

I Digitaalisen historian tutkimuksen rajat ja mahdollisuudet

Tutkijoiden digitaaliset tutkimusaineistot ja datan avoin jakaminen digitaalisen humanismin kontekstissa

KATJA FÄLT

Digitalisoitumisen myötä akateemisessa tutkimuskulttuurissa yliopistojen, rahoittajien ja julkaisijoiden edellyttämä tutkimusaineistojen, menetelmien ja tulosten avoin saatavuus nousee jatkuvasti tärkeämpään asemaan. Tutkimuksessa käytettyjen aineistojen avoin saatavuus on tieteen kehityksen ja tutkimustulosten luotettavuuden arvioinnin kannalta olennaista. Se mahdollistaa lisäksi aineistojen monipuolisen käytön. Tutkijoiden omien digitaalisten aineistojen avoin jakaminen ja niiden mieltäminen kulttuuriperintöaineistoksi ei kuitenkaan ole suomalaisessa tutkimuskulttuurissa saanut vielä riittävästi tuulta alleen. Digitaalisten ihmistieteiden parissa käydyssä keskustelussa avoimen saatavuuden kysymykset kohdentuvat usein ensisijaisesti tutkimuksessa käytettävien erilaisten aineistojen ja aineistokokoelmien avoimen saatavuuden puutteisiin, ei niinkään tutkijoiden itsensä tuottamien aineistojen avoimeen jakamiseen.

Tietoisuus tutkimusaineistojen avoimuuden merkityksestä humanistiselle tutkimuskulttuurille on silti vähitellen lisääntymässä. Suomessa tutkijoiden tuottamia ja keräämiä tutkimusaineistoja arkistoi ja luovuttaa jatkokäyttöön ensisijaisesti Tampereen yliopiston erillisyksikkö Tietoarkisto.¹ Artikkelini toimii katsauksena, joka esittelee historian tutkijoiden käytettävissä olevia aineistoresursseja. Artikkelissa tarkastellaan Tietoarkistoa ja sen humanistisille aloille laajentuvia palveluja ensisijaisesti historian tutkimuksen näkökulmasta sekä pohditaan tutkimusaineistojen avaamiseen ja jakamiseen liittyviä mahdollisuuksia sekä rajoitteita digitaalisen humanismin kontekstissa. Tavoitteena on lisätä tietoisuutta tutkimusaineistojen avaamiseen liittyvistä näkökulmista sekä sen ympärillä käytävästä keskustelusta. Lisäksi artikkeli avaa tutkimusaineistojen avaamiseen liittyvää käsitteistöä yleisemmällä tasolla.

1 Tässä artikkelissa aikaisemmin nimellä Yhteiskuntatieteellinen tietoarkisto tunnettua erillisysikköä kutsutaan sen lyhemmällä nimellä Tietoarkisto. Tietoarkiston tieteenalalajentumisesta johtuen lyhennettyä muotoa käytetään erityisesti markkinoidessa palveluita uusille tieteenaloille.

Aineistojen ja avoimuuden lähteillä

Määritelmällisiä polkuja datasta aineistoon

Ihmistieteiden puolella suomalaisessa tutkimuspuheessa sanaa data käytetään enenevässä määrin viittaamaan tutkimuksessa käytettäviin aineistoihin. Big data vyöryy retoriikassa päälle yhtä massiivisena kuin millaiseksi se mielletään; samaa päälle vyöryvyyttä vihjaa puhe, jossa data vertautuu tulvaksi tai vedenpaisumukseksi.² Big dataa voidaan kuitenkin pilkkoa myös pienempiin osiin, ”pieneen” *big dataan* (*small big data*).³ Dataa on verrattu myös öljyyn,⁴ ja tämä vertaus pyrkii liittämään dataan kallisarvoisuuden ja taloudellisen hyödyn tuntua. Yleisen määritelmän mukaan data viittaa informaatioon tai tietoon, joka on jollakin tavoin faktuaalista. Se mielletään usein merkeistä tai symboleista koostuvaksi ”raaka-aineeksi”, jolla ei itsessään välttämättä ole merkitystä. Merkitykselliseksi data muuttuu vasta perustellun tulkinnan kautta.

Mikä on humanistisen tutkimuksen kontekstissa dataa? Sanoja data ja tutkimusaineisto käytetään tiedemaailmassa usein viittaamaan käytännössä samaan asiaan, mutta eri tieteenalojen erilaiset käsitykset sanojen määritelmällisestä tarkkuudesta vaihtelevat. Luonnontieteilijä saattaa ymmärtää datan eri tavalla kuin kulttuuriperinnön tutkija. Kun esimerkiksi luonnontieteissä dataa tuotetaan itse usein kokeellisesti ja laskennallisesti, on humanistien hyödyntämä data tuotettu yleensä jo jossain historian vaiheessa. Humanistisen tutkimuksen pääasiallinen data on erilaista ihmisen kulttuurista aineistoa, tekstiä, kuvia, esineitä, ääniä ja numeroita, jotka on alun perin tuotettu tai kerätty jotakin muuta kuin käsillä olevaa tutkimustarkoitusta varten. Ihmistieteissä käsiteltävä data on siten tyypillisesti jo jonkinlaisen tulkinnan tai valinnan läpikäynyttä aineistoa, jonka kulloinenkin käyttötarkoitus valikoituu ja muokkautuu tutkijan ja tutkimuskysymyksen mukaan.

Informaatiotutkimuksen professori Christine Borgmanin mukaan kysymys ”mitä data on” on humanistisilla aloilla tärkeä, sillä vastaus

2 Borgman 2012, 1059–1078.

3 Gray ym. 2015.

4 Palmer 2006; Throp 2012.

määrittää, millaista aineistoa tuotetaan, miten sitä kerätään ja kuinka sitä huolletaan jatkokäyttöä varten. Hänen mukaansa humanistiset alat tarvitsevatkin digitaalisia resursseja, työkaluja, palveluita sekä datan säilytyspaikkoja. Borgman itse määrittelee datan formaalissa muodossa olevaksi, uudelleentulkittavaksi tiedolliseksi esitykseksi, joka soveltuu tiedonvälitykseen, tulkintaan tai prosessointiin. Borgman tarkastelee dataa myös alkuperän mukaan, jolloin se voidaan jakaa havaintodataan, laskennalliseen dataan, kokeelliseen dataan ja asiakirjoihin.⁵ Toisaalla Borgman huomauttaa, että kysymyksen ”mitä data on” sijasta olisi tärkeämpää kysyä ”milloin data on”. Tämä kysymys johdattaa pohtimaan niitä prosesseja, jonka kautta muotoutuu jotakin, jota voidaan kutsua dataksi. Datan määrittelyyn sisältyy siten Borgmanin mukaan aina luomista, valintaa ja käyttötapoja.⁶ Digitaalisen humanismin tutkija Christof Schöch linjaa humanististen alojen datan digitaalseksi, valikoidusti rakennetuksi ja koneellisesti käsiteltäväksi abstraktioksi/käsitteellisyudeksi, joka esittää jonkin osan tietyn humanistisen tutkimuksen kohteesta.⁷

Digitaalisuuden tutkija Trevor Owens on päättänyt artikkelissaan ”Defining Data for Humanists: Text, Artifact, Information or Evidence” (2011) käsittelemään humanististen alojen dataa neljän eri asiakategorian (*things*) kautta: artefakteina, teksteinä, informaationa ja todisteina. Datan artefaktuaalisuutta, esineellisyyttä korostaa sen rakennettu, ihmisen tuottama olemus. Sama tuotettu esineellisyys tekee datasta myös tekstin kaltaisen. Artefaktina, kuten esimerkiksi tekstinä data voi siten informaation potentiaalisena lähteenä sisältää Owensin mukaan samaa todistuksellista arvoa kuin mikä tahansa muu artefakti. Owens herättelee humanististen alojen tutkijoita hyödyntämään oman alansa metodiikkaa datan tulkinnessa. Owensin mukaan humanistiset lähestymistavat datan tekstuaalisuuteen ja artefaktuaalisuuteen avaavat uudenlaisia mahdollisuuksia tulkita dataa.⁸

5 Borgman 2009.

6 Borgman 2015, 17–18.

7 ”Digital, selectively constructed, machine-actionable abstraction representing some aspects of a given object of humanistic inquiry.” Schöch 2013.

8 Owens 2011; ks. esimerkiksi Drucker 2012.

Owensin esittelemä materialistis-tekstuaalinen konkretisoiminen pyrkii eräällä tavalla vapauttamaan humanistit tulkitsemaan dataa aivan kuten mitä tahansa kulttuurisesti tuotettua aineistoa. Sanan data käyttäminen humanistisessa tutkimuksessa on nimittäin osoittautunut paikoin ongelmalliseksi ja sitä onkin kyseenalaistettu. Johanna Drucker on huomauttanut, että sanana data on itse asiassa riittämätön termi kuvaamaan humanististen alojen tietoaineistoja sen etymologisten juurien vuoksi. Druckerin mukaan data viittaa johonkin valmiiksi annettuun (pohjalla latinan *do*-verbin supiini datum, joka merkitsee muun muassa antamista ja tarjoamista). Schöch puolestaan tulkitsee sanan alkuperäisen merkityksen viittaavan johonkin havainnoinnista riippumattomaan faktaan, jota ei itsessään voida kiistää.⁹ Drucker sen sijaan ehdottaa kompleksisen data-sanan tilalle niin ikään latinaan pohjaavaa sanaa *capta*, joka merkitsee muun muassa ottamista, tarttumista ja vangitsemista. Etymologiansa mukaisesti *capta* on siten jotakin, joka on aktiivisesti otettu ja jonka ottamista ohjaa jokin päämäärä. Druckerin mukaan sanojen määrittelyllä on epistemologinen ulottuvuus, joka liittyy olennaisesti humanistiseen tutkimukseen. Sen mukaisesti tutkimus tunnustaa tiedon tuottamisen osittaisen, tilannesidonnaisen ja tulkinnallisen luonteen sekä sen, että tieto on tuotettua, otettua, eikä niinkään valmiiksi annettua.¹⁰

Humanistisen tutkimuksen kontekstissa data voidaan siis ymmärtää muun muassa monipuolisiksi kulttuurisiksi aineistoiksi tai erilaisiksi tietovarannoiksi, joita hyödynnetään tutkimuksessa.¹¹ Datalla voidaankin viitata suoraan tiettyssä tutkimuksessa käytettävään pohja-aineistoon, johon tutkimuksen tulokset perustuvat. Tässä artikkelissa tällaisesta teollisen tutkimuksen pohja-aineistona käytettävästä materiaalista käytetään selvyiden vuoksi nimitystä tutkimusaineisto. Opetus- ja kulttuuriministeriön rahoittaman Avoin tiede ja tutkimus -hankkeen (ATT) käsikirjassa määritellään tutkimusaineistoiksi ne resurssit, joita tutkija tuottaa tai käyttää tutkimusprosessin aikana.¹² Tampereen teknillisen

9 Schöch 2013.

10 Drucker 2011; Schöch 2013. *Capta*-sanon pohjalla on latinan *capio*-verbin supiiniimuodon (*captum*) monikko.

11 Poikola ym. 2010, 14.

12 ”Tutkimuksen tietoaineistot”. Avoin tiede ja tutkimus -hankkeen [www-sivut](#).

yliopiston (TTY) määritelmän mukaisesti tutkimusaineisto on laajempi kokonaisuus, joka koostuu lisäksi metatiedoista eli kontekstia, sisältöä ja rakennetta sekä niiden hallintaa ja käsittelyä kuvaavista tiedoista.¹³ Tässä artikkelissa käsite tutkimusaineisto pohjaa ATT:n ja TTY:n määritelmiin. Lisäksi termiä data käytetään artikkelissa viittaamaan ensisijaisesti muihin kuin tutkimustarkoituksissa tuotettuihin ja kerättyihin resursseihin. Termiä raakadata käytetään ensisijaisesti viittaamaan prosessoimattomaan ja käsittelemättömään dataan, joka voi koostua esimerkiksi alkuarvoista, luvuista tai kirjaimista sellaisinaan.

Aineistot tutkimuspoliittisessa valokeilassa

Vuonna 2014 käynnistyi Opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus -hanke (ATT), joka pyrkii tiedon saatavuuden ja avoimen tieteen edistämiseen. Hankkeen tavoitteena on nostaa Suomi vuoteen 2017 mennessä yhdeksi tieteen ja tutkimuksen avoimuuden sekä avoimen tieteen johtomaiksi.¹⁴ Yliopistomaailmaan laajentunut avoimen tiedon vaatimus pohjaa julkishallinnon pyrkimyksiin avata julkishallinnolle, organisaatioille, yrityksille ja yksityishenkilöille kertynyttä tietoa ulkopuolisten maksutta ja vapaasti käytettäväksi.¹⁵ Suomessa avointa tietoa kansalaisten käyttöön ovat ensimmäisinä julkaisseet esimerkiksi ministeriöt ja niiden alaiset asiantuntijaorganisaatiot. Tutkimusaineistojen avoimuus liittyy olennaisena osana avoimen tieteen ja tutkimuksen tavoitteisiin.

