

E n n e n j a n y t

Historian tietosanomat

Marino González, [CC BY-NC-ND 2.0](#)

2015/2

Monitieteinen historiantutkimus - Historiografia ja tulevaisuus

REFERE-ARTIKKELIT

AATEHISTORIA JA DIGITAALISTEN AINEISTOJEN MAHDOLLISUUDET

11.8.2015 | MIKKO TOLONEN JA LEO LAHTI

Aatehistoria on yleensä kiinnostunut laadullisista kysymyksistä.^[1] Alan tutkijat pohtivat esimerkiksi miten ihmisen moraalisuus tai yhteiskuntaa koossa pitävät mekanismit ymmärrettiin menneisyydessä. Tässä kontekstissa digitaalisten aineistojen merkitys on nähty ensisijaisesti aineistojen parantuneen saatavuuden kautta. Viime aikoina avoimen tiedon "open access"-periaate on kuitenkin alettu nähdä laajempänä kokonaisuutena, joka tarjoaa runsaasti uusia välineitä, lähestymistapoja ja mahdollisuuksia aatehistorian tutkimukseen. Yksinkertaistaen voisi puhua vain avoimesta tieteestä. Kotimainen osaaminen aatehistorian ja digitaalisten ihmistieteiden alalla on jo kansainvälisesti vertailtuna korkeatasoista. Vahvistamalla uusia hyväksi koettuja tutkimuskäytäntöjä voidaan löytää tapoja nostaa entisestään tekemisen tasoa.

Erityisesti vahvan laadullisen tutkimustradition ja modernin tieteellisen laskennan yhdistäminen tarjoaa lupaavia ja ajankohtaisia mahdollisuuksia uusiin, kansainvälisesti merkittäviin avauksiin aatehistoriassa ja yleisemminkin ihmistieteiden alalla. Merkittävien tieteellisten tulosten saavuttaminen vaatii kuitenkin uteliasta, ennakkoluulotonta ja pitkäjänteistä tieteiden välistä yhteistyötä yli perinteisten laitosrajojen. Avoimet verkostomaiset yhteistyömallit tarjoavat mahdollisuuksia digitaalisten aineistojen hyödyntämiseen tavalla josta voivat hyötyä monet eri osapuolet.

Digital Humanities – digitaaliset ihmistieteet

Digital humanities (DH) on maailmalla vakiintunut käsite, jolla tarkoitetaan erityisesti modernien tietojenkäsittelymenetelmien käyttöä humanistis-yhteiskunnallisessa tutkimuksessa koskien laajoja historiallisia, kirjallisia tai esimerkiksi kieltä ja kuvaa yhdistäviä digitaalisia aineistoja.^[2] Käsite on juurtunut tieteelliseen kielenkäyttöön 2000-luvun aikana. Tällä nimikkeellä operoivat tieteelliset lehdet ja professorit ovat pääasiallisesti 2010-luvulla perustettuja. Vastaavan suomenkielisen termin puuttuessa Suomen Akatemia on päättänyt keväällä 2015 viittaamaan ”digitaalisiin ihmistieteisiin”, joka kattaa sekä humanistisen alan että yhteiskuntatieteet.^[3] Termi on kuvaava ja lähellä englanninkielistä vastinettaan.

Kevään 2015 lehtien tulviessa digitalisaatiota, big dataa ja Suomen tulevaisuutta, kehitystä on seurattu rauhallisemmin historian alalla ja humanistisessa tutkimustraditiossa. Alan kotimaisessa tutkimuksessa on kuitenkin tapahtumassa merkittävä käänne ja humanisteihin kohdistuu digitalisoitumispaineita niin alan sisältä kuin ulkopuoleltakin. Tietotekniikan käytöllä ei kuitenkaan ole humanistisessa tutkimustraditiossa itseisarvoa, eikä esimerkiksi pelkän aineistojen digitoinnin sinänsä katsota vielä kuuluvan alan varsinaisen tutkimuksen piiriin. Jaottelu on haastavaa, sillä toisaalta tutkimusaineistojen keruu ja järjestäminenkin on kiistämättä keskeinen osa tieteellistä tutkimusprosessia.

Digitaalisia ihmistieteitä tehdään usein tekstiaineistojen pohjalta, mutta tutkimusaineistoksi käy siis kaikki kulttuuriperintö kuten esineet, kuva, ääni tai liike.^[4] Aineistojen kuvailutietojen avulla on mahdollista tutkia esimerkiksi kirjoja materiaalisina objekteina tai kirjapainotoiminnan kehittymistä ja leviämistä historiallisena ilmiönä. Tutkimusaineistot voidaan jaotella digitaalisena (esim. sosiaalinen media) tai muuten syntyneisiin (esim. digitoidut kirjat) aineistoihin. Keskitymme tässä kirjoituksessa ensisijaisesti ei-digitaalisina syntyneiden aineistojen käyttöön joita aatehistorian tutkimustraditio erityisesti koskee. Digitaaliset ihmistieteet käsitteen piiriin katsotaan kuuluvan myös digitaalisuuden tutkimus ilmiönä. Voidaan tutkia mi-

ten kaupunkien toimintaan liittyvää avointa ja suljettua lähes reaaliaikaistakin dataa voidaan ymmärtää ja hyödyntää kaupunkien kehityksen kontekstissa. Kaupunkeihin ja yhteiskuntaan liittyvää dataa on saatavilla kasvavassa määrin. Kyseessä on uudenlaisten tutkimusaineistojen ja -menetelmien käyttöönotto laadullisen tutkimuksen tueksi jo vakiintuneiden menetelmien rinnalle.^[5] Humanistien panos voi tuottaa tuoreita näkökulmia, kun tutkitaan esimerkiksi kaupunkien dynamiikkaan liittyvien ilmiöiden kehittymistä ja yhteyksiä ihmisten käyttäytymiseen, hyvinvointiin, vuorovaikutukseen ja maailmankuvaan.

Digital humanities -käsite rinnastetaan toisinaan myös laskennalliseen yhteiskuntatieteeseen (Computational Social Science, CSS).^[6] Humanistis-yhteiskuntatieteellisen tutkimuksen kenttä kattaakin laajan kirjon erilaisia tutkimuskohteita ja -menetelmiä. Tulee kuitenkin sisäistää, että digitaaliset ihmistieteet -käsite viittaa tutkimukseen jossa uusia informaatiotieteiden menetelmiä hyödynnetään nimenomaan perinteisten humanistis-yhteiskuntatieteellisten kysymysten tutkimukseen muun muassa historiassa, lingvistiikassa ja muilla ihmistieteen aloilla, kun taas laskennallinen yhteiskuntatiede on keskittynyt vahvemmin esimerkiksi yhteiskunnan ja sosiaalisten verkostojen dynamiikkaa. Nuori, mutta jo paikkansa vakiinnuttanut laskennallisen yhteiskuntatieteen ala soveltaa ennakkoluulottomasti esimerkiksi tilastollisen fysiikan menetelmiä sosiaalisten järjestelmien tutkimuksessa.^[7] Samalla sosiaalisuuteen liittyvät kysymykset ovat nousemassa entistä vahvemmin osaksi "kovien" luonnontieteiden tutkimuskenttää. Tämä avaa näköaloja niin ihmistieteiden kuin informaatiotieteidenkin alalla ja uudistaa tutkimuskulttuuria. Samalla perinteiset tulkinnat sosiaalisuudesta saattavat kuitenkin suoraviivaistua tavalla joka on humanistis-yhteiskuntatieteen perinteelle vierasta. Kasvaessaan nuori luonnontieteestä ponnistava laskennallinen yhteiskuntatiede elää ja monipuolistuu. Tähän liittyviä mahdollisuuksia tulee tunnistaa ja löytää yhteisiä kosketuspintoja muihin tieteenaloihin. Digitaalisten ihmistieteiden kysymyksenasettelut ovat kuitenkin osin lähtökohdiltaan erilaisia kuin vaikkapa laskennallisen yhteiskuntatieteen pyrkimykset ymmärtää yhteiskunnan toimintaa ja sosiaalista dynamiikkaa vaikkapa sovelletun fysiikan keinoin. Gadamer on esittänyt, että luonnontieteen menetelmiä ei voida sellaisenaan suoraan soveltaa ihmistieteisiin, koska näissä tulee ymmärtää inhimillisen toiminnan kulttuurisidonnaisuus sekä jatkuva dialogi menneen, nykyisen ja tulevan välillä.^[8] Tähän sisältyy ripaus viisautta mistä kannattaa pitää kiinni.

