

ARTIKKELI

Kansalliskirjaston aineistojen jatkokäyttö tutkijakyselyssä

Liisa Näpärä

Kansalliskirjasto, Helsingin yliopisto

liisa.napara@helsinki.fi

<https://orcid.org/0000-0002-6473-5896>

Johanna Lilja

Kansalliskirjasto, Helsingin yliopisto

johanna.e.lilja@helsinki.fi

<https://orcid.org/0000-0002-3905-0035>

This paper examines the practices of data preservation and data sharing among the customers of the National Library of Finland (NLF). The research is based on a survey, which was made to gather information for developing the research services of the NLF as a part of Digital Open Memory project (funded by ERDF). The focus of the survey was on the users of digital material. Among many other questions, respondents were asked how they preserved their data during the research and what they did with their data after the research. The respondents were given an opportunity to comment on their willingness to share their data and to discuss the questions that affected their choices with their data. Mostly, the respondents preserved the data on their own personal use even though they were usually willing to share the data with others either from the beginning of the research or whenever there was a change for sharing. The results found in this survey indicated that the practices of data sharing among the customers of the NLF were quite similar to the practices found in studies concerning social sciences and humanities scholars internationally: only approximately 20 per cent of researchers deposited their material to a data repository after their research. Verbal answers and comments gave valuable information on the obstacles of data sharing, e.g. that the traditional methods of humanities studies produced data which was understandable only to the researchers themselves, or that the data was gathered from many different sources which were difficult to combine. The survey indicated that there is an obvious need to provide more concrete support for the researchers who are working with the digital data collected from the NLF services.

Asiasanat: tutkimusaineisto, ihmistieteet, menetelmät, avoimuus, kansalliskirjastot, avoin tiede, avoin data



Artikkeli on lisensoitu Creative Commons Nimeä-EiKaupallinen-JaaSamoin 4.0 Kansainvälinen -lisenssillä

Pysyvä osoite: <https://doi.org/10.23978/inf.99425>

Johdanto

Tutkijoiden halukkuudesta tallentaa ja avata tutkimusaineistonsa on tehty useita kyselytutkimuksia viime vuosina. Osa tutkimuksista antaa viitteitä siihen, että ihmistieteissä on tekijöitä, joiden perusteella aineiston jakaminen ei ole luontevaa – osin aineistoon itseensä liittyvistä, osin tutkimusmenetelmiin liittyvistä syistä. (Enwald 2018; Tenopir et al. 2015; 2020; Zhu 2019.) Tässä artikkelissa tarkastelemme Kansalliskirjaston tarjoamien aineistojen käyttäjien näkemyksiä aineiston tallentamisesta ja pohdimme, onko julkaistulla kulttuuriperintöaineistolla ja siihen perustuvalla tutkimuksella ominaisuuksia, jotka estävät tutkimusaineiston tallentamisen.

Digitaalisen murroksen tuomista haasteista tutkimusorganisaatioille ja datan määrän nopeasta kasvusta on keskusteltu koko 2000-luvun ajan (Laney 2001; Pinfield, Cox & Smith 2014, 2). Keskustelun painopiste siirtyi 2010-luvulla tieteen avoimuuteen, erityisesti avoimeen julkaisemiseen ja tutkimusaineiston hallintaan (Perrier et al. 2017, 9–10). Kansainväliset ja kansalliset linjaukset osoittavat nykyisen etenemisen suunnan (Plan S; Avoimen tieteen ja tutkimuksen julistus). Erityisesti tutkimusrahoittajien vaatimukset ovat merkittäviä kannustajia avoimen tieteen käytäntöihin. Sekä tutkimustuloksien että aineiston avoimuus kuuluvat esimerkiksi Euroopan Unionin Horizon 2020 -rahoituksen edellytyksiin (Féret & Cros 2019). Suomen Akatemia ja Tekes ovat suosittaneet samaa vuodesta 2015 lähtien (Räisänen 2016). Avoimen tieteen toteuttaminen alkaa vähitellen toimia myös yhtenä tutkijan meritoitumisen mittarina, mutta tutkimusaineistojen avaaminen ei vielä tapahdu samassa mittakaavassa kuin tutkimusjulkaisujen. Vuoden 2020 maaliskuussa *Tutkimuseettisen neuvottelukunnan* (TENK) tutkijan ansioluettelumalliin nostettiin aiempaa näkyvämmiin esille tutkimuksen tuotosten avaaminen vastuullisesti muutenkin kuin pelkkien julkaisujen muodossa (Tutkijan ansioluettelomallin 9. kohta: kehitetyt menetelmät, ohjelmat, infrastruktuurit, aineistot, oppaat ja työkalut).

Avoimen tieteen käytäntöjen vakiinnuttamiseksi on viimeisimpien vuosien aikana annettu koulutusta tutkijoille sekä aineiston hallintaa opettaville tahoille (Hazlett 2016; Saarti 2019, 11). Tutkimusaineistoja tallentavia ja niiden jatkokäytön mahdollistavia arkistoja on perustettu useita. Ihmistieteille merkittävimpiä kotimaisia palveluita ovat Tietoarkisto ja Fin-Clarín, mutta saatavilla on myös lukuisia muita ja kansainvälisiä palveluita. (Borg & Kuula 2007; Fält 2016).

Käytettävissä olevien resurssien ja palvelujen lisäksi ihmisten asenteet vaikuttavat halukkuuteen jakaa tutkimusaineistoja (Frederick 2016, 10). Avoimeen tieteeseen liittyy paljon oletuksia, ennakkoluuloja ja jopa yli-

varovaisuutta esimerkiksi henkilötietojen käsittelyyn liittyen (Järviluoma & Ruotsalainen 2018). Myös tutkijan persoonallisuuspiirteillä on havaittu olevan vaikutusta tutkimusaineiston jakamiseen (Abele-Brehm et al. 2019). Avoimen tieteen koulutusta ei ole koettu vielä riittävänä tai siihen osallistuminen ei ole ollut pakollista (Rice 2019, 266; Tenopir et al. 2020, 13). Laajasta markkinoinnista ja edistämisyrittämisistä huolimatta avoin tiede vakiinnut-taakin asemaansa vähitellen ja hitaasti tieteen tekemisen käytäntöihin.

Tutkimusaineistojen avaamisessa muiden käyttöön on havaittu tieteenala-kohtaisia eroja. Aineistojen jakamisella on pitempi perinne monilla luonnon-tieteellisillä aloilla ja yhteiskuntatieteissä, kun taas humanisteille käytäntö on vieraampi. Syitä tutkimusaineistojen vähäiseen avaamiseen on arveltu olevan useita. Joillakin humanistisilla tieteenaloilla tutkijan tuottamat aineistot ovat sellaisia, että muiden on vaikea hyödyntää niitä. On myös arveltu, että humanisteilta puuttuu välineitä ja osaamista linkittää arkistoissa olevaa aineistoa uudelleenkäytettäväksi, eikä aineistojen tallentamiseen muuten-kaan ole humanistisilla aloilla yhtä selkeitä velvoitteita kuin muilla tieteen-aloilla. Lisäksi eettiset kysymykset, kuten tietosuojat, saattavat estää aineiston tallentamisen. (Tenopir et al. 2011; Zhu 2019, 43.) Suomalaisia tutkijoita ja väitöskirjantekijöitä käsittelevässä tutkimuksessa kuitenkin osoitettiin, että halukkuudessa jakaa tutkimusaineistoa tieteenalakohtaisilla eroilla ei ollut tilastollista merkitystä (Enwald 2018, 35).

Suurin osa Kansalliskirjaston aineistojen käyttäjistä on ihmistieteilijöitä, erityisesti humanisteja, joten tarkastelemme tutkimusaihetamme luonnon-tieteiden ja ihmistieteiden välisen dikotomian kautta pohtien, vastaavatko aineiston tallentamista ja avoimuutta ohjaavat periaatteet riittävästi humanististen tieteiden tieteenfilosofisia periaatteita ja niiden metodologiaa. Kirjas-saan *Explanation and Understanding* Georg Henrik von Wright pohtii tieteen-historian kahta perinnettä: positivistista ja hermeneuttista tutkimussuuntaa. Positivistinen perinne on korostanut kausaalista päättelyä ja tieteellisen menetelmän yhdenmukaisuutta, hermeneuttinen perinne puolestaan teleologi-suutta ja intentionaalisuutta. Tässä kieli ja sen kautta ilmenevät toiminnan päämäärät ja tarkoitukset ovat tulkinnan kohteena. Jälkimmäiseen suun-taan on yleensä katsottu kuuluvaksi humanistiset tieteet, kun taas yhteiskunta-tieteet toimivat kahden tieteellisen kulttuurin välimaastossa. (von Wright 1971, 3–33.)

Humanistitutkijoiden tutkimuksen teon käytäntöjä on tarkasteltu mones-sakin tutkimuksessa, joissa on havaittu, että humanistilla oppialoilla tutki-musprosessi ei ole yhtä selkeästi määritelty kuin luonnontieteissä tai kysely-tutkimuksiin perustuvilla yhteiskuntatieteiden alueilla. Tutkimuksen lähteet ovat moninaisia ja niitä löydetään koko ajan uusia, samoin menetelmiä ja

teknologioita kehitetään tai lainataan muilta tieteenaloilta. Aineiston kerääminen, tallentaminen, sen muokkaaminen ja rakenteistaminen analyysiä varten sekä päätelmien tekeminen opitaan ennemminkin seuraamalla kollegoja kuin suoraan menetelmäoppaista. Tekemisen tapa eri tutkijoilla voi olla hyvinkin erilainen. Humanistinen tutkimus on iteratiivinen prosessi, jossa tutkija, tutkijayhteisö ja data tai tutkimusaineisto ovat vuorovaikutuksessa keskenään. (Borgman 2015, 162–164; Trace & Karadkar 2017, 491–493, 504; Lönnqvist 2003, 253–260.)

Data ymmärretään usein humanistisessa tutkimuksessa eri tavoin kuin luonnontieteissä, joissa sillä viitataan enimmäkseen koneluettavaan dataan. Humanistisilla aloilla datasta ei aina edes puhuta tai se käsitetään lähinnä alkuperäisaineistoiksi. Joskus jopa sanotaan, etteivät humanistit edes käytä dataa tai tutkimusaineistoja jo olemassa jossakin arkistoissa. Niistä kuitenkin koostetaan oma erillinen tutkimusaineistokokonaisuus, jonka säilyttäminen tutkimuksen teon jälkeen tulee usein pohdittavaksi. (Masson 2017, 25; Schöch 2013.)

Tämä artikkeli sai alkunsa Kansalliskirjaston tutkijakyselystä, joka toteutettiin keväällä 2020. Kyselyssä Kansalliskirjasto keräsi tietoa käyttökokemuksista sekä kehitysideoita sen omia digitaalisia aineistoja ja palveluita käyttäviltä tutkijoilta. Kysely toteutettiin osana *Digitaalinen avoin muisti* -hanketta, jonka tarkoituksena on ensisijaisesti Kansalliskirjaston tutkijapalveluiden kehittäminen. Kyselystä saatiin tietoa Kansalliskirjaston omien digitaalisten aineistojen säilyttämisestä tutkimuksen aikana ja tallentamisesta tutkimuksen jälkeen avoimen tieteen periaatteiden mukaisesti, vaikka näkökulma olikin ensisijaisesti Kansalliskirjaston omien digitaalisten aineistojen hyödyntäminen, ei avoimuuden kysymykset. Kyselyn analyysi perustuu tutkimusaineistojen avaamiseen liittyvien kysymysten tarkastelun aineistolähtöiseen perspektiiviin. Aineiston perusteella hahmotimme neljä tutkimuskysymystä:

1. Miten Kansalliskirjaston aineistosta kerättyä tutkimusaineistoa säilytetään tutkimuksen aikana?
2. Kuinka suuri osa Kansalliskirjaston aineiston käyttäjistä arkistoi aineistonsa data-arkistoon tai muuhun julkiseen palveluun tutkimuksen jälkeen?
3. Vaikuttaako tutkimuksessa käytetty tutkimusmenetelmä aineiston arkistointiin?
4. Mitkä aineistoon tai kirjaston palveluihin liittyvät ominaisuudet voivat muodostua esteeksi arkistoinnille?

Vaikka kyselyn tavoite oli käytännöllinen – tutkijapalvelujen kehittäminen – katsomme, että kysely avaa myös tutkimuksellisen näkökulman ihmis-tieteilijöiden mahdollisuuksiin avata aineistoaan. Vertaamme Kansalliskirjaston kyselyn vastauksia kansainväliseen tutkimukseen ja pohdimme aineiston luonteen ja käytettyjen menetelmien vaikutusta aineistojen tallentamiseen.

