as a result of COVID-19 and over 500 million confirmed cases, with many COVID-19 survivors going on to experience long-term effects weeks, months, or years after their illness. Despite this staggering toll, those who work with pandemic-relevant data often face significant systemic barriers to accessing, sharing or re-using this data. In this paper we report results of a study, where we interviewed data professionals working with COVID-19-relevant data types including social media, mobility, viral genome, testing, infection, hospital admission, and deaths.

These data types are variously used for pandemic spread modelling, healthcare system strain awareness, and devising therapeutic treatments for COVID-19. Barriers to data access, sharing and re-use include the cost of access to data (primarily certain healthcare sources and mobility data from mobile phone carriers), human throughput bottlenecks, unclear pathways to request access to data, unnecessarily strict access controls and data re-use policies, unclear data provenance, inability to link separate data sources that could collectively create a more complete picture, poor adherence to metadata standards, and a lack of computer-suitable data formats.

Results expected: This set of tasks is particularly useful for people exploring the challenges around data access. It is likely to generate further more detailed questions and avenues for investigation.

Tools provided: interview guide for semi-structured interviews.

GUIDELINES

See ethics statement - this research protocol touches sensitive subjects and should be verified with your ethics board before initiating the research.

Protocol status: Working
We use this protocol and it's working

Created: Nov 28, 2023

Last Modified: Nov 28, 2023

PROTOCOL integer ID: 91515

Keywords: interview guide, qualitative research, semi structured, data access, data barriers, data stewardship, data sharing, data re-use

Funders Acknowledgement:
EPSRC
Grant ID: EP/S021779/1

BEFORE START INSTRUCTIONS

Prepare a participant information sheet. This is likely to vary based on your institutional needs, so we do not provide a template, but the content from https://covid19-data-sharing-study.github.io/assets/COVID-19%20participant%20information%20sheet.pdf might be helpful for inspiration.

Initial interview set-up

1 Start by preparing the materials you'll need for your interviews.

If interviews are **in-person**, prepare an audio and/or video recording setting which is quiet uninterrupted.

If interviews are **online** (e.g. via Zoom or other video conferencing tool), ensure that you have a quiet area to take the call, a microphone, and the ability to record and a subscription so the meeting will last long enough if relevant. (Free Zoom meetings end after 40 minutes).

2 **Participant information sheet and consent forms**
These will vary from institute to institute, so we do not provide templates. Consider the following elements:

1. How long the interview will be, and what method (online/in-person)
2. What you will do with participant data (including cloud storage, GDPR, and privacy)
3. Will you share this data? The recordings? Quotes only? None of the above? Will participants have the chance to redact phrases if needed?

4. If a participant wants to remain anonymous, what steps can you take for them to feel safe?

If your interviews are **online**, a survey tool might be good. Make sure you mention it in your ethics application, as your institute might have a preferred tool for managing personal/experimental data, such as Qualtrics. You can use this method at face-to-face interviews, too - get them to fill in the form on your computer!

Consider batch-running your interviews

3 If you can, consider doing the interviews in batches. Go to a conference, or find a friend or colleague at a local friendly institute, book a meeting room, and set up a few slots throughout the day (no shorter than one hour!). This saves effort on set-up time.

Make sure you have quite a few copies of your information sheet and consent form printed out for participants. Send these to participants in advance, but be prepared to administer them on the day if they didn't read them before the appointment.

Whether you're doing your interviews online or offline, a calendar-booking tool like https://calendly.com/ (Google Workspace and Microsoft 365 both have equivalents) can help your book slots with your interviewees with very low effort.

Interview time!

4 **Consent and information sheets**
Before going any further, make sure the participant understands what the purpose of the study is, roughly what will happen, and if there are any consent or information forms to share with them - now is the time to get participants to read and sign them, so they know what you'll be doing with their data and recordings.

5 **Start recording**
Once the paperwork is done, it's time to kick off! Make sure the participant is happy with your recording method (audio or video), and start the recording.

6 **Run the interview**
This section is relatively informal – various clarifying
questions may need to be asked depending on what the participant answers.

“Amazing – let’s get on with the questions, then!
Occasionally I may pause or you may hear typing while I’m making notes.”

6.1 Tell me a little about the data sources you’ve been working with so far – have any been restrictive or hard to access?

6.2 a. What did you do when you encountered the access restriction?

6.3 b. Did you ever consider going to another source, or encouraging data creators and curators to share their information somewhere less restrictive?

6.4 c. Did others you work with share your views and opinions? Can you provide any examples?

6.5 What about good experiences – are there any data sources that are exemplary?

6.6 a. What did they do right?

6.7 b. Was there anything they could do better?

6.8 If you could design your own “dream” COVID data source, would there be anything else you’d like to see?

6.9 Are there any ethical issues you’re aware of, or have considered, with regards to this data sharing?

6.10 Is there anyone else working in this domain who you think might be interested in participating in this study?

7 **Wrap-up**

"That was brilliant – thank you very much! That's all I had to ask you today. Do you have any questions, or was there anything you'd like me to explain further?"

Once the participant has asked any questions they might have:

"Awesome. For my next steps, I'll transcribe the interview. If I find any quotes I'd like to quote in the write-up, I'll run them by you first to make sure you're okay with the phrasing and any anonymisation that may be needed."

Post-interview

8 After the interview, make sure to comply with the promises you made around data privacy, reviewing quotes, etc.