Perspectiva de los Investigadores sobre la Publicación de Datos de Investigación: Entrevistas Semiestructuradas de Perú

ReadMe (English)

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

Transcription: Christine Schnichels

Files:

- data_dictionary.csv
  # Overview of the data sets
- ReadMe_ES.pdf
  # Description of the data set in Spanish
- os_017.docx and os_017.pdf
  # Semi-structured interview with a researcher from learning analytics in Peru (in Spanish)
- os_018.docx and os_018.pdf
  # Semi-structured interview with a researcher from learning analytics in Peru (in Spanish)
- os_019.docx and os_019.pdf
  # Semi-structured interview with a researcher from bio-medicine in Peru (in Spanish)
- os_020.docx and os_020.pdf
  # Semi-structured interview with a researcher from medicine in Peru (in Spanish)
- os_21.docx and os_021.pdf
  # Semi-structured interview with a researcher from medicine in Peru (in Spanish)
- os_22.docx and os_022.pdf
  # Semi-structured interview with a researcher from medicine in Peru (in Spanish)

- os_23.docx and os_023.pdf
  # Semi-structured interview with a researcher from climate impact research in Peru (in Spanish)
- preguntas.pdf
  # Questions used for the semi-structured interviews (in Spanish)
- transciption_norms.pdf
  # Norms and rules that have been applied for transcription (in Spanish)

1. Data collection

The semi-structured interviews included in this data set were collected in October 2019 with scientists from Peru. 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 **preguntas.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 in person and recorded with a TASCAM DR-40 recorder. The audio files 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.

2. Data processing

The produced audio files were transcribed by an assistant 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 **normas_transcripcion.pdf** were adopted.

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 "Peru" (instead of the exact institute name or city name). Where it was not possible otherwise (because the risk of de-identification 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.