Kansalliskirjasto ja tutkimusaineistojen hallinta

Tieteellisten kirjastojen tehtävänä on tukea tutkimusta ja avointa tiedettä, koska kirjastot ovat yksi tärkeimmistä tutkimuksen kumppaneista (Räisänen 2016; Southall & Scutt 2017, 313). Perinteisesti kirjastot ovat avustaneet tutkijoita tutkimusaineistojen hallinnassa, avoimessa julkaisemisessa ja tutkimusjulkaisujen bibliometrisessä mittaamisessa. Aineistohallinnassa kirjaston tehtävä on yleensä neuvonnallinen ja ohjaava. Edellytyksenä ovat kirjastossa toimivien tietoasiantuntijoiden omat taidot sekä tietämys aineiston säilyttämisen periaatteista ja erilaisten aineistojen säilyttämiseen sopivista paikoista (Antell et al. 2014; Rice 2019, 264–265). Erityisesti digitaalisten aineistojen kohdalla kirjastoilta vaaditaan nopeita reagoiteja sekä toisaalta aikaa kestävien linjauksien ja toimintamallien kehittämistä tai haltuun ottamista (Heidorn 2011, 668). Tutkimusaineistojen lukuisat digitaaliset säilytyspaikat ovat johtaneet siihen, että tieteelliset kirjastot joutuvat pohtimaan omaa positiotaan tutkimuksen tukijana ja aineistohallintapalvelujen tarjoajina (Bardyn, Resnick & Camina 2012, 275; Corral, Kennan & Afzal 2013; Southall & Scutt 2017, 303). Amerikkalaisessa kontekstissa kirjastojen tarjoamien aineiston hallintapalvelujen määrä on kasvanut vuosien 2013 ja 2016 välillä (Yu 2017, 784, 791). Kuitenkin vain harva kirjasto on ottanut tehtäväkseen toimia tutkimusaineistojen säilytyspaikkana (Shen 2016, 500).

Aineistohallintaan liittyvien ohjeistusten antaminen tutkijoille on haastavaa, koska paitsi tieteenalojen käytännöt myös tutkijoiden yksilölliset tavat toteuttaa tutkimusta vaikuttavat aineiston tuottamiseen ja sen hallintaan. (Lönngqvist 2003, 41–44, 253–260). Toisaalta myös tutkijalla itsellään on vastuuta ohjeistuksien noudattajana. Kirjastojen on ymmärrettävä, miten tutkijat toimivat sekä tekijät, jotka vaikuttavat aineiston hallintaan ja tutkimusaineiston hallintaa koskevan ohjeistuksen kehittymiseen. Tekijöitä ovat aineistohallinnan kansainväliset ja kansalliset strategiat, erilaiset tieteenalojen aineistohallintapolitiikat, paikalliset ohjeet, teknologiat, paikallisesti aineiston hallintaan tarjottavat palvelut ja käytännön prosessit. Niillä pyritään turvaamaan aineiston säilyttämiseen tarvittavat tilat, turvallisuus ja laatu. Yhdessä ne muodostavat aineiston hallintaan tarvittavan infrastruktuurin,

jonka toimivuutta arvioidaan tieteellisessä keskustelussa sekä varsinaisissa tieteen käytännöissä (Pinfield, Cox & Smith 2014, 4, 9, 25–26; Steinerová 2019, 12–13; Rice 2019, 266.)

Kansalliskirjaston palvelutehtävä on kansallinen, eikä sillä ole yliopisto-kirjastojen tapaan velvoitetta tukea emo-organisaationsa tutkijoiden aineistonhallintaa. Sen sijaan se tarjoaa runsaasti aineistoja, jotka soveltuvat tutkimuksen tarpeisiin. Kansalliskirjaston tarjoama aineisto on suurimmalta osin julkaistua kulttuuriperintöä: painettua, digitaalisena syntynyttä ja digitoitua. Kirjastolla on myös arkistokokoelmia, jotka ovat edelleen digitointamatta. Avoimen tieteen edistäminen ja FAIR-periaatteiden noudattaminen ovat keskeisessä osassa Kansalliskirjaston vuosien 2021–2030 strategiassa (Kansalliskirjaston strategia, 2020). FAIR-periaatteiden neljä kriteeriä ovat löydettävyys (findable), saavutettavuus (accessible), yhteentoimivuus (interoperable) ja uudelleenkäytettävyys (reusable) (EU-neuvoston linjaus 2016).

Kansalliskirjaston mahdollisuuksia edistää omien aineistojensa tallentamista ja jakamista tutkimustarkoituksissa rajoittaa vahvasti lainsäädäntö. Kirjaston kansalliskokoelman aineistot on saatu ja talletettu kulttuuriaineistolain (1433/2007) perusteella, ja niiden jakamista rajoittaa tekijänoikeuslaki (404/1961). Vapaakappaleina luovutetut tai haravoidut verkkoaineistot sekä digitoidut aineistot ovat käytettävissä vain tekijänoikeuslain puitteissa vapaakappalekirjastoissa ja ne voidaan avata vasta, kun kaikkien tekijöiden kuolemasta on kulunut 70 vuotta. Toinen mahdollisuus avaamiseen on tekijänoikeusjärjestöjen kanssa solmittu sopimus aineistojen käytöstä. Nämä käyttönsaattamissopimukset ovat maksullisia, joten kirjasto joutuu harkitsemaan, mitä aineistoja voidaan avata. Tällä hetkellä sopimuksilla on avattu sanoma- ja aikakauslehtiaineisto vuoteen 1939 kaikille käyttäjille, sanoma- ja aikakauslehtiaineisto vuoteen 2018 tutkijoille Tutkain-hankkeessa sekä tuhat tietokirjaa yleisölle ja tuhat kaunokirjaa tutkijoille 2000 kirjaa -hankkeessa. Kaikki nämä sopimukset ovat määräaikaista, eikä niiden mahdollistama käyttöoikeus ole yhtä laaja kuin esimerkiksi Creative Commons, CC-lisenssillä avatussa aineistossa (<https://creativecommons.fi/lisenssit/>). Datan tallennus ja jatkokäyttö tutkimustarkoituksissa on kuitenkin tuoreimmissa sopimuksissa pyritty mahdollistamaan. (Pääkkönen & Lilja, 2019, 40–41; Hakkarainen 2019; Lilja & Pääkkönen 2019, 6–7; Kansalliskirjasto avasi 1930-luvun lehdet asiakaskäyttöön). Kansalliskirjasto voi jakaa CC-lisenssillä esimerkiksi itse tuottamaansa kuvailutietoa ja ontologioita, kun taas vanhaa, tekijänoikeuksista vapautunutta aineistoa Kansalliskirjasto voi jakaa Public domain -lisenssillä. Avoimien rajapintojen avulla tutkijan on mahdollista ladata kansalliskirjaston bibliografia Fennica ja muut kirjaston ylläpitämät tietokannat, sekä sanastot ja ontologiat, muokata niitä ja jakaa edelleen. (Open data service. 5.7.2020)

Kansalliskirjaston digitaaliset aineistot ovat lähes aina löydettävissä metatietojen ansiosta kunkin digitaalisen aineiston omasta hakupalvelusta (Hyvönen 2015). Aineistojen saatavuutta rajoittaa se, etteivät kaikki digitaaliset aineistot ole haettavissa hakupalvelu Finnasta. Lisäksi ne eivät välttämättä löydy yleisistä internetin hakukoneista, kuten Googlesta. Ari Häyrinen (2012, 42) huomauttaa, ettei tällaisissa tapauksissa voi puhua oikeasti löydettävästä tai saavutettavissa olevasta aineistosta, koska tavallinen käyttäjä ei välttämättä osaa käyttää aineiston omaa hakuliittymää ja hyödyntää olemassa olevaa metatietoa. Kansalliskirjasto pyrkii edistämään tutkijoiden kanssa tehtävällä yhteistyöllä digitaalisten aineistojensa laatua ja meneillään olevassa kehityshankkeessa se hahmottelee erilaisia prosesseja, joiden avulla myös tutkijoiden rikastamia aineistoja voisi integroida sen omiin järjestelmiin aiempaa paremmin.

Kansalliskirjaston tutkijakyselyn toteutus

Kansalliskirjasto kysyi tutkijakäyttäjiltään keväällä 2020 näiden kokemuksista digitaalisten aineistojen ja palvelujen käytöstä. Kysely toteutettiin osana Digitaalinen avoin muisti -hanketta, jonka tarkoituksena on ensisijaisesti Kansalliskirjaston tutkijapalveluiden kehittäminen. Kysely koostui 20 kysymyksestä ja siihen vastasi yhteensä 130 henkilöä. Se tarjottiin vastaajille suomeksi, ruotsiksi ja englanniksi. Tässä artikkelissa keskitytään analysoimaan ensisijaisesti kysymyksiä 10, 13–15 (Liite 1), joista kolme ensimmäistä on tutkimusmenetelmiin ja tutkimusaineiston säilyttämiseen liittyviä monivalintakysymyksiä. Kysymys 15 on avokysymys, johon jätettiin 73 vastausta. Se on analysoitu sisällönanalyttisesti. Aineiston arkistointi ja avaaminen tapahtuu soveltuvin osin DAM-hankkeen päättymisen jälkeen.

Kyselyn tavoitteena oli tietää, miten Kansalliskirjaston nykyiset palvelut toimivat tutkimuksen eri vaiheissa. Kysely rakennettiin noudattelemaan tutkimuksen elinkaarta tutkimuksen suunnittelusta tulosten jakamiseen. Tutkimuksen elinkaariajattelu auttaa kiinnittämään huomiota tutkijan omaan panostukseen aineiston hallintaan ja suunnitteluun (Perrier et al. 2017, 10). Sopivaksi lähestymistavaksi tähän pidettiin tutkimusaineiston hallintasuunnitelman (data management plan, DMP) mukaista tarkastelua aineistojen keruun, analysoinnin ja säilyttämisen osalta. Aineistojen hallintasuunnitelmassa huomioidaan usein aineiston keruun menetelmät, aineiston analysointi, aineiston säilyttäminen, tutkimuksen dokumentaatio, metadata, aineiston arkistointi, löydettävyys sekä aineiston jaettavuus (Yu 2017, 784).

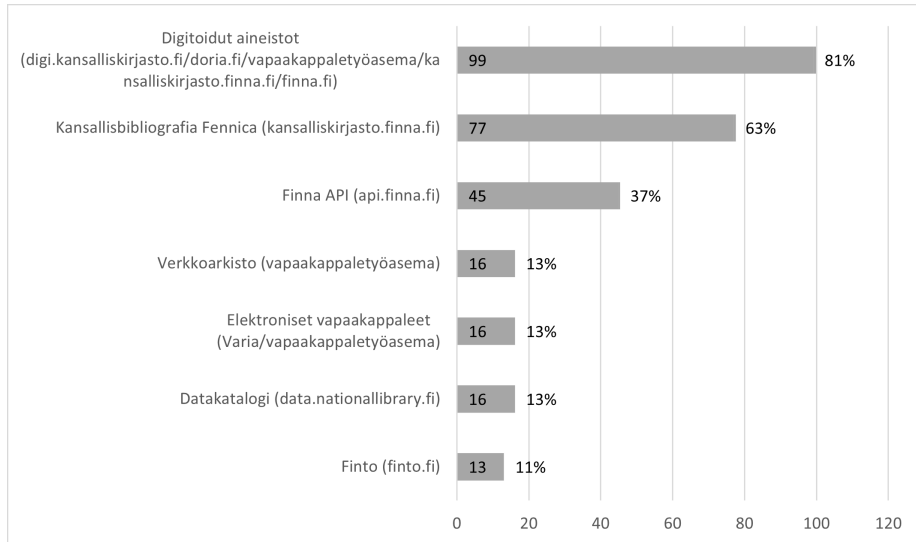
Kyselyssä DMP:n mukainen tulokulma näkyi erityisesti siinä, että vastaajilta kysyttiin, missä he säilyttävät aineistoaan tutkimuksen teon ajan ja tutkimuksen teon jälkeen. Kysymyksiä pohjustettiin seuraavalla tekstillä: ”Tutkimusaineistojen säilyttäminen on tiedekentällä iso puheenaihe, ja moni rahoittaja vaatii tekemään aineiston säilyttämissuunnitelman. Myös erilaisten tutkimustuotosten säilyttäminen ja eteenpäin jakaminen voivat olla joissain tutkimuksissa olennaisia pohdinnan paikkoja.” Kyselyyn vastaajat saivat kysymyksessä 15 kuvailla omin sanoin tilannetta ja ehtoja, joiden mukaan he olisivat valmiita jakamaan tutkimustuoksensa. Tutkimustuoksilla tarkoitettiin kysymyksessä laajasti tutkimuksessa kehitettyjä menetelmiä, ohjelmia, aineistoja, oppaita ja työkaluja. Kysymys oli asettelultaan johdatteleva, koska sen tavoitteena oli myös kannustaa niitä tutkijoita, jotka eivät olleet aiemmin tehneet aineistohallintasuunnitelmaa, pohtimaan avoimen tieteen käytäntöjä.

