

Data management plan for the project “Knowledge networks in transnational higher education policymaking”

PI: Professor Jaakko Kauko, University of Tampere

- Application number: 340267
- Funder: Academy of Finland
- Date of data management plan: 2.6.2021

1. General description of data

We will collect three kinds of data: (1) quantitative data from public sources; (2) qualitative interview data; and (3) qualitative data from observations in public and private meetings.

1) The quantitative data for the network analysis will be collected via desktop research on working-group memos, government online project archives, the European Commission’s Register of expert groups and other sources. An important source will be the Finnish government’s project management service, *Valtioneuvoston Hankeikkuna* (<https://api.hankeikkuna.fi/>), providing data on membership of government working group and committees as well as their memos with membership information. The register does not allow downloading of data in database format, so we will make a request for information to the registry office of the Ministry of Education and Culture (MEC). We expect to analyse hundreds of documents to form a database for the knowledge network: the network will be based on the people involved in the working groups, projects, and registers concerned. The quantitative network data will be stored in .csv and .xml files.

We already have an existing database acquired the registry office of the MEC. Through the registry office, we received two separate reports for the period from 1 January 2010 to 14 November 2018, the first of which lists 643 projects in the database in the MEC’s field in tabular form and the second of which contains the members of these projects in text form. The database includes basic identification information and the names, roles, and terms of office for working group members. The tabular report includes the name of the working group, summary of their mandate, type (organ, pre-legislative project, strategy, or project), more specific type (e.g., management team), status of the project (e.g., ongoing or planned), term of office, information about responsible or contact persons as well as some technical identification information about the working group. We aim to request an update of this data starting from 14 November 2018 and merge this new data with the existing MEC data and the new data that we will collect from other public sources.

2) The qualitative interview data will be gathered in thematic interviews, the interviewees being policy actors (policymakers, officials, policy advisers, consultants, experts, and other stakeholders) in Finland and the EU in the field of higher education policy. We plan to conduct 20–30 interviews, each lasting approximately one hour, and they will be transcribed. We therefore expect around 30–40 hours of audio files (stored in WAV or MP3 format) and 500–800 pages of transcribed text in plain text format.

3) The qualitative observation data will be collected at policymaking occasions in working groups and committee meetings in Finnish and EU organs, both public (e.g., EU committee meetings) and non-public. The observations will be based on a loose framework and written down by hand. The amount of observation time will depend on the timetables and agendas of the working groups and committees, but we expect to produce a few hundred pages of fieldnotes and diaries. These texts will be electronically transcribed and stored in plain text format.

2. Ethical and legal compliance

The project is committed to complying with all rules, regulations, and directives that have come into force in Finland (ex. EU’s GDPR regulations) to protect personal data. The data management plan will be updated on a regular basis (every year or if major legislation is passed) to track any changes and increase transparency and the quality of the research. The PI will be responsible for data use in the

project. Additionally, all members of the project will go through a training program (if they have not already) covering EU's GDPR regulations, Finnish law and regulations, and university regulations relating to the collection, use, and storage of personal data. This is to increase the probability all data will be handled responsibly throughout and after the research process.

With regard to the quantitative network data (1), most of it is of public record, for instance the government database archive is registered with CC-BY 4.0 (<https://api.hankeikkuna.fi/api>). Further coding might bring additional layers in the description files that can include results from the interviews. What can be archived will be decided with specialists from the Finnish Social Science Data Archive (FSD <https://www.fsd.tuni.fi/en/>). For the parts of the data, where further coding is done, the researchers in the project will retain ownership and copyright of the archived data, which will be preserved and accessed under established guidelines and regulations of FSD. Data will be used in accordance with standard academic practice by referring to the original source of the data.

Qualitative data (2, 3) is inherently high risk in its further use for identification purposes and therefore our starting point is that this will not be shared via FSD. The network analysis makes the identification of the interviewees relatively easy. Some of the intended interviewees are in prominent positions and have high visibility and can easily be exposed by their word use. In addition, and given the nature of the researched network, controversial opinions might result in discrimination and the exclusion of individuals. During the project, however, we will explore the nature of the data, the possibilities and resources for anonymisation, and the way in which some parts could be published with a long embargo, for instance.