Merkittävät tieteelliset läpimurrot vaativat syvällistä ymmärrystä uusien menetelmien avaamista mahdollisuuksista ja niihin liittyvistä rajoitteista. Ihmistieteiden tutkimuskysymysten on pohjauduttava perustutkimukseen ja tutkimustraditioon. Esimerkiksi kieliteknologiaa ei sellaisenaan joidenkin näkemysten mukaan lasketa digitaaliseksi ihmistieteeksi, mutta toisaalta

kieliteknologian sovelluksilla on merkittävää potentiaalia muun muassa aatehistoriallisen tutkimuksen keskeisenä tutkimusvälineenä ja alalta löytyy jo valmiiksi kotimaista osaamista.^[9] Muita esimerkkejä ihmistieteissä jo hyödynnettävistä informaatiotekniikan menetelmistä ovat esimerkiksi erilaiset visualisointitekniikat ja muut datan kuvaamismenetelmät kuten semanttiset verkostot. On selvää, että nousevan tutkimusalan alkuhuumassa tehtävistä kokeiluista vain osa kestää ja vakiintuu osaksi tutkimuskulttuuria. Näyttää kuitenkin varmalta, että modernien informaatiotekniikan menetelmien käyttö tulee uudistamaan ihmistieteiden tutkimusta sillä se tarjoaa uudenlaisia välineitä ja mahdollisuuksia systemaattisen kokonaiskuvan muodostamiseen. Laajojen tietoaaineistojen analyysi auttaa hahmottamaan yksittäisiä havaintoja laajemmissa konteksteissa ja tällä tavoin saatavaa tietoa voidaan vertailla laadullisen tutkimuksen tarjoamaan kuvaan. Liikkuminen eri näkökulmien välillä ja erilaisten tutkimusperinteiden kohtaaminen voi synnyttää innovatiivisia tutkimusnäkökulmia ja tieteellisiä läpimurtoja. Samalla kun uusia menetelmiä käytetään klassisten kysymysten tutkimiseen, tulevat ne väistämättä muokkaamaan perustutkimuksen metodologista pohjaa ja avaavat uusia näkökulmia perustavan laatuisiin kysymyksiin.^[10]

“Digital humanities” -liikehdintä on lähivuosina saanut tuulta purjeisiin myös historian laitosten porteilla eri puolilla Suomea. Yhtäältä käsitettä käytetään arkipäiväisesti kuvaamaan kaikkea digitaalisiin aineistoihin liittyvää tutkimusta ja toisaalta esimerkiksi Turussa, Porissa ja Tampereella on tehty tärkeitä aloitteita digitaalisen kulttuurin tutkimuksen alalla. Aiheen parissa toimivat saattavat kantaa humanistisen tutkimustradition valitettavaa taakkaa jossa asetetaan suuri paino yksittäisen sankaritutkijan tekemisille laajemman yhteistyön sijasta. Näin tutkimuskentässä ei synny luontevasti kokemusta ja ymmärrystä monitieteisyydestä ja sen mahdollisuuksista, koska tutkimuskysymykset, aineistot ja menetelmät eivät ole kohdanneet tavalla jossa erilaisia taustoja omaavien tutkijoiden yhteistyö tuottaisi lisäarvoa. Kun yksittäinen tutkija pyrkii samanaikaisesti omaksuma ja soveltamaan menetelmiä ja tietoa useilta tieteenaloilta, tekemisen määrä ja monimuotoisuus eivät voi korvata laatua. Usein sorrutaan liian helposti hätäisiin päätelmiin tai päädytään selvittämään yksittäisen tieteenalan näkökulmasta marginaalisia ongelmia kykenemättä edes kysymään – saatika vastaamaan – sellaisiin kysymyksiin, joilla olisi aitoa pitkän tähtäimen vaikuttavuutta tieteenalan sisällä. Nykyaikana on lapsellista olettaa, että yksittäinen tutkija muuten kuin poikkeustapauksissa voisi hallita syvällisesti useiden tieteenalojen tutkimusperinnettä. Tarvitaan monitieteistä yhteistyötä, jonka tuottamaa lisäarvoa ei voida saavuttaa pelkästään jakamalla tehtäviä ja kasvattamalla tekijöiden lukumäärää.

Alan kehitykselle on keskeistä, että eri tieteenalojen tutkijat pystyvät ymmärtämään toisten-

sa tarpeita ja viestimään sujuvasti keskenään. Tämä vaatii kaikilta osapuolilta perustietämystä yhteistyötahojen tutkimusperinteestä ja käsitteistöstä, sekä aitoa mielenkiintoa oman tieteenalan ulkopuolisiin kysymyksiin ja eri tieteenalojen perinteistä ammentavan tutkimuksen tekoon. Ihmistieteiden tutkijat voivat osallistua laskennallisten menetelmien ideointiin ja soveltamiseen. Informaatiotieteilijät puolestaan voivat osallistua ihmistieteiden kysymyksenasettelujen muotoiluun ja ratkomiseen, mutta tutkimuksen laadun varmistamiseksi mukana täytyy olla vankka asiantuntemus tutkimuksen kannalta keskeisiltä aloilta. Sujuvan yhteistyön ja parhaiden tulosten saavuttamiseksi on ehdottoman tärkeää, että tekijöillä on jaettu näkemys tutkimushankkeen tavoitteista. Tätä tuetaan parhaiten tekemällä tiivistä ja pitkäjänteistä yhteistyötä. Kokemuksia ja parhaita käytäntöjä voidaan ammentaa myös muilta tieteenaloilta, kuten laskennallisen biologian tutkimuksesta, joissa vastaavaa monitieteistä yhteistyötä on tehty jo pidempään.

Aatehistorian ja kielitieteen lähestymistapojen yhtäläisyyksistä ja eroista

Kielentutkija Franco Moretti on käyttänyt etäluennan (distant reading) termiä kuvatessaan maailmankirjallisuuden liikkeiden tutkimusta, jossa painopiste on yksittäisten dokumenttien sijasta laajempien trendien hahmottamisessa.^[11] Tämä kuvaa hyvin myös tavoitteita joihin pyritään monissa nykyisissä digitaalisten ihmistieteiden tutkimushankkeissa, kun etsitään laajoista tekstimassoista yhdistäviä tekijöitä ja säännönmukaisuuksia vastauksena tiettyihin tutkimuskysymyksiin. Monimuotoisten digitaalisten aineistojen tulkinta vaatii selkeän asiantuntijanoksen ja räätälöityjen laskennallisten työkalujen kehittämistä tutkimuksen tueksi.

”Big datan” käsitteelle ei ole vakiintunut suomenkielistä vastinetta, joten puhumme tässä tekstissä suurten aineistojen tiedonlouhinnasta ja etäluennasta. Nykyään jo mantran lailla hoetaan, että dataintensiivisten tieteenalojen näkökulmasta humanistisen tutkimuksen data on harvoin aidosti ”suurta” verrattuna esimerkiksi tähtitieteeseen, hiukkasfysiikkaan tai molekyylibiologiaan. Tekstikorpuksia joissa yhdistetään ja editoidaan esimerkiksi kaunokirjallisuutta tutkijakäyttöön sekä muut vastaavat digitoidut aineistokokoelmat ovat usein kuitenkin niin laajoja ettei yksittäinen tutkija kykene mitenkään lukemaan tai tulkitsemaan koko aineistoa manuaalisesti. Samalla ihmistieteiden erityisluonne verrattuna luonnotieteen tutkimuskohteisiin tuo omat lisähaasteensa aineistojen tilastolliseen analyysiin ja tulkintaan. Voitaissiinkin Morettia mukaillen puhua esimerkiksi suurten aineistojen kvantitatiivisesta analyysistä. Korpus-tutkimuksessa on perinteisesti tähdätty lähes täydellisten tutkimusaineistojen koamiseen. Siinä teksti on mahdollisimman tarkasti käsitelty ja rikastettu tutkimuksen vaati-

maan muotoon. Etäluennassa taas aineiston ei tarvitse olla lähelläkään täydellistä, jotta sillä voisi jo olla merkittävää tutkimuskäyttöä.^[12] Sen sijaan pyrkimyksenä on, että aineisto sisältää tutkittavasta aiheesta riittävästi tutkimuksen kannalta käyttökelpoista informaatiota joka voidaan kaivaa esiin sopivilla menetelmillä. Samalla kvantitatiivisen analyysin avulla aineistosta saadaan uutta tietoa esimerkiksi julkaisumääristä ja sanafrekvensseistä. Karkeakin aineisto voi näin olla tutkimuksellisesti arvokasta pyrittäessä tunnistamaan laajoja tilastollisia säännönmukaisuuksia. Menetelmät voivat sietää huomattaviakin määriä satunnaisia epätarkkuuksia joiden vaikutus laajempien trendien havaitsemiseen on usein rajattu. Laajojen aineistojen kvantitatiiviseen analyysiin perustuvassa tutkimuksessa datasta voidaan tehdä myös odottamattomia havaintoja. Nämä voivat suunnata tutkimusta tavalla jota ei alunperin ole osattu ennakoida. Molemmilla lähestymistavoilla – huolella laadittujen pienempien aineistojen ja laajempien karkeiden aineistojen tutkimuksella – on oma paikkansa. Lähestymistavasta riippumatta aineiston siistiminen käyttökelpoiseen muotoon on tutkimuksessa suhteellisesti yksi eniten aikaa vievistä vaiheista.