Kyselyn luotettavuus ja tuloksien yleistettävyys

Kyselyn tavoittaman vastausprosentin tarkka määrittely on haasteellista, sillä Kansalliskirjaston aineistoja käyttävät tutkijat eivät muodosta asiakaskunnassa mitään selkeästi eriteltyä ryhmää, jonka lukumäärä olisi tunnistettavissa. Esimerkiksi kirjastokortin haltijat eivät sovi viiteryhmäksi, sillä heidän joukossaan on tutkijoiden lisäksi muita ryhmiä. Toisaalta verkkoaineistojen käyttämiseen ei tarvita kirjastokorttia eikä kaikissa palveluissa muutakaan tunnistautumisvälinettä. Vastausprosentin edustavuuden hahmottamiseksi on olennaista reflektoida vastaajamäärää vertailemalla muista lähteistä saataviin lukuihin. Yleisesti tutkijakyselyn vastaajamäärää (130) voi pitää Kansalliskirjaston omassa kontekstissa tyypillisenä, koska se noudattelee vuosien 2018 ja 2019 aikana digitaalisten sanomalehtiaineistoon kohdistuvien kyselyiden vastanneiden määrää. Vuoden 2018 digi.kansalliskirjasto.fi -palvelun käyttäjäkyselyssä vastaajia oli 140 (Pääkkönen & Lilja 2018) ja vuonna 2019 Digin käyttäjäkyselyyn vastanneiden määrä oli noussut 172:een (Pääkkönen 2019). Nämä kyselyt oli kohdennettu kaikille digi.kansalliskirjasto.fi -palvelun käyttäjille.

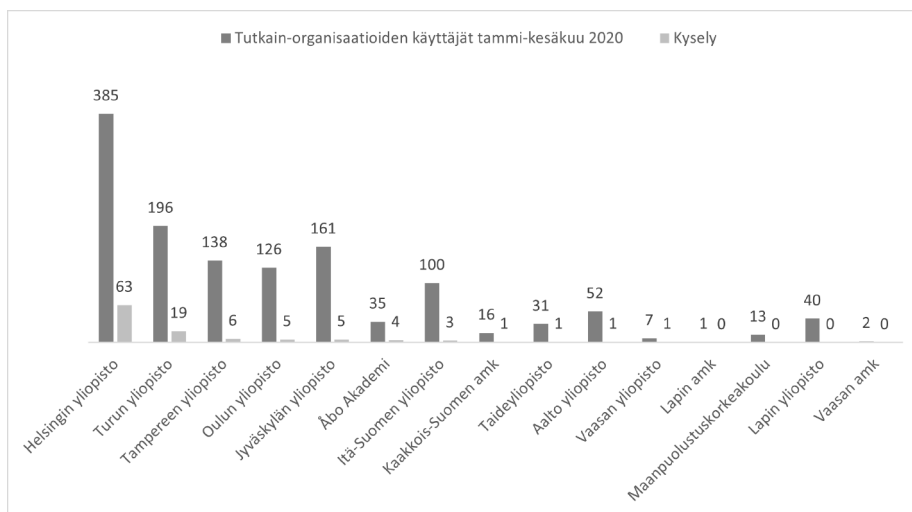
Kyselyn vastausprosentti olisi 10 prosenttia, jos vastanneiden määrää verrataan Tutkain-hankkeessa digi.kansalliskirjasto.fi -palvelua vuoden 2020 tammi- ja kesäkuun välillä käyttäneisiin 1303 tutkijaan. Tutkain-hankkeella tarkoitetaan digi.kansalliskirjasto.fi -palvelussa tutkimuskäyttöön avattua sanomalehtiaineistoa, joka on käytössä hankkeeseen liittyneille 15 korkeakoululle. Tutkijat tunnistautuvat palveluun ja sitoutuvat aineistojen käyttöön tutkimustarkoituksissa.

Vertailu Tutkain-hankkeen käyttäjiin ei kuitenkaan tuota täysin tarkkaa vastausprosenttia, sillä osa tutkijakyselyyn vastanneista oli käyttänyt muita kuin Tutkain-hankkeessa avattuja digitoituja lehtiaineistoja, kuten Kuva 1 osoittaa.



Kuva 1: Vastaajien käyttämät aineistolajit, vastaajia 122, valintoja 282, prosentit suhteessa kysymykseen vastanneisiin, kun vastaajat saivat valita useita vaihtoehtoja.

Lisäksi osa kyselyyn vastaajista edusti muita organisaatioita kuin Tutkain-hankkeeseen liittyneitä korkeakouluja. Kyselyyn vastanneet edustivat yhtätoista Tutkain-hankkeen 15 korkeakoulusta. Kyselyn vastanneista 19 tutkijaa edusti muuta kuin Tutkain-organisaatioita. Näihin muihin organisaatioihin kuului esimerkiksi Luonnonvarakeskus ja Sodan ja rauhan keskus Muisti. Asiaa havainnollistetaan Kuvassa 2.



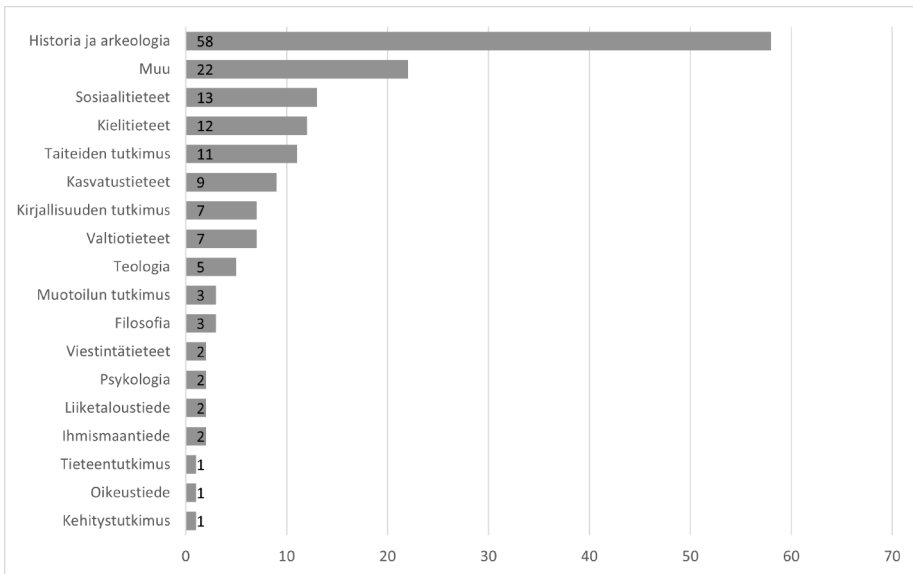
Kuva 2: Tutkain-käyttäjien ja kyselyn vastanneiden affiliaatio.

Eroavaisuudet Tutkain-hankkeen käyttäjissä ja kyselyyn vastanneiden suhteessa ovat verrattain pieniä. Tämän perusteella arvoimme, että vastausprosentti Kansalliskirjaston aineistoa käyttävien tutkijoiden osalta liikkuu jossain 5–10 prosentin välillä.

Akateemisten tutkijoiden osuus vastanneista

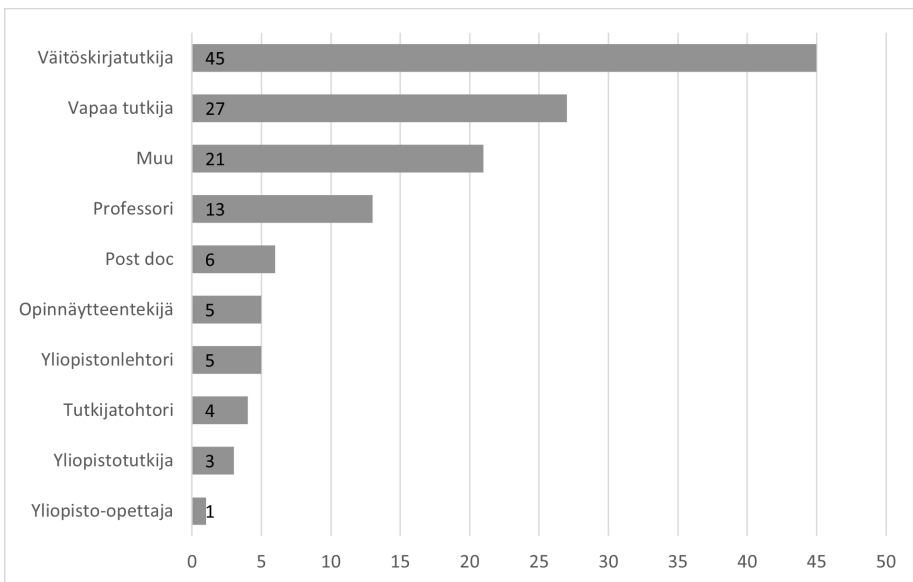
Seuraavaksi pohdimme, kuinka suuri osuus vastanneita edusti akateemisten tutkijoiden ryhmää, jolle tutkimusaineistojen tallentaminen ja arkistointi ensisijaisesti kuuluu. Viestimme kyselystä yliopistojen suuntaan ja ennestään Kansalliskirjastolla oleville tutkijakontakteille. Kysely suunnattiin erityisesti humanististen ja yhteiskuntatieteellisten alojen tutkijoille, koska Kansalliskirjaston aineistot palvelevat perinteisesti näitä aloja.

Vastausvaihtoehtoina tieteenalalle käytettiin Suomen Akatemian tieteenalaluokitusta, joka sopi melko hyvin vastaajille. Vain 22 vastaajaa kertoi edustavansa jotakin muuta kuin valmiiksi annettua vaihtoehtoja (Kuva 3: Tieteenalat). Jo kyselyn testausvaiheessa huomattiin, että oman tieteenalan sijoittaminen johonkin kategoriaan tuotti haasteita. Sen vuoksi kyselyssä annettiin mahdollisuus valita useita vaihtoehtoja sekä mahdollisuus tarkentaa omaa tieteenalaa avovastauksella. Avovastauksissa moni kuitenkin tarkensi jonkin historiaan kuuluvan erityisalan. Varsinaisten humanististen ja yhteiskuntatieteellisten alojen ulkopuolelle jäi alle 10 vastausta.



Kuva 3: Tieteenala, Suomen Akatemian pääluokat, vastaajia 130, valintoja 161, kun vastaajat saivat valita useita vaihtoehtoja.

Kuvassa 4 nähdään, että 21 vastaajaa ilmoitti edustavansa jotakin muuta kuin vastausvaihtoehtoissa annettua tutkijapositiona ja 27 kertoi olevansa vapaa tutkija.



Kuva 4: Tutkijan uran vaihe.

Avovastauksissa muu-kategorian valinneista vain pieni osa vastaajista kertoi edustavansa yliopiston ulkopuolista tutkimusorganisaatiota, yritystä tai olevan affilioitumaton tutkija. Vapaat tutkijat sen sijaan affilioituvat enimmäkseen johonkin yliopistoon, mikä avovastausten perusteella tarkoitti pääosin sitä yliopistoa, josta he olivat valmistuneet. Kahdeksan vastaajaa vapaista tutkijoista edusti jotakin muuta kuin jotakin yliopistoa tai he eivät tarkentaneet avovastaukseen mitään organisaatiota. Avovastauksia tarkasteltaessa sukututkijat ja harrastelijat jäivät selvästi vähemmistöön, mutta vastaajien joukkoon mahtui muutamia eläköityneitä tutkijoita ja yksittäisiä kylähistorian kerääjiä ja sukututkijoita. Nämä kuitenkin muodostavat selvästi alle kymmenen henkilön ryhmän, jos otetaan huomioon myös vastaajat, jotka eivät tarkentaneet vastausvaihtoehtoa muu tai vapaa tutkija siihen ollessa mahdollisuus. Vaikka päädyimme lopulta melko väljään määritelmään vastaajan tutkijan positioista ja luotimme osittain vastaajan subjektiiviseen kokemukseen omasta tutkijudesta, voidaan arvioida, että kysely tavoitti tutkijat melko hyvin ja suurin osa vastauksista tuli nimenomaan akateemista tutkijuutta edustavilta henkilöiltä.

