Plan Overview

A Data Management Plan created using DMPTuuli

Title: Tohtorikoulutettavien tutkimusdatanhallinta: tärkeys, osaaminen ja koulutus tieteenalojen, tutkimusmenetelmien ja datatyyppien episteemisissä kulttuureissa

Creator: Jukka Rantasaari

Principal Investigator:

Data Manager:

Affiliation: University of Turku

Template: Tuuli DMP

ORCID iD: 0000-0001-5927-3781

Project abstract:

Tavoite – Tutkimuksen tavoitteena on selvittää, 1) kuinka tohtorikoulutettavat ja tiedekunnan jäsenet kokevat tutkimusdatanhallinnan tärkeyden ja millaisina he näkevät tohtorikoulutettavien tämänhetkiset datanhallinnan taidot ja osaamistarpeet, 2) miten moniammatillisella koulutuksella pystytään vastaamaan nuorempien tutkijoiden datanhallinnan osaamistarpeisiin sekä lisäämään heidän ymmärrystään hyvien käytäntöjen merkityksestä tutkimuksen läpinäkyvyydelle, eheydelle ja luotettavuudelle, 3) miten datan elinkaari sekä datanhallinnan tärkeiksi koetut toimenpiteet ja osaamiset poikkeavat tieteenaloittain, tutkimusmenetelmittäin tai datatyypeittäin, 4) millä tavalla tämän tutkimuksen empiirisissä tuloksissa näkyvät eri tieteenalojen, tutkimusmenetelmien tai datatyyppien episteemiset kulttuurit.

Rakenne - Tutkimus koostuu neljästä artikkelista ja ne yhteen kokoavasta yhteenvedosta.

Ensimmäisessä artikkelissa keskitytään kysymys 1:een. Toisessa artikkelissa vastataan kysymykseen 2, erityisesti datanhallinnan koulutuksen rakenteen, palautteiden ja kyselyvastausten osalta. Myös kolmannessa artikkelissa vastataan kysymykseen 2 arvioimalla datanhallinnan koulutuksessa laadittujen datanhallintasuunnitelmien laatua. Neljännessä artikkelissa vastataan kysymykseen 3. Tutkimuksen yhteenvedossa vastataan kysymyksiin 1-4.

Toteutus – Artikkelissa 1 tutkimusmenetelmänä oli kvantitatiivinen strukturoitu haastattelu sekä avointen vastausten kvalitatiivinen sisällönanalyysi. Tutkimus julkaistiin vuonna 2021. Artikkelissa 2 vuosien 2019-2021 datanhallintakurssin toteutusten jälkeisen survey-kyselyn tulokset analysoitiin kuvailevaa tilastoanalyysiä ja tilastollista päättelyä käyttäen. Lisäksi palautteet analysoitiin grounded theory-inspired -menetelmällä. Tutkimus on parhaillaan vertaisarvioitavana. Artikkelissa 3 analysoidaan kuvailevan tilastoanalyysin ja tilastollisen päättelyn avulla em. kurssilla laaditut datanhallintasuunnitelmat vuosilta 2020-2022. Käsikirjoitus valmistuu loppuvuoden 2022 aikana. Artikkeli 4 on teoreettinen ja sen tutkimusmenetelmänä on kirjallisuuskatsaus ja kvalitatiivinen ja kvantitatiivinen sisällönanalyysi. Käsikirjoitus valmistuu n. 1 vuoden kuluttua rahoituksen alkamisesta. Tutkimuksen yhteenveto toteutetaan kvalitatiivisena ja kvantitatiivisena sisällönanalyysinä, jossa artikkeleissa 1-4 raportoituja tuloksia peilataan episteemisten kulttuurien viitekehyksen kautta. Yhteenveto valmistuu 2 vuoden kuluessa rahoituksen alkamisesta.