Aatehistoriassa laskennallisten menetelmien merkitys liittyy erityisesti niiden käyttöön tutkimuksen apuvälineenä jolla voidaan tunnistaa laajoista aineistoista mielenkiintoisia trendejä. Aineistojen analyysi usein liittyy laajempiin ja kielen käytön näkökulmasta ulkopuolisiin konteksteihin. Tämä poikkeaa sellaisesta lingvistiksestä tutkimuksesta jossa kielen muutoksen tutkimuksessa automatisoiduilla menetelmillä tehdään jo nyt pitkälle vietyjä johtopäätöksiä liittyen esimerkiksi pronominiin historialliseen käyttöön ja merkitykseen.^[13] Tämä esitys korostaa tarkoituksella tätä lingvistiikan ja historian tutkimuksen eroa. Kielitieteellä on automatisoitujen tutkimusmenetelmien käytössä selkeä etumatka. Tässä kritiikki kohdistuu erityisesti historian tutkimukseen itseensä ja kyvyttömyyteen ottaa vastaavaa roolia teknologian kehityksessä mihin kielitieteissä on toisenlaisin kysymyksenasetteluin pystytty.

Aatehistoria on keskittynyt käsitteelliseen muutokseen josta tunnettu esimerkki on Quentin Skinnerin vaatimus keskittyä kirjoittajan retoriseen intention, kun tämä tekee jonkun tietyn lingvistisen teon joka tulkitaan sen poliittisessa kontekstissa.^[14] Näin poliittisen kontekstin tunnistamisesta ja tulkitsemisesta tulee keskeinen osa käsitteen käytön analyysiä. Skinner korostaa historiallisen kielen ymmärtämistä sen käyttökontekstissa ja ettei käsite välttämättä ole kytköksissä tiettyyn sanaan.^[15] Tämä erilainen ymmärrys kontekstista erottaa aatehistoriallisen tutkimuksen intressit useista muista sosio-lingvistiksestä tutkimusasetelmista. Etäluennan merkitystä ei kuitenkaan voida vähätellä edes skinneriläisen tutkimustradition jyrkimpienkään linjausten puitteissa. Skinneriläisen ajattelun ytimessä on jo vuosikaudet ollut erilaisten keskeisten poliittisten termien käytön laaja-alainen tunnistaminen ja tulkinta eri-

laisten poliittisten kontekstien kautta. Laajojen aineistojen käytössä korostuu mahdollisuus löytää uusin menetelmin historiallisia kiinnekohtia, jotka ovat aikaisemmin jääneet huomaamatta. Keskeinen rajoittava tekijä tässä on ollut pragmaattinen: yksittäisen tutkijan kyvyttömyys hallita suuria tekstimassoja. Aikaisemmin laajat kontekstit on rakennettu usein yksittäisen tutkijan tuntemusten varaan (kuten historian tutkimuksessa valitettavan usein muutenkin). Nyt tämän rinnalle on mahdollista saada kvantitatiivista tietoa esimerkiksi sanojen ja fraasien laajemmasta käytöstä.

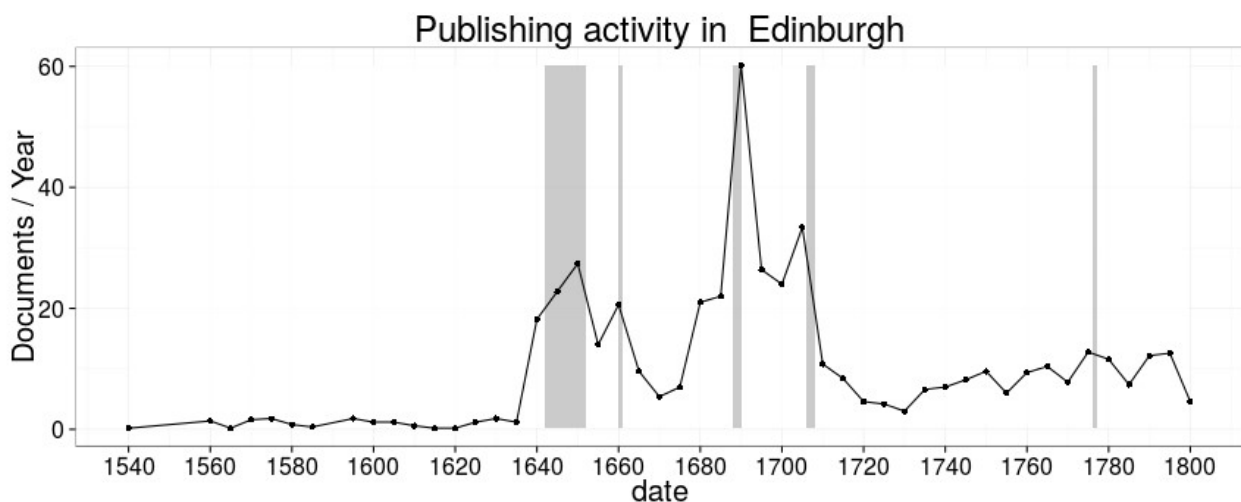
Suhteessa Skinneriin, Reinhart Koselleckin esimerkin mukainen ja Suomessa vahvasti edustettu käsitehistoria on suuntaus jolla on erinomaista potentiaalia kasvaa ja kehittyä laajojen digitaalisten aineistojen käytön myötä.^[16] Käsitehistoria (*Begriffsgeschichte*) tutkii käsitteiden käyttöä, niiden muutosta sanastossa ja mikä sosiaalinen vaikuttavuus tietyllä käsitteellä on ollut. Avainkäsite ei koostu ainoastaan sen suhteista muihin termeihin, vaan tulee ymmärtää kiistanalaisena ja poliittisten kamppailujen myötä sosiaalista maailmaa muokkaavana tekijänä. Näin sitä tutkitaan myös hieman eri lähtökohdista kuin skinneriläisiä intentioita. Tähän asti kuitenkin jopa diakronisen tutkimuksen tekeminen liittyen laajoihin historiallisiin aineistoihin on ollut vähintäänkin työlästä, mutta mallintamismahdollisuuksien kehittyessä käsitehistorian traditio pystynee hyödyntämään hyvän teoreettisen lähtökohtansa. Automatisoidun etäluen-
nan avulla voidaan paremmin kontrastoida tunnettujen ajattelijoiden kielenkäyttöä arkipäiväisempään julkiseen keskusteluun kuten sanomalehtiteksteihin. Näin voidaan paremmin hahmottaa sellaisia julkisen keskustelun konteksteja joissa klassikkotekstit ovat syntyneet. Voidaan olettaa, että tulevaisuudessa aatehistoriaa tehdäänkin kasvavassa määrin erilaisten sana- ja käsiteryppäiden semanttisena analyysinä. Tätä kautta päästään tarkastelemaan uudestaan aatehistorian metodologista ydinkysymystä siitä miten "aate" tai "idea" voidaan määritellä ja tunnistaa. Syntyy myös tarve miettiä uudestaan myös kuopatuiksi luultuja ajattelijoi-
ta, kuten vaikkapa Lovejoyta.^[17]

Uusista mahdollisuuksista huolimatta ei ole syytä vähätellä perinteistä lähiluentaan perustuvaa aatehistorian tai filosofian historian tutkimusta, vaikka perinteisten metodologisten kysymysten rinnalle nousee uudenlaisia käytännön (sekä syvällisiä metodologisia) ongelmia, kun pystymme tutkimaan huomattavasti aiempaa laajempia aineistoja. Toisaalta laajemman kontekstin kytkeminen aineistojen analysointiin on haastavaa ja kohtaa väistämättä rajoituksia. Näin uudet informaatiotieteen menetelmät voivat tukea ja täydentää laadullista tutkimusta. On hedelmätöntä inttää vastaan kun todetaan, että laajojen aineistojen analysointi tuo lähiluvun rinnalle tarkempaa tietoa julkisen keskustelun määrästä ja trendeistä, jotka ovat välttämättömiä uusien tulkintojen tekemiseksi.

