How can you increase the FAIRness of your qualitative data?

Katrine Utaaker Segadal, Open Access Week 27.10.2021
How can you increase the FAIRness of your qualitative data?
FAIR

FINDABLE  ACCESSIBLE  INTEROPERABLE  REUSABLE
As open as possible

- Cost efficiency
- Knowledge sharing
- Trust

As closed as necessary

- Security
- GDPR
Think about archiving and sharing early in the research process

- Preferably before the data collection
- Use a Data Management Plan
- What archive suits my data?
  - [https://www.re3data.org/](https://www.re3data.org/)
  - Domain
  - Degree of openness
  - Quality control
  - Persistent identifier
Are you allowed to archive the data?

- Intellectual Property Rights
- Are you going to archive personal data?
  - Obtain consent for archiving and sharing
  - Specify archiving and sharing in the consent form and information letter to the participants
Is it possible to archive anonymised qualitative data?

- Anonymising can be time-consuming and costly
- Might reduce the value of the data
- Providing assurances of 'complete anonymity' is often not possible to achieve in practice
- NSD selection criteria: For qualitative data, consent to archive and share the data with indirect personal identifiers must have been obtained.
How to ensure that data is as reusable as possible?

- **Clear access conditions**
- **Documentation to provide context**
  - Without documentation, data has little value for reuse and verification
  - Document before, during and after data collection
  - Interview and observation guides and the like
Study level documentation

- Purpose
- Content
- How the data were collected
- Who collected the data, where and when
- How data can be accessed
Data level documentation

- Each item should have an unique identifier
- Key biographical characteristics and features of interviewees
- Customised to the study
- Data file name, number of pages in the file and date relevant to the file
- Balance: protecting confidentiality and giving enough contextual information
How can you increase the FAIRness of your qualitative data?

Think about archiving and sharing early in the research process.

Obtain informed consent to archive and share your data with indirect personal identifiers.

Document your data before, during and after data collection.
The ethics and responsibility of creating an open-access, multimedia language corpus

Lindsay Ferrara
Norwegian University of Science and Technology
27 October 2021; Open Access Week
Outline

• Introduce the Norwegian Sign Language Corpus
• Collecting non-anonymized data in the age of GDPR
• Gaining consent and building trust with research participants/language users
• Creating open access resources that benefit both the academic community and the Norwegian deaf community
The Norwegian Sign Language Corpus
(under development, 2019-2024)

• The corpus is a group of four datasets:
  • 2012: Ph.D. dissertation research (Halvorsen, 2012)
  • 2015: Pilot corpus project (Ferrara & Bø, 2015)
  • 2017-2018: Project investigating spatial language (Ferrara & Ringsø, 2017-2018)
  • Ongoing: Project to collect a more representative corpus

• Current size of the corpus:
  • 41 signers
  • 20.5 hours of signing
  • 169 video clips
  • 104 associated ELAN files
Collecting non-anonymous + personal data
Consent process

• Consent is one important way that we can protect the interests of research participants/language users, in a respectful and ethical way.

• Providing information in accessible language formats

• Re-consent process for participants of earlier projects.

https://youtu.be/TpWbVTCuUSQs
Open data: benefitting research and language communities

Language community
• The data with a CC license can be used to create teaching materials for various deaf and hearing groups.
• Support the development of other language resources, like a dictionary.
• Documentation of one part of Norway’s rich linguistic heritage

Research community
• The time-consuming and resource intensive annotation of signed language data is mitigated and minimized.
• Language users avoid research fatigue.
• Provides a sound, empirical base to a body of research studies.
Thank you

lindsay.n.ferrara@ntnu.no
Challenges and opportunities

Sharing Pseudonymized Interview Data

Live Håndlykken Kvale
Academic Librarian and PhD Candidate
Oslo University Library and OsloMet

27. October 2021
My PhD Project

Infrastructure for data sharing
- Sharing of research data in Norway as scope

Delphi study
- 24 participants

Mixed methods
- Interview x2 and questionnaire
Writing about research data sharing and having worked with data sharing at UiO

---

I felt morally obliged to find out how the data I collected could be shared, as open as possible
The Data I Collected

25 interviews; approximately 24 hours
215 pages of interview transcripts
24 responses to 9 questions
24 interviews; approximately 12 hours
98 transcribed pages
and all the other stuff that contextualise this material
How can I share the data?
Does it make sense to make interviews interoperable at the datapoint level?

Which file formats?

Should the consent form be machine-readable in RDF?
Does it make sense to make interviews interoperable at the datapoint level? NO

Which file formats? XML and Python rather than Word and Nvivo

Should the consent form be machine-readable in RDF? NO
Can the data be anonymous?

Should I let participants review the transcripts?

Can sharing be compatible with the right to be forgotten and the right to redraw a consent at any time?
Personal Privacy

Can the data be anonymous? NO

Should I let participants review the transcripts? YES

Can sharing be compatible with the right to be forgotten and the right to redraw a consent at any time? Explicit language Separate consent form for data sharing
Clear Language in the Consent Form

After having read and analyzed the interviews multiple times, I see that several interviews contain information that makes some participants identifiable for someone with knowledge of the participant or by combining it with other available information.

If you are not comfortable with the possibility to be identified I advise you to redraw previously given consent to share the interviews (below).

The results from the questionnaire are less identifiable, as the results are aggregated in groups (funder, researcher, librarian, infrastructure provider etc.) Every group contains 3-4 participants, with 24 participants in total.
Clear Language in the Consent Form

1. I have had the possibility to review the material shared and remove any information that I do not wish to share openly.

2. I understand that publication of the data implies that it will not be possible to remove the material post publication.

3. I understand that despite removal of directly identifiable information such as name and workplace, it could still be possible to identify me based on the information in the interviews.

I consent to publication of the following data:

• Interviews from winter 2018
• Questionnaire from autumn 2018
• Interviews from spring 2019
• This consent (pseudonymized) from summer 2019
“Even if the information I have provided is in no way sensitive or provoking, I feel some hesitation in accepting publication of the interview material, even in pseudonymised form. This is interesting and will be a useful experience.”

(Research support)

“Selv om informasjonen jeg har gitt ikke på noen måte er sensitiv eller provoserende, kjenner jeg en viss nøling med å akseptere publikasjon av intervjumateriale, selv under pseudonymisert form. Det er interessant og vil være en svært nyttig erfaring.”

(Forskerstøte)
Pros

• Forces me to plan
• Gives me an overview of what I want to keep and what to delete
• I have access and possibility to use and refer to the data after my projects ends
• Learned a lot and met my own ideals in the door sometimes
Cons

• Multiple versions, as I did not get permission to share everything.
• Time consuming
• No credit for sharing data
• Plenty of work, but not sure if the data are relevant to anyone
“There is something with qualitative interviews where the whole essence is about the specific details, so if you suddenly start removing all that is identifiable then you have kind of also removed the object, many people do not understand this.” (Researcher)

“Det er noe med kvalitative intervjuer hvor hele essensen handler om det spesifikke, så hvis du plutselig begynner å fjerne alt som er identifiserbart så har du egentlig også fjernet gjenstanden, det er det mange som ikke forstår.” Forsker
Take aways

Sharing means documenting and spending extra time on organising data

It is possible to balance privacy and data sharing, but it might not always be right towards the participants to push towards open sharing.

Privacy is everything
“If you have nothing to hide you are nothing”
(Zuboff 2019)
References