Avoimen tieteen edistämiseen liittyviä linjanvetoja ryhdyttiin laatimaan 2000-luvulla. Vuonna 2003 julkistettiin *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, jonka Suomen yliopistojen rehtorien neuvosto allekirjoitti vuonna 2006. Julistuksen mukaan avoimen saatavuuden piiriin on ulotettava alkuperäiset tutkimustulokset, prosessoimaton raakadata, metatiedot, lähdeaineistot, digitaalisessa muodossa olevat kuvalliset ja graafiset esitykset sekä tieteellinen

13 ”Tutkimusdata”. Tampereen teknillisen yliopiston www-sivut.

14 ATT-hankkeen www-sivut.

15 Mm. Open Knowledge Foundation.

multimedia-aineisto.¹⁶ Julistuksen fokuksena on ensisijaisesti tutkimustulosten avoimen julkaisemisen lisääminen ja mahdollistaminen. Berliinin julistus määrittelee avoimen saatavuuden tiedon ja kulttuuriperinnön perustavanlaatuisiksi lähteeksi. 2007 OECD julkisti julkisrahoitteisten tutkimusaineistojen saatavuuden lisäämiseksi ohjeistuksen (*OECD Principles and Guidelines for Access to Research Data from Public Funding*), jonka mukaan keskeisten julkisesti rahoitettavien tutkimusaineistojen tulee olla avoimessa käytössä. OECD:n suositus koskee erityisesti digitaaliseen muotoon tallennettuja aineistoja, jotka on alun perin koottu tutkimuskäyttöön.¹⁷ Suosituksen piiriin kuuluu siten suurin osa yliopistoissa ja tutkimuslaitoksissa kerättävistä ja hallinnoitavista tutkimusaineistoista. Ensimmäinen yliopistokohtainen datapolitiikka, Tutkimuksen tietoa-aineistojen hallinnan periaatteet Jyväskylän yliopistossa, hyväksyttiin kuitenkin rehtorin päätöksellä vasta 9.9.2014 Jyväskylän yliopistossa. Sen mukaan tutkimusaineistojen hallinta ja uudelleenkäytön edistäminen pyritään integroimaan osaksi tutkimushankkeiden, tutkimusryhmien ja tutkijoiden työtä. Päätöksessä kiinnitetään huomiota myös tutkimusaineistojen pitkäaikaistallentamiseen ja avaamiseen.¹⁸ Helsingin yliopistossa puolestaan hyväksyttiin tutkimusdatapolitiikka 11.2.2015 ja Itä-Suomen yliopistossa 1.12.2015. Vuoden 2016 ensimmäisellä puoliskolla datapolitiikkansa ovat julkaisseet Tampereen teknillinen yliopisto (27.1.2016), Oulun yliopisto (5.2.2016), Turun yliopisto (9.2.2016) ja Aalto-yliopisto (10.2.2016).¹⁹ Tampereen yliopiston datapolitiikka on tekstin kirjoitusajankohdalla (9.6.2016) lausuntokierroksella.

Yliopistojen ohella kansalliset rahoittajat ovat kiinnittämässä huomiota tutkimusaineistojen avoimeen saatavuuteen. Aineistojen avaamista suosittelevat Suomen Akatemia²⁰, Koneen Säätiö²¹, Alkoholi-

16 Berlin Declaration 2003.

17 OECD Principles and Guidelines 2007; ks. myös Borg ja Kuula 2007.

18 Tutkimuksen tietoa-aineistojen hallinnan periaatteet Jyväskylän yliopistossa 2014.

19 Tutkimusorganisaatioiden avoimuuden linjaukset löytyvät kootusti Avoin tiede ja tutkimus -hankkeen sivuilta.

20 ”Avoin tiede”. Suomen Akatemian www-sivut.

21 ”Apurahan saajalle”. Koneen Säätiön www-sivut.

tutkimussäätiö²² ja Työsuojelurahasto. Suomessa tutkimusaineistojen avaamista siis vasta suositellaan, kun esimerkiksi Iso-Britanniassa ja USA:ssa linja on joillain rahoittajilla selvästi tiukempi; tutkimusaineistojen avaaminen on pakollinen ehto rahoituksen saamiseksi.²³ Rahoittajien ja yliopistojen linjaukset vaikuttavat tutkijoiden ja tutkimusryhmien jokapäiväiseen työhön ja tutkimusaineistojen avaamiseen liittyviä päätöksiä on historia-aloillakin tehtävä viimeistään rahoitusvaiheessa.

Tutkimusaineistojen äkinäiseltä tuntuva nouseminen tutkimuspoliittiseen valokiilaan ei kuitenkaan ole tapahtunut yhdessä yössä. Jo vuonna 1985 julkaistussa *Sharing Research Data* -raportissa pyrittiin kohdentamaan huomiota tutkimusaineistojen jakamisen ja huoltamisen käytäntöjen kehittämiseen. Raportissa muun muassa suositellaan, että tutkimusaineistojen jakaminen ja avoimeksi saattaminen tulisi tehdä suunnitelmallisesti osana tutkimusprosessia.²⁴ 1980-luvun puolivälin jälkeen tutkimusaineistojen avaamisen propagointi on ollut tyypillisintä luonnontieteellisillä ja yhteiskuntatieteellisillä aloilla.²⁵ Keskeisiä tutkimusdatan säilytyspaikkoja ovat luonnontieteissä muun muassa Dryad (lääketiede) ja Pangea (geotieteet). Poliitiikan tutkimus ja yhteiskuntatieteet ovat vastaavasti kunnostautuneet kattavan, maailmanlaajuisen tutkimusaineistopalveluverkoston luomisessa. Amerikassa vuonna 1962 perustettu Inter-University Consortium for Political and Social Research (ICPSR) pyrkii toimimaan tutkimusta palvelevana monipuolisena aineistoresurssina. Euroopassa puolestaan vuonna 1976 perustettu Consortium for European Social Science Data Archives (CESSDA)²⁶ toimii kansallisten data-arkistojen kattojärjestönä ja tarjoaa mittavia,

22 ”Mitä hakemus sisältää”. Alkoholitutkimussäätiön www-sivut.

23 Borg 2014, 16–17; Kvalheim ja Kvamme 2014, 16, 18.

24 *Sharing Research Data* 1985, 27–28.

25 Luonnontieteissä ja yhteiskuntatieteissä datan jakamiseen liittyvää keskustelua on käyty 1980-luvulta asti enenevässä määrin. Esimerkiksi Craig ja Reese 1973; *Sharing Research Data* 1985; Ceci 1988; Stanley ja Stanley 1988; Sieber 1989; Sterling ja Weinkam 1990; *Sharing Social Science Data* 1991; Sieber 1991; Fienberg 1994; McCain 1995; Arzberger ym. 2004; Costello ja Berghe 2006; Piowar ym. 2008; Piowar ja Chapman 2010; Costello 2009; *Preparing Data for Sharing* 2010; Tenopir ym. 2011.

26 Aikaisemmalta nimeltään Council for European Social Science Data Archives.

integroituja ja vakaalla pohjalla olevia aineistopalveluita yhteiskuntatieteellisen tutkimuksen käyttöön. Humanistisilla aloilla täysin vastaavan laajuista arkistoverkostoa ei ole vielä luotu; toimivia tutkimusinfrastruktuureja on tosin kehitetty. Esimerkiksi kielen tutkimuksen verkosto CLARIN, johon Suomikin kuuluu (FIN-CLARIN), on rakentamassa kieliaineistoja vastaanottavien data-arkistojen verkostoa. CLARINin palveluihin kuuluu humanististen tieteiden tutkimusaineistojen käytön, jalostamisen, säilytyksen ja jakamisen edistäminen. DARIAH puolestaan on humanistisen ja taiteiden tutkimuksen eurooppalainen digitaalinen tutkimusverkosto, joka muun muassa tuottaa erilaista tutkimus- ja opetusmateriaalia digitaalisen humanismin tarpeisiin. Lisäksi CESSDA-verkoston arkistoista humanistista tutkimusaineistoa vastaanottavat Britannian perinteikäs UK Data Archive (UKDA), Ruotsin Svensk Nationell Datatjänst (SND), alankomaalainen Data Archiving and Networked Services (DANS) ja Suomessa Tampereen yliopiston erillisyksikkö Tietoarkisto. Arkeologista tutkimusta palvelee puolestaan vuonna 1996 perustettu, Englannin Yorkissa sijaitseva Archaeology Data Service (ADS), jonne on mahdollista arkistoida arkeologista aineistoa.

Arkistoinnin luontevuudesta ja kulttuuriperintöorganisaatioiden digitointiprojekteista huolimatta useimmat humanististen alojen tutkijat eivät vielä luontevasti miellä omien tutkimusaineistojensa potentiaalia avaamisen ja jatkokäytön näkökulmasta.²⁷ Historia-alojen tutkijoiden on varsin luontevaa käyttää jo valmiiksi arkistoituja aineistoja, joten arkistoitujen digitaalisten tutkimusaineistojen käytön ei pitäisi suuresti poiketa tutkijantyön perustoimista. Historia-alojen tutkijoiden suosima arkistotutkimus on perinteisesti kohdistunut monipuolisiin paperiaineistoihin, mutta myös digitaalisiin aineistokategorioihin kuuluvia kokonaisuuksia on hyödynnetty paljon. Tällaisia ovat esimerkiksi kielikorpukset sekä muistiorganisaatioiden aineistojen hakupalvelut ja digitaaliarkistot.²⁸ Monelle tutkijalle muiden tuottamien aineistojen jatkokäyttö ajatuksena voikin olla helpompaa kuin omien kerättyjen ja koostettujen tutkimusaineistojen luovuttaminen arkistoitavaksi. Tutkimusaineistot

27 Esimerkiksi Broom ym. 2009, 1163–1180.

28 Parland-von Essen 2015.

joko koetaan liian ”hankaliksi”, monipuolisiksi, jo valmiista arkistotutkimusaineistoista ”jatkojalostetuiksi” tai tutkijan henkilökohtaiseksi omaisuudeksi. Tutkimusaineistot koetaan usein myös liian kontekstuaalisiksi ja moneen suuntaan linkittyviksi kokonaisuuksiksi. Alalla onkin tyypillistä, että kerätyt tutkimusaineistot jäävät tutkijan itsensä haltuun. Tutkimusaineistojen potentiaalinen jatkokäyttäjä on myös yleisimmin tutkija itse. Siksi puhe avaamisesta saattaa hämmentää ja jopa pelottaa. Tutkimusaineistojen avaaminen ei kuitenkaan ole mikään Pandoran lippaan avaamiseen verrattavissa oleva teko, vaan yksinkertaisimmillaan kyse on oman kerätyn tai koostetun tutkimusaineiston tarjoamisesta muun tiedemaailman käyttöön maksutta jonkinlaisen jakelukanavan kautta. Tutkimusaineistojen avaamisen peruserätyksen mukaisesti tutkimuksessa käytettävät tausta-aineistot tai raakadata eli prosessoimaton data²⁹ tulee olla muun tiedeyhteisön käytettävissä viimeistään kyseisen tutkimuksen päätyttyä. Pohjalla on ajatus tieteen läpinäkyvyydestä ja sitä myöten tutkimusmenetelmien, -analyysien ja -tulosten paremmasta todentamisesta ja toistettavuudesta – siis avoimen tieteen perusperiaatteista.

Avaamiseen liittyä olennaisena osana ajatus tutkimusaineistojen jatkokäytöstä. Tutkimusaineistot ovat arvokkaita nimenomaan käytettynä. Tämän mukaisesti tiedeyhteisön muut jäsenet voivat hyödyntää omissa tutkimuksissaan jo avattuja tutkimusaineistoja. Avaamiseen ei kuitenkaan kuulu ajatus holtittomuudesta tai täysin vapaasta leviämisestä, jota ei voi kontrolloida. Tutkimusaineistoja ei ole tarkoitus tallentaa epämääräisesti pilveen, josta ne voivat pahimmassa tapauksessa hengenkäyskenkevyesti levitä tai kadota kokonaan. Vaikka erilaiset pilvipalvelut tai alustat voivat toimia tutkimuksen ajan näppäränä ja tehokkaanakin säilytyspaikkana, on aineistojen avaamisen ja luotettavan pitkäaikais-säilytyksen ehtona jonkinlainen vakiintunut jakelukanava, esimerkiksi data-arkisto. Seuraavassa kappaleessa esitellään kansallisen data-arkiston, Tietoarkiston toimintaa humanististen tieteenalojen, erityisesti historia-alojen näkökulmasta. Tavoitteena on esitellä historia-alojen

29 Raakadatalle voidaan viitata myös ns. primääriseen dataan tai lähdedataan. Tyypillistä raakadatalle on, että se on käsittelemätöntä, prosessoimatonta ja validoimatonta, tyypillisesti numeroista tai biteistä koostuvaa ainesta.

tutkijoiden käytössä olevan, valmiin palveluinfrastruktuurin mahdollisuuksia historia-alojen tutkimusaineistojen säilyttämisen ja jatkokäytön näkökulmasta.