Kuvailutietojen merkitys ja aatteiden maantiede

Laajojen tekstitietokantojen ohella aineistoja voidaan kartoittaa myös laajojen kuvailutietojen pohjalta. Näin voidaan tarkastella esimerkiksi kirjapainotoiminnan ja muun tiedontuotannon kehittymistä yhdistelemällä taustatietoa kirjojen painovuosista, -paikoista, -määristä, kirjoittajista ja aiheista. Lähteiksi käyvät monenlaiset aineistot arkistoista rekisteriaineistoihin ja kirjastoluetteloihin. Yhtenäisen rakenteen omaavien kuvailutietojen automatisoitu jäsentely on usein helpompaa ja niitä voidaan tarkastella suoraviivaisemmilla tutkimusmenetelmillä kuin laajempia kokotekstitietokantoja, koska kuvailuaineistojen koko on rajatumpi ja rakenne selkeä. Tällainen lähestymistapa täydentää merkittävästi tekstipohjaisen tutkimuksen mahdollisuuksia ja koneluettavien kuvailutietokantojen hyödyntäminen alan tutkimuksessa on saatettu tähän mennessä merkittävästi aliarvioida.

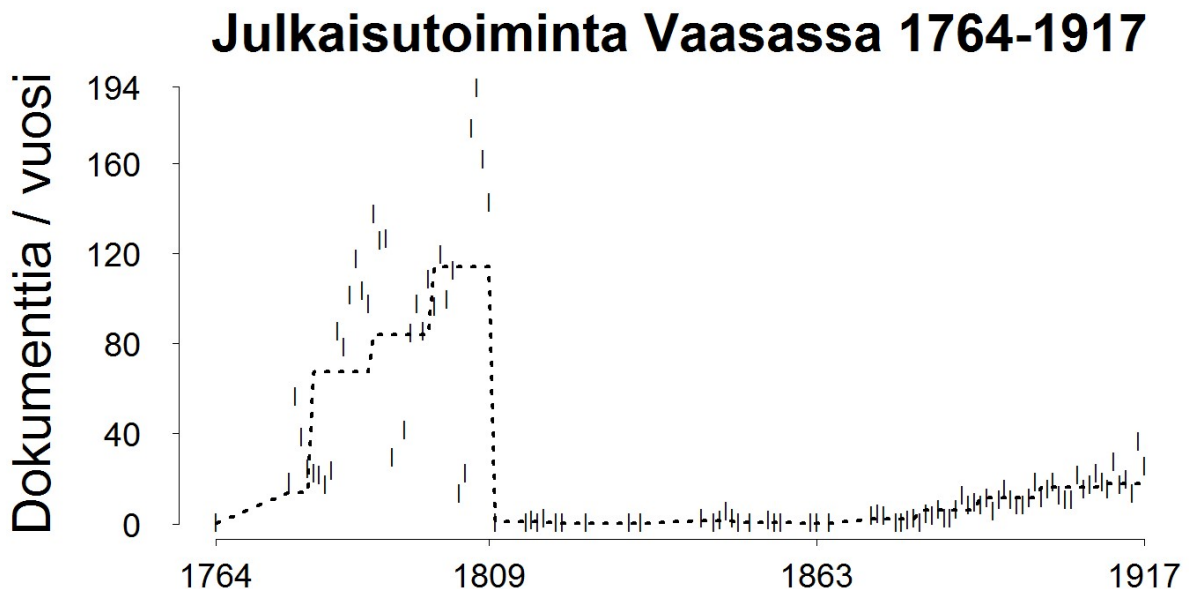
Olemme tutkineet kuvailutietojen käyttöä muun muassa brittiläisen English Short-Title Catalogue (ESTC) -kirjastokokoelman pohjalta, jonka tiedot kattavat valtaosan Britanniassa ja Pohjois-Amerikassa 1477-1800 painetuista kirjallisista dokumenteista. Tämän pohjalta pystymme tekemään suoraviivaista analyysiä aktuaalisen sosiaalisen muutoksen vaikutuksesta tiedontuotantoon (Kuva 1).



Kuva 1. Historia-aiheiset julkaisut Edinburghissa, 1540-1799. Kuvasta näkyy kuinka Edinburghissa julkaistuissa historia-aiheisissa dokumenteissa keskeisimmät yhteiskunnalliset muutokset vaikuttavat kirjastoluettelosta ilmeneviin julkaisumääriin vuosina 1540–1800. Näitä muutoksia, jotka näkyvät kaaviossa tummennettuina, ovat: Englannin sisällissota (1642–1651), Stuart-suvun paluu valtaistuimelle (1660), Mainio vallankumous (1688–1689), Ison-Britannian kuningaskunnan syntyminen eli unioni (1706–1707). Huomionarvoista on myös ettei Yhdysvaltojen itsenäisyys (1776) aiheuta samanlaista julkaisupiikkiä.

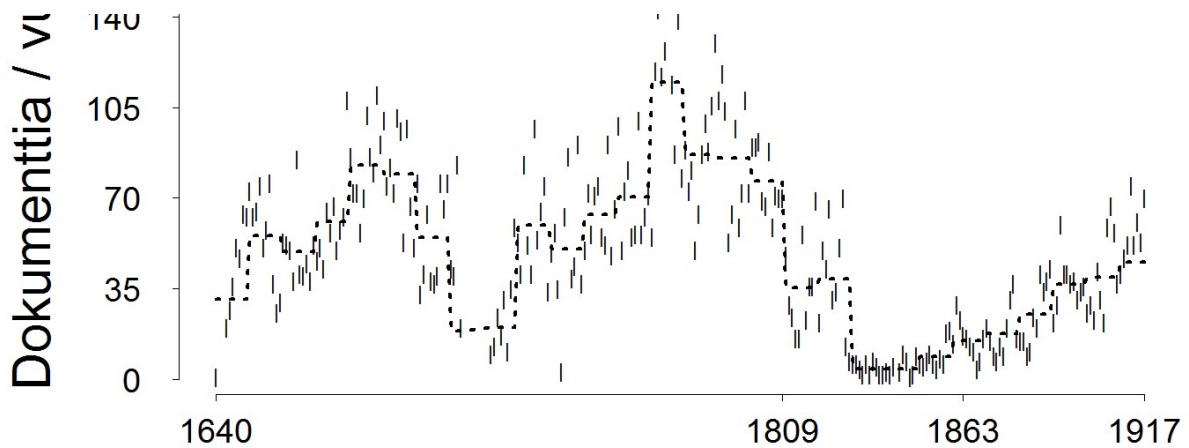
Kotimaisen historiantutkimuksen kannalta kiintoisan esimerkin tarjoaa tieteellisten laskenta-

kirjastojen laadinta suomalaisen kansallisbibliografian Fennican tutkimiseen, vaikka samalla tuleekin tiedostaa – ja korjata – Fennican nykyiset puutteet. Aineiston alustava analyysi osoittaa esimerkiksi suoraan historiallisiin tapahtumiin kytkeytyvät tiedontuotannon huiput ja aallonpohjat Vaasassa, Helsingissä, Turussa ja Suomessa vuosina 1764-1917 (Kuvat 2-5).