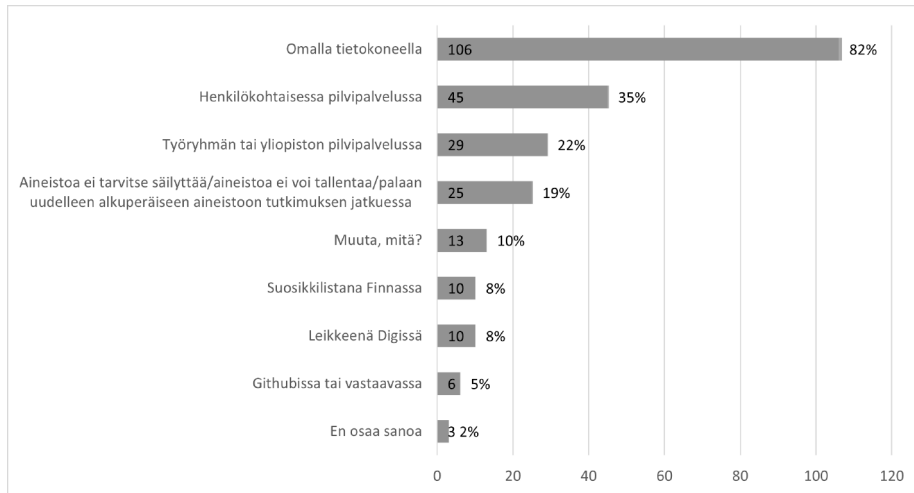
Tulokset ja vertailu

Aineiston säilyttäminen tutkimuksen ajan

Tutkimusta tehdessä aineisto muodostetaan jonkin tutkimusasetelman mukaisesti. Usein tutkimusaineisto kerätään konkreettisesti johonkin paikkaan, jotta se muodostaa yhden loogisen ja riittävän kattavan kokonaisuuden ollakseen tutkimukseen soveltuva tutkimusaineisto. Tutkimuksen teon aikana aineistoa käsitellään ja analysoidaan erilaisilla menetelmillä, jotta siitä saadaan aikaiseksi tutkimustuloksia ja tutkimustuotoksia myös laajemmin. Tutkimuksen analyysin mahdollistamiseksi se usein siirretään johonkin sen alkuperäisestä paikasta, jotta sen työstäminen jollakin menetelmällä onnistuu. Usein tässä vaiheessa aineisto on enimmäkseen vain tutkijan tai tutkimusryhmän käytössä, mutta tämäkään vaihe ei sulje pois sitä, etteikö aineisto voisi olla avoimesti kaikkien muidenkin tutkijoiden käytössä. Suomen Akatemian aineistohallintasuunnitelman ohjeistus kehottaa säilyttämään aineiston sen luonteen edellyttämällä tavalla oman organisaation ylläpitämässä palvelussa tietoturvallisuus huomioiden ja välttämään ulkoisia kovalevyjä pääasiallisena säilytysratkaisuna. (Suomen Akatemia: Aineistohallintasuunnitelma)

Kyselyssämme tiedusteltiin, missä tutkijat säilyttävät tutkimusaineistoaan tutkimuksen aikana eli tutkimuksen ollessa vielä kesken. Kuvasta 5 käy ilmi, että suurin osa kyselyyn vastaajista säilytti tutkimusaineistonsa omalla

tietokoneellaan tai henkilökohtaisessa pilvipalvelussa. Ulkoinen kovalevy mainittiin avovastauksissa seitsemän kertaa. Toisille ulkoinen kovalevy toimi todennäköisesti varmuuskopioiden paikkana ja varsinainen työstettävä tutkimusaineisto saattoi sijaita samaan aikaan myös muualla, mutta vain harvoin muiden kuin yksittäisen tutkijan tai tutkimusryhmän käytössä.



Kuva 5: Aineiston säilyttäminen tutkimuksen teon ajan, 129 vastaajaa 247 valintaa, prosentit suhteessa kysymykseen vastanneisiin, kun vastaajat saivat valita useita vaihtoehtoja.

Yli 90 prosentilla vastaajista tutkimusaineisto oli koottu useista eri lähteistä – paitsi digitaalisista myös Kansalliskirjaston painetuista kokoelmista, mikrofilmeiltä tai arkistoista sekä muiden arkistojen ja museoiden kokoelmista, myös ulkomailta. Sanallisissa vastauksissa mainittiin myös audiovisuaalinen aineisto, sosiaalinen media ja Tietoarkistosta noudettu kyselyaineisto. Aineiston pitämiseen itsellä tutkimuksen ajan saattaa siis vaikuttaa myös se, että sitä kootaan eri lähteistä ja sitä rakennetaan tutkimuksen eri vaiheissa – toisin kuin vaikkapa survey-tutkimusta, jossa aineisto valmistuu yhdellä kertaa. Lisäksi erilaisista digitaalisista lähteistä koostettu aineisto ei välttämättä ole kovin yhteentoimivaa. Eri lähteistä koostetun aineiston analyysi saattaa vaatia erilaisia välineitä, kokoamista ja työstämistä ennen kuin yhdenmukainen analyysi on mahdollista. Tällöin tutkimuksen aikana aineiston käsittely on kesken lähes tutkimuksen loppuun asti, eikä se ole riittävän itsenäinen ja toimiva kokonaisuus jaettavaksi ennen kuin se vastaa riittävällä tasolla tutkimusasetelmaa.

19 prosenttia vastaajista vastasi vaihtoehdon, ettei aineistoa tarvitse tai voi säilyttää tutkimuksen aikana. Valinta tarkentui vain yhdessä avovastauksessa,

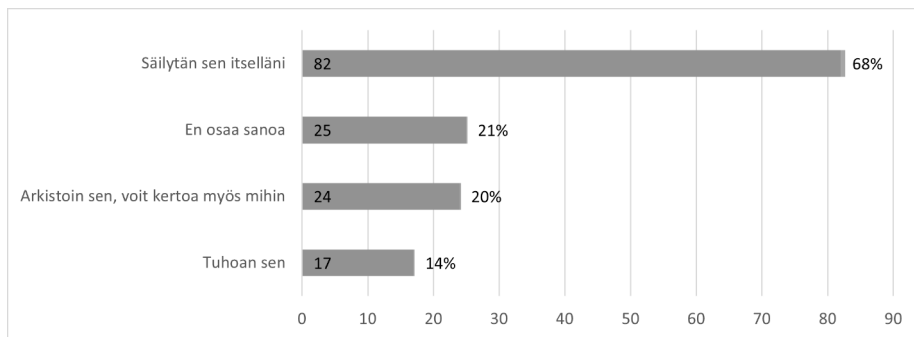
jossa vastaaja totesi, että siteeraa aineistoa suoraan tekstiinsä eikä tallenustarvetta synny.

On myös syytä huomioida, että kaikki Kansalliskirjaston palvelut eivät aina mahdollista aineiston tallentamista kovin helposti. Esimerkiksi vapaakappaletyöasemilla aineisto on suojattu kopioinnilta, jolloin ainoa tapa ottaa siitä tallenne on tulostaa tai kuvata se omalle laitteelle. Lisenssisopimuksilla verkon kautta käytettävissä aineistoissa yhteentoimivuus muiden aineistojen kanssa on helpommin tavoitettavissa, koska osa niistä toimii usein käytettyjen taulukointiohjelmien kanssa. Nämäkään eivät aina tarjoa aineistoa tutkijalle ladattavaksi suoraan tutkijan tarpeisiin soveltuvassa tiedostomuodossa.

Aineiston säilyttäminen tutkimuksen jälkeen

Tutkimusprosessin alussa, sen aikana tai viimeistään sen jälkeen on ratkaistava, mitä aineistolle tapahtuu tutkimuksen jälkeen (Järviuoma & Ruotsalainen 2018, 145). Aineiston säilyttämisessä on huomioitava ainakin sen tietoturva, henkilötietojen käsittely ja aineiston säilytysaika. Esimerkiksi Suomen Akatemian aineistohallintasuunnitelma edellyttää, että tutkija määrittelee jo ennen tutkimusta, mitä osia aineistosta säilytetään ja missä, mitä aineistoja voidaan avata ja mitä hävitetään. (Suomen Akatemia: Aineistohallintasuunnitelma)

Kuva 6 osoittaa, että Kansalliskirjaston kyselyyn vastanneista suurin osa säilytti tutkimusaineistonsa itsellään tutkimuksen valmistuttua. Tämä 68 prosenttia vastaajista on vain hieman suurempi osuus kuin tuoreessa Carol Tenopirin ja hänen tutkijakumppaneidensa tutkimuksessa, jossa tarkasteltiin pohjoisamerikkalaisten ja eurooppalaisten tutkijoiden tutkimusaineiston tallennuskäytäntöjä useiden tieteenalojen kesken. Noin 20 prosenttia vastaajista edusti ihmistieteellisempää otetta (*social sciences*) suurimman osan ollessa luonnontieteilijöitä. Tutkimuksessa erilaisiin vain omaan käyttöön tarkoitettuihin säilytyspaikoihin (oma tietokone, pilvipalvelu ja USB-tikku) prosentit vaihtelivat 30–60 välillä ja erilaisiin arkistoihin tallentaminen oli 20 prosentin molemmin puolin. (Tenopir et al. 2020.)



Kuva 6: Aineiston säilyttäminen tutkimuksen jälkeen, vastaajia 121, valintoja 148, prosentit suhteessa kysymykseen vastanneisiin, kun vastaajat saivat valita useita vaihtoehtoja.

Tutkimusaineistonsa arkistoi viidennes Kansalliskirjaston kyselyn vastaajista. Tämäkin vastaa melko hyvin kansainvälistä tilannetta. Tuoreessa Ison Britannian tutkijoita koskevassa tutkimuksessa 21 prosenttia tutkijoista arkistoi aineistonsa, joskin humanisteilla osuus oli hieman alhaisempi 19 prosenttia (Zhu 2019, 45, 48). Arkistointipaikoiksi Kansalliskirjaston kyselyn vastaajat ilmoittivat Kielipankin, Tietoarkiston ja yliopistojen arkistot. Osa vastaajista tarkensi kuitenkin, että arkistointipaikka valittiin tapauskohtaisesti aineiston laadun ja arkistointipaikan omien kriteerien mukaan. Tutkimusaineiston arkistointipaikan valinta ei kansainvälisestikään tarkastellen ole itsestään selvää ja ongelmatonta. Tenopirin ryhmän tutkimus osoittaa, että tutkijat eivät olleet aina kovin tyytyväisiä kaikkiin arkistointipaikkoihin ja niiden ohjeistukseen, eivätkä siksi kokeneet niitä sopivaksi omaksi aineistolleen (Tenopir et al. 2015, 9).

Kansalliskirjaston kyselyaineistossa 21 prosenttia kaikista vastaajista ei tiennyt, mitä aineistolle tapahtuu tutkimuksen jälkeen. Tämä saattoi johtua vastaajien uravaiheesta, jolloin aineistohallintasuunnitelma ei ollut tullut vielä tutuksi – olihan suurin osa kyselyyn vastanneista vielä väitöskirjatutkijoita. Esimerkiksi Heidi Enwaldin (2018) tutkimuksessa erityisesti uransa alkuvaiheessa olevat suomalaiset tutkijat olivat epävarmoja jakamaan tutkimusaineistonsa. Toisaalta asenteet avointa tiedettä kohtaan voivat olla tutkijan uran alussa positiivisemmat, mutta käytännöt juurtuvat vasta kokemuksen myötä (Abele-Brehm et al. 2019; Tenopir 2011, 11). Lisäksi syynä voivat olla ihmistieteiden ominaispiirteet, kuten kyselyn sanallisissa avovastauksissa kuvatut tilanteet, joissa aineistoa siteerataan suoraan tekstiin tai aineistosta kootaan teksti- tai kuvailutietokanta. Osa tutkijoista myös tuottaa Kansalliskirjaston metadatan pohjalta rikastettua kuvailutietoa, jonka mielellään

jakaisi muille tutkijoille, mutta jolle on vaikea löytää luontevaa säilytyspaikkaa. Teksti- ja kuvatietokantoja on tuotettu Kansalliskirjaston aineistoista esimerkiksi COMHIS-hankkeessa (Rantala et al. 2019) ja Translocalis-hankkeessa (Kokko 2019).

14 prosenttia kyselyyn vastanneista valitsi aineiston säilyttämistavakseen aineistonsa tuhoamisen. Aineiston tuhoajissa (n=17) ei ole havaittavissa kyselyn vastaajien keskuudessa tieteenalakohtaisia eroja, koska historia ja arkeologia -kategorian ulkopuolella eri tieteenalojen edustus on lähinnä yksittäisiä vastaajia. On kuitenkin huomattava, että kyselyn lähtökohdat ohjaavat ennalta määrättyjen alojen edustajia vastamaan siihen, eikä yleistettäviä johtopäätöksiä tieteenaloista voi tämän perusteella tehdä. Lisäksi aineiston tuhoajia löytyi niin yliopistoista kuin muista edustetuista organisaatioista, joten yliopistojen ulkopuolinen tutkijuus ei riitä selittämään kyselyn vastauksissa tutkimusaineiston loppusäilytyspaikan puutetta. Tarkasteltaessa yliopiston ulkopuolisten vastaajien edustusta (n=21, 16,1 % vastaajista) yli puolet säilyttää sen itsellään, neljännes tuhoaa, kaksi arkistoi ja yksi ei osaa sanoa. Avovastauksien mukaan kuitenkin moni heistä kirjoittaa olevansa valmis jakamaan aineistonsa tai muut tutkimustuotokset ”heti” (Vastaja 105) tai ”kaikissa tilanteissa” (Vastaja 2).

Kyselyn avovastauksien perusteella voidaan päätellä, että osa tutkijoista ei kokenut omalla aineistollaan olevan hyötyä muille tutkijoille ja päätyi siksi omaan ratkaisuunsa aineiston tuhoamisesta tai sen säilyttämisestä itsellään. Oma aineisto saattoi olla muille tutkijoille vaikeasti ymmärrettävissä tai omaa tutkimusaineistoa pidettiin vain omaan tutkimukseen kertaluontoisesti sopivana. Lisäksi avovastauksissa todettiin, että tutkimusaineiston viimeistely muiden käyttäjien kannalta ymmärrettäväksi kokonaisuudeksi vie liikaa aikaa ja vaatii lisää resursseja.