Kansallinen infrastruktuuri tutkijoiden apuna – tapaus Tietoarkisto

Tietoarkisto on vuonna 1999 opetus- ja kulttuuriministeriön aloitteesta perustettu Tampereen yliopiston erillisyksikkö, jolla on tutkimuksen ja opetuksen valtakunnallinen palvelutehtävä. Sen keskeisin tehtävä on tieteellisen tutkimuksen lähteinä käytettävien digitaalisten tutkimusaineistojen dokumentointi, luettelointi, käytettävyyden ylläpito, turvallinen jakaminen ja pitkäaikaissäilytys.³⁰ Käytännössä Tietoarkisto on monipuolinen, kansallista tutkimusta tukeva palveluinfrastruktuuri, joka tietoaaineistojen vastaanottamisen, säilyttämisen, jakamisen ja hoitamisen ohella tarjoaa tutkimusaineistoihin liittyvää tietopalvelua. Tietoarkisto ottaa vastaan tieteelliseen tutkimukseen tuotettuja tutkimusaineistoja. Tutkimukseen, opetukseen ja opiskeluun on mahdollista ladata arkistoituja tutkimusaineistoja palveluportaali Ailan kautta rekisteröitymällä käyttäjäksi suomalaisten yliopistojen ja korkeakoulujen Haka-tunnuksilla. Tietoarkiston arkistonmuodostussuunnitelman mukaan aineistokokoelmaa kartutetaan aktiivisesti ja selektiivisesti. Arkistoitavien tutkimusaineistojen tulee täyttää soveltuvin osin laadulliset, tekniset ja lainsäädännölliset seulontakriteerit.³¹

Tietoarkiston tutkimusaineistoja voidaanakin eräällä tapaa tarkastella digitaalisten kokoelman näkökulmasta. Digitaalisen humanismin kiinnostuskenttä käsitetään useasti erilaisten digitaalisten (kulttuuriperintö) kokoelmien kokoamiseksi, huoltamiseksi ja ylläpitämiseksi. Tyypillisesti tällaisissa digitaalisissa kokoelmissa, varsinkin kulttuuriperintöaloilla, on kyse analogisen aineiston muuntamisesta digitaaliseen muotoon. Tietoarkiston aineistot ovat jo valmiiksi digitaalisessa muodossa, eivätkä

30 Tampereen yliopiston johtosääntö 1.11.2011, 5 §.

31 Tietoarkiston arkistonmuodostussuunnitelma.

aineistojen kartuttamis-, järjestämis- ja säilyttämisperiaatteet³² suuresti poikkeaa perinteisten arkistojen vastaavista.

Perinteisestä arkistotyöstä tietoarkistotyö poikkeaa erityisesti siinä, että aineistojen metatiedot ovat tyypillisesti laajempia kuin perinteisissä arkistoissa. Metatiedoista selviää tietoa aineiston synnystä, alkuperäisestä käyttötarkoituksesta, tekijöistä, saatavuudesta, käyttöehdoista ja niin edelleen. Tutkimusaineistot kuvaillaan suomeksi ja englanniksi. Jane Greenbergin mukaan metatiedoilla tarkoitetaan strukturoitua tietoa tietystä kohteesta, joka tukee kyseisen kohteen toimintoja. Kohde on tässä tapauksessa mikä tahansa entiteetti tai arvo, josta on mahdollista kirjata kontekstoivaa tietoa.³³ Priscilla Caplan on jaotellut metatiedot kuvaileviin, hallinnollisiin tai rakenteellisiin metatietoihin ja tämä jaottelu viittaa metatietojen funktionaaliseen tarkoitukseen ja käyttöön.³⁴ Metatietojen hallintaan on lisäksi kehitetty standardeja kuten Dublin Core, joka on kansainvälinen, erityisesti digitaalisten aineistojen kuvailuun kehitetty formaatti.³⁵ Metatiedot ovat oleellinen osa humanististen alojen tutkimusaineistojä; jopa siinä määrin, että aineistoa ei voi välttämättä lainkaan tulkita ilman taustoittavaa tietoa. Metatiedot ovat siis aina kontekstoivia, eivät pelkästään sisältöjä kuvailevia. Metatietojen ominaisuudet voidaan ulottaa jopa laajemmalle. Historiantutkija Michael J. Kramer on todennut, että metatiedot sisältävät itsessään myös tulkinnallisen ulottuvuuden. Kramerin mukaan arkistoinnin voidaan siten ymmärtää lisäävän

32 Säilyttämisen osalta varsinkin pitkäaikaissäilyttäminen voi aiheuttaa ongelmia. Tietoarkiston arkistonmuodostussuunnitelmassa todetaan: ”Laitteistojen, ohjelmistojen ja tallennusmenetelmien vanheneminen johtaa aika ajoin migraatiotarpeeseen eli tarpeeseen muuntaa aineistovaranto toiseen laite- ja ohjelmistoympäristöön sopiviksi. Migraatio ei muuta aineistojen tietosisältöä. Pitkäaikaisesti säilytettävien aineistojen tiedostoformaattit on valittu siten, että ne ovat mahdollisimman helposti muutettavissa. Ensisijaisesti suositaan avoimia, standardien mukaisia ja useiden valmistajien tukemia formaatteja suljettujen ja valmistajariippuvien formaattien sijaan.” Tietoarkiston arkistonmuodostussuunnitelma.

33 ”Metadata as structured data about an object that supports functions associated with the designated object.” Greenberg 2003, 1876; Greenberg 2005, 20.

34 Caplan 2003, 3–5.

35 Muunlaisista metatietostandardeista mm. Caplan 2003, 5–8.

eräänlaisen uuden tulkinallisen kerroksen tutkimusaineistoon. Hän haastaa kiinnostavalla tavalla pohtimaan arkistoitavan tutkimusaineiston rajoja, ja sitä leikkauspistettä, jossa alkuperäinen tutkimusaineisto ja metatiedot liittyvät yhteen.³⁶

Vastaanotettuja tutkimusaineistoja muokataan lisäksi jatkokäyttöön soveltuvaksi. Digitaalisessa tiedostomuodossa oleva Tietoarkistoon arkistoitava tutkimusaineisto käydään aina lävitse huolellisesti. Tietoarkiston arkistointiprosessissa kiinnitetään lisäksi erityistä huomiota tietosuojaan ja tarvittaessa tutkimusaineisto anonymisoidaan. Tieto on tunnisteteetonta eli anonyymiä silloin, jos suoria ja epäsuoria tunnisteita yhdistelemällä yksittäisiä tutkittavia ei ole mahdollista tunnistaa yksilöiden aineistosta. Anonymisointiprosessissa tutkimusaineistosta poistetaan tai muunnetaan yksilöön viittaavat tunnistetiedot kuten nimet, osoite- ja yhteystiedot tai henkilötunnukset sekä taustamuuttujat kuten ikä, sukupuoli, koulutus, ammatti, kieli tai kansalaisuus.³⁷ Tunnisteellisuutta on mahdollista vähentää esimerkiksi antamalla tutkittaville pseudonyymit sekä kategorisoimalla eli karkeistamalla tunnistetietoja yleisemmälle tasolle. Tämä on erityisen hyödyllistä taustamuuttujien käsittelyssä. Henkilötietojen käsittely pohjaa henkilötietolakiin ja sen tarkoituksena on yksilön yksityisyyden suojan turvaavien perusoikeuksien takaaminen.³⁸ Historian ja kulttuurintutkimuksen aloilla on kuitenkin usein perusteltua esittää tutkittavat omilla nimillään, mutta tunnistettavuudesta tulee aina sopia tutkittavien kanssa. Tunnisteiden poisto on aina tapauskohtaista, eikä anonymisointi siten ole jokaisen tutkimusaineiston kohdalla automaattinen tai samanlainen prosessi.

36 Kramer 2014. Termi metadata juontuu alun perin tietojenkäsittelyopin puolelta. Termi itsessään on kehitetty 1960-luvulla. Metatiedoista on kirjoitettu runsaasti varsinkin kirjasto- ja informaatioalalla, mutta myös muun muassa asiakirjahallinnon ja kulttuuriperintöalan näkökulmista. Katso esimerkiksi *Introduction to Metadata* 1998; Hudgins ym. 1999; Saarela 1999; Caplan 2003; *Metadata in Practice* 2004; *Metadata Best Practices and Guidelines* 2011; *Handbook of Metadata* 2014; *Cultural Heritage Information* 2015.

37 ”Tunnisteellisuus ja anonymisointi”. Tietoarkiston aineistohallinnan käsikirja. Ihmistieteellisessä tutkimuksessa tutkittavilta kerätyt tiedot harvoin sisältävät esimerkiksi henkilötunnuksia.

38 Henkilötietolaki 523/1999.

Tietoarkisto on alun perin perustettu ensisijaisesti yhteiskunta-tieteellisten alojen digitaalisten tutkimusaineistojen arkistointi- ja jatkokäyttöpalveluksi. Tästä johtuen tyypillisin aineistotyyppi on kvantitatiivinen tutkimusaineisto, jota on kerätty kyselyillä ja haastatteluilla sekä tallennettu havaintomatriiseiksi, joiden analysointi helpottuu jonkin tilastomatematiikan ohjelmiston käytöllä. Laadullisia eli kvalitatiivisia tutkimusaineistoja on artikkelin kirjoitushetkellä 173 kappaletta.³⁹ Tietoarkistoon arkistoidut laadulliset tutkimusaineistot sisältävät pääsääntöisesti tekstiä ja kuvaa. Tyypillisiä aineistomuotoja ovat haastattelulitteraatiot, etnografiset kenttäpäiväkirjat ja erilaiset kirjoitusaineistot.

Historia-alojen tutkimuksen näkökulmasta Tietoarkiston tutkimusaineistovarannolla on tällä hetkellä vielä muutamia rajoitteita. Ensinnäkin puhtaasti historia-alojen tutkimukseen liittyviä jatkokäytettäviä tutkimusaineistoja on rajallinen määrä. Artikkelin kirjoitusajankohtana keväällä 2016⁴⁰ tietoarkiston aineistohaku antaa ensisijaisen tieteenalasanan historia perusteella kuusi aineistoa: Elina Haavio-Mannilan, Museoviraston ja Marttaliiton tuottaman aineiston Naiset ja sota 1987–1988⁴¹, Helsingin yliopiston yhteiskuntahistorian laitoksen (nykyään politiikan ja talouden tutkimuksen laitos) tutkijoiden haastatteluaineistot Suomalaisten historiatietoisuus 2009 ja 2010⁴², Kauhajoen yritysyritystutkimuksen haastattelut 2013⁴³, Helsingin syyttäjänviraston historiikkihaastattelut 2013⁴⁴ sekä Suomen Kuurosokeat ry:n, Kuurojen palvelusäätiön, Kuulonäkövammaisten lasten vanhemmat ry:n ja Jyväskylän yliopiston Kasvatustieteiden tiedekunnan aineiston Kuulonäkövammaisten lasten vanhemmat 2007⁴⁵. Osa historiaa käsittelevistä tutkimusaineistoista on kuitenkin luokiteltu ensisijaisesti jonkin muun tieteenalan tutkimukseksi, kuten kulttuurintutkimuksen, miestutkimuksen, politiikan tutkimuksen tai sosiologian. Täten historian

39 Viimeisin aineistohaku suoritettu 13.6.2016.

40 Viimeisin haku suoritettu 23.5.2016.

41 Aineistonumero FSD2551.

42 Aineistonumerot FSD2932 ja FSD2867.

43 Aineistonumero FSD2882.

44 Aineistonumero FSD2920.

45 Aineistonumero FSD3064.

tutkimuksen kohteeksi soveltuvia jatkokäytettäviä tutkimusaineistoja on määrällisesti enemmän kuin mitä tieteenalaluokitus antaa ymmärtää. Tämä ei kuitenkaan rajoita tutkimusaineistojen käyttömahdollisuutta. Kaikki tutkimusaineistot ovat kaikkien tieteenalojen tutkijoiden käytettävissä ja siten myös historia-alojen tutkijoille on mahdollisuus laaja-alaisesti hyödyntää muuhun kuin alun perin historiantutkimukseen kerättyä tutkimusaineistoa.