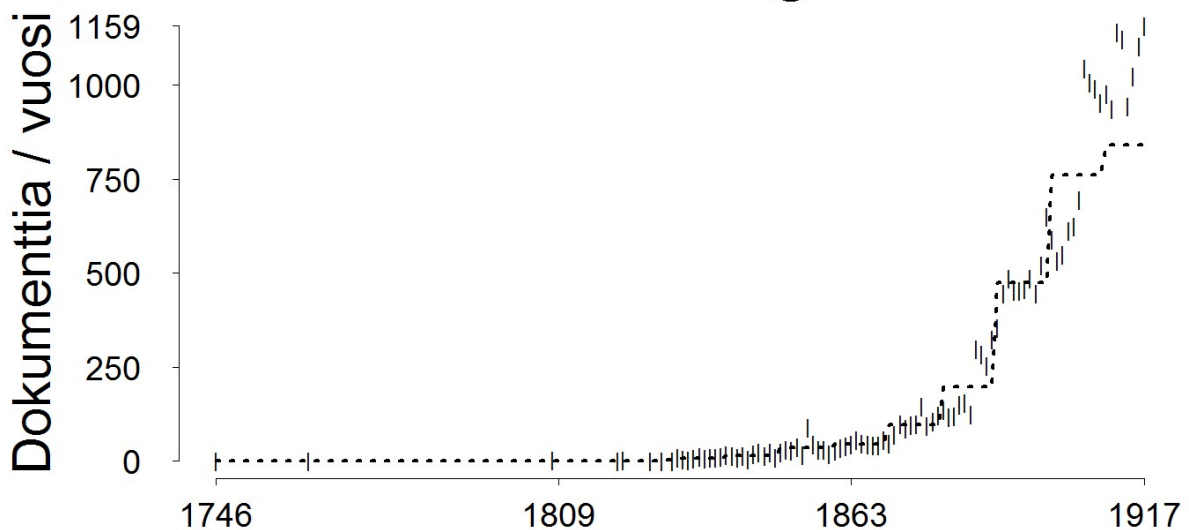
Kuva 2. Julkaisutoiminta Vaasassa 1764-1917. Oheinen kuva näyttäisi viittaavan tiedontuotannon romahdukseen Vaasassa välittömästi Venäjän vallan alussa^[18]. Tämä saattaa osittain selittyä sillä, että painotoimintaa harjoitti Vaasassa 1776-1804 yksi suurempi toimija (Georg Wilhelm Londicer) ja näyttäisi myös siltä, että hovioikeuden asiakirjojen julkisuusperiaate ja painaminen muuttui Venäjän vallan myötä. Tämä kuvaa yksinkertaisella tavalla Suomen historiassa tapahtuneita julkisuuteen ja julkiseen keskusteluun liittyviä käännteitä. Suomen kansallisbibliografian Fennican perusteella julkisuuden kehitys painotuotannon muodossa näyttäisi olevan tilastollisesta näkökulmasta Venäjän vallan aikana hitaampaa kuin mitä yliopiston siirron ja Helsingin kehittämisen merkitystä korostanut historiantutkimus on aiemmin esittänyt.^[19] On kuitenkin alleviivattava, että mukana on muitakin selittäviä tekijöitä, kuten systemaattisia keisarikunnan aikaiseen painotuotantoon liittyviä puutteita kirjastoluettelossa (on tunnettua, että Fennica ei ole täydellinen, vaan luetteloinnista puuttuu runsaasti erityisesti 1800-luvun ruotsinkielistä tuotantoa). On kuitenkin selvää, että aineistojen kvantitatiivinen analyysi tarjoaa tehokkaan tavan tunnistaa tiedontuotannon historiaan liittyviä trendejä. Lopulliset johtopäätökset tulee kuitenkin varmistaa myös muilla keinoilla. Yksi syy miksi tahdomme tässä julkaisussa tuoda nämä vielä vajavaiset tilastolliset näkökulmat esiin on, että emme yksinkertaisesti voi jäädä odottelemaan kansallisbibliografian täydentymistä enää seuraaviksi vuosikymmeniksi (ks. esim. huomio Jyrki Hakapää, Kirjan tie lukijalle. Kirjakauppojen vakiintuminen Suomessa 1740-1960, SKS, 2008, s. 380), vaan tämä tärkeä luettelointityö olisi tutkimuksellisista syistä kyettävä viedä loppuun mahdollisimman pian.





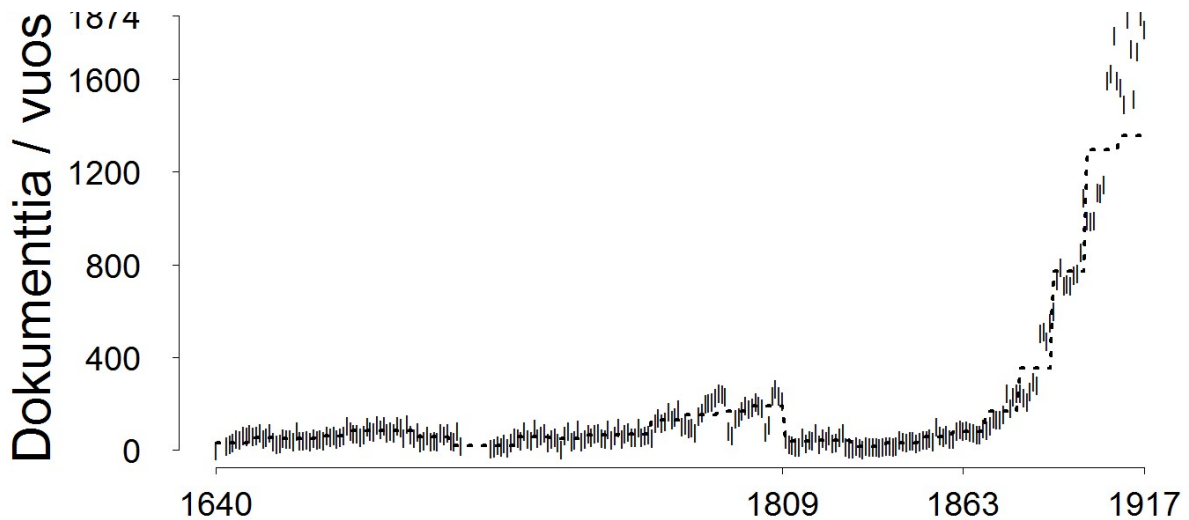
Kuva 3. Julkaisutoiminta Turussa 1640-1917. 1700-luvun kehitys näyttäisi seuraavan eurooppalaista kasvavaa trendiä, kirjapainotoiminnan kehitys 1820-luvulta lähtien on taas nykyisten luettelointitietojen valossa hidasta.

Julkaisutoiminta Helsingissä 1746-1917



Kuva 4. Julkaisutoiminta Helsingissä 1746-1917. Visuaalinen esitys Helsingissä tapahtuneesta kirjapainotoiminnan kehityksestä viittaa siihen, että jos Fennican luottelointiin on ollenkaan luottamista, niin ennakkosensuurin osittaisen poistumisen vaikutukset 1850-luvulla ovat yllättävän vähäiset ja varsinainen painotoiminnan kehitys lähtee käyntiin vasta huomattavasti myöhemmin. Yliopiston siirtäminen Helsinkiin näyttää olleen julkaisutoiminnan kannalta, jos ei lamauttava, niin ainakin hidastava toimenpide.

Julkaisutoiminta Suomessa 1640-1917



Kuva 5. Julkaisutoiminta Suomessa 1640-1917 kertoo siitä, että 1700-luvun lopun ruotsalainen kehitys näyttäisi suunnilleen samalta kuin yleinen eurooppalainen painotoiminnan nousu. Vuoden 1809 jälkeen toiminta Suomen alueella taantuu ja kestää aina 1890-luvulle asti, että lukumääräisesti painotoiminnassa nousee samalle tasolle kuin Ruotsin vallan aikana. Tämä kuva korjautuu ainakin osittain kansallisbibliografian täydentyessä.

Kartoittamalla lisäksi vaikka ruotsalaisiin tai venäläisiin paikannimiin tehtävien viittausten määrää ja laatua suomalaisissa sanomalehdissä, voimme saada uutta tietoa suomalaisen julkisen keskustelun rakenteesta ja kehityksestä historian eri vaiheissa.^[20] Samaa voidaan tehdä paikallisessa mittakaavassa tutkimalla viittauksia naapurikuntien ja suurempien kaupunkien välillä.^[21] Kansallisbiografian historiallisten henkilöiden nimien paikantaminen ja kartoitus voi myös osaltaan kehittää suomalaisen julkisen sfäärin tutkimusta. Näin uudet informaatiotekniikan menetelmät voivat antaa aatehistorian tutkimuksen käyttöön uudenlaista tutkimusmateriaalia ja tutkimuksellisia näkökulmia.