”Todennäköisesti näille ei muilla erityistä käyttöarvoa yleensä ole.” (Vastaja 85)

”Tutkimuksessani ei synny omaa erillistä aineistokokonaisuutta, joka voisi olla muille tutkijoille avuksi.” (Vastaja 108)

”Jos jostain saisi resursseja aineiston viimeistelyyn. Minun tutkimukseni viimeistelty materiaali ei ole mikään tilasto tai tietokanta, vaan sellaisen raakaver[s]ion perusteella kirjoitettu teksti. Se tietokanta on usein oman tutkimukseni päätyttyä aika fragmentaarinen ja kaikkea muuta

kuin muidenkin käytettäväksi soveltuva – sen toimittaminen, muotoileminen ja puhdistaminen jakamiskuntoon vaatisi resursseja, joita minulla ei ole (aikaa tai rahaa ostaa jonkun muun aikaa).” (Vastaja 31)

Kansalliskirjaston aineistojen käyttäjien vastaukset mukailevat Ciaran B. Tracen ja Unmil P. Karadkarin kuvaamaa arkistotutkijan työtappaa, jonka tuloksena syntyvä data koostuu muistiinpanoista, aineistosta otetuista paperikopioista, digitaalisista kuvista ja erilaisista linkeistä ja luetteloista, jotka usein ovat ymmärrettäviä lähinnä vain tutkijalle itselleen. (Trace & Karadkar 2017, 499–502.)

Kansainvälisessä datan tallentamista koskevassa tutkimuksessa on huomioitu aineiston jakamiseen tarvittavat resurssit suhteessa ajateltuihin hyötyihin, joita tutkimusaineiston jakaminen muille tarjoaisi. Humanistille hänen kokoamallaan datalla on suuri arvo paitsi omissa tulevilla tutkimuksissa myös työnhaussa tai uusien tutkimuskumppanien etsinnässä. Siksi aineistoja ei haluta jakaa, ei ainakaan ennen tutkimuksen julkaisua eikä välttämättä julkaisun jälkeenkään. Lisäksi tutkimuksissa vedottiin tutkijan etiikkaan ja laillisuuteen, jotka asettavat aineiston jakamiselle ehtoja. (Borgman 2015, 211–212; Tenopir et al. 2020; Abele-Brehm et al. 2019.) Joskus aineiston tuhoaminen voi olla perusteltua, jos esimerkiksi tutkittavien kanssa sovitaan näin. Tuhoaminen voi olla myös osittaista ja koskea esimerkiksi henkilötietoa. On myös mahdollista, että tutkijalla ei ole mahdollisuuksia säilyttää aineistoaan siihen tarvittavan tilan vuoksi. Tutkimusaineistoa saatetaan kuitenkin tarvita tutkimuksen verifointia varten, eikä tuhoaminen ole suositeltavaa kovin pian tutkimuksen valmistumisen jälkeen.

Perrier tutkijakumppaneineen (2017) huomauttaa, että tutkijoiden ja tutkimuksen pitäisi kiinnittää enemmän huomiota tutkimuksen aineistonhallinnan alkuvaiheeseen. Kyselymme avovastauksissa tutkijat toivatkin esiin, että he tarvitsevat lisää tietoa aineiston hallinnasta ja avoimesta tieteestä, vaikka he olivat jo pitkällä tutkimuksessaan. Muutama heistä kertoi olevansa kyselyyn vastaamishetkellä avoimen tieteen kurssilla, ja he toivoivat siellä käsiteltävän aineiston säilyttämistä ja tutkimustuotosten jakamista. Avovastauksista näkyi, että aineiston hallinnan kysymykset olisivat heidän urallaan vasta edessä, ja muutamat vastaajat kertoivat, etteivät olleet ajatelleet vielä koko asiaa. Myös hollantilaisia maisteriopiskelijoita tutkineiden Daen Smit-sin ja Marta Teperekin (2020) mukaan opiskelijat olivat usein hämmentyneitä aineiston hallinnan kysymysten äärellä ja moni tutkimukseen osallistuneista oli kertonut, ettei ollut saanut minkäänlaista kokemusta aineistonhallinnasta. Aineistonhallinta pitäisi kuitenkin olla selvää jo jatko-opintoihin siirtyvillä väitöskirjan tekijöillä ennen kuin he alkavat aineistonkeruunsa. Se tarkoittaa,

että yliopistojen ja tieteellisten kirjastojen on tehtävä työtä nykyistä enemmän jo perustutkintovaiheessa aineistojen hallinnan opetuksessa.

Tutkimuksessa käytetyn tutkimusmenetelmän ja arkistoinnin välinen yhteys

Tässä luvussa tarkastelemme, onko aineiston itsellä säilyttämisessä, tallentamisessa, arkistoisemisessa tai tuhoamisessa eroavaisuuksia suhteessa tutkimuksessa käytettyyn menetelmään. Kyselyssä tiedustelimme aineiston käyttötapaa. Yleisimmin valittu vaihtoehto *Selaillut ja lueskellut* ei tarkoita varsinaista tutkimusmenetelmää vaan se voi liittyä uusien ideoiden haeskeleluun, taustatietojen etsintään tai tutkimusaiheen kontekstin hahmottamiseen. Vaihtoehdolla *Lähiluku ja tulkinta* tarkoitetaan perinteiseen ihmistieteelliseen tutkimukseen liittyvää hermeneuttista menetelmää. *Tiedonlouhinnalla* tarkoitetaan laskennallisia tai muihin koneellisiin analyyseihin perustuvia menetelmiä, joilla suurista tietomassoista voidaan etsiä tutkimuksen kannalta kiinnostavia rakenteita ja säännönmukaisuuksia. (Tolonen & Lahti 2018, 249–250)

Taulukko 1 tuotettiin ristiintaulukoimalla kysymykset *10 Mitä olet tehnyt tai aiot tehdä Kansalliskirjaston aineistoille tutkimuksessasi?* ja *14 Mitä Kansalliskirjaston digitaalisille aineistosta muodostetulle tutkimusaineistolle tapahtuu tutkimuksen jälkeen?*

Taulukko 1: Tutkimusmenetelmän suhde aineiston tallentamiseen. Vastaaja sai valita useita vaihtoehtoja. Prosenttiosuudet on määritelty kunkin vaihtoehdon osalta suhteessa vastaajiin.

Aineiston käyttötapa	Säilytän itselläni	Arkistoin	Tuhoan	EOS
Selaillut ja lueskellut (n=93)	65,6 %	15,1 %	12,9 %	19,4 %
Lähiluku ja tulkinta (n=72)	68,1 %	18,1 %	13,9 %	15,3 %
Tiedonlouhinta (n=35)	71,4 %	45,7 %	22,9 %	14,3 %
Muuta, mitä? (n=14)	85,7 %	28,6 %	7,1 %	7,1 %

Taulukko 1 osoittaa, että tiedonlouhintaa käyttäneet tutkijat olivat aktiivisempia aineiston tallentajia kuin perinteisillä ihmistieteiden menetelmillä tutkimusta tekevät, joilla arkistojien osuus jäi alle viidennekseen tutkijoista. Avovastauksissa tiedonlouhijat jättivät kysymykseen *tutkimusaineiston*

säilyttämisestä tutkimuksen jälkeen kaksi kolmasosaa kaikista vastauksista, mikä myös kuvaa heidän aktiivisuuttaan aiheen pohtijoina. He mainitsivat yksinään lähes kaikki vastauksissa mainitut tutkimusaineistoa vastaanottavat arkistointipaikat. Tiedonlouhijat asettavat kuitenkin avovastauksissa ehtoja tutkimustuotosten (esim. aineisto, työkalut, julkaisut) jakamiselle, kuten muutkin vastaajat ja toteavat, etteivät ole aina valmiita jakamaan tuotoksiaan automaattisesti muiden kanssa.

Muuta, mitä kysymykseen vastanneista 13/14 vastauksesta voi pitää laadullisten menetelmien tarkennuksena tai aineistoon tutustumista kuvaavana ilmaisuna. Tutkijat mainitsivat tässä esimerkiksi sisällön erittelyn, kuva-analyysin, erilaisen tutkimusta varten tarvittavan taustatyön tekemisen. Myös vastauksissa mainitun tunneanalyysin voidaan jossain tapauksia katsoa kuuluvan perinteisiin laadullisiin menetelmiin, mutta useimmiten se kuitenkin liitetään sanojen tietokoneavusteiseen luokitteluun.

Tutkimustuotosten jakaminen

Kyselyssä pyydettiin sanallisia vastauksia tilanteista, joissa vastaajat olisivat valmiita jakamaan tutkimusaineistoa laajemmin tutkimustuotoksensa toisten kanssa. Vastauksien (n=73) pituudet ja sisällöt tähän vaihtelivat varsin laajasti. Vastausten mukaan toiset tutkijat olivat valmiita jakamaan tuotoksensa tutkimuksensa alusta alkaen toisten tutkijoiden tai esimerkiksi Kansalliskirjaston kanssa milloin tahansa tutkimuksen aikana. Muutama vastaaja myös rakensi tietokantaa, jota he avasivat muiden käyttöön sitä mukaan kuin se valmistui. Monet vastaajat olivat valmiita jakamaan tekemänsä tuotokset ja tekniset analyysivälineet aina sitä erikseen pyydettyäessä esimerkiksi työpajoissa tai missä tahansa yhteistyötilanteissa henkilökohtaisesti saamistaan hyödyistä riippumatta. He myös pitivät avoimia lisenssejä ensisijaisena vaihtoehtona kaikille tutkimustuotoksille, jotta kaikki halukkaat pääsisivät hyödyntämään niitä.

"Lähes poikkeuksetta ja tämä on toimintamalli, johon tulisi pyrkiä. Mieluiten vielä avoimin lisenssein." (Vastaaja 25)

"Milloin vaan ovat halukkaita käyttämään niitä." (Vastaaja 28)

"Lähtökohtaisesti olen valmis jakamaan kaiken." (Vastaaja 30)

"In any situations where this would be an appropriate output." (Vastaaja 125)

Vastauksista heijastuu tutkimusaineiston ja muiden tutkimustuotosten jakamiseen liittyvä yleinen positiivinen asenne, vaikka tutkimusaineistoa ei varsinaisesti tarjotakaan muiden tutkijoiden käyttöön. Kansainvälisissä tutkimuksissa on osoitettu samanlainen ristiriita avoimuutta suosivan asenteen ja avaamisen edellyttämien käytännön toimien välillä (Tenopir et al. 2011; Smits & Terek 2020)

Kyselyssämme tutkijat asettivat tutkimustuotosten jakamiselle ennalta määriteltäviä ehtoja, joita voidaan pitää yleisesti eettisesti kestävinä toimintamalleina. Esimerkiksi vaatimukset tekijyyden näkymisestä ja tekijänoikeuksien noudattamisesta ovat sellaisia, jotka pitää ottaa huomioon joka kerta tutkimustuotoksia jaettaessa ja niihin viitattaessa, olipa kyse sitten artikkeleista tai aineiston analyysityökaluista (myös Tenopir et al. 2015; Volk, Lucero & Barnas 2014). Muutamat vastaajat korostivat myös sitä, ettei tutkimustuotosten jakaminen saa olla haitallista tutkittaville.

"I think I would be more willing to contribute if authorship is visible." (Vastaaaja 111)

"Useimmiten, tekijänoikeuden salliessa." (Vastaaaja 86)

"I alla situationer där inte något av det empiriska materialet kan påverka forskningssubjekten (interjuude) negativt." (Vastaaaja 130)

Myös jatkorahoituksen saaminen saattoi olla ehto sille, että tutkimustuotoksia oltaisiin valmiita jakamaan nykyistä enemmän. Kyse oli yhteistyön syventämisestä ja laajentamisesta sekä osittain rahoittajan asettamista ehdoista. Osalle vastaajista ei ollut selvää, sallisiko rahoittaja tutkimustuotosten jakamista.

"Voisin jakaa, jos se toisi esimerkiksi rahoitetun jatkotutkimuksen yhden tai useamman toisen tutkijan kanssa tai jonkin myös taloudellisesti kannattavan tutkimussopimuksen Kansalliskirjaston kanssa." (Vastaaaja 24)

"Jos tutkimushanke esimerkiksi rahoittajan osalta sen [tutkimustuotosten jakamisen] sallii." (Vastaaaja 34)

Eräs vastaajista pohti tutkimustuotosten yleistä jakamisalustaa, jolle olisi valmis jakamaan tuotoksensa. Hän halusi alustalle jonkinlaista ulkoista kontrollointia ja kuratointiasiantuntemusta, mutta toivoi sen olevan avoin kaikelle datalle ja kaikille käyttäjille. Lisäksi hän toivoi, että käyttäjät voisivat

itse rikastaa alustalla olevaa aineistoa esimerkiksi lisäämällä aineistoihin omia tunnisteita, joiden avulla aineiston löydettävyyttä onnistuttaisiin parantamaan.