Toiseksi historiantutkimuksen näkökulmasta tutkimusaineistojen ajallinen ulottuvuus on verrattain lyhyt: vanhimmat kerätyt aineistot ovat 1960-luvun alusta ja sisällöllisesti aineistojen aikajänne kaartuu 1810-luvulta nykyaikaan.⁴⁶ Tällä hetkellä Tietoarkiston historia-alojen tutkimukseen soveltuvat aineistot mahdollistavat siten vain lähihistorian tutkimuksen. Tilanne toivon mukaan paranee lähitulevaisuudessa, sillä Tietoarkisto on laajentamassa palveluitaan humanistisille aloille ja pyrkii jatkossa vastaamaan myös näiden alojen erityiskysymyksiin. Osana tätä laajentumista Tietoarkisto pyrkii profiloitumaan myös humanistisia aloja täysipainoisesti palvelevana infrastruktuurina, jonne humanististen alojen tutkijoiden olisi luontevaa arkistoida tutkimusaineistojaan.

Kolmanneksi historia-alojen tutkimusaineistojen luonne rajoittaa jossain määrin niiden avaamista muiden tutkijoiden käyttöön. Tutkimusaineistot ovat perinteisesti monipuolisia ja voivat sisältää muun muassa arkisto-, rekisteri- ja kirjallisuuslähteistä koostettuja tekstiaineistoja, haastatteluja, kyselyvastauksia, teemakirjoituksia, seuranta-päiväkirjoja, visuaalisia aineistoja, esineitä, bibliografioita, blogitekstejä, kenttämuistiinpanoja, taulukoita, listauksia, kartoja ja niin edelleen.⁴⁷ Historia-alojen tyypillisin tutkimusaineisto on usein ”jatkojalostettu” arkistoaineistosta tai koostettu kirjallisuudesta. Digitaalisen humanismin tutkimuksen piirissä tuotetaan lisäksi erinäistä ei-tekstuaalista tutkimusaineistoa kuten vaikkapa visualisointeja tai 3D-mallinnuksia, joiden prosessointi arkistointi- ja jatkokäyttökelpoiseksi vaatii enemmän

46 Vanhanen, Tatu: Demokratisoitumisen mittarit 1810–2012 [elektroninen aineisto]. FSD1289, versio 6.0 (2014-01-31).

47 Yleisemmin humanististen alojen aineistojen piirteistä mm. Borgman 2015, 161–202

resursseja kuin puhtaasti tekstipohjaisen aineiston käsittely.⁴⁸ Erilaisista aineistotyypeistä koostuvasta tutkimusaineistokokonaisuudesta voi olla mahdollista arkistoida vain osa tai osia. Laadullisia tutkimusaineistoja voidaan silti monessa tapauksessa käsitellä varsin helposti arkistointi- ja jatkokäyttökelpoiseksi. Esimerkiksi Iso-Britannian data-arkisto UKDA on menestyksekkäästi avannut laadullisia tutkimusaineistoja 2000-luvun alusta lähtien. Myös Alankomaissa toimiva DANS sekä Ruotsin SND arkistivat laadullisia tutkimusaineistoja. Varsinkin tekstiaineistojen arkistointiin on luotu toimivat käytännöt. Myös kuvien (esimerkiksi tutkijoiden itsensä ottamat valokuvat)⁴⁹, karttojen⁵⁰ ja audiovisuaalisten⁵¹ tutkimusaineistojen arkistointiin on kehitetty toimivia ratkaisuja. Laadullisten tutkimusaineistojen arkistoinnille ei siten ole varsinaisia esteitä, joskin eräät aineistotyyppit vaativat arkistoinnilta kehittyneempiä teknisiä ratkaisuja tai esimerkiksi tekijänoikeudellisten seikkojen selättämistä.

Tietoarkiston palvelut ovat kuitenkin avoinna kaikkien alojen tutkijoille. Täten myös historia-alojen tutkijoilla on mahdollisuus sekä tallentaa omia digitaalisia tutkimusaineistojaan Tietoarkistoon että jatkokäyttää muiden tutkimusaineistoja. Käytettävissä ovat myös tieto- ja aineistohallintapalvelut kuten Aineistohallinnan käsikirja, menetelmäopetuksen tietovaranto, poliittisten ohjelmien tietovaranto Pohtiva sekä kotimaisten ja ulkomaisten aineistojen hakemiseen liittyvä neuvonta. Tavoitteena 2010-luvun loppuun mennessä on humanististen alojen aineistojen erityistarpeiden parempi huomioiminen. Esimerkiksi tekijänoikeudella suojattu materiaali asettaa erityisiä haasteita arkistoinnille. Tekijänoikeudella suojatun aineiston uudelleen julkaiseminen on edelleen sallitumpaa painetussa muodossa kuten tutkimusjulkaisuissa, mutta vaatii usein sekä lupia ja joissain tapauksissa erinäisiä maksuja.

48 Terras 2009.

49 UKDAn suosittelemat aineistoformaatit ”Recommended formats”. UKDAn [www](http://www.ukda.ac.uk)-sivut.

50 Esimerkiksi SND:n arkeologinen GIS-data sekä UKDAn GIS-, CAD- sekä vektori- ja rasteridata.

51 Esimerkiksi UKDA ja Kielipankki.

Elisabeth Borgmanin mukaan aineistojen jakamisen linjaukset olettavat edelleen, että aineiston tuottajilla on kaikki sitä koskevat oikeudet luovuttaa aineisto arkistointiin ja jatkokäyttöön.⁵² Humanistisen tutkimuksen kohdalla näin ei kuitenkaan aina ole. Tällä hetkellä tutkijoiden on kuitenkin mahdollista arkistoida Tietoarkistoon jo arkistoidusta Arkistolaitoksen aineistosta tutkimustarkoituksiinsa ottamat digitaaliset valokuvat.⁵³ Kapioston ja Tietoarkiston välisen sopimuksen mukaan tutkimusaineistoihin kuuluvia, tekijänoikeuslain mukaisesti teossuojaa saavia sanomalehti-, aikakauslehti- ja kuva-aineistoja on mahdollista arkistoida jatkotutkimuskäyttöön Tietoarkistoon. Tällainen kokoelma, joka on koottu tietyistä näkökulmasta tiettyä tutkimuskysymystä vasten, on usein jo itsessään tulkintaa ja analysointia olemassa olevasta laajemmasta kokoelmasta. Jatkokäytettävänä tällainen tulkintakokonaisuus on avoinna muiden tutkijoiden analysoinnille. Tämä mahdollistaa tutkijoiden itsensä tekemän arkistokokoelmien tulkinnan, analysoinnin ja dokumentoinnin laajemman leviämisen.

Tutkimusaineistojen turvallista pitkäaikaissäilytystä tulisi edesauttaa; kartuttavathan tutkijoiden itsensä koostamat ja keräämät tutkimusaineistot tutkimuksellisesti tärkeää kulttuuriperintövarantoa. Osana tätä prosessia Tietoarkisto toteutti keväällä 2015 kyselyn, jonka avulla kartoitettiin historia-alojen tutkijoiden mielipiteitä arkistoinnista ja jatkokäytöstä.⁵⁴ Sen perusteella suurin osa tutkimusaineistoista jää alkuperäisten tutkijoiden tai tutkimusryhmien haltuun tutkimuksen päätyttyä. Mikäli aineistoja arkistoidaan, tyypillisin arkistointipaikka oli vastaajien mukaan oma yliopisto tai tutkimuslaitos. Toisaalta historia-alojen tutkijoista 77 prosenttia oli ainakin kuullut Tietoarkistosta ja sen palveluita oli valmis käyttämään 62 prosenttia. Vastaajista lisäksi 26

52 Borgman 2009; tekijänoikeudesta myös mm. Parry ja Mauthner 2004, 141–143.

53 Arkistointilupaa pyritään laajentamaan myös yksityisarkistoja koskevaksi vuoden 2016 aikana.

54 Fält ja Okuloff 2016. Kyselyn avulla pyrittiin kartoittamaan lisäksi terveys- ja lääketieteen alojen tutkijoiden tietoutta ja käytäntöjä digitaalisten tutkimusaineistojen käsittelystä ja jatkohyödyntämisestä. Kysely toteutettiin internet-kyselynä ja lähetettiin 1 428 tutkijalle. Vastaajista oli historia-alojen tutkijoita 37.2 % ja terveys- ja lääketieteen tutkijoita 62.8 %.

prosenttia ilmoitti jo käyttäneensä Tietoarkiston palveluita. Kyselyn vastausprosentti oli kuitenkin vain 13,9, joten tämän perusteella pitkälle vietyjä tai syvällisiä päätelmiä on hankala tehdä.⁵⁵ Tulokset ovat kuitenkin suuntaa-antavia.

Lähitulevaisuudessa selvitystyötä vaativat muun muassa historian ja arkeologian alojen digitaaliset paikkatietoaineistot ja niiden asettamat vaatimukset Tietoarkiston arkistointiprosessille sekä teknisille alustoille. Tämän kaltaiset tutkimusaineistot ovat arkistointi- ja jatkokäyttökelpoisia kuten pohjoismainen esimerkki osoittaa: Ruotsin tietoarkistolla (SND) on kattava kokoelma arkeologista GIS-materiaalia.⁵⁶ Myös muunlaisten, digitaalisessa muodossa olevien kuva-aineistojen vastaanottaminen vaatii työtä. Kuvien kohdalla erityisesti tekijänoikeudet saattavat rajoittaa arkistointia, mutta tutkijoiden itsensä tuottamien kuva-aineistojen kohdalla tätä ongelmaa ei yleensä ole. Ylipäätään arkistoitavien aineistotyyppien repertuaarin lisääminen tietoarkistotyössä on merkittäväällä sijalla, jos tutkimusaineistojen arkistoinnista ja jatkokäytöstä halutaan luoda vakiintuneet kansalliset käytännöt myös humanistisille aloille.

Käytännöstä kehitykseen

Tutkimusaineistojen arkistointi ja jatkokäyttö digitaalisen humanismin käytäntönä

Tämän artikkelin loppuosan pohdinta kohdentuu digihumanismin piirissä käydyssä keskustelussa vähemmälle huomiolle jääneeseen kysymykseen tutkimusaineistojen avoimuudesta digitaalisen humanismin kontekstissa. Digitaalinen humanismi mielletään alan toimijoiden mukaan usein eräänlaiseksi katto- tai sateenvarjokäsitteeksi, jonka alle asettuu moninainen määrä erilaisia humanistisiin tieteenaloihin ja teknologisiin

55 Kyselytutkimusten vastausprosentit ovat yleisesti laskeneet sekä kansallisissa että kansainvälisissä tutkimuksissa. Tietoarkiston kyselyn kohdalla matalaa vastausprosenttia voi selittää lisäksi se, että tutkimusaineistojen avaamiseen ja jatkokäyttöön liittyvät asiat ovat verrattain uusia kohdistetuilla tieteenaloilla.

56 "Archaeological GIS material freely available at SND". SND:n [www-sivut](http://www.snd.se).

ratkaisuihin nojaavia toimintoja.⁵⁷ Osa liittää digihumanismin tiukemmin laskennalliseen tai ohjelmoinnilliseen tietoteknologiapainotteisuuteen. Toiset taas liittävät sen monipuolisemmin kaikenlaisiin menetelmiin ja työkaluihin, jotka ovat tekstien ja teknologioiden liitoskohdassa.⁵⁸ Digihumanismi korostaakin voimakkaasti nimenomaan tekstuaalisuutta; tosin kiinnostus muun muassa visuaalisuuteen on lisääntymässä.⁵⁹

Digitaalisen tutkimuksen apulaisprofessori Mark L. Sample on yksinkertaistanut digitaalisen humanismin seuraavasti. Hänen mukaansa se perustuu pohjimmiltaan digitaaliseen objektiin, joko alun perin digitaaliseen tai analogiseen objektiin, joka on jollakin tavalla skannattu, valokuvattu, kartoitettu tai mallinnettu digitaalisessa ympäristössä.⁶⁰

Samplen esittämän laajan määrittelyn mukaisesti lähes jokainen humanistisen alan nykytutkimus voisi jollain tapaa liittyä digitaaliseen humanismiin; onhan digitaalisuus jo suorastaan erottamaton osa tämän päivän tutkimuskäytäntöjä. Osa tutkijoista näkeekin digitaalisuuden ja humanistisen tutkimuksen liittyvän saumattomasti yhteen. Median ja viestinnän tutkijan Dave Parryn mukaan digitaalisuutta ei itse asiassa enää voi erottaa humanismista, vaan se on sulautunut perustavanlaatuisena osana tieteenalan käytäntöihin.⁶¹ Englannin kielen professori Alexander Reid puolestaan viittaa niin sanottuun kolmanteen digihumanismiin, jonka mukaan humanistiset tieteenalat kokonaisuudessaan ovat siirtymässä tekstiparadigmasta digitaaliseen.⁶² Myös digitaalista humanismia tutkineen Lisa Spiron mukaan ala eräällä tapaa uudelleen järjestää humanistista tutkimusta aikana, jolloin ihmiskunnan koko kulttuurinen perintö on liikahtamassa vääjäämättä kohti digitaalista olomuotoa. Se pyrkii työntämään humanistisen tutkimuksen kiivaammin kohti yhteis-

57 Gibbs 2013, 293–294; Gavin ja Smith 2012, 61; Svensson 2012, 36–49; Terras 2013, 263–270.

58 Mm. Ramsay 2011; Ramsay ja Rockwell 2012, 77–81; Haverinen ja Suominen 2015; Gibbs 2013, 293–294; Gavin ja Smith 2012, 61.