Alkuperäiset kuvailutietueet eivät sellaisenaan sovellu tilastolliseen analyysiin, vaan ne pitää ensin jalostaa sopivaan muotoon. Tämä sisältää mm. maantieteellisten ja henkilönimien erilaisten kirjoitusmuotojen yhtenäistämistä, samannimisten kirjoittajien erottelua, kirjoitusvirheiden käsittelyä, tai kirjojen sivumäärien arviointia bibliografisten standardimerkintöjen pohjalta. Tietoja voidaan myös täydentää ja rikastaa muiden lähdeaineistojen nojalla. Kirjoittajista voidaan esimerkiksi selvittää puuttuvat elinvuodet, sukupuoli, muuta taustatietoa, paionapaikkojen perusteella painomaa ja niin edelleen. Valtaosa työstä voidaan ja tulee automatisoida, kun tutkittavien aineistojen määrät kasvavat kymmeniin tai satoihin tuhansiin dokumentteihin. Samalla analyysien tilastollinen voima kasvaa ja satunnaisten virheiden vaikutus

johtopäätöksiin hälvenee. Kaikkea ei voida kuitenkaan automatisoida. Juuri tästä syystä joustavaan lähdekoodiin perustuvat rätälöidyt laskentakirjastot tarjoavat erinomaisia mahdollisuuksia alan tutkimukselle verrattuna valmiisiin ohjelmistokokonaisuuksiin, jotka usein olettavat että aineisto on jo valmiiksi saatettu tilasto-ohjelmiston vaatimaan muotoon. Tämä on merkittävä puute, sillä käytännön kokemuksen perusteella aineistojen siistiminen ja laadunvarmistus ohjelmistojen ja tilastollisten analyysien vaatimaan muotoon on yksi työläimmistä ja eniten aikaa vievistä työvaiheista. Tieteellisen ohjelmoinnin hyödyntäminen auttaa automatisoimaan kaiken mikä automatisoitavissa on. Tieteellinen laskentakirjasto tarjoaa nimensä mukaisesti kokoelman hyvin dokumentoituja menetelmiä tutkimuskohteen erityispiirteet huomioiviin tietojenkäsittelytehtäviin. Laskentakirjastot auttavat automatisoimaan merkittäviä osia aineistojen tilastollisesta analyysistä, mutta valmisohjelmistoista poiketen antavat samalla tutkijalle merkittävästi laajemmat vapaudet yhdistellä ja täydentää menetelmiä haluamallaan tavalla, tutkimuskohteen erityispiirteet huomioiden. Tämä tarjoaa tutkijalle sujuvan pääsyn sekä aineistojen jalostamiseen että tilastollisen analyysin ja visualisointien kannalta keskeisiin työkaluihin. Laskentakirjastoja ylläpidetään usein avoimen lähdekoodin palveluissa, jolloin koko työvirta voidaan toteuttaa läpinäkyvällä ja toistettavalla tavalla, tieteellisen avoimuuden periaatteita kunnioittaen (ks. kuva 6).

Tiedonlouhinnan vaihtoehtoja aatehistorian näkökulmasta

Laajojen aineistojen tekstilouhinta

- **Tavoite:** ymmärtää käsitteellistä muutosta ja kielen käyttöä
- **Aineistot:** tekstitietokannat (ECCO, EEBO, digitoidut sanomalehdet jne.)
- **Potentiaali:** kvantitatiivinen evidenssi laadullisen tutkimuksen tukena ja uudet havainnot; etäluenta
- **Käytäntö:** alkuperäisaineistot harvoin avoimesti saatavilla tai sidottu rajoitettuihin käyttöliittymiin
- **Skaalautuvuus:** rajattu; data on kooltaan massiivista ja voi vaatia huomattavaa jalostamista; aineistosta ja menetelmistä riippuvainen;
- **Metodologinen näkökulma:** aatehistorian sovelluksia toistaiseksi hyvin vähän

Kuvailutiedot tilastollisena työkaluna

- **Tavoite:** ymmärtää yleisiä trendejä historiallisten ilmiöiden dokumentoinnissa
- **Aineistot:** kirjastot ja arkistot tulvivat erilaisia kuvailutietokantoja (metadata)
- **Potentiaali:** tutkimuskäyttö aliarvioitu
- **Käytäntö:** saatavuudessa ongelmia, mutta vähemmän kuin tekstiaineistojen kohdalla
- **Skaalautuvuus:** erinomainen; data on rakenteista ja sen koko on rajattu
- **Metodologinen näkökulma:** tarjoaa selkeän lähtökohdan uusien käytäntöjen soveltamisessa humanistiseen tutkimukseen.

Raakadata



Muut tietolähteet
(avoimet & suljetut)



Kuva 6. Datasta tietoon. Alkuperäisen tutkimusaineiston jalostaminen tiedoksi ja lopulliseksi julkaisuksi on monivaiheinen prosessi, jossa tarvitaan välineitä aineistojen järjestämiseen, yhdistelyyn, ja tilastolliseen analysointiin. Merkittäviä osia tästä työstä voidaan automatisoida laatimalla tarkoitukseen suunnattuja tieteellisiä laskentakirjastoja. Laskentakirjasto on hyvin dokumentoitu kokoelma tutkimuskohteen mukaan räätälöityjä laskenta-algoritmeja, joilla voidaan automatisoida monia rutiininomaisia työvaiheita ja luoda tähän yhtenäisiä käytäntöjä. Samalla lähdekoodin avoin julkaisu verkossa tekee tutkimusprosessista toistettavan ja läpinäkyvän, mahdollistaa niin datan kuin välineiden tehokkaan jatkokäytön muissakin yhteyksissä, ja edistää laajemman tutkijaverkoston osallistumista kehitystyöhön.

Ideaalitilanteessa digitaalisten ihmistieteiden tutkimushankkeita toteuttavat tutkimusryhmät kootaan vastaamaan selkeisiin tutkimuskysymyksiin ja tutkimusmenetelmien kehitys ja soveltaminen tapahtuu ryhmän jäsenten yhteistyönä. Tärkeää on työn suunnittelu niin, että muualta lainattuja parhaita tieteellisiä käytäntöjä voidaan siirtää tehokkaasti humanistis-yhteiskunnallisen tutkimuksen piiriin. Hankkeeseen osallistuvien humanistien tulee omata riittävä ymmärrys alan vaatimuksista ja itseluottamusta niihin vastaamiseen. Tekstiaineistojen tutkimukseen liittyy kaksi toisiaan täydentävää lähestymistapaa, joita ovat rakenteettoman tekstin louhinta sekä erilaisten aineistojen kuvailutietojen tutkimuskäyttö joka pitää sisällään esimerkiksi kirjasto-, arkisto- kuin myös museoluetteloinnit (ks. ylläoleva kaavio). Näiden hyödyntäminen tutkimuksessa ei poista perinteisen lähiluennan merkitystä, vaan täydentää sitä ammentamalla aiheita laajemmasta perspektiivistä.

Tekstilouhinnan potentiaali on valtava. Jos aatehistorioitsija on esimerkiksi kiinnostunut tutkimaan oikeudenmukaisuuden käsitteen muuttumista brittiläisessä julkisessa keskustelussa, sitä voidaan periaatteessa tutkia esimerkiksi louhimalla kansainvälisiä aineistoja (esim. Early Modern English Books Online (EEBO) ja Eighteenth-Century Collections Online (ECCO)), jotka yhdessä käsittävät käytännössä kaikki varhaisella uudella ajalla, 1470-1800, Britanniassa

painetut kirjat. Hyvä esimerkki muista vastaavista aineistoista on ranskalainen kansallinen Gallica-projekti. Lisäksi erityisesti antiikin aatehistorialliseen tutkimukseen on tarjolla kasvavassa määrin).^[22] Samalla tähän tekstilouhintaan voidaan yhdistellä luovalla tavalla käsittehistorian metodeja. Käsitteen muutosta voidaan täten ryhtyä tutkimaan aivan eri tavoin kun käytössä on automatisoituja, laajojen tietoaaineistojen käsittelyyn skaalautuvia mallintamismenetelmiä. Kotimaisena esimerkkinä on kaikkien historiallisten suomalaisten sanomalehtien tekstilouhinta, jossa aineistoista haetaan automatisoidusti paikannimiä ja henkilöitä.^[23] OCR-menetelmä jonka avulla kuvattu aineisto käännetään koneluettavaan muotoon ei ole täydellinen; parhaimmassakin tapauksessa käännostarkkuus on noin 90%.^[24] Edellä esitellyn etäluennan kannalta potentiaaliset hyödyt ovat suurempia kuin mahdolliset puutteet, jos olemme kiinnostuneita yleisistä säännönmukaisuuksista ja muutoksista käsitteiden käytössä. Tekstilouhinnan sovellukset ovat kuitenkin vasta muotoutumassa, joten menetelmien tulevaa kehitystä on vielä vaikea arvioida. Varmaa on kuitenkin, että uusien menetelmien käyttö yleistyy ja muokkaa alan tutkimuskäytäntöjä.