Odotetut tulokset – Tutkimus tuottaa uutta tietoa datanhallinnan koetusta tärkeydestä, tämänhetkisestä osaamisesta ja osaamistarpeista sekä moniammatillisen koulutuksen vaikutuksesta osaamiseen. Kontekstoimalla datan elinkaaren, datanhallinnan koetun tärkeyden ja osaamistarpeiden geneeristä analyysiä tieteenala-, tutkimusmenetelmä- ja datatyyppikohtaisesti, edistetään datanhallinnan hyvien käytäntöjen laajempaa ja syvällisempää omaksumista ja siten läpinäkyvämmän ja toistettavamman tutkimuksen tekemistä. Tämä mahdollistaa tutkimustuotosten entistä monipuolisemman hyödyntämisen, tehostaa tiedettä ja tutkimusta ja lisää yhteistyömahdollisuuksia mm. isojen kv. ongelmien kuten ilmastonmuutoksen ja pandemioiden torjumisessa.

ID: 7520

Start date: 01-08-2019

Last modified: 30-04-2022

Tohtorikoulutettavien tutkimusdatanhallinta: tärkeys, osaaminen ja koulutus tieteenalojen, tutkimusmenetelmien ja datatyyppien episteemisissä kulttuureissa

General description of data

What kinds of data are collected or reused?

Doctoral students are interviewed and asked about their research project, what kind of data they are collecting and managing, what are their data management principles, practices, needs, skills, trainings, etc.
Faculty members (supervisors with a few biostatisticans) are interviewed and asked about their students' data management practices, needs, skills, trainings, etc.
Faculty members and doctoral students are asked to evaluate importance of the RDM competencies / doctoral students' current RDM competencies.
Informed by the interview and previous research literature, a basic, multi-stakeholder RDM course (BRDM) is built together with academic and research support professionals.
During the course, participants are asked to fill in module based feedback.
After the course, participants of the BRDM Course (The Basics of Research Data Management, 3 ECT credit, UTUGS -courses 2019-2020) are asked to fill in the survey in which they self-assess their RDM skills before and after the course. They will also express their needs for further education, rate the course, and give a course feedback.
Participants of the BRDM 2020 Course return their personal DMP:s that they have prepared in the Course. PI analyses and assesses returned DMPs for to assess and develop the curriculum using criteria developed for this purpose by PI.
Please find more accurate, updated and precise information of the data types and other specifications in the data table.

What file formats will the data be in?

Please find accurate, updated and precise information of the data types, formats, documentation and metadata, and other specifications in the data table.

Documentation and Quality

How will the data be documented?

CSV files that have been created in Webropol and REDCap software, have variables and descriptive metadata (Interview questions work as variables and description). Interview and feedback forms have been prepared in Webropol (2019), survey form in REDCap (2020). There are more descriptions in the forms concerning different data lifecycle stages. Different versions of the data (CSV files with interview/feedback/survey data plus visualitions) are stored in Seafile during research process. Empty interview/feedback/survey forms will also be saved as pdf documents. Answers to extra questions like other notes made by interviewer are first written down in Word document, after which the answers will be transferred to the same CSV files where the main questions and the answers of the interviewees are.
Research Diary will be kept in IOS Notes during the data collecting, processing, analysing, storing and preserving.
Readme.tab will be added in excel-files covering information of the:
- name of the researcher
- license
- name of the project
- description of the coded answers (scale of numerical answers) in interview form
- questions of the interview form

When preserving, excel files will be converted to csv-files and variable level metadata information will be written in readme.txt -files. Besides of variable level metadata, also study level metadata will be added (DDI in FSD repository; free keywords in Zenodo repository.
Please find more accurate, updated and precise information of the data types, formats, documentation and metadata, and other specifications in the data table.

How will the consistency and quality of data be controlled and documented?

I ask fixed questions in the same order from all the interviewees. Interviewees answer the questions and fill themselves the interview form during interview. Principal Investigator, with the help of research assistant, will transfer the answers to extra questions and other notes concerning the interview from word document to webropol / REDCap based excel sheet after every single interview. During transferring PI makes sure that all the information has been correctly and consistently coded.
There is a difference between the two questionnaires 1) Doctoral Students' and 2) Supervisors'. Likert scale goes from 1 to 5 in DS questionnaire, but from 2 to 6 in SV questionnaire. This difference in scales have been taken account and have been calibrated before getting results: We have converted supervisors' answers to go from 1 to 5 scale, before analysing answers (e.g through tables and visualisations).
The master files will be named and stored separately from modified and processed files. Besides in Webropol and REDCap, master files will be stored in their own folders in Seafile and My personal UTU Net Drive. Folders and Files will be named in a systematic way using acronyms, version ID:s and dates, e.g. "Diss_Data_table_v02_2020_09".
PI updates research diary in IOS Notes, e.g. marking down the decisions and methods concerning different phases of data processing and analysing.