”Heti jos tälle [tutkimustuotosten jakamiselle] vain saadaan jokin tarpeeksi yleiskäyttöinen alusta. Suljetuista ja keinotekoisesti rajoitetun käyttäjäkunnan alustoista en ole kiinnostunut. Tärkeimmät toiveeni:

– mielivaltaisten tutkimusaineistojen lataaminen mahdollista, riippumatta siitä mistä lähteistä ne ovat peräisin; en halua joutua pohtimaan, mihin palveluun minkäkin aineistoni saa ladata.

– aineistoja on oltava mahdollista asettaa halutessaan täysin julkisiksi

– aineistoja on oltava mahdollista ainakin jonkun kuratoida, s.o. niiden löydettävyyys ei saa riippua ainoastaan tekijän lisäämistä kuvailu- ja metatiedoista. Mieluiten palvelussa oltava kaikille käyttäjille avoin tägäysjärjestelmä.”

(Vastaja 68)

Jotkut vastaajat olivat valmiita jakamaan tuotoksensa muille vasta tutkimuksen valmistuttua. Useat vastaajat ymmärsivät tutkimustuotosten jakamisen tarkoittavan vain tutkimusartikkelia eivätkä he olleet kovin innokkaita jakamaan muita tutkimuksensa tuotoksia. Osa asiaan kriittisesti suhtautuvista jakaisi tuotoksensa vain selkeästi asiasta sovittaessa muiden tutkijoiden tai Kansalliskirjaston kanssa. Vastaajat puhuivat runsaasti erilaisista neuvotteluista ja pelisäännöistä, joita pitäisi luoda muiden tutkijoiden tai organisaatioiden kanssa suullisesti tai kirjallisesti.

”Olin valmis ja kiinnostunut jakamaan tutkimustuotoksia, jos ehdoista sovittaisiin selkeästi ja mielellään myös mikäli se olisi oman tutkimukseni kannalta mielekästä.” (Vastaja 40)

Osa vastaajista nimesi rajaksi tutkimustuotosten jakamiselle sen, ettei enää itse tarvitse aineistoa omassa tutkimuksessaan. Lisäksi ehtona pidettiin sitä, etteivät muiden tutkijoiden tutkimusprojektit esimerkiksi samalla aineistolla menisi päällekkäin ensin tehdyn tutkimuksen kanssa. Riskiksi tutkimusaineistojen jakamiselle koettiin oma asema tutkijana sekä tutkimuksen asema, koska tutkimusideoiden varastaminen oli joillekin vastaajille todellinen huoli. Pelot eivät tieteen kentällä ole täysin aiheettomia, sillä joskus toisten tutkimusideoiden ja -tuotoksien väärinkäyttöön tarjolla oleva mahdollisuus konkretisoituu, vaikka lähtökohtaisesti tiedeyhteisö tuomitsee toisten ideoiden varastamisen ja viittauksiltaan puutteellisten sisältöjen hyödyntämisen epäeettisenä toimintana. (Moore 2020.)

Yksi tavallisimmista kysymyksistä, joita digitaalisten aineistojen käyttäjät esittävät tavallisimpia palautekanavia käyttäen Kansalliskirjastolle on huoli siitä, miten aineistoja saa käyttää. Se kuvaa tutkijoiden orientoitumista eettisiin työskentelytapoihin, mutta myös tekijänoikeuskysymysten monimutkaisuutta ja vaikeaselkoisuutta. Tällä kertaa tekijänoikeuskysymykset näkyivät kyselyssä enemmänkin vain ihmettelynä, miksei aineistoja voi käyttää tutkimuksessa nykyistä laajemmin tai miksi ne eivät ole avoimia ilman Tutkainorganisaatiota. Tietoisuus tekijänoikeudellisten aineistojen käytöstä jää helposti henkilötietojen käsittelyn varjoon esimerkiksi tutkimusaineistojen hallinnan kurseilla, vaikka molemmat ovat olennaisia aineiston arkistointia ja säilyttämistä ohjaavia tekijöitä, ja ne saavat tutkimuskirjallisuudessa toisilleen rinnakkaisen aseman (esim. Enwald 2018, 27).

Kyselyn vastauksissa olleet näkemykset tutkimusaiheiden varastamisesta eivät ole täysin yllätyksellisiä, koska asia on huomattu myös kansainvälisiä tutkijoiden aineistonhallintaa käsittelevissä tutkimuksissa (Tenopir et al. 2015, 7). Toisaalta tämän päivän avoimen tieteen aikakaudella kyselyssä esitetyissä näkemyksissä pidetään edelleen kiinni oikeudesta itse tuotettuihin tutkimusaineistoihin, tutkimuksen työkaluihin ja tutkimustuloksiin. Kysymykset tutkimusaineistojen tai tutkimustuotosten omistajuudesta ovat kuitenkin yleisesti isoja kysymyksiä ja niihin liittyy runsaasti eettistä pohdintaa, joka pitää ottaa huomioon kunkin tutkijan omassa toiminnassa. Nämä näkemykset, joissa vastaajat eivät olleet valmiita jakamaan tutkimustuotoksiaan juuri lainkaan muussa kuin tutkimusartikkelin muodossa, jäivät kuitenkin lopulta vähemmistöön. Kyselyssä ei huomioitu julkaisijoiden vaatimuksia aineistojen avoimuuteen, eivätkä vastaajatkaan kiinnittäneet huomiota siihen.

Johtopäätökset: Tutkimusdatan tallennuksen ja jakamisen ala- ja aineistokohtaisia erityispiirteitä

Vastaukset tutkimuskysymyksiin 1 ja 2 osoittivat, että Kansalliskirjaston aineistojen käyttäjillä aineistojen säilyttäminen tutkimuksen aikana ja niiden tallentaminen ja avaaminen tutkimuksen valmistuttua oli varsin samankaltaista kuin tutkijoilla kotimaassa ja kansainvälisesti. Kyselyn tulosten perusteella voidaan päätellä, että aineistonhallintasuunnitelmia ei laadita vielä kattavasti, sillä viidennes vastanneista ei tiennyt, mitä tekee aineistollaan tutkimuksensa jälkeen. Avovastauksissa korostettiin aineiston säilytyspaikan tapauskohtaisuutta, joka ratkaistaan vasta tutkimuksen teon aikana ja lopullisen tutkimusaineistokokonaisuuden muotouduttua.

Nykyisten avoimen tieteen vaatimusten perusteella olisi odottanut, että aineiston arkistointi olisi ollut kyselyyn vastanneiden tutkijoiden joukossa yleisempää. Toisaalta kyselyn tulokset heijastelevat kansainvälisten tutkimuksien tuloksia, joissa tutkimusaineiston arkistoinnin käytännöt näyttävät jäävän vähäisemmäksi kuin rahoittajien ja emo-organisaatioiden luomat avoimen tieteen odotukset ja tutkijoiden yksittäiset asenteet edellyttäisivät. Selitystä vähäiselle tallentamiselle on syytä etsiä humanistisen tutkimuksen erityispiirteistä, tutkimusprosessien vaihtelevuudesta ja iteratiivisuudesta, tutkimusaineiston monimuotoisuudesta ja sen ymmärrettävyydestä ainoastaan tutkijalle itselleen sekä tutkijan tarpeesta käyttää keräämäänsä aineistoa uudelleen omissa tulevilla tutkimuksissa. Hallintasuunnitelma ei ole sen vuoksi humanistisilla aloilla useinkaan muuttumaton kokonaisuus, vaan enemmänkin osoitus siitä, että niihin liittyviä asioita on huomioitu eettisesti kestäväällä tavalla. On myös huomattavaa, että tieteen kentällä puhutaan usein vain tutkimusaineiston hallinnan suunnittelusta, eikä esimerkiksi tutkimuksessa syntyneiden työkalujen hallinnasta osana tutkimustuotoksia. Niidenkin säilyttäminen ja ylläpito pitää ottaa huomioon jo tutkimusta suunniteltaessa. Erityisesti digitaalisissa ihmistieteissä kehitettävät ja datan käsittelyyn tarvittavat työkalut ovat olennaisia tutkimuksen tuotoksia.

Kysymyksessä 3 analysoimme, vaikuttaako tutkimuksessa käytetty menetelmä tutkijan halukkuuteen arkistoida aineisto. Vastaus oli ilmeinen: tiedonlouhintaa menetelmänä käyttävät tutkijat olivat selkeästi aktiivisempia tallentajia kuin pelkästään lähilukuun ja tulkintaan päätelmänsä rakentavat tutkijat. Vain harvat vastasivat menetelmäkseen pelkästään tiedonlouhinnan, mutta kyselyaineistomme perusteella louhinta menetelmänä lisäsi valmiutta tallentaa aineistoja. Tiedonlouhinnassa data kerätään tietyn suunnitelman mukaisesti ja on siten helpommin esitettävissä muille ymmärrettävässä taulukkomuodossa tai visualisoituna kuin perinteisessä tutkimuksessa koottu monimuotoinen aineisto. On myös mahdollista, että usein tutkimusryhmässä tehtävä tiedonlouhinta on usein erillistä aineistohallintasuunnitelmaa edellyttävien suurten julkisten rahoittajien rahoittamaa, kun taas perinteisiin menetelmiin perustuvaa tutkimusta tehdään useammin yksin, ja ehkä useammin pienten säätiöiden rahoittamana tai omalla ajalla. Tällöin erillisen aineistohallintasuunnitelman laadintaan ei välttämättä ole rahoittajan asettamaa velvoitetta.

Kysymyksellä 4 halusimme selvittää, onko kirjaston aineistoissa tai palveluissa ominaisuuksia, jotka muodostuvat esteeksi arkistoinnille. Kansalliskirjaston aineistojen käyttöön ei liity tutkittavien informointia tutkimusaineiston säilyttämisestä tai suostumuksen kysymistä. Sen vuoksi muissa tutkimuksissa esiin noussut ihmistieteille tyypillisten tietosuojakysymysten pohdinta

jää vähäisemmäksi Kansalliskirjaston tarjoamia aineistoja käytettäessä. Osa vastaajista tiedosti julkaistulle aineistolle ominaisen tekijänoikeussuojan tallentamisen ja avaamisen ongelman. Kovin laajamittaisesti tekijänoikeuskysymykset eivät kuitenkaan herättäneet huolta, jos vastauksia suhteuttaa siihen, kuinka paljon tekijänoikeuksiin liittyviä palautteita tulee esimerkiksi verkkoaineistojen käyttäjäkyselyissä. Merkittävä este tutkimusaineistojen jakamiselle on aineistojen kerääminen useasta lähteestä – Kansalliskirjaston eri kokoelmista ja muista arkistoista tai museoista. Seurauksena on heterogeeninen aineisto tai useita aineistoja, joiden analyysi vaatii erilaisia välineitä ja ulkopuoliselle ymmärrettäväksi tekeminen vie aikaa. Toisaalta jaettavaa aineistoa ei aina synny lainkaan, jos kootut tiedot tai sitaatit viedään suoraan tutkimustekstiin. Tällöin läpikäyty kokonaisuus, jota tutkimuksessa on käytetty, ei välttämättä avaudu muille kuin tutkijalle itselleen.

Monilla vastaajilla oli myönteinen asenne aineistojen avaamiseen. Avo-vastauksista ilmeni selvästi, että tutkijat kaipaavat nykyisten rinnalle uusia datapalveluja: avoimia tallennuspalveluita ja tukea datan kuratointiin. Nykyiset vaihtoehdot koettiin yhtäältä riittämättömiksi, toisaalta tuntemattomiksi tai hankalasti ymmärrettäviksi. Lisäksi huoli tutkimusaineistojen ja -työkalujen tekijyyden säilymisestä näkyi vastahakoisuutena luottaa avoimen tieteen eettisiin periaatteisiin.

Pohdinta

Kysely osoitti, että huolimatta myönteisestä asenneympäristöstä ihmistieteilijöillä on monia esteitä datan tallentamiselle ja jakamiselle. Tilanteissa, joissa aineisto rakentuu tutkimuksen edetessä polveillen ja koostuu etukäteen ennakkoimattomista osista ja keskenään yhteentoimimattomista palasista, tutkijat tarvitsevat tukea joustavaan aineistonhallinnan suunnitteluun ja tietoa omalle tieteenalalle potentiaalisten aineistojen ominaisuuksista ja jatkokäytöstä. Tutkijoiden on tunnettava, mihin aineistoa voi arkistoida ja millaisia aineistotyyppisiä eri arkistot ottavat vastaan. Tämän edistämiseen tarvitaan lisää yhteistyötä tutkimusaineistoa tarjoavien ja vastaanottavien tahojen kesken.

On syytä pohtia, voidaanko ihmistieteiden menetelmäopetuksella muodostaa yhteistä näkemystä siitä, millaisia ominaisuuksia muille tutkijoille jaettavalla aineistolla on ja näin ohjata opiskelijoita tuottamaan tutkimusaineistoja, jotka avautuvat muillekin kuin tutkimuksen tekijälle itselleen. Tämä tuskin on mahdollista toteuttaa kattavasti ihmistieteellisen tutkimuksen tekemisen luonteen vuoksi, joten tarvitsemme myös selkeämmän näkemyksen siitä,

millaisissa tapauksissa on perusteltua tuottaa aineistoa, jota ei voi jakaa ja miten tämä ilmaistaan aineistonhallintasuunnitelmissa.