59 Svensson 2013, 178; Terras 2006, 236; Alvarado 2012, 52–53.

60 Sample 2012, 189.

61 Parry 2012, 436; myös Drucker 2012, 85–86.

62 Reid 2012, 357.

työtä, avoimuutta ja kokeilua, jotka Spiron mukaan muodostavat digitaalisen humanismin arvopohjan.⁶³

Digitaalisen humanismin digitaalisuus voi kohdentua digitaalisessa muodossa oleviin työkaluihin, menetelmiin, tutkimuskohteisiin, -ympäristöihin tai tutkimusaineistoihin. Tämän artikkelin kiintopisteen mukaisesti tarkastelu keskittyy tutkijoiden itsensä tuottamiin, digitaalisessa muodossa oleviin tutkimusaineistoihin. Digihumanismi on usein riippuvainen digitaalisista aineistoista jopa siinä määrin, että sitä on paikoin syytetty datafetisistisestä lähestymistavasta.⁶⁴ Tutkimusaineistojen avoimuus ei kuitenkaan ole vielä noussut erityisen keskeiseen asemaan digihumanismin kentällä. Avoimuus ja *open access* kyllä kiinnostavat, mutta avoimuudella viitataan tyypillisimmin tutkimusjulkaisujen avoimuuteen tai kulttuuriperintölaitosten avoimiin (ja usein digitoituihin) aineistokokoelmiin. Erilaisten tutkimusaineistojen avoimuus ymmärretään keskeisenä muun muassa tutkimuksen läpinäkyvyyden, uudenlaisten näkökulmien ja tutkimuksen toistettavuuden edistämiseksi, vaikka humanistisessa kontekstissa tutkimusaineistoilla ei tyypillisesti viitata nimenomaisesti tutkijoiden itsensä tuottamiin tai keräämiin tutkimusaineistoihin.⁶⁵ Aineistojen avoimuus mielletäänkin usein yksisuuntaisena käytäntönä; voidakseen luoda ja kehittää digitaalisia käytänteitä tutkijat tarvitsevat pääsyn erilaisiin avoimiin aineistoihin.⁶⁶

Tutkimusaineistojen avoimen saatavuuden vaatimus linkittyy olennaisena osana digihumanismin korostamiin aiheisiin. Digitaalisten tutkimusaineistojen arkistoinnin, mutta myös arkistoitujen tutkimusaineistojen jatkokäytön edistämisen pitäisikin selkeästi kuulua digihumanismin propagoimiin käytäntöihin. Alan tutkijoiden parissa jatkokäyttö – tai täsmällisemmin jakaminen – määrittyy tyypillisesti joko laajemmin tiedon jakamisen tai rajatumminkin työkalujen jakamisen tärkeytenä. Laajaa jakamista tavoittelee Mark L. Sample, jonka mukaan digihumanismin ytimessä on tiedon tuottaminen, esittäminen ja jakaminen

63 Spiro 2012, 21–23.

64 Hall 2013, 788.

65 Tolonen ja Lahti 2015; Hall 2013, 6.

66 Spiro 2012, 24.

mahdollisimman oikeudenmukaisesti ja laajasti. Digitaalisuuteen ei siten liity olennaisena osana vain tiedon tuottaminen vaan sen uudelleen tuottaminen ja lisääntyminen (*reproduction*). Digitaalisen humanismin tulisi Samplen mukaan pyrkiä haastamaan ne tavat, jolla tietoa tuotetaan ja jaetaan. Sample näkee perinteiset humanistiset tutkimuskohteet ja julkaisemismallin tiedon tuottamista sekä jakamista rajoittavina portinvartijoina, ja digitaalisuuden puolestaan vapauttavana ja mahdollistavana toimintana.⁶⁷ Tutkija Charlie Edwards viittaa tekstissään digihumanisti Jeremy Boggsin basaari-ostelun käsitteeseen, jonka mukaan digihumanismin parissa on taitoa poimia ja tuottaa erilaisia avoimuuteen perustuvia työkaluja, koodeja ja sovelluksia, mutta vähemmän kiinnostusta jakaa niitä. Boggsin toteamus liittyy ensisijaisesti koodin jakamiseen ja Edwards rakentaakin jakamiseen perustuvan vaatimukseen lähinnä avoimen lähdekoodin varaan rakentuvaksi jakamiseksi. Edwards kuitenkin nostaa jatkokäytön kaikkien rakennelmien tai systeemien ensisijaiseksi hyveeksi.⁶⁸ Laajasti ajatellen Edwardsin ajatusta voi aivan hyvin soveltaa myös itse digihumanismin tieteenalan ja sen käytänteiden rakentamisen näkökulmasta, kuuluuhan rakentaminen olennaisena osana digitaalisen humanismin keskusteluun ja käytäntöihin.⁶⁹ Jotta digitaalista humanismia voitaisiin kehittää ja rakentaa, tiedon mahdollisimman avoin jakaminen tulisi nostaa alan keskiöön.

Digihumanismi liittyy kiinteästi kulttuurisen aineksen luomiseen, siirtämiseen ja säilyttämiseen. Tällaista kulttuurista aineista ovat tutkijoiden digitaaliset tutkimusaineistot, jotka talletettuina ja jatkokäytettyinä

67 Sample 2013, 256–257.

68 Edwards 2012, 224–225.

69 Digihumanismin pitkään linkittyneen tekemisen korostaminen lienee selkein oppialan suuntauksessa, joka mieltää sen ensisijaisesti rakentamiseksi (*building*). Digihumanismi rakentamisena pohjaa Stephen Ramsayn (kärjistettyyn) ajatukseen digihumanistista koodarina ja ohjelmoijana. Ramsay 2012, 244. Määrittelyissä esiintyy usein myös polarisaatioajattelu ”making things” ja ”theorising things” tekemisen ja teorisoinnin välillä. Theorising things korostaa teoreettisia lähestymistapoja digitaalisessa humanismissa ja tämä lähestymistapa pohtii, voiko esimerkiksi erilaisia digitaalisia objekteja, sovelluksia tai ohjelmistoja tulkita teorioiden. Spiro 2012; Ramsay ja Rockwell 2012, 77–81; Haverinen ja Suominen 2015.

muodostavan merkittävän kulttuuriperintöaineksen. Digitaalisten tutkimusaineistojen arkistoinnin tulisi siten nivoutua olennaisena, jopa paradigmaattisena osana tieteenalan peruskäytäntöihin,⁷⁰ joiden puitteissa metatiedolla rikastetut digitaaliset tutkimusaineistot – sekä analogisesta aineistosta digitoidut että syntymästään asti digitaaliset (*born-digital*) aineistot – linkittyvät digitaaliseen ympäristöön, esimerkiksi data-arkistoon. Kun aineistojen avoimuuteen kiinnitetään huomiota jo tutkimuksen alkuvaiheessa, tutkimusprosessin avaaminen on helpompaa ja luontevampaa tutkimuksen lopussa tai jo sen kuluessa. Tiede- ja tutkimuskentän korostaessa alati laajentuvaa avoimuuspolitiikkaa ei tutkimusaineistojen avaamatta jättäminen voi tässä kontekstissa enää olla vallitseva toimintamalli. Tieteen avoimuutta ylistävä triumfi jää myös helposti irralliseksi ja pahimmillaan vain tiedepoliittisen ilmiömaailman tasolle, mikäli aineistojen avointa saatavuutta ei saada sidottua tiukemmin osaksi tutkijoiden omaa todellisuutta.

Quo vadis eli mihin avoimet tutkimusaineistot ja humanistinen tutkimus voisivat olla matkalla

Humanistisen tutkimuksen ydin, monipuoliset aineistot, ovat tuloksellisuudessa ja arvostuksessa jääneet yllättävän toispuoliseen asemaan verrattuna esimerkiksi tutkimusjulkaisuihin. Tutkimuksessa hyödyllisimpänä yksikkönä pidetään edelleen julkaistua tutkimustulosta, ei julkaistua tai avointa tutkimusaineistoa tai tutkimusprosessia. Digitaalisen humanismin tutkijoiden Mikko Tolosen ja Leo Lahden arvion mukaan viimeksi mainitut kuitenkin saattaisivat sujuvoittaa ja edistää tutkimuksen, tutkimusetiikan ja -metodiikan kehittymistä sekä tulosten laajempaa liikuvuutta.⁷¹ Humanististen alojen tutkimusaineistoihin kohdistuu enenevässä määrin kiinnostusta ja avoimuuden etuja propagoidaan monelta suunnalta. Avoimuuden korostamisessa pyritään tyypillisesti nostamaan esille konkreettisia hyötyjä sekä itse tutkijalle että laajemmin tieteen ja

70 Edwards 2012, 219.

71 Tolonen ja Lahti 2015.

tutkimuksen kehittyemiselle.⁷² Tutkimusaineistojen avaaminen meritoi tutkijaa, sillä jatkokäyttäjien on aina hyvän tieteellisen käytännön mukaisesti viitattava käyttämiinsä aineistoihin. Mikäli tutkimusaineisto on julkaistu ja sille on annettu lähdeviite, noudattaa viittaaminen tyypillisesti aineiston julkaisijan merkintätapaa.⁷³ Avoimuus saattaa myös useassa tapauksessa parantaa tutkijan ja tutkimuksen näkyvyyttä. Laajemmin avoimuus ja tutkimusaineistojen jakaminen auttaa todentamaan tutkimustuloksia, kysymään uudenlaisia kysymyksiä jo olemassa olevien tutkimusaineistojen pohjalta, levittää ja laajentaa julkisesti rahoitettavien tutkimusten tuloksia, ja täten edistää sekä tutkimusta että innovaatioita. Tutkimusaineistojen, mutta myös analyysien ja menetelmien avaaminen auttaa lisäksi raottamaan verhoa tutkimusprosessin edestä ja lisäämään siten tutkimuksen ja tieteen tekemisen avoimuutta ja läpinäkyvyyttä.⁷⁴ Arkistointi ja jatkokäyttö tehostavat lisäksi aineistojen tutkimuskäyttöä, kun vaivalla kerätyt aineistot eivät jää yhden ainoan tutkimuksen pohja-aineistoksi.

Humanististen alojen teorian ja metodiikan kehittyminen edellä kuvattuun suuntaan edellyttää digitaalisia resursseja, työkaluja, palveluita sekä säilytyspaikkoja.⁷⁵ Tähän listaan voinee lisätä myös kaikki sellaiset toimenpiteet, jotka lisäävät ja tehostavat tutkimusaineistojen avaamista ja jatkokäyttöä. Tutkimusaineistojen käytön edistäminen ja tehostuminen edellyttää kuitenkin sekä laadukkaita tutkimusaineistoja että toimivia infrastruktuureja, jotka huolehtivat aineistojen käyttökelpoisuu-

72 Aineistojen avaamiseen liittyvät mahdolliset haitat nousevat keskustelussa esiin erityisesti silloin, kun kartoitetaan tutkijoiden mielipiteitä tutkimusaineistojen avaamisesta. Ks. esimerkiksi Kuula ja Borg 2008; Van den Eynden ja Bishop 2014; Fält ja Okuloff 2016.

73 Mikäli aineisto on julkaistu ja sille on annettu lähdeviite, noudattaa viittaaminen tyypillisesti aineiston julkaisijan merkintätapaa. Tietoarkisto esimerkiksi tarjoaa malliviittaustiedon, jota on mahdollista muokata julkaisun tai tieteenalan käytäntöjen mukaisesti. ”Aineistoon viittaaminen”. Tietoarkiston [www-sivut](#); ATT-hankkeen ohjeistus: ”Tunnisteet ja viittaaminen”. ATT-hankkeen [www-sivut](#); Digital Curation Centerin ohjeistus: ”How to Cite Datasets and Link to Publications. DCC:n [www-sivut](#); Borg 2015.

