Researchers' Perspective on the Publication of Research Data: Semi-structured Interviews from Germany

ReadMe

These semi-structured interview data are part of the project "Research Integrity and Privacy" of the HEADT Centre (https://headt.eu), a study on the presumed conflict between publishing of research data and the issues of data protection.

Researcher:

> Katarzyna Biernacka
> Humboldt-Universität zu Berlin
> Department of Computer Science
> Computer Science Education | Computer Science and Society
> biernack@hu-berlin.de
> ORCID: 0000-0002-6363-0064

Project members:

> Principal Investigator: Prof. Niels Pinkwart
> Student assistants: Malte Smits and Rudi Rudiak

Files:

- data_dictionary.csv
  # Overview of the data sets
- os_001.docx and os_001.pdf
  # Translated semi-structured interview with a researcher from climate impact research in Germany
- os_002.docx and os_002.pdf
  # Translated semi-structured interview with a researcher from learning analytics in Germany
- os_003.docx and os_003.pdf
  # Semi-structured interview with a researcher from bio-medicine in Germany
- os_004.docx and os_004.pdf
  # Translated semi-structured interview with a researcher from bio-medicine in Germany
- os_005.docx and os_005.pdf
  # Translated semi-structured interview with a researcher from climate impact research in Germany
- os_009.docx and os_009.pdf
  # Translated semi-structured interview with a researcher from climate impact research in Germany

- os_010.docx and os_010.pdf
  # Translated semi-structured interview with a researcher from medicine in Germany
- os_011.docx and os_011.pdf
  # Translated semi-structured interview with a researcher from learning analytics in Germany
- os_012.docx and os_012.pdf
  # Translated semi-structured interview with a researcher from medicine in Germany
- os_013.docx and os_013.pdf
  # Translated semi-structured interview with a researcher from learning analytics/medicine in Germany
- os_014.docx and os_014.pdf
  # Translated semi-structured interview with a researcher from learning analytics in Germany
- os_015.docx and os_015.pdf
  # Translated semi-structured interview with a researcher from climate impact research in Germany
- os_024.docx and os_024.pdf
  # Translated semi-structured interview with a researcher from learning analytics in Germany
- questions.pdf
  # Questions used for the semi-structured interviews
- transciption_norms.pdf
  # Norms and rules that have been applied for transcription

1. Data collection

The semi-structured interviews included in this data set were collected between June 2019 and October 2019 with scientists from Germany. Researchers from learning analytics, medicine and climate impact research were interviewed in order to assess their behaviour regarding the publishing of their research data (see questions.pdf for more details). Both junior (2-4 years of experience) and senior scientists (more than 5 years of experience) were included in the semi-structured interviews.

The semi-structured interviews were conducted 1) in person and recorded with a TASCAM DR-40 recorder, or 2) via DFNconf video conference system with the camera switched off. In the latter case the audio was extracted to an m4a-file in the next step. The audio files, in both cases, are preserved at the TSM service of the Humboldt-Universität zu Berlin, where data access is protected and regulated via a login. Third parties have no access to the data.

The semi-structured interviews were mostly conducted in German, but were translated into English for the purpose of the research project. The transcriptions published in this data set are all in English.

2. Data processing

The produced audio files were transcribed by two student assistants (see project members) using the denaturalized technique in order to preserve the purest and least altered form of the interview. In this way the transcription is as transparent as possible. The transcription norms and rules defined in the transcription_norms.pdf were adopted. The same project members translated the interviews into English.

The anonymisation was carried out manually by the researcher and subsequently checked by the project members. In most cases, it was decided to aggregate the information, e.g. "university" or "Germany" (instead of the exact institute name or city name). Where it was not possible otherwise (because the risk of identification of the participants was too high) strong abstraction was used (e.g. "name").

3. Data evaluation

The qualitative evaluation of the semi-structured interviews is based on the Grounded Theory methodology. These hypotheses shall be verified and quantified via a wide spread online survey that will follow.