

# Kokemuksia datanhallintasuunitelmista kirjaston opetuksissa

Mika Sarlin, informaatikko, Turun yliopiston kirjasto



# Kirjaston datanhallintaopetukset

- Tuuli-työpajat Akatemian haun yhteydessä
- Kirjaston UTUGS-kurssi *Information Resources and Tools for Research*
- Kirjasto mukana myös UTUGS-kurssissa *Tutkimusdatan hallinnan perusteet*
- sähköpostineuvonta

# Mitä tutkimusdata on?

- Selkeää, yleispätevää vastausta ei ole!
- Esim. Google-haku ”what is research data”
  - *”Defining ‘research data’ is challenging. The challenge is: There is not a consensus on the definition. It varies according to discipline It varies according to the research funder “ (University of Leicester)*
  - *Research data is any information that has been collected, observed, generated or created to validate original research findings. (University of Leeds)*
  - *any information collected, stored, and processed to produce and validate original research results (Dewitt Wallace Library – Macalester College)*

# Selvät sävelet

- Samalla tutkijalla voi olla sekä kvalitatiivista että kvantitatiivista dataa
- Selviä pehmeitten tieteitten dataja ovat mm.
  - Kyselytutkimusten vastaukset
  - Haastattelut (äänitiedostot, videotiedostot)
  - Mittaustulokset (esim. foneettinen data, arkeologiset mittaustulokset)
  - Eli ylipäätään tutkijan itse keräämä (haastattelemalla, mittaamalla, jne.) data
  - Digitaalinen data usein selkeämpää kuin analoginen (digital humanities)

# Minulla ei ole dataa

- ”Luen vain kirjoja ja artikkeleja ja sitten kirjoitan niistä.”
- Kirjoitatko oikeasti tekstiäsi suoraan ilman mitään välivaiheita?
- Etkö tee mitään muistiinpanoja?
  - Miten järjestät muistiinpanosi? Ovatko ne post-it-lapuilla?
- Teetkö mitään taulukoita tms.?
- ”Ei minulla ole kuin noita Exceleitä...”
- Aineistonhallintasuunnitelmaan voi toki kirjoittaa, että dataa ei ole

- Datanhallintasuunnitelman tarkoitus on, että tutkija kiinnittää huomiota mahdollisiin datanhallintansa ongelmakohtiin.
  - Miten huolehdit muistiinpanojen varmuuskopioinnista?
  - Miten varmistat, että kahvikupin kaatuminen, vesivahinko tai vaikka tulipalo ei tuhoa koko tutkimustasi?
  - Jos olisit itse jakamassa tutkimusrahoitusta, antaisitko rahaa projektille, joka on näin haavoittuvainen?
- Jos työsi aikana syntyvä ainoa data todella on vain kirjoittamasi teksti, miten huolehdit sen tallentamisesta, varmuuskopioinnista ja suojaamisesta? (ei datanhallintaa)

# Digitaalinen data vrs. analoginen data

- Akatemian haussa ei selvästi rajata datan käsitettä vain digitaalisen dataan. Sen sijaan Turun yliopiston datapolitiikassa puhutaan vain ja ainoastaan digitaalisesta datasta.
  - Kun puhutaan datasta, ota aina selvää, mitä sillä tarkoitetaan!
- Jos keräät analogista aineistoa, miten järjestät sen, miten huolehdit sen säilyvyydestä, tietoturvasta, jne.?

# Oma data vrs. Muiden data

- Datahallintasuunnitelmassa tutkija kertoo, miten **hän itse** huolehtii tutkimukselle tärkeästä datastaan.
  - Usein näkee pitkiä selityksiä siitä, miten joku toinen on kerännyt datan ja miten hän sen säilyttää.
  - Toisen keräämää dataa voi toi käyttää (ja on suositeltavaakin käyttää), mutta se ole tutkijan omaa datanhallintaa. Tutkija ei välttämättä edes tiedä, miten kolmas osapuoli teknisesti huolehtii datasta.
  - Kerro sen sijaan, miten itse käsittelet ko. dataa, miten siitä syntyy uutta data, miten huolehdit sen varmuuskopioinnista, tietoturvasta jne.



- Sama koskee analogista dataa. Tarkat kuvaukset esim. siitä, miten jokin museo huolehtii aineistostaan, ei ole sinun datanhallintaasi, se on museon ”datanhallintaa”.

# Tutkimusdata vrs. Lähdekirjallisuus

- Lähdekirjallisuus (kirjat, lehtiartikkelit) ovat yleensä lähdekirjallisuutta, eivät tutkimusdataa
  - Kuuluvat lähdeluetteloon, eivät aineistohallintasuunnitelmaan
    - Paitsi jos tutkimuksen tarkoitus on nimenomaan luoda bibliografia (esim. tieteen historiaa)
- Lähdekirjallisuudesta tulee dataa, jos tutkija muuttaa sitä esim. toiseen muotoon (esim. digitoi ja luo OCR:n avulla tekstitiedostoja). Selvää tutkijan luomaa dataa on esim. kaunokirjallisuuden pohjalta luotu kielitieteellisesti annotoitu, sähköisessä muodossa oleva tekstikorpus.

- Kaunokirjallisuus: lähdekirjallisuutta vai tutkimusdataa??
  - Teokset ovat (yleensä) olleet olemassa jo ennen ja ne ovat useimmiten helposti saatavissa (julkaistuja teoksia).
  - Tutkijan ei ole tarvinnut tehdä mitään varsinaista aineiston keruuta.
  - Kyseessä ei ole myöskään tutkimuksen aikana syntyvästä datasta.
  - Kaunokirjallinen teos on tutkimuksen kohde
- Datanhallintasuunnitelmassa ei kerrota, miten huolehdit lähdekirjallisuuden tallentamisesta (“pidän kirjoja hyllyssäni”).

- Lähdekirjallisuus muiden datana??
  - Julkaistut teokset (lehdet, artikkelit), oli ne sitten primääri- tai sekundäärilähteitä, voidaan nähdä muiden tekemänä datana.
    - Vertaa toisen keräämän datan käyttö: datan alkuperä mainittava (lähdeviitteet), mutta data ei ole tutkijan omaa (itse keräämää tai itse tuottamaa) dataa, vaan lähde. Kerro datanhallintasuunnitelmassa, mitä konkreettista teet datalle (miten siitä syntyy uutta dataa)
  - Jos tutkija ei tee datalle mitään (esim. koneellisesti), eikä uutta (digitaalista) dataa synny, tutkija vain tulkitsee dataa. Tulkinnan tulokset julkaistaan esim. artikkelina tai kirjana (eikä esim. jossain datarepositoryssa)



**TURUN  
YLIOPISTO**