74 Borgman 2012; Sinnemäki ja Tolonen 2015, 32.

75 Friedlander 2009; Borgman 2009.

desta ja pitkäaikaissäilytyksestä. Elizabeth Borgmanin mukaan tällaista huoltamiseen, jatkokäyttöön ja jakamiseen sopivaa laadukasta aineistoa voidaan saada vain silloin kun tutkimuksessa käytettävä pohja-aineisto tallennetaan mahdollisimman virheettömästi jo tutkimuksen elinkaaren alkuvaiheessa.⁷⁶

Tutkimusaineistojen jakamiseen – ja samalla myös aikaisemmin mainittuun rakentamisen käsitykseen – liittyy olennaisena osana kysymys infrastruktuureista. Olivatpa aineistot miten merkittäviä tai urauurtavia tahansa, säilytyspaikkojen tai sopivien infrastruktuurien puute heikentää tulosten ja menetelmien leviämistä. Digitaaliseen säilyttämiseen erikoistuneen Arjun Sabwarhalin mukaan ilman käyttövarmoja aineistojen säilytyspaikkoja digitaalista humanismia ei voi pitää luotettavana tieteenalana.⁷⁷ Luotettavat säilytyspaikat eivät vain säilytä ja suojaa erilaisia digitaalisia aineistoja, vaan takaavat niiden käyttökelpoisuuden hoitamalla aineistoja oikein. Avoimet säilytyspaikat haastavat lisäksi tieteenalojen rajoja silloin kun ne arkistovat monipuolisesti eri alojen tutkimusaineistoja, ja voivat siten edistää merkittävällä tavalla tieteidenvälisyyttä. Digitaalisia kirjastoja ja arkistoja tutkinut Mark Hedges on ehdottanut, että digitaaliset säilytyspaikat ovat tulevaisuudessa avainasemassa kun humanistilla aloilla pyritään luomaan toimivia e-infrastruktuureja. Hedgesin mukaan säilytyspaikat voisivat tarjota ensinnäkin keinoja tehdä akateemista tukimusta saataville ilman, että tutkimus olisi ensin julkaistu esimerkiksi jossain tutkimusjulkaisussa. Toisekseen säilytyspaikkojen avulla olisi mahdollista julkistaa sellaista aineistoa, joka ei taivu perinteisen julkaisemisen piiriin.⁷⁸

Humanististen alojen ja digitaalisen humanismin näkökulmasta katsottuna tällaisten omien, tieteenalakohtaisten data-infrastruktuurien puute on yksi keskeinen ongelma digitaalisten tutkimusaineistojen käytön tehostamisessa ja edistämisessä.⁷⁹ Kulttuuriperintölaitokset ja muut humanistista aineistoa hallinnoivat tahot toki tuovat oman lisänsä

⁷⁶ Borgman 2009.

⁷⁷ Sabharwal 2015, 25.

⁷⁸ Hedges 2009.

⁷⁹ Borgman 2015, 171, 175; Marker 2015, 29.

tutkimusinfrastruktuurien kentälle, mutta niiden toimenpiteet aineistojen avoimuuden edistämiseksi saattavat vaihdella. Mikko Tolonen ja Leo Lahti ovat kritisoineet merkittäviä aineistoja ylläpitäviä laitoksia (määrittelemättä näitä laitoksia tarkemmin) siitä, että ne vaikuttavat joskus välinpitämättömiltä edistämään aineistojen avointa tutkimuskäyttöä.⁸⁰ Digitaalisia tutkimusaineistoja arkistoivilla, säilyttävillä ja eteenpäin luovuttavilla tahoilla tavoite on kuitenkin täysin päinvastainen ja kentälle on nousemassa myös humanistisia aloja palvelevia toimijoita. Tutkimusta tukevien laitosten kuten CSC:n, FIN-CLARINin ja Tietoarkiston rooli yhteistyön ja monitieteisen tutkimuksen kehittämisessä on huomioitu myös historian tutkijoiden parissa.⁸¹ Alalla toimivien tahojen välinen yhteistyö onkin elintärkeää.

Toimivaan infrastruktuuriin liittyy olennaisesti sen kyky huoltaa aineistoja oikeaoppisesti ja ammattitaitoisesti. Digitaalisen humanismin piirissä tuotetaan paljon humanistisen tutkimuksen kontekstissa uudentyyppistä, digitaalisessa muodossa olevaa aineistoa, joka vaatii erityistä huoltamista sekä suunnitelmallista strategiaa, jonka puitteissa aineiston huoltaminen on mahdollista toteuttaa. Huoltamisella tarkoitetaan yleisesti aineistojen keräämistä, järjestämistä, vahvistamista ja säilyttämistä niiden koko elinkaaren ajan. Varsinkin digitaalisessa muodossa oleva tutkimusaineisto on haavoittuvaista, joten se edellyttää huolellista hoitoa ja hallintaa pysyäkseen käyttökelpoisena pitkään. Nopeasti vanheneva teknologia voi osaltaan edesauttaa aineistojen tuhoutumista, mikäli niiden säilytys jää vain yhden tietoteknisen formaatin varaan. Hoidettu aineisto on lisäksi aina arvokkaampi kuin hoitamaton; esimerkiksi metatietojen käsittelyyn tulisi luoda standardisoituja käytäntöjä, jotta aineistojen koneellinen käsittely ja yhteisten luetteloiden laatiminen helpottuu.⁸² Standardisoinnilla voidaan välttää myös päällekkäisen työn tekemistä.

Arkistoja ja digitaalista dataa tutkineen Alex H. Poolen mukaan digitaalinen humanismi ei itse asiassa voi edes kukoistaa ilman datan

80 Tolonen ja Lahti 2015.

81 Sinnemäki ja Tolonen 2015, 29–32; Parland-von Essen 2016.

82 Standardisoituja formaatteja ovat muun muassa digitaalisten aineistojen kuvailuun kehitetty Dublin Core ja aineistojen kuvailutyökalu Data Documentation Initiative (DDI).

ja tutkimusaineistojen huoltamista. Hän hahmottelee ideaalitilan, jossa digitaalisen humanismin ja digitaalisten aineistojen huoltamisen välillä vallitsee symbioottinen ja synerginen kumppanuus, jota itsessään toki tulee myös huoltaa. Poole korostaa eri tahojen – sekä kuratoinnin ammattilaisten että aineistoja tuottavien – yhteistyön merkitystä tarvittavien aineistopalvelujen ja -infrastruktuurien luomiseksi, jotta vaadittavat käytännöt on mahdollista luoda.⁸³ Arjun Sabharwal lähestyy aihetta lisäksi abstraktimmalla tasolla hahmotellessaan digitaalisen huoltamisen laajemman kontekstin, jossa aineistojen hoitaminen koostuu linkittyvästi sekä teoriasta että käytännöstä. Sen puitteissa on mahdollista tuottaa erilaisia digihumanismin projekteja. Sabwarhalin ajatuksen mukaisesti aineistojen huoltaminen toimii eräänlaisena kriittisenä kontekstina, jonka sisällä humanistista dataa ja aineistoja voi ja tulisi analysoida sekä tulkita.⁸⁴ Tutkimusaineistojen huoltamista ei siten voi ajatella puhtaasti mekanistisena, infrastruktuurisidonnaisena velvoitteena vaan huomattavasti laajempana kontekstointina.

Palveluinfrastruktuureilta vaaditaan herkkyyttä huomioimaan eri tieteenalojen aineistojen tarpeet. Valmiiseen muottiin pakottamisen tai vain valmiiden työkalujen käytön sijasta pitäisi olla joustava ja herkkä yksittäisen aineiston tarpeille ja mahdollisuuksille. Erilaisten digitaaliseen säilyttämiseen keskittyneiden organisaatioiden rooli niin toimivan infrastruktuurin kuin aineistojen huoltamiseen liittyvien toimien luomiseksi ja edistämiseksi on siten merkittävä.⁸⁵ Digihumanisti Trevor Muñozin mukaan aineistojen huoltaminen pitäisi ymmärtää julkaisemiseen vertautuvana toimintana. Erilaisten avointen aineistojen ja datajoukkojen julkaiseminen sopivilla alustoilla muuntaa tutkimusaineiston luonnetta samanlaisen ”lopputuotteen” kaltaiseksi kuin millaiseksi tutkimusjulkaisu mielletään. Tällaisen aineiston huoltaminen vaatii lisäksi aikaa, vaivaa ja rahaa, jotta lopputulos on laadukas.⁸⁶ Myös pitkäaikainen strategia huoltamisen suhteen on oleellinen, jotta aineistot eivät rapistu.

83 Poole 2013; Kornbluh 2008.

84 Sabharwal 2015, 25.

85 Choudhury 2010, 195; myös mm. Friedlander 2009, 1–15.

86 Muñoz 2013; Choudhury ym. 2009, 479–481.

Pitkäaikaissäilytykseen erikoistuneilta muistiorganisaatioilta sekä data-arkistoilta tällainen strategia pääsääntöisesti löytyy. Aineistojen huoltamisen ja julkaisemisen välisen yhteyden korostaminen voi toimia strategisena linjanvetona, jonka avulla ajatusta akateemisen tutkimuksen lopputuotteesta muutetaan.⁸⁷ Muñoz korostaa artikkelissa nimenomaisesti kirjastojen merkitystä tutkimusaineistojen julkaisemisessa yhdessä tutkijan kanssa, mutta data-arkistojen rooli tässä kehityksessä voisi nousta samankaltaiseen asemaan.

Tutkimusaineistojen korkeatasoinen huoltaminen vaikuttaa sekä yksittäisiin aineistoihin ja aineistokokoelmiin, mutta myös laajemmin tieteenalan kehitykseen. Mediatutkimuksen professori Kathleen Fitzpatrick on kiinnostavalla tavalla nostanut esiin mahdollisen muutoksen humanistisen tutkimuksen eetoksessa. Hänen mukaansa akateemisessa tutkimuksessa pitäisi ylipäätään luopua ajatuksesta, että vain ainutlaatuisuus on tutkimuksellisesti merkittävää. Fitzpatrick ehdottaa, että tutkijoiden tärkein työ saattaakin tulevaisuudessa muistuttaa editoinnin tai kuratoinnin käytäntöjä, jotka korostavat, sekoittavat ja tuovat yhteen merkittäviä ideoita. Hän korostaa siten uudenlaista tekijyyttä, jonka ytimessä on prosessikeskeisyys, yhteistyö ja ”remiksaus”, toisin sanoen jo olemassa olevien tekstien uudelleen yhdisteleminen.⁸⁸ Tutkijoiden pitäisi siten Fitzpatrickin mukaan keskittyä kehittämään alan instituutioita sekä edistämään uudenlaisia akateemisen ajattelujen tapoja, joita digitaalisuus ja sen tulkinnat mahdollistavat. Ongelmana ei hänen mukaansa siis ole digihumanismin teknologialuonne vaan sosiaalinen luonne.⁸⁹ Myös Suomessa alan tutkijat ovat korostaneet yhteistyön voimaa.⁹⁰ Open Knowledge Foundationin antamassa Tiedon saatavuus -työryhmän lausunnossa todetaan, että varsinkin humanistilla aloilla kaikki kulttuuri-perintö on potentiaalista tutkimusdataa. Tällöin muun muassa aineistojen linkittäminen ja yhteistyö tietoarkistojen välillä korostuu.⁹¹

87 Siddiqui 2015.

88 Fitzpatrick 2011, 12–13.

89 Fitzpatrick 2010.

90 Haverinen ja Suominen 2015; Sinnemäki ja Tolonen 2015, 31–32.

91 *Lausunto* 2013.

