

Tampereen yliopiston dataselvitys 2017

Tomi Toikko ja Saira Huuskonen
Informaatiotutkimuksen päivät 2.11.2018

TAMPEREEN YLIOPISTON KIRJASTO
TAMPERE UNIVERSITY LIBRARY



Dataselvityksen tausta

- Keväällä 2017 Tampereen yliopiston avoimen tieteen kyselyssä dataosuus, kysymyksinä esim. aineiston avaaminen ja esteet avaamiselle.
 - Alle 30 % tallentanut aineistojaan arkistoihin.
 - Noin 50 % hyödyntänyt muiden kokoamaa tutkimusaineistoa opetuksessa tai tutkimuksessa.
 - Avaamisen esteinä mm.
 - Eettiset seikat
 - Tietämättömyys
 - Avaamisen vaatima työmäärä
 - Sopimusten puute tutkimusryhmässä
- Avoin tiede -kyselyn data: <http://urn.fi/urn:nbn:fi:fsd:T-FSD3221>



Tampereen yliopiston dataselvitys 2017

- Tutkimuskysymykset
 - Millaisia tutkimusaineistoja yliopistossamme käytetään?
 - Millaisia määriä tutkimusaineistoa on?
 - Millaista työmäärää aineiston kerääminen vaatii?
 - Mitä aineistoon liittyviä eettisiä seikkoja mainitaan?
 - Onko aineisto tallennettu jatkokäyttöä varten?
- Toteutus
 - Tutkimustietojärjestelmästä vuonna 2017 rekisteröidyt A1-julkaisut (n=307)
 - Data järjestettiin OKM:n päätieteenalaluokkien mukaan
 - Ensimmäisen vaiheen otos 20 % ja toiseen vaiheen otos 20 % päätieteenaloista
 - Tiedot kerättiin Excel-pohjaan



Selvityksen aineisto

- 129 A1-artikkelia
- Rajoitukset:
 - Kieli: suomi/englanti
 - Artikkelit saatavana avoimesti tai TaY:n tilausten kautta
- OKM:n päätieteenaloittain:
 - Lääketiede ja terveystieteet (n=72)
 - Yhteiskuntatieteet (n=35)
 - Luonnontieteet (n=12)
 - Humanistiset tieteet (n=8)
 - Tekniikka (n=2)
- Kansainvälisiä yhteisjulkaisuja n=56 (43%)
- JUFO-tasoinen:
 - Taso 0 (n=5)
 - Taso 1 (n=82)
 - Taso 2 (n=29)
 - Taso 3 (n=13)



Miten paljon aineistoa on ja mikä on sen vaatima työmäärä?

- Esimerkkejä aineiston määrästä
 - "970 kudosnäytettä tietokannasta"
 - "Cunninghamin kirja By Nightfall"
 - "Aiempi tutkimus (European Social Survey), kulttuuritilastot (Eurostat, Unesco ja Council of Europe). n=24,887 (21 maata)"
- Muutkin kuin tutkija osallistuu
 - "20 haastateltavaa"
 - "119 vastaajaa"
 - "31872 miestä koeryhmässä ja 48295 miestä kontrolliryhmässä"
- Esimerkkejä työstä
 - "Soluviljelmiä haudutettiin 72 tuntia ennen kokeita, koe-eläimet hiiriä, näytteitä valmisteltiin yön yli"
 - "Kenttätyö 09/2009-06/2012"
 - "Havainnointia 2 vuoden ajan päivittäin"
 - "Työpajoja tutkijoiden ja yrityksen edustajien kesken, monta erilaista haastattelua...syksyllä 2012, videoiden teko ja niiden jakaminen YouTuben kautta...v. 2014 loppupuolella, haastatteluja...lokakuussa 2015."

Mitä aineistoon liittyviä eettisiä seikkoja mainitaan?

- Eettiset periaatteet, esim. Helsingin julistus
- Yli puolessa (n=69) mainintoja eettisistä luvista, esim.
 - PSHP, KYS, HUS
 - THL
 - National Animal Experiment Board of Finland
 - Luvat tutkittavilta tai lasten vanhemmilta
- Anonymiteetti
 - ”Yhtiöiden nimet pseudonymisoitu.”
- Yli puolessa (n=66) mainintoja kaupallisista kytköksistä, eturistiriidoista tai niiden puutteesta
- Koe-eläinten inhimillinen kohtelu
 - ”Hiirillä ruokaa ja vettä, ne lopetettiin inhimillisesti.”



Onko aineisto tallennettu jatkokäyttöä varten?

- Ei mainintaa (n=119)
- Maininta (n=10)
 - Aineisto saatavana kirjoittajalta pyydettäessä (n=6)
 - Data-arkisto (n=2) (GEO, UK Biobank)
 - Tutkimusyhteisön verkkosivut (n=1)
 - Lehden verkkosivut (n=1)



Mihin katosi aineistojen avaaminen?

- Ei mainittu, mutta kuitenkin tallennettu?
- Oma tutkimus vielä kesken, ei siksi tallennettu?
- Selvityksen aineistossa paljon lääketieteen tutkimuksia
→sensiitivinen aineisto →ei voi tallentaa?
- Ei aineiston omistajuutta (esim. rekisterit, tilastot, dokumentit)
→ei voi tallentaa?
- Aineisto muutenkin saatavilla (esim. kaunokirjallisuus)
→ei kannata tallentaa?
- Keskustelu aineiston jakamisesta ollut pinnalla vasta muutaman vuoden → aineistoa kerätessä ei ”ajankohtaista”?



Entä sitten – Datapalvelut käytännössä?

- Tunnistettava, että tutkimusdata on hyvin heterogeenista
 - Erilaisia tarpeita aineistonhallinnalle ja datapalveluille.
 - Aineistonhallintasuunnitelmalle ehkä yksi rakenne ok, mutta valmiita vastauksia ei ole olemassa.
 - Monta relevanttia data-arkistoa (esim. FSD, biopankit, geenipankit) ja aina dataa ei ole mahdollista avata.
- Datanhallinnassa ja sen suunnittelussa tarvitaan monenlaista osaamista
 - Parhaat palvelut erilaisten toimijoiden yhteistyöllä.



Entä sitten – Data tutkimusaiheena?

- Tieteenalojen trendien seuraaminen pitkällä aikavälillä
 - Esim. lisääntykö datan uudelleenkäyttö tai avaaminen?
- Datan elinkaari aineistonhallintasuunnitelmasta projektin päättymiseen
 - Käytännön prosessin seuraaminen
 - Toteutuuko suunnitelma todellisuudessa?
- Tutkimusaineistona aineistonhallintasuunnitelmat
 - Esim. Suomen Akatemialle lähetettyjen hakemusten analysointi.
 - Millaiset heikkoudet ja vahvuudet? Esim. millaiset eettiset asiat nousee esiin? Mahdollistavatko aineistonhallinnan käytännön ratkaisut aineiston avaamisen?



Lopuksi

- Kiitokset
 - Samppa Penttinen
 - Erno Juvonen
- Kysymyksiä?
- Yhteystiedot:
 - saila.huuskonen@staff.uta.fi
 - tomi.toikko@staff.uta.fi