Kansalliskirjastolla ei ole perinteisesti ollut kovin vahvaa roolia tutkimusaineistojen tai -tuotosten hallinnassa neuvojana tai arkistojana. Jatkossa roolin vahvistamiselle on kuitenkin kysyntää, koska tutkijat ovat ilmaisseet tarpeensa opastukseen Kansalliskirjaston omien aineiston käytöstä sekä toiveensa tutkimuksessa rikastettujen aineistojen säilyttämisestä. Vaikka suurin osa datanhallinnan opastuksesta tehtäisiin yliopistojen kirjastoissa, Kansalliskirjasto tarvitsee aineistonhallinnan opastustaitoja, kun se tarjoaa omia aineistojaan tutkimukseen. Sen on myös itse löydettävä vastaukset kysymyksiin, miten ja missä sen tarjoamasta datasta rikastettu digitaalinen aineisto ja tutkijoiden kehittämät työkalut säilytetään mahdollisimman avoimesti ja miten tutkimusdataa hallinnoidaan silloin, kun se sisältää tekijänoikeudenalaista aineistoa.

Kiitokset

Digitaalinen avoin muisti -hanketta on rahoittanut Euroopan aluekehitysrahasto Vipuvoimaa EU:lta 2014–2020. Digitaalinen avoin muisti on Kansalliskirjaston tutkimuspalvelujen kehityshanke, jossa suunniteltavat ja myöhemmin rakentuvat tutkimuksen palvelut ovat digitaalisten palveluratkaisujen ytimessä ja hyödyttävät kaikkia digitaalisen aineiston käyttäjiä.

Kiitokset kyselyn suunnitteluun osallistuneelle hankkeen projektiryhmälle ja kyselyn testivaiheessa mukana olleille: Jussi-Pekka Hakkarainen, Kristiina Hormia-Poutainen, Minna Kaukonen, Mikko Lappalainen, Marko Oja, Tuula Pääkkönen, Juha Rautiainen, Irma Reijonen, Petteri Veikkolainen, Osma Suominen, Jouni Ahmajärvi, Jaakko Tahkokallio ja Kimmo Kettunen.

Lähteet

- Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes toward Open Science and public data sharing: A survey among members of the German Psychological Society. *Social Psychology, 50*(4), 252–260. <https://doi.org/10.1027/1864-9335/a000384>
- Antell, K., Bales Foote, J., Turner, J., & Shults, B. (2014). Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries, 75*(4), 557–574. <https://doi.org/10.5860/cr1.75.4.557>
- Avoimen tieteen ja tutkimuksen julistus 2020–2025 (2020). *Vastuullisen tieteen julkaisusarja 1:2020*. <https://doi.org/10.23847/isbn.9789525995237>

- Bardyn, T. P., Resnick, T., & Camina, S. K. (2012). Translational Researchers' Perceptions of Data Management Practices and Data Curation Needs: Findings from a Focus Group in an Academic Health Sciences Library. *Journal of Web Librarianship*, 6(4), 274–287. <https://doi.org/10.1080/19322909.2012.730375>
- Borg, S., & Kuula, A. (2007). *Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa*. Tampereen yliopisto: Yhteiskuntatieteellisen tietoarkiston julkaisuja, 6. <http://urn.fi/urn:isbn:978-951-44-6999-2>
- Borgman, C. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press.
- Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends*, 61(3), 636–674. <https://doi.org/10.1353/lib.2013.0005>
<https://creativecommons.fi/lisenssit/> [Viitattu 7.1.2021]
- Enwald, H. (2018). Tutkimusaineiston avoin jakaminen – tutkimusorganisaatioiden jäsenten käsityksiä, kokemuksia ja mielipiteitä. *Informaatiotutkimus*, 4(37), 26–47. <https://doi.org/10.23978/inf.77411>
- EU-neuvoston linjaus (2016). <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf> [Viitattu 16.6.2020].
- Féret, R., & Cros, M. (2019). The embedded research librarian: a project partner. *LIBER Quarterly*, 29(1), 1–20. <http://doi.org/10.18352/1q.10304>
- Frederick, D. E. (2016). Data, Open Science and libraries – The Data Deluge Column. *Library Hi Tech News*, 33(8), 11–16. <https://doi.org/10.1108/LHTN-09-2016-0040>
- Fält, K. (2016). Tutkijoiden digitaaliset tutkimusaineistot ja datan avoin jakaminen digitaalisen humanismin kontekstissa. Teoksessa Kimmo Elo (toim.), *Digitaalinen humanismi ja historia-tieteet* (s. 39–67). Turun Historiallinen yhdistys.
- Hakkarainen, J.-P. (2019). Lehtiaineistojen tutkimuskäyttö laajenee Tutkain-hankkeessa. *Tietolinja*, 2019(2). <http://urn.fi/URN:NBN:fi-fe2019120445611>
- Hazlett, D. R. (2016). Libraries develop open data training program. *Library Journal*, 141(4), 24–26.
- Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*, 51(7–8), 662–672. <https://doi.org/10.1080/01930826.2011.601269>
- Hyvönen, N. (2015). Kansalliskirjasto avaa dataansa. *Tietolinja*, 2015(1). <http://urn.fi/URN:NBN:fi-fe201506029972>
- Häyrynen, A. (2012). *Open sourcing digital heritage: digital surrogates, museums and knowledge management in the age of open networks*. University of Jyväskylä. <http://urn.fi/URN:ISBN:978-951-39-4908-2>
- Järviluoma, H., & Ruotsalainen, J. (2018). Betoniporsaita avoimen tiedon valtatiellä: SENSOTRA-hanke tietosuojalainsäädännön murroksessa. *Elore*, 25(2), 141–151. <https://doi.org/10.30666/elore.77216>
- Kansalliskirjasto avasi 1930-luvun lehdet asiakaskäyttöön. <https://www.kansalliskirjasto.fi/fi/uutiset/kansalliskirjasto-avasi-1930-luvun-digitoidut-lehdet-asiakaskayttoon> [Viitattu 17.1.2021]

- Kansalliskirjaston strategia 2021–2030* (2020). <http://urn.fi/URN:NBN:fi-fe2020040210186>
- Kokko, H. (2019). Suomenkielisen lehdistön paikalliskirjekulttuuria tallentava digitaalinen Translocalis-tietokanta. *Ennen ja nyt*, 2019(2). <https://www.ennenjanyt.net/2019/08/suomenkielisen-lehdiston-paikalliskirjekulttuuria-tallentava-digitaalinen-translocalis-tietokanta/> [Viitattu 2.7.2020].
- Laney D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group*. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [Viitattu 1.7.2020].
- Lilja, J., & Pääkkönen, T. (2019). Sisältöä verkossa kirjan ystäville, tutkijoille ja historian harrastajille. *Bibliophilos*, 78(3).
- Lönnqvist, H. (2003). *Humanistiska forskningsprocesser och informationssökare. Typologier för informationssökande forskare*. Tampere University Press. Acta Electronica Universitatis Tamperensis, 921. <http://urn.fi/urn:isbn:951-44-5630-0>
- Masson, E. (2017). Humanistic data research – An encounter between epistemic traditions. Teoksessa Mirko Tobias Schäfer & Karin van Es. (toim.), *The datafied society: Studying culture through data* (s. 25–37). Amsterdam University Press.
- Moore, E. (2020). Inconsistent Responses to Notifications of Suspected Plagiarism in Finnish Higher Education. *Journal of Academic Ethics*, 18(2), 207–222. <https://doi.org/10.1007/s10805-020-09366-1>
- Open data service*. <http://data.nationallibrary.fi/> [Viitattu 5.7.2020].
- Perrier, L., Blondal, E., Ayala, A. P., Dearborn, D., Kenny, T., Lightfoot, D., . . . MacDonald, H. (2017). Research data management in academic institutions: A scoping review. *PLoS ONE*, 12(5), 1–14. <https://doi.org/10.1371/journal.pone.0178261>
- Pinfield, S., Cox, A. M., & Smith, J. (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE*, 9(12), 1–28. <https://doi.org/10.1371/journal.pone.0114734>
- Plan S: Making Full and Immediate Open Access a Reality. Principles and Implementation*. <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/> [Viitattu 4.10.2020].
- Pääkkönen, T. (2019). Digi.kansalliskirjasto.fi - Käyttäjäkysely 2019. Kansalliskirjaston sisäinen julkaisu.
- Pääkkönen T., & Lilja, J. (2018). Hieno palvelu, mutta sisältöä lisää – Kansalliskirjasto kyseli Digi-palvelun käyttökokemuksia. *Tietolinja*, 2018(2). <http://urn.fi/URN:NBN:fi-fe2018092336401>
- Pääkkönen, T., & Lilja, J. (2019). Digitaalisten aineistojen sisältökäyttötoiveita tutkijoilta. Teoksessa M. Kosonen (toim.), *Ratkaisuja digitaalisten aineistojen käytettävyyden parantamiseen* (s. 40–45). Kaakkois-Suomen ammattikorkeakoulu. (XAMK Kehittää; Vuosikerta 80).
- Rantala, H., Salmi, H., Nivala, A., Paju, P., Sippola, R., Vesanto, A., & Ginter, F. (2019). Tekstien uudelleenkäyttö suomalaisessa sanoma- ja aikakauslehdistössä 1771–1920. *Historiallinen Aikakauskirja*, 117(1), 53–67.
- Rice, R. (2019). Supporting Research Data Management and Open Science in Academic Libraries: a Data Librarian's View. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, 72(2), 263–273. <https://doi.org/10.31263/voebm.v72i2.3303>

- Räisänen, S. (2016). Tieteellisten kirjastojen tulevaisuus – datanhallintaa ja avointa julkaisemista. *Tietolinja*, 2016(1). <http://urn.fi/URN:NBN:fi-fe201602105316>
- Saarti, J. (2019). Kohti avoimempaa ja vaikuttavampaa tiedettä. Teoksessa Jarmo Saarti ja Tuulevi Ovaska (toim.), *Avoin tiede - avoin mieli: Itä-Suomen yliopiston kirjaston vuosikertomus 2018* (s. 11–16). Itä-Suomen yliopisto. <http://urn.fi/URN:ISBN:978-952-61-3084-2>
- Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2(3). <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/> [Viitattu 11.1.2021]
- Shen, Y. (2016). Strategic Planning for a Data-Driven, Shared-Access Research Enterprise: Virginia Tech Research Data Assessment and Landscape Study. *College & research libraries*, 77(4), 500–519. <https://doi.org/10.5860/cr1.77.4.500>
- Smits, D. A. B., & Teperek, M. (2020). Research Data Management for Master’s Students: From Awareness to Action. *Data Science Journal*, 19(1), 1–11. <http://doi.org/10.5334/dsj-2020-030>
- Southall, J., & Scutt, C. (2017). Training for Research Data Management at the Bodleian Libraries: National Contexts and Local Implementation for Researchers and Librarians. *New Review of Academic Librarianship*, 23(2–3), 303–322, <https://doi.org/10.1080/13614533.2017.1318766>
- Steinerová, J. (2019). Challenges of Information Infrastructures for Open Science and Academic Libraries. *Zagadnienia Informacji Naukowej - Studia Informacyjne*, 57(1), 12–27. <https://doi.org/10.36702/zin.11>
- Suomen Akatemia. *Aineistonhallintasuunnitelma*. <https://www.aka.fi/tutkimusrahoitus/hae-rahoitusta/nain-haet-rahoitusta/ohjehakemisto/aineistonhallinta/aineistonhallinta-suunnitelma/> [Viitattu 18.10.2020].
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6), 1–21. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE*, 10(8), 1–24. <https://doi.org/10.1371/journal.pone.0134826>
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., . . . Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, 15(3), 1–26. <https://doi.org/10.1371/journal.pone.0229003>
- Tolonen, M., & Lahti L. (2018). Digitaaliset ihmistieteet (”Digital Humanities”) ja historiantutkimus. Teoksessa Matti O. Hannikainen, Mirkka Danielsbacka & Tuomas Tepora (toim.), *Menneisyyden rakentajat: Teoriat historiantutkimuksessa* (s. 235–258). Gaudeamus.
- Trace, C. B., & Karadkar, U. P. (2017). Information management in the humanities: scholarly processes, tools, and the construction of personal collections. *Journal of the Association for Information science and technology*, 68(2), 491–507. <https://doi.org/10.1002/asi.23678>
- Tutkimuseettinen neuvottelukunta (2020). *Tutkijan ansioluettelomalli*. <https://www.tenk.fi/tutkijan-ansioluettelomalli> [Viitattu 11.6.2020]

- Volk, C. J., Lucero, Y., & Barnas, K. (2014). Why is Data Sharing in Collaborative Natural Resource Efforts so Hard and What can We Do to Improve it? *Environmental Management*, 53, 883–893. <https://doi.org/10.1007/s00267-014-0258-2>
- Von Wright, G. H. (1971). *Explanation and understanding*. Cornell University Press.
- Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision: Opportunities and challenges. *The Electronic Library*, 35(4), 783–797. <http://dx.doi.org/10.1108/EL-10-2016-0233>
- Zhu, Y. (2019): Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, 46(1), 41–52. <https://doi.org/10.1177%2F0165551518823174>

Liite 1: kysely



Vipuvoimaa
EU:lta
2014–2020

Hei tutkija!