Storage and Backup

How will the data be stored and backed up?

All the materials are stored in PI:s personal net-drive and back-upped regularly in Seafile (seafile.utu.fi) secured cloud service.

How will you control access to keep the data secure?

Access to net-drive is personal: Principal Investigator only. PI is the only one who is responsible for collecting, processing, storing, analysing and preserving the data. Access to Seafile, if needed, is granted individually by principal investigator. All data requires user ID and password.

Ethics and Legal Compliance

How will ethical issues be managed?

The interviews:

PI has described the research process, conducting interviews and target of the interviews in the beginning of the questionnaire ("Information on the research project and the interview"). The rights of the interviewees (e.g participation is voluntary and intervieweee can interrupt his/her participation any time) have also been told in the beginning of the questionnaire.

In questionnaires there is also included informed consent ("Confidentiality in the research publication and in the research data. Preservation and potential reuse of the research data") in which interviewees are told e.g, how the interview data (answers to questions) will be processed, that all the identifiable data will be deleted, and only statues (doctoral student, supervisor or biostatistician) of the interviewees will be told. There is also told the long term preservation repository, where the data will be preserved after the project, to what purposes the research data can be utilised after this research project, etc.

Before interview we go all the afore mentioned details through together with the interviewee, after which he/she can either accept it as such, or (s)he can accept it with certain conditions that (s)he want to define.

The above chapters ("Information on the research project and the interview" and "Confidentiality in the research publication and in the research data. Preservation and potential reuse of the research data") have been read and accepted by Head of the Legal Services of University of Turku.

Basics of Research Data Management (BRDM)-course:

PI will conduct pre survey and after survey before and after the course in 2019, and after survey in 2020, in which participants are asked to fill their evaluation of the importance / skills concerning different stages of research data management (RDM). Survey is anonymous and voluntary: PI only ask their study programme and their faculty, so there will not be created any kind of registry of personal information. I have asked the data security officer of the University of Turku read and comment our survey questionnaire and she has approved it.

In Teacher survey (after BRDM 2020 Course) I ask the name of the teacher. The basis for the handling of the personal information is scientific research. I have asked permission to share the anonymised data from the participants of the survey. See the privacy notice.

I don't handle sensitive information in this research project, so there is no need to risk analysis.

PI is the data controller.

Please find more accurate, updated and precise information of the data types, formats, documentation and metadata, ethics and privacy, and other specifications in the data table.

How will ownership, copyright and Intellectual Property Right (IPR) issues be managed?

Principal Investigator (Jukka Rantasaari) is the owner of the above described data.

Please find more accurate, updated and precise information of the data types, formats, documentation and metadata, ethics and privacy, ownership, and other specifications in the data table.

Data Sharing and Long-Term Preservation

How, when, where and to whom will the data be made available?

Anonymised research data sets will be preserved and made accessible in FSD and Zenodo, after each article, where applied, and after the research project, in more full format. Before preservation I will go carefully through all the data (CSV files) and anonymise the data. Metadata will be DDI in FSD, and readme.csv/txt files in Zenodo. Concerning the data sets, that cannot be shared,please check the data table.

How and where will data with long-term value be made available?

The CSV files, which contain answers to research questions, additional questions and metadata (readme tab) and which are created by Webropol or REDCap and supplemented by interviewers will be archived and shared (CC BY 4.0) for studying, teaching and research purposes on Finnish Social Sciences Data Archive (FSD) in Tampere and in Zenodo repository.

Please find more accurate, updated and precise information of the data types, formats, documentation and metadata, ethics and privacy, preserving, sharing, and other specifications in the data table.

Have you estimated costs in time and effort for preparing the data for preservation and sharing?

The data will be documented, metadated and curated all the time during the research process, so it's hard to separate, which part of the process is preparing data for preservation and sharing, which is e.g. processing and analysing the data to get good quality results. Taking care of data, which is called data curation, is natural and essential part of the sound research process to get the coherent and good quality results. How could we otherwise have reliable results?