3. Documentation and metadata

Data 1 will be made available via FSD once the project is completed and will follow FSD metadata standards. Concerning the quantitative data (1), the metadata will identify the nodes and edges connecting within the networks. Data files in .CSV format are allowed by software in use (UCINET, R, Python, SQL). The qualitative interview data (2) will include metadata information in the format of basic interview identifiers (e.g. time, project, date, and interviewer) and the codes used in the analysis. The transcribed interviews will most likely be coded with Atlas.ti and the metadata of codes and quotations will be stored in the Atlas.ti Hermeneutic Unit (.hu) in copy bundle format (.atlcB). The observation data will have the same type of metadata, when applicable, as the interviews and its analysis are also similar.

The research team will ensure the quality and consistency of the qualitative data (2, 3) by planning, drafting, testing, and finalising the interview themes and observation templates collaboratively; the Advisory Board will be consulted as needed. Recorders (where possible, two) will be used and tested for interviews (data 2). Regarding publicly available quantitative data (1), the team agrees on principles of inclusion and exclusion to ensure its consistency for further use.

4. Storage and backup during the research project

During the project the data will be stored electronically on a university secure network drive (owned physically by the university) and will be password-protected: when it is moved it will be protected by means of additional strong file encryption (Trusted Platform Module hardware encryption chip or software encryption). Identifier codes (on the data files) will be used and participant information will be stored separately. Data sharing for the purposes of the project will be done internally via the university network drive, which can be accessed via the virtual private network (vpn). Parts of the data can be temporarily stored and accessed electronically on a GDPR-compliant Microsoft-provided university secure cloud Onedrive for Business administered by the Tampere University.

Members of the project's main research group will have access to the data and may carry an encrypted copy of it. Only the manually written observation notes for the day in question will be carried unencrypted. Any other access to the data during the project, such as for master's thesis work or

Advisory Board use, will be dependent on the PI's approval following what has been agreed with the participants.

The observation diaries (data 3) will be stored in a locked cabinet at the University premises or in a hotel safety deposit box when travelling. As soon as possible, they will be stored in electronic plain text format to allow digital back-up.

Data backup files will be stored automatically at the University of Tampere on the network drive, protected by the strong safety protocols of the IT Administration.

5. Opening, publishing, and archiving the data after the research project

Co-authors outside the research team, such as members of the Advisory Board, will have limited access to small parts of the data during the research period only through the team, which will evaluate any anonymity risks on a case-by-case basis.

We will use the services of the FSD to ensure the storage of quantitative network data 1 for further use after the project. The storability of this dataset will be handled by the research team and the practical work done by a research assistant in co-operation with the FSD.

Based on the concerns raised above (section 2), the publication of interview data and observational data (2, 3) would pose too high a risk for the interviewees' anonymity and the nature of the data may also lead to discrimination and exclusion. Part of transcripts of data 2 may be public via FSD if agreed with participants. Selected metadata of data 2 and 3 will be public (e.g. professional position and used data coding), and reported in the research outputs. After the project, the permanent storage of data 2 and 3 transcripts is handled by the PI in the Tampere university network drive. The expected lifespan for longitudinal analysis is in principle decades. After the nature of the research material is clearer and with the permission of the participants, we will explore the possibilities of opening parts of this data (2, 3) via the FSD with a long embargo.

Dispensable data files and temporary files created when programs are used will be erased when they are no longer needed. The audio files and observation diaries will be destroyed after the original research has been completed to protect privacy.

6. Data management responsibilities and resources

The project PI will have responsibility for the data management of the project. All team members have agreed to follow the guidelines and regulations set forth in this document and by the Academy of Finland and FSD to ensure the responsible, legal, and ethical collection, use, and storage of data for this project. Additional resources can be found below relating to specific aspects of the plan.

Academy of Finland: <https://www.aka.fi/en/funding/apply-for-funding/az-index-of-application-guidelines/data-management-plan/data-management-plan/>

FSD: <https://www.fsd.tuni.fi/en/>

Tämä dokumentti on allekirjoitettu sähköisesti TUNI Sign-järjestelmällä

This document has been electronically signed using TUNI Sign

Päiväys / Date: 02.06.2021 14:08:53 (UTC +0300)

Päivi Pahta

dekaani Kasvatustieteiden ja kulttuurin tiedekunta

Organisaation varmentama (TUNI käyttäjätunnus)

Certified by organization (TUNI user account)

Certified by organization