Oletko joskus käyttänyt Kansalliskirjaston digitaalisia aineistoja tutkimuksessasi tai miettinyt, miten voisit hyödyntää niitä? Haluaisitko mukaan kehittämään Kansalliskirjaston tarjoamia tutkimuksen palveluja? Jos vastasit kyllä, nyt on mahdollisuus osallistua kyselyyn, jolla kehitämme Kansalliskirjaston uudenlaisia tutkimuksen palvelukonseptia Louhosta. Kyselyn tulokset julkistetaan analysoinnin valmistuttua.

Kysely sisältää 20 kysymystä ja vastaaminen vie noin 10-15 minuuttia. Kysely on auki 15.5.2020 asti. Vastanneiden on mahdollista osallistua arvontaan, jossa palkintona on museokortti ja Kansalliskirjaston julkaisuja.

Voit osallistua myös Kansalliskirjaston tutkimuksen palveluiden kehittämiseen, vaikka ne tai Kansalliskirjaston aineistot eivät olisikaan ennestään tuttuja tai kyselyyn vastaamiseen ei tällä kertaa olisi aikaa. Jos haluat vaikuttaa asiaan myöhemmässä vaiheessa, ota yhteyttä liisa.napara@helsinki.fi

Tutkimuksen palveluita kehitetään Digitaalinen avoin muisti (DAM) -hankkeessa. Tavoitteena on kehittää tutkijoiden mahdollisuuksia käyttää Kansalliskirjaston aineistoja ja luoda yhtenäisempi palvelupolku tutkimuksen palveluihin kansainvälisiä Library Lab -palvelukonsepteja mukaillen. Lab-tyyppisten palvelujen tarkoituksena on liittää tutkijat, kirjastosta saatavat tutkimuksen palvelut ja tutkimusaineistotyhteen ja edistää vuorovaikutusta näiden välillä.

Lue lisää DAM-hankkeesta ja tietosuojaselosteesta [https://www.kansalliskirjasto.fi/fi/projektit /digitaalinen-avoin-muisti-dam](https://www.kansalliskirjasto.fi/fi/projektit/digitaalinen-avoin-muisti-dam)

1. Tieteenala, jota edustan

- Ihmismaantiede
- Talousmaantiede, alueellisen kehittämisen ja matkailun tutkimus
- Yhteiskunta- ja kulttuurimaantiede, kaupunkitutkimus, aluetiede
- Kansantaloustiede
- Kasvatustieteet
- Aikuiskasvatustiede
- Ammattikasvatuksen tutkimus
- Erityispedagogiikka
- Koulutuksen arvioinnin tutkimus
- Varhais- ja alkukasvatus
- Yleinen kasvatustiede ja opettajankoulutus
- Kehitystutkimus
- Liiketaloustiede
- Nais- ja sukupuolentutkimus
- Oikeustiede
- Psykologia
- Kehitys- ja kasvatopsykologia
- Psykoterapia
- Terveyspsykologia
- Työ- ja organisaatiopsykologia
- Sosiaalitieteet
- Antropologia ja etnologia
- Sosiaalipsykologia
- Sosiologia, väestötiede
- Yhteiskunta- ja sosiaalipolitiikka, sosiaalityö
- Tieteentutkimus
- Valtiotieteet
- Hallintotiede
- Kansainvälinen politiikka
- Politologia
- Viestintätieteet

2. Muu mikä?

Tieteenalaluokitus noudattaa Suomen Akatemian pääluokitusta humanististen ja yhteiskunnallisten tieteiden osalta <https://www.aka.fi/tutkimusalaluokitus/>

Jos omaa tieteenalaa ei löydy, valitse kohta muu. Tällöin sinulle aukeaa avovastausmahdollisuus. Huomaathan, että aineistomme sopivat myös humanistisen ja yhteiskunnallisen tutkimuksen ulkopuoliseen käyttöön.

3. Olen

- Professori
- Yliopistotutkija
- Yliopistonlehtori
- Yliopisto-opettaja
- Tutkijatohtori
- Apurahatutkija (post doc)
- Väitöskirjatutkija
- Opinnäytteentekijä
- Vapaa tutkija
- Muu

4. Mikä?

5. Kotiyliopisto/kotiorganisaatio

- Aalto-yliopisto
- Helsingin yliopisto
- Itä-Suomen yliopisto
- Jyväskylän yliopisto
- Kansallisarkisto
- Kansalliskirjasto
- Kotimaisten kielten keskus (KOTUS)
- Lapin yliopisto
- Lappeenrannan–Lahden teknillinen yliopisto (LUT)
- Maanpuolustuskorkeakoulu
- Museovirasto
- Oulun yliopisto
- Suomen Kirjallisuuden Seura (SKS)
- Svenska handelshögskolan
- Svenska litteratursällskapet (SLS)
- Taideyliopisto
- Tampereen yliopisto
- Turun yliopisto
- Vaasan yliopisto
- Åbo Akademi
- Jokin ammattikorkeakoulu
- Muu

6. Mikä?

7. Digitaaliset aineistot ja palvelut

Valitse sopivin vaihtoehto

Täysin eri mieltä, eri mieltä, ei samaa eikä eri mieltä, samaa mieltä, täysin samaa mieltä (1-5)

- A) Tiedän, millaisia palveluita Kansalliskirjasto tarjoaa tutkijalle ja tutkimuksen tueksi
- B) Tiedän, mitä digitaalisia aineistoja Kansalliskirjasto tarjoaa tutkimuskäyttöön
- C) Kansalliskirjaston digitaaliset aineistot sopivat tutkimusintresseihini
- D) Tarvitsen Kansalliskirjaston tarjoamien digitaalisten aineistojen rinnalle muuta aineistoa, tarkenna halutessasi millaista aineistoa

8. Yhteistyö

Valitse sopivin vaihtoehto

Ei lainkaan, vähän, jonkin verran, paljon, todella paljon

- A) Olen tehnyt tutkimusyhteistyötä Kansalliskirjaston kanssa
- B) Tiedän, miten voin ryhtyä tutkimusyhteistyöhön kansalliskirjaston kanssa
- C) Tutkimusyhteistyö Kansalliskirjaston kanssa kiinnostaa minua tulevaisuudessa

9. Mitä seuraavista digitaalisista aineistoista tai aineistopalveluista olet käyttänyt viimeisen vuoden aikana tai suunnittelet lähiaikoina käyttäväsi? Kukin aineisto on osa jotakin palveluja, joka on merkitty (sulkuihin)

- Datakatalogi (data.nationallibrary.fi)
- Digitoitut aineistot (digi.kansalliskirjasto.fi/doria.fi/vapaakappaletyoasema/ / kansalliskirjasto.finna.fi/finna.fi)
- Elektroniset vapaakappaleet (Varia/vapaakappaletyoasema)
- Finna API (api.finna.fi)
- Finto (finto.fi)
- Kansallisbibliografia Fennica (kansalliskirjasto.finna.fi)
- Verkkoarkisto (vapaakappaletyoasema)
- Jokin erityinen kokoelma, mikä?
- Muuta, mitä tai tarkenna edellisiä halutessasi

Kansalliskirjaston kokoelmia voit katsella tästä <https://www.kansalliskirjasto.fi/fin/collections/search> osa aineistoista on digitoituina ja merkitty erikseen

10. Mitä olet tehnyt tai aiot tehdä Kansalliskirjaston aineistoille tutkimuksessasi?

- Selailut ja lueskellut
- Lähilukenukset ja tulkinnut (lähiluku metodina)
- Tiedonlouhintaa
- Tiedonlouhinta metatiedoista
- Tiedonlouhinta tekstistä
- Tiedonlouhinta kuvista
- Ennustava mallintaminen
- Kuvaava mallintaminen
- Klusterointi
- Korrelointi
- Nimien poiminta (NER)
- Aiheiden/rakenteiden etsintä (topic modelling)
- Muuta, mitä?

11. Millaisia a) hyviä puolia b) haasteita olet huomannut Kansalliskirjaston digitaalisten aineistojen käytössä?

12. Millaisia tarpeita tai vaatimuksia sinulla olisi Kansalliskirjaston digitaalisille aineistoille? Mitä ei vielä ole?

Tutkimusaineistojen säilyttäminen on tiedekentällä iso puheenaihe, ja moni rahoittaja vaatii tekemään aineiston säilyttämissuunnitelman (DMP). Myös erilaisten tutkimustuotosten säilyttäminen ja eteenpäin jakaminen ovat joissain tutkimuksissa olennaisesti pohdittavia asioita.

13. Säilytän tutkimusaineistoni tutkimuksen ajan (valitse sopivimmat vaihtoehdot)

- Omalla tietokoneella
- Henkilökohtaisessa pilvipalvelussa
- Työryhmän tai yliopiston pilvipalvelussa
- Githubissa tai vastaavassa
- Leikkeenä Digissä
- Suosikkilistana Finnassa
- Aineistoa ei tarvitse säilyttää/aineistoa ei voi tallentaa/palauttaa uudelleen alkuperäiseen aineistoon tutkimuksen jatkuessa
- En osaa sanoa
- Muuta, mitä?

14. Mitä Kansalliskirjaston digitaalisille aineistosta muodostetulle tutkimusaineistolle tapahtuu tutkimuksen jälkeen?

- Säilytän sen itselläni
- Arkistoin sen, voit kertoa myös mihin
- Tuhoan sen
- En osaa sanoa

15. Missä tilanteissa olisit valmis jakamaan tutkimustuotoksesi (esim. rikastetun/prosessoidun tutkimusaineiston tai tutkimuksessa kehittämäsi analyysityökalut) muiden Kansalliskirjaston aineistoja käyttävien tutkijoiden kanssa?

16. Millaista tukea olet saanut tutkimusaineiston muodostamiseen Kansalliskirjastolta?

- Olen osallistunut ryhmäperehdytykseen Kansalliskirjastossa
- Olen saanut tietopalvelua (esim. apua tutkimusaineiston tai -kirjallisuuden valintaan)
- Minua on neuvottu sähköpostilla/verkkopalvelussa
- Olen saanut opetusta dataklinikalla
- Kansalliskirjasto on osallistunut tutkimussuunnitelmani laadintaan
- En ole tarvinnut tukea
- En osaa sanoa
- Muuta, mitä?

17. Haluan/tarvitsen tuekseni Kansalliskirjastolta lisää

- Tietopalvelua (esim. neuvontaa aineistojen sisällöistä ja käytöstä)
- Teknisiä analyysivälineitä
- Tietoa tekijänoikeuksista tai muista juridisista kysymyksistä
- Mahdollisuuksia jakaa tutkimustuotoksiani muille Kansalliskirjastojen aineistoa käyttäville tutkijoille
- En tarvitse mitään lisää
- Muuta, mitä?

18. Yhteistyötoiveet Kansalliskirjaston kanssa

a) Minkälaista tutkimustyötäsi tukevaa yhteistyötä haluaisit tehdä Kansalliskirjaston kanssa?

b) Miten Kansalliskirjasto voisi auttaa tutkimusaineiston muodostamisessa?

19. Kehitysehdotuksia Kansalliskirjaston digitaalisiin aineistoihin ja tutkimuksen palveluihin

- a) Mitä asioita haluat kehittää Kansalliskirjaston tutkimuksen palveluissa?
- b) Mitkä asiat toimivat mielestäsi hyvin?
- c) Mikä ei mielestäsi toimi?
- d) Millaisia ominaisuuksia tarvitsisit/haluaisit tutkimusta tukeviin palveluihin?
- e) Millaisia teknisiä apuja haluaisit/tarvitsisit aineistojen käyttöön?
- f) Mitä palveluja haluaisit erityisesti kehitettävän tutkimustarpeisiisi?

20. Halutessasi voit jättää yhteystietosi, jos haluat meidän ottavan yhteyttä jatkohaastatteluja varten. Tietoja ei yhdistetä aiempiin vastauksiin

- En halua jättää yhteystietojani
- Osallistun museokortin ja Kansalliskirjaston julkaisujen arvontaan
- Suostun yhteydenottoihin, joiden tarkoituksena on olla mukana kehittämässä Kansalliskirjaston tutkimuksen palveluja

Voit myös ottaa suoraan yhteyttä Digitaalinen avoin muisti (DAM) -hankkeen suunnittelijaan liisa.napara@helsinki.fi ja kertoa, minkälaisia tutkimuksellisia tarpeita sinulla Kansalliskirjaston tutkimuksen palveluille, aineistoille ja tutkimuksen työkaluille.