Erilaisten aineistojen toisiinsa linkittäminen onkin nousemassa keskeiseen asemaan, kun pohditaan humanistisen tutkimuksen tulevaisuudennäkymiä. Linkittämisen tarkoituksena on yhdistää eri lähteissä sijaitsevia aineistoja. Käytännössä aineistossa esiintyvät resurssit (esimerkiksi henkilö, paikka, tapahtuma, asiaana) on linkitetty toisessa aineistossa oleviin resursseihin ja tämän yhteyden merkitys on koneluettavassa muodossa. Linkityksessä on lisäksi olennaista yksilöivä tunnistetieto, jolla aineistoon voidaan viitata.⁹² Käytännössä kansallisella tasolla ajatus aineistojen linkittämisestä koskee ensisijaisesti toimia, joilla eri kulttuuriperintölaitosten (kirjastot, museot, arkistot) aineistojen käyttöä voisi tehostaa luomalla käyttökelpoisia kytköksiä aineistojen ja aineistokokonaisuuksien välille.⁹³ Digitaalisia tutkimusaineistoja voisi linkittää esimerkiksi jo olemassa oleviin relevantteihin kulttuuriperintökokoelmiin. Tämä tosin edellyttää, että linkitettävien aineistojen sisällöt tai temaattiset suhteet ovat toisiinsa liittyviä tai toisiaan täydentäviä. Tällaisissa tapauksissa linkittämisen avulla on mahdollista avata sekä täydentää erilaisten kokoelmien ja aineistojen välisiä suhteita. Näin myös erilaiset aineistoinfrastruktuurit voisivat linkittyä tiiviimmin toisiinsa. Jo aikaisemmin on mainittu, kuinka esimerkiksi tutkijan itsensä ottamat digitaaliset valokuvat Arkistolaitoksen asiakirja-aineistosta voivat muodostaa esimerkiksi data-arkistossa säilytettävänä itsenäisen, arkistoitavan ja linkitettävän tutkimusaineistokokelman. Digitaalisia tutkimusaineistoja voisi myös linkittää keskenään tai siten, että tutkimusaineistot olisivat myös muun kuin akateemisen yhteisön saatavilla.⁹⁴ Koska humanististen alojen tutkimusaineisto on tyypillisesti useista eri lähteistä ja arkistoista kerätty, niin tulevaisuudessa voisi olla mahdollista arkistoida myös ”virtuaalisia tutkimusaineistoja”, toisin sanoen aineistoja tai niiden osia, jotka ovat linkitettyjä erilaisiin aineistoihin säilyttäviin laitoksiin tai infrastruktuureihin. Virtuaaliaineisto olisi siten kooste tutkimuksessa käytetyistä aineistoista ja siihen linkitetyistä kokonaisuuksista. Virtuaaliaineiston tallentaminen ja jatkokäyttö edellyttää kuitenkin

92 Frosterus 2013, 26–27.

93 Esim. Kornbluh 2008.

94 Kolowich 2011.

pysyviä tunnisteita sekä mahdollisesti jonkinlaista virallista tahoja, jonka vastuulla on aineistojen ylläpito ja kuvailu.⁹⁵ Linkittäminen helpottaa myös tutkijan työtä, kun eri laitoksiin ja palveluinfrastruktuureihin säilötyt aineistot on mahdollista tavoittaa kootusti linkitysten kautta.⁹⁶

Lopuksi

Digitaalisuus muokkaa hitaasti tutkimuksen tekemisen tapoja sekä tutkimusympäristöjä, kun erilaiset aineistokokoelmat ja aineistoja hallinnoivat tahot on mahdollista tavoittaa digitaalisten infrastruktuurien avulla. Vaikka jalkautuminen arkistoon on historia-aloilla edelleen tutkimuksen perustyötä ja monelle tutkijalle hyvin mieluisaa, lisääntyvät digitaaliset kokoelmat aiheuttavat muutoksia työskentelytavoissa. Moni tutkija käyttäisi mielellään digitaalisia aineistoja mukavasti omalla kotisohvallaan istuen jonkin käyttöliittymän tai etäkäyttöpalvelun kautta. Asiaan liittyy myös puhtaasti käytännöllisempi näkökulma. Monelle tutkijalle relevantti arkisto saattaa sijaita oman asuinkunnan ulkopuolella. Tällaisissa tapauksissa tutkijat voivat hyödyntää esimerkiksi Tietoarkistoon arkistoitua, alun perin Arkistolaitoksen aineistosta koottua tiiviimpää osa-aineistoa arvioidessaan laajemman kokoelman tarpeellisuutta omalle tutkimukselleen. Tämä voi säästää sekä aikaa että rahaa kun alkuperäisen aineiston äärelle ei kaikissa tapauksissa tarvitse matkata.

Teknologiahistorian tutkija professori Cathy N. Davidson on esittänyt, että arkistot saattavat tulevaisuudessa toimia pelkkää aineistovarantoa laajempänä entiteettinä. Hänen mukaansa uuden sukupolven humanistinen arkisto voisi olla sellainen, joka on avoin sen käyttäjien aktiivisemmalle panokselle kuten digitaalisen sisällön tuottamiselle ja kustomoinnille.⁹⁷ Davidson ulottaa visiointinsa muualle kuin data-arkistomaailmaan, mutta tieteenalakehityksen kannalta olisi mielekästä pohtia, mikä tutkimusaineistoja säilyttävien, huoltavien ja jakavien

95 Pysyvä tunniste on koneluettava sekä ainutkertainen, yksilöityyn resurssiin johtava linkki. Hakala 2010.

96 Sabharwal 2015, 23–24.

97 Davidson 2012, 480.

tahojen rooli tulevaisuudessa on, ja kuinka ylipäätään olisi mahdollista luoda toimiva ja mahdollisimman kattava data-infrastruktuuri humanistisille aloille.

Vaikka tutkimuksen tekemisen tavat, tiedon esittäminen, jakaminen ja keskustelu muuttuvatkin digitaalisuuden myötä, digitaalisuus itsessään ei lopulta olennaisesti muuttane historian tutkimuksen perimmäisiä päämääriä.⁹⁸ Kyseessä on edelleen humanistisen ja historian tutkimuksen ydintehtävän eli kulttuurisen ymmärryksen välittäminen, joka tapahtuu enenevässä määrin digitaalisen teknologian avustuksella.⁹⁹

Lähteet

- ”Aineistoon viittaaminen”. Tietoarkiston www-sivut. <<http://www.fsd.uta.fi/fi/aineistot/jatkokaytto/viittaaminen.html>>. 9.9.2016.
- Alvarado, Rafael C: The Digital Humanities Situation. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 50–55.
- ”Apurahan saajalle”. Koneen Säätiön www-sivut. <<http://www.koneensaatio.fi/apurahat/apurahan-saajalle/>>. 9.9.2016.
- ”Archaeological GIS material freely available at SND”. SND:n www-sivut. <<http://snd.gu.se/en/search-and-order-data/gis-data>>; <<http://snd.gu.se/en/theme/history-and-archaeology>>.
- Arzberger, Peter; Schroeder, Peter; Beaulieu, Anne; Bowker, Geoff; Casey, Kathleen; Laaksonen, Leif; Moorman, David; Uhlig, Paul; Wouters, Paul: *An International Framework to Promote Access to Data*. Science 303(5665) 2004, 1777–1778, <<http://science.sciencemag.org/content/303/5665/1777>>. 13.6.2016.
- ATT-hankkeen www-sivut. <<http://avointiede.fi/home>>. 9.9.2016.
- ”Avoin tiede”. Suomen Akatemian www-sivut. <<http://www.aka.fi/fi/rahoitus/hyva-tutkimustapa/tieteen-avoimuus/>>. 9.9.2016.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Open Access Max-Planck-Gesellschaft 2003. <http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf>. 16.12.2015.

98 Sample 2013, 256.

99 Berry 2012, 2.

- Berry, David M.: Understanding the Digital Humanities. Teoksessa *Understanding Digital Humanities*. Toim. David M. Berry. Basingstoke, Palgrave Macmillan 2012, 1–20.
- Borg, Sami, Kuula, Arja: *Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa*. Yhteiskuntatieteellisen tietoarkiston julkaisuja 6, Tampere 2007.
- Borg, Sami: Kansainvälinen raportti IFDOlta. Rahoittajilta vahvaa tukea aineistojen avoimelle saatavuudelle. *Tietoarkisto* 40(2) 2014, 16–18. <<http://www.fsd.uta.fi/lehti/fi/40/tietoarkisto2014.pdf>>.
- Borg, Sami: Miten tutkimusdataan viitataan? *Portti-verkkolehti* 24.02.2015. <<http://portti.avointiede.fi/tutkimusdata/miten-tutkimusdataan-viitataan>>. 26.5.2016.
- Borgman, Christine L.: The Digital Future is Now: A Call to Action for the Humanities. *Digital Humanities Quarterly* 3(4) 2009, <<http://digitalhumanities.org:8081/dhq/vol/3/4/000077/000077.html>>.
- Borgman, Christine L.: The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63(6) 2012, 1059–1078.
- Borgman, Christine L.: *Big Data, Little Data, no Data. Scholarship in the Networked World*. The MIT Press, Cambridge MA, London 2015.
- Broom, Alex; Cheshire, Lynda; Emmison, Michael: Qualitative Researchers' Understandings of Their Practice and the Implications for Data Archiving and Sharing. *Sociology* 43(6) 2009, 1163–1180.
- Caplan, Priscilla: *Metadata Fundamentals for All Librarians*. American Library Association, Chicago 2003.
- Ceci, Stephen J.: Scientists' Attitudes Towards Data Sharing. *Science, Technology, & Human Values* 13(1/2) 1988, 45–52. <http://www.jstor.org/stable/pdf/690052.pdf?_=1465817906977>.
- Choudury, Sayeed; Furlough, Mike; Ray, Joyce: *Digital Curation and E-Publishing: Libraries Make the Connection*. Proceedings of the Charleston Library Conference. Purdue University, Purdue E-Pubs 2009.
- Choudhury, Sayeed: Data Curation. An Ecological Perspective. *College & Research Libraries News* 71(4) 2010. <<https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/34025/Data%20curation.pdf>>.
- Craig, James R.; Reese, Sandra C.: Retention of Raw Data: A problem revisited. *American Psychologist* 28(8) 1973, 723. <<http://dx.doi.org/10.1037/h0035667>>.
- Costello, Mark. J.; Berghe, Edward Vanden: Ocean Biodiversity Informatics:

- A New Era in Marine Biology Research and Management. *Marine Ecology Progress Series* 316(1) 2006, <http://www.int-res.com/articles/meps_oa/m316p203.pdf>.
- Costello, Mark J.: Motivating Online Publication of Data. *BioScience* 59(5) 2009, 418–427, <<http://dx.doi.org/10.1525/bio.2009.59.5.9>>.
- Cultural Heritage Information: Access and Management*. Toim. Ian Ruthven ja G. G. Chowdhury. Facet Publishing, London 2015.
- Davidson, Kathy N.: Humanities 2.0: Promise, Perils, Predictions. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 476–489.
- Drucker, Johanna: Humanities Approaches to Graphical Display. *Digital Humanities Quarterly* 5(1) 2011, <<http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>>.
- Drucker, Johanna: Humanistic Theory and Digital Scholarship. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 85–95.
- Edwards, Charlie: The Digital Humanities and Its Users. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 202–232.
- Fienberg, Stephen E.: Sharing Statistical Data in the Biomedical and Health Sciences: Ethical, Institutional, Legal, and Professional Dimensions. *Annual Review of Public Health* 15(1) 1994, 1–18.
- Fitzpatrick, Kathleen: Reporting from the Digital Humanities 2010 Conference. *The Chronicle of Higher Education*, 13.7.2010. <<http://chronicle.com/blogs/profhacker/reporting-from-the-digital-humanities-2010-conference/25473>>. 14.1.2016.
- Fitzpatrick, Kathleen: *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York, New York University Press 2011.
- Friedlander, Amy: Asking Questions and Building a Research Agenda for Digital Scholarship. Teoksessa *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*. Washington, Council on Library and Information Resources No. 145 2009, 1–15. <<http://www.clir.org/pubs/resources/promoting-digital-scholarship-ii-clir-neh/index.html/friedlander.pdf>>.
- Frosterus, Matias: Avoin linkitetty data: mitä se on ja miksi? *Kansalliskirjasto* 3 2013, 26–28.
- Fält, Katja; Okuloff, Annaleena: *Avoin tiede ja tutkimusaineistot –kysely 2015. Raportti terveystieteiden ja humanististen alojen tutkijoille suunnatun kyselyn tuloksista*. Yhteiskuntatieteellisen tietoarkiston julkaisuja 12, Tampere 2016.

- Gavin, Michael; Smith, Kathleen Marie: An Interview with Brett Bobley. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 61–67.
- Gibbs, Fred: Digital Humanities Definitions by Type. Teoksessa *Defining Digital Humanities: A Reader*. Toim. Melissa Terras, Julianne Nyhan ja Edward Vanhoutte. Ashgate, Farnham and Burlington 2013, 289–297.
- Gray, Emily; Jennings, Will; Farrall, Stephen; Hay, Colin: Small Big Data: Using multiple data-sets to explore unfolding social and economic change. *Big Data & Society* 1 2015, 1–6. DOI: 10.1177/2053951715589418.
- Greenberg, Jane: Metadata and the World Wide Web. Teoksessa *Encyclopedia of Library and Information Science*. Second Edition, Volume 3. Toim. Miriam A. Drake. Marcel Dekker Inc. New York, Basel 2003, 1876–1888.
- Greenberg, Jane: Understanding Metadata and Metadata Schemes. Teoksessa *Metadata: A Cataloguer's Primer*. Toim. Richard P. Smiraglia. Routledge, New York and London 2005, 17–36.
- Hall, Gary: Towards a Post-Digital Humanities: Cultural Analytics and the Computational Turn to Data-Driven Scholarship. *American Literature* 85(4) 2013, 781–809. <<http://dx.doi.org/10.1215/00029831-2367337>>.
- Hakala, Juha: Persistent Identifiers: an Overview. *Technology Watch Report (TWR). Standards in Metadata and Interoperability*, <<http://www.metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/>>. 26.5.2016.
- Handbook of Metadata, Semantics and Ontologies*. Toim. Miguel-Angel Sicilia. World Scientific Pub. Co, Singapore 2014.
- Haverinen, Anna; Suominen, Jaakko: Koodaamisen ja kirjoittamisen vuoropuhelu? Mitä on digitaalinen humanistinen tutkimus. *Ennen ja Nyt* 2/2015. <http://www.ennenjanyt.net/2015/02/koodaamisen-ja-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/#identifier_50_1502>.
- Hedges, Mark: Grid-Enabling Humanities Datasets. *Digital Humanities Quarterly* 3(4) 2009. <<http://digitalhumanities.org:8081/dhq/vol/3/4/000078/000078.html>>.
- Hudgins, Jean; Agnew, Grace; Brown, Elizabeth: Getting Mileage Out of Metadata: Applications for the Library. ALA Books, Chicago 1999.
- Introduction to Metadata: Pathways to Digital Information*. Toim. Murtha Baca. Getty Information Institute, Los Angeles 1998.
- Kolowich, Steve: The Promise of Digital Humanities. *Inside Higher Ed*. <https://www.insidehighered.com/news/2011/09/28/national_endowment_for_the_humanities_celebrates_digital_humanities_projects>. 13.1.2016.

- Kornbluh, Mark Lawrence: From Digital Repositories to Information Habitats: H-Net, the Quilt Index, Cyber Infrastructure, and Digital Humanities. *First Monday* 13(8) 2008. <<http://firstmonday.org/ojs/index.php/fm/rt/prINTER-Friendly/2230/2019>>.
- Kramer, Michael J.: Going Meta on Metadata. *Journal of Digital Humanities* 3(2) 2014. <<http://journalofdigitalhumanities.org/3-2/going-meta-on-metadata/>>.
- Kvalheim, Vigdis; Kvamme, Trond: *Policies for Sharing Research Data in Social Sciences and Humanities. A Survey about Research Funders' Data Policies*. International Federation of Data Organizations for Social Science 2014. <<http://www.ada.edu.au/documents/ifdo-report-on-policies-for-data-sharing>>.
- Kuula, Arja; Borg, Sami: *Open Access to and Reuse of Research Data: The State of the Art in Finland*. Finnish Social Science Data Archive 7, Tampere 2008.
- Lausunto Tiedon saatavuus -työryhmän raporttiin. Open Knowledge Foundation 2013. <<https://avointiede.fi/documents/10864/37972/Open+Knowledge+Foundation+lausunto.pdf/ffco6fa-614b-4396-b237-2fcb9a44538f>>. 15.1.2016.
- Marker, Hans-Jørgen: Humanistinen tutkimus kaipaa omia data-infrastruktuureja. *Tietoarkisto* 43(3) 2015, 29.
- McCain, Katherine: Mandating Sharing: Journal Policies in the Natural Sciences. *Science Communication* 16(4) 1995, 403-431.
- Metadata Best Practices and Guidelines. Current Implementations and Future Trends*. Toim. Jung-ran Park. Routledge, London and New York 2011.
- Metadata in Practice*. Toim. Diane I. Hillmann ja Elaine L. Westbrook. American Library Association, Chicago 2004.
- "Mitä hakemus sisältää". Alkoholitutkimussäätiön www-sivut. <<http://www.alkoholitutkimussaatio.fi/alkoholitutkimussaatio-apurahat/alkoholitutkimus-ja-rahapelihaitta-apurahan-hakeminen/mita-hakemus-sisaltaa/>>. 9.9.2016.
- Muñoz, Trevor: Data Curation as Publishing for the Digital Humanities. *Journal of Digital Humanities* 2(3) 2013. <<http://journalofdigitalhumanities.org/2-3/data-curation-as-publishing-for-the-digital-humanities/>>.
- OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD Publications, Paris 2007. <<http://www.oecd.org/sti/scitech/38500813.pdf>>.
- Owens, Trevor: Defining Data for Humanists: Text, Artifact, Information or Evidence? *Journal of Digital Humanities* 1(1) 2011. <<http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/>>. 30.11.2015.

- Palmer, Michael: *Data is the New Oil*. Blogikirjoitus 3.11.2006. <http://ana.blogs.com/maestros/2006/11/data_is_the_new.html>. 20.11.2015.
- Parland-von Essen, Jessica: Miten humanistien aineistot avautuvat? *Portti-verkkolehhti* 01.12.2015. <<http://portti.avointiede.fi/tutkimusdata/miten-humanistien-aineistot-avautuvat>>. 17.12.2015.
- Parland-von Essen, Jessica: *Tankar kring den snabba utvecklingen i Finland*. Blogikirjoitus 5.1.2016, <<http://digihist.se/>>. 14.1.2016.
- Parry, Dave: The Digital Humanities or a Digital Humanism. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 249–437.
- Parry, Odette; Mauthner, Natasha S.: Whose Data Are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology* 38(2) 2004, 139–152.
- Piowar, Heather; Becich, Michael; Bilofsky, Howard; Crowley, Rebecca: Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers. *PLoS Medicine* 5(9) 2008. <<http://dx.doi.org/10.1371/journal.pmed.0050183>>.
- Piowar, Heather; Chapman, Wendy W.: Public Sharing of Research Datasets: a Pilot Study of Associations. *Journal of Informetrics* 4(2) 2010, 148–156. <<http://www.sciencedirect.com/science/article/pii/S1751157709000881>>.
- Poikola, Antti; Kola Petri; Hintikka Kari A.: Julkinen data. Johdatus tietovarantojen avaamiseen. Liikenne- ja viestintäministeriö, Helsinki 2010, <http://www.lvm.fi/c/document_library/get_file?folderId=964902&name=DLFE-10617.pdf&title=Julkinen%20data>.
- Poole, Alex H.: Now is the Future Now? The Urgency of Digital Curation in the Digital Humanities. *Digital Humanities Quarterly* 7(2) 2013. <<http://digital-humanities.org:8081/dhq/vol/7/2/000163/000163.html>>.
- Preparing Data for Sharing: Guide to Social Science Data Archiving*. DANS Data Guide 8. Pallas Publications, Amsterdam 2010.
- Ramsay, Stephen: On Building. Online-kirjoitus 11.1.2011. <<http://stephen-ramsay.us/text/2011/01/11/on-building/>>. 17.11.2015.
- Ramsay, Stephen; Rockwell, Geoffrey: Developing Things: Notes toward an Epistemology of Building in the Digital Humanities. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 75–85.
- ”Recommended formats”. UKDA:n [www-sivut](http://www.ukdataservice.ac.uk/manage-data/format/recommended-formats). <<https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>> 9.9.2016.
- Reid, Alexander: Graduate Education and the Ethics of the Digital Humanities.

- Toeksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 350–367.
- Saarela, Janne: *The Role of Metadata in Electronic Publishing*. Finnish Academy of Technology, Espoo 1999.
- Sabharwal, Arjun: *Digital Curation in the Digital Humanities. Preserving and Promoting Archival and Special Collections*. Chandos Publishing, Waltham and Kindlington 2015.
- Sample, Mark. L.: Unseen and Unremarked On: Don DeLillo and the Failure of the Digital Humanities. Toeksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 187–201.
- Sample, Mark: The Digital Humanities is not about Building, it's about Sharing. Toeksessa *Defining Digital Humanities: A Reader*. Toim. Melissa Terras, Julianne Nyhan ja Edward Vanhoutte. Ashgate, Farnham and Burlington 2013, 255–257.
- Schöch, Christof: Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities* 2 (3) 2013, <<http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>>.
- Sharing Research Data*. Toim. Stephen E. Fienberg, Margaret E. Martin ja Miron L. Straf. National Academy Press, Washington 1985.
- Sharing Social Science Data. Advantages and Challenges*. Toim. Joan E. Sieber. Sage Publications, London 1991.
- Siddiqui, Nabeel: Data Assemblages: A Call to Conceptualize Materiality in the Academic Ecosystem. *Digital Humanities Quarterly* 9 (2) 2015. <<http://digitalhumanities.org:8081/dhq/vol/9/2/000215/000215.html>>.
- Sieber, Joan E.: Openness in the Social Sciences: Sharing Data. *Ethics & Behavior* 1(2) 1991, 69–86.
- Sieber, Joan E.: Sharing Scientific Data I: New Problems for IRBs. *IRB: Ethics & Human Research* 11(6) 1989, 4–7. <<http://www.jstor.org/stable/3564184>>.
- Sinnemäki, Kaius; Tolonen, Mikko: Digitaaliset ihmistieteen tutkimuskartalle. *Tieteessä tapahtuu* 33(4) 2015, 29–32.
- Spiro, Lisa: “This Is Why We Fight”: Defining the Values of the Digital Humanities. *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 16–34.
- Stanley, Barbara; Stanley, Michael: Data Sharing: The Primary Researcher's Perspective. *Law and Human Behavior* 12(2) 1988, 173–180.
- Sterling, Theodor D.; Weinkam, James J.: Sharing Scientific Data. *Communications of the ACM* 33(8) 1990, 112–119. <<http://dx.doi.org/10.1145/79173.79182>>.

- Svensson, Patrik: Beyond the Big Tent. Teoksessa *Debates in the Digital Humanities*. Toim. Matthew K. Gold. University of Minnesota Press, Minneapolis 2012, 36–49.
- Svensson, Patrik: Humanities Computing as Digital Humanities. Teoksessa *Defining Digital Humanities: A Reader*. Toim. Melissa Terras, Julianne Nyhan ja Edward Vanhoutte. Ashgate, Farnham and Burlington 2013, 159–186.
- Tampereen yliopiston johtosääntö 1.11.2011. 5 §. <http://www.uta.fi/hallinto/tiedostot/TaY_Johtosaanto.pdf>. 9.9.2016.
- Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev Umur; Wu, Lei; Read, Eleanor; Manoff, Maribeth; Frame, Mike: Data Sharing by Scientists: Practices and Perceptions. *PLoS One* 6(6) 2011. <<http://dx.doi.org/10.1371/journal.pone.0021101>>.
- Terras, Melissa: Disciplined: Using Educational Studies to Analyse ‘Humanities Computing’. *Literary and Linguistic Computing* 21(2) 2006, 229–246.
- Terras, Melissa: The Potential and Problems in using High Performance Computing in the Arts and Humanities: the Researching e-Science Analysis of Census Holdings (ReACH) Project. *Digital Humanities Quarterly* 3(4) 2009. <<http://digitalhumanities.org:8081/dhq/vol/3/4/000070/000070.html>>.
- Terras, Melissa: Peering Inside the Big Tent. Teoksessa *Defining Digital Humanities: A Reader*. Toim. Melissa Terras, Julianne Nyhan ja Edward Vanhoutte. Ashgate, Farnham and Burlington 2013, 263–270.
- Tietoarkiston arkistonmuodostussuunnitelma. <http://www.fsd.uta.fi/fi/hallinto/asiakirjat/AMS/ams_index.html>.
- Throp, Jer: Big Data is Not the New Oil. *Harvard Business Review* 30.11.2012. <<https://hbr.org/2012/11/data-humans-and-the-new-oil/>>.
- Tolonen, Mikko; Lahti, Leo: Aatehistoria ja digitaalisten aineistojen mahdollisuudet. *Ennen ja Nyt* 2/2015. <http://www.ennenjanyt.net/2015/08/aatehistoria-ja-digitaalisten-aineistojen-mahdollisuudet/#identifier_32_1758>.
- ”Tunnisteellisuus ja anonymisointi”. Tietoarkiston aineistohallinnan käsikirja. <<http://www.fsd.uta.fi/aineistohallinta/fi/tunnisteellisuus-ja-anonymisointi.html>>. 9.9.2016.
- ”Tunnisteet ja viittaaminen”. Avoin tiede ja tutkimus -hankkeen www-sivut. <<http://avointiede.fi/tunnisteet-ja-viittaukset>>. 9.9.2016.
- ”Tutkimusdata”. Tampereen teknillisen yliopiston www-sivut. <<http://science-port.tut.fi/c.php?g=217025?p=2944899>>. 9.9.2016.
- ”Tutkimuksen tietoaaineistot”. Avoin tiede ja tutkimus -hankkeen www-sivut. <<http://avointiede.fi/tietoaaineistot>>. 9.9.2016.

Tutkimuksen tietoaaineistojen hallinnan periaatteet Jyväskylän yliopistossa 9.9.2014. Rehtorin päätös 9.9.2014. Jyväskylän yliopisto. <https://www.jyu.fi/hallinto/strategia/periaatteet/tutkimuksen_tietoaaineistojen_hallinnan_periaatteet>. 9.9.2016.

Van den Eynden, Veerle; Bishop, Libby: *Sowing the Seed: Incentives and Motivations for Sharing Research Data. A Researcher's Perspective*. Knowledge Exchange, Bristol 2014.