Aatehistorian kannalta relevanttien tutkimusaineistojen saatavuuden ongelmat

Tutkimusaineistojen avoimen saatavuuden puutteet muodostavat tutkimuksessa merkittävän pullonkaulan. Tähän liittyy humanistis-yhteiskuntatieteellisessä tutkimuksessa erityisesti kolme keskeistä ongelmaa. Ensinnäkin digitoidut lähdeaineistot ovat usein saatavilla vain kaupallisten lisenssien tai luottamuksellisten yhteistyösopimusten kautta. Toiseksi monet aineistot ovat saatavilla ainoastaan rajatun verkkoliittymän kautta, eikä tutkija välttämättä pääse käsiksi koko aineistoon ja alkuperäisiin raakadatoihin joita tarvittaisiin kattavan kokonaiskuvan muodostamiseen tutkimusaineistosta. Nykytilanteessa tutkija voi esimerkiksi käyttää edellämainittua, kaikki 1700-luvun painetut brittiläiset kirjat sisältävää ECCO-tietokantaa valmiin käyttöliittymän kautta, mutta kallis lisenssi tarjoaa vain helpotettuja välineitä yksittäisten dokumenttien sanahakuun tai yksinkertaisiin tilastollisiin yhteenvetoihin, ei pääsyä alkuperäiseen raakatekstiin. Tästä on hyötyä lähinnä käyttäjille joilla ei ole varsinaisia tekstilouhinnan tarpeita. Vaikka käyttöliittymien työkaluvalikoima monipuolistuisi – esimerkiksi GALE-niminen yritys, joka on digitoinut merkittävän määrän historiallisia kirjoja, on kehittänyt 1800-luvun kokoelmiensa (Nineteenth-Century Collections Online, NCCO) selaamiseen suunnattuja tilastollisia välineitä – tämä merkitsee viime kädessä perinteisten tapojen viilaamista, joka ei uudista merkittävästi aineistojen kokonaiskäyttöä ja alan tutkimuskulttuuria. Rajoitettujen hakukoneiden käyttö ei konkreettisen lukemisen kannalta eroa yksittäisten kirjojen manuaalisesta selaamisesta, vaikka selaaminen ohjelmistojen avulla onkin helpottunut

merkittävästi. Tulevaisuuteen katsovilla tutkijoilla on kuitenkin nopeasti kasvava tarve monipuolisempaan tekstien analyysiin. Yksityiskohtaisempi ja alaa uudistava tekstilouhinta vaatii pääsyä koko alkuperäisaineistoon ja mahdollisuuksia sen vapaaseen käsittelyyn ja jatkojalostamiseen. Tähän liittyy kolmas aineistojen saatavuuden ongelma, aineistojen käsittelyyn räätälöityjen joustavien tieteellisten laskentakirjastojen puute. Valmiit graafiset käyttöliittymät eivät voi koskaan tarjota samaa joustavuutta aineistojen analyysissä kuin tutkijan itse kirjoittama ohjelmakoodi, jonka laadinnassa voidaan hyödyntää laskentakirjastojen kautta saatavia valmiita rakennuspalikoita.

Edistystä aineistojen avoimuuden ja laskentamenetelmien suhteen on kuitenkin tapahtumassa. Aikaisemmin mainitusta EEBO-kokoelmasta on tehty eri yliopistojen yhteistyössä EEBO-TCP (text creation partnership), jossa OCR-menetelmällä käännettyjä tekstejä on korjattu, editoitu ja annotoitu eri tavoin. Vuoden 2015 alusta kyseinen EEBO-TCP tietokanta avattiin yleiseen käyttöön.^[25] Samalla myös tutkijalähtöisesti esimerkiksi avoimen lähdekoodin rOpenGov-hankkeessa kehitetään tieteellisiä laskentakirjastoja nimenomaan humanistis-yhteiskuntatieteellisen tutkimuksen tarpeisiin.^[26] Myös valmiita analyysiohjelmistoja tarjotaan yhä useammin avoimen lähdekoodin toteutuksina. Nähtäväksi jää mikä näiden avausten merkitys tulee olemaan alan kehitykselle. Suljetut EEBO- ja ECCO-tietokannat ovat olleet Suomessakin kaupallisesti rajoitetun käyttöliittymän kautta saatavilla jo melkein 15 vuotta, mutta tämän merkitys on jäänyt aatehistorian metodiikan kannalta vähäisemmäksi kuin mitä voisi toivoa. Samalla, totta kai, alkuperäislähteiden saatavuus on helpottunut moninkertaisesti, mutta aineiston jatkojalostaminen varsinaista kvantitatiivista tutkimusta kohden ei ole useiden aineistojen kohdalla vielä edes alkanut. Yksi perusteltu väite onkin, että digitaalisten ihmistieteiden ensimmäinen aalto ei ole vielä ehtinyt rantaan asti ennen kuin innokkaimmat ovat alkaneet puhua uudesta. EEBO-TCP:n avaaminen on merkittävä askel, mutta samalla alalla tarvittaisiin laajempaa pääsyä kokonaisiin alkuperäisaineistoihin ja varsinaiseen raakadataan (tässä tapauksessa koneluettavaan tekstiin). Esimerkiksi suomalaisten sanomalehtien avaaminen yleiseen käyttöön on aatehistoriankin kannalta merkittävää.^[27] Toivottavasti yhä useammat aineistonhaltijat tällaisista institutionaalisista avauksista rohkaistuneina auttavat pilotoimaan uudenlaisia tutkimustapoja.

Huolellisella editoinnilla varustetuilla tietokannoilla on merkittävää lisäarvoa, mutta lisäksi olisi tärkeää tutkia editoimattomien alkuperäisaineistojen käytön mahdollisuuksia tutkimuksessa. Nykyaikaiset tietojenkäsittelymenetelmät tarjoavat mielenkiintoisia vaihtoehtoja alkuperäisaineistojen automatisoituun siistimiseen, rikastamiseen ja jalostamiseen samalla kun hiomatonta, mutta tutkimuksen kannalta käyttökelpoista ja potentiaalisesti arvokasta dataa

on olemassa huomattavasti suurempia määriä kuin huolella viimeisteltyjä aineistokokoelmia. Käytännössä kokemuksemme mukaan on kuitenkin usein erittäin haastavaa saada pääsyä aatehistorian tutkimuksessa keskeisiin alkuperäisaineistoihin edes luottamuksellisen tutkimusyhteistyön kautta, puhumattakaan samojen aineistojen avoimesta julkaisemisesta verkossa.^[28] Keinotekoisien kaupallisten ja teknisten rajoitteiden yleisyys digitoitujen aineistojen saatavuudessa poikkeaa merkittävästi luonnontieteiden tutkimuksessa nykyisin yleisestä käytännöstä, jossa mittausaineistojen raakadatat tallennetaan julkaisun yhteydessä avoimiin tietokantoihin.^[29] Selittävänä tekijänä ovat erot tutkimuskulttuurissa sekä aineistojen jakelun järjestämisessä ja rahoitusmalleissa.

Luonnontieteellinen tutkimus on ollut edelläkävijä aineistojen avoimessa jakelussa, ja alalla on tiedostettu alkuperäisaineistojen sujuvan saatavuuden merkitys tutkimuksen läpinäkyvyydelle ja avoimuudelle joka on yksi tieteellisen toiminnan kulmakivistä. Aineistojen jakeluun esimerkiksi hiukkasfysiikassa ja molekyylibiologiassa on perustettu keskitettyjä julkisia tietokantoja, joiden pitkän tähtäimen rahoitus on turvattu.^[30] Humanistis-yhteiskuntatieteellisen tutkimuksen lähdeaineistot on usein koottu kaupallisen mallin puitteissa tai esimerkiksi kirjastojen toimesta sillä ajatuksella, että niiden käyttöoikeuksia voidaan myydä. Tämä johtuu osittain siitä, ettei avoimen julkaisemisen ole perinteisesti nähty tuovan merkittävää lisäarvoa. Tulevaisuudessa tekstimassoista voitaisiin kuitenkin esimerkiksi tunnistaa ohjaamattomasti erilaisia käsitteitä ilman ennalta asetettuja hierarkkisia kategorioita, jotka ovat aina tavalla tai toisella mielivaltaisia. Näin perustavanlaatuiset aatehistorian metodologiset kysymykset, kuten skinneriläinen kontekstualismi ja koselleckilainen käsitehistoria voivat tulevina vuosikymmeninä näyttäytyä uudenaikaisessa perspektiivissä. Tämä vaatii, että substanssi- ja menetelmäosaaminen löytävät toisensa ja aatehistorian tutkimuskysymysten ehdoilla etenevä tutkimus voi ammentaa kieliteknologiasta, tilastollisesta koneoppimisesta ja muista informaatiotieteiden traditioista. Samalla olisi äärimmäisen tärkeää, että yhteistyötä tutkimuksen ja raakadataa useimmissa tapauksissa hallinnoivien kirjastojen sekä muiden muistiorganisaatioiden kesken syvennetään huomattavassa määrin.

Tieteellisen laskennan ja aineistojen jakeluun suunnattujen teknologioiden ja palveluiden kehittymisen myötä tutkimusaineistojen suljetut ja kaupalliset jakelumallit ovat nopeasti vanhentumassa samalla kun uudet menetelmät tarjoavat tehokkaita mahdollisuuksia digitaalisten aineistojen raakadatan hyödyntämiseen tutkimuksessa. Tutkimuksen avoimuudelle esitetyt vaatimukset ovat myös vahvistuneet. Niin tutkimuksen rahoittajat, julkaisijat, suuri yleisö, kuten myös valveutuneet tutkijat vaativat yhä useammin koko tutkimusprosessin avaamista tavalla joka sisältää tutkimuksen yleisen kuvaamisen lisäksi myös pääsyn alkuperäiseen raa-

kadataan ja analyysikoodeihin.^[31] Myös digitaalisten aineistojen jakeluun on tarjolla tehokkaita avoimen lähdekoodin tietokanta- ja rajapintaratkaisuja tutkimuslaitosten tarpeisiin. Yksittäiset tutkijat voivat julkaista dataa hajautetuissa versionhallinta- ja jakelupalveluissa. Myös tavanomaisia kirjoja tai tutkimusartikkeleita kirjoitetaan erityisesti yhteistyöprojekteissa kasvavassa määrin tällaisten palveluiden kautta, josta esimerkin tarjoavat vaikkapa Authorea ja GitBook-palvelut.^[32]

Avoimempi tutkimuskulttuuri onkin vauhdilla leviämässä myös humanistis-yhteiskuntatieteelliselle aloille. Jos tahtoa löytyisi, alan lähdeaineistojen julkaiseminen voisi olla monessa tapauksessa helpompaa kuin luonnontieteen aloilla, koska aineistot ovat kooltaan huomattavasti pienempiä ja esimerkiksi historiallisten tekstiaineistojen kohdalla niihin liittyy vähemmän yksityisyydensuojasta tai tekijänoikeuskysymyksistä nousevia rajoitteita tai salausvaatimuksia kuin vaikkapa biolääketieteen aineistoihin. Digitaalisten alkuperäisaineistojen julkaiseminen onnistuisi pienemmällä taloudellisella ja teknisellä panoksella kuin luonnontieteellisten aineistojen kohdalla. Aineistojen tutkimuskäytön tarpeeton kaupallinen tai tekninen rajaaminen onkin ongelma, jonka ratkaisemiseen alalla tulisi kiinnittää erityistä huomiota. Ei riitä, että aineistojen hallinnasta vastaavilla tahoilla on retoriikka kunnossa ja tutkijat kopioivat toisiltaan lauseet jotka kirjoittaa rahoitushakemuksen aineistojen ja tutkimusdatan avoimuutta koskevaan kohtaan. Tarvitaan myös konkreettisia toimia.

Avoin data^[33] viittaa digitaalisten aineistojen osalta siihen, että aineisto on saatavilla kokonaisuudessaan, maksutta tai minimaalisin tuotantokustannuksin koneluettavassa muodossa, eikä sen jatkokäytölle ole asetettu teknisiä, juridisia, kaupallisia tai muita rajoitteita. Aineistoa voi siis jalostaa, yhdistellä muihin aineistoihin, jakaa edelleen ja käyttää haluamallaan tavalla. Ainoa ehto avoimen aineiston käytölle on viittaaminen alkuperäiseen lähteeseen tavanomaisen tieteellisen käytännön mukaisesti. Tämä avoimen datan määritelmä on sopusoinnussa tieteen avoimuuden kanssa ja siitä tehtävät poikkeamat heikentävät samalla koko tutkimusprosessin avoimuutta ja arvoa. Aineistojen avoimen saatavuuden puolesta voidaan argumentoida erityisesti silloin kun aineistojen keruu ja ylläpito on toteutettu julkisin varoin, esimerkiksi tutkimuslaitosten tai kirjastojen toimesta. Joillakin muilla aloilla tehtyjen arvioiden mukaan avoimuuden tuottama lisäarvo tutkimukselle ja yhteiskunnalle saattaa olla merkittävästi suurempi kuin avoimuuden edistämiseen käytettävä taloudellinen panos.^[34] Tämä on suora vastaus kysymykseen kuka maksaa. Avoin datan hyödyt tieteen kehitykselle ovat ilmeisiä – eräässä äskettäin julkaistussa katsausartikkelissa pyydettiin osuvasti pohtimaan vaikutuksia molekyylibiologian kehitykselle, mikäli ihmisen geenikartasta olisi päätetty julkaista pelkästään 'kiinnostavat' osat.^[35] Sama pätee digitaalisten ihmistieteiden alalla – tällaisten

perustavan laatuisten kysymysten pohdinta kuuluu koko tutkijayhteisölle – ei yksittäiselle aineiston tarjoajalle. Tutkimuksen kannalta arvokkaita aineistoja on toki myös kaupallisesti saatavilla. Esimerkiksi Google Books tarjoaa merkittäviä tekstiresursseja, mutta sen sisältämä aineisto ei ole avoimesti saatavilla. Kaupallisten toimijoiden osalta avoimuuden vaatiminen on kuitenkin ongelmallisempaa kuin julkisten toimijoiden kohdalla. Tutkimuksen kannalta keskeisten digitaalisten aineistojen avaaminen tutkimuskäyttöön alkuperäisten raakatekstien ja täydellisten kuvailutietojen osalta on joka tapauksessa tärkeää humanistis-yhteiskuntatieteellisen alan uusien tutkimusmahdollisuuksien edistämiseksi.

Avoimen tieteen ja tutkimusprosessin läpinäkyvyyden merkitys aatehistorian tutkimukselle

Avoimen datan lisäksi keskeistä tutkimuksen läpinäkyvyydelle, toistettavuudelle ja edelleen käytölle on lisäksi tilastollisten analyysien taustalla olevien lähdekoodien avoin julkaiseminen.^[36] Avoimesta lähdekoodista keskusteleminen saattaa tuntua aatehistorian kontekstissa vielä tänään vieraalta, mutta erilaisten laskennallisten menetelmien käytön lisääntyessä tulee olemaan merkittävää, jos tutkimuksessa keskitytään alusta lähtien avoimiin toimintatapoihin sen sijaan, että käytettäisiin pelkästään valmiita ja rajoitettuja ohjelmistoja. Tämäkin käytäntö on nopeasti yleistynyt erityisesti luonnontieteen alalla ja vertaisarvioitujen tutkimusten laadintaan liittyviä lähdekoodeja julkaistaan hajautetuissa versionhallintapalveluissa normaalina käytäntönä.^[37] Tämä takaa parhaassa tapauksessa, että jokainen tutkimuksessa tehty tilastollisen analyysin yksityiskohta alkuperäisestä raakadatasta lopullisiin yhteenvetoihin ja visualisointeihin on täydellisen läpinäkyvä. Se myös vähentää mahdollisuuksia virheiden piilotteluun ja 'tilastokalasteluun', jossa tilastollisen merkitsevyyden kunnollinen testaus jätetään vahingossa tai jopa tahallisesti huomiotta seulottaessa laajoja tietoaaineistoja.^[38] Tämä vähentää myös rutiininomaisten työvaiheiden toistoa ja potentiaalisesti skaalautuu äärimmäisen hyvin muiden vastaavien aineistojen tutkimukseen. Aatehistoriallisiin lähteisiin liittyen esimerkiksi kirjastoluetteloiden kuvailutiedot ovat varsin yhtenäisiä ylläpitäjästä riippumatta. Tämä tarjoaa mahdollisuuksia eri maiden kirjastoluetteloiden yhdistämiseen ja tiedontuotannon tutkimiseen kansainvälisellä tasolla. Avoimen toimintamallin sivutuotteena syntyy perinteiset rajat ylittävää yhteistyötä ja uudenlaisia tutkimusnäkökulmia.

Tutkimusaineistojen ja menetelmien avoimuus parantaa tehokkuuden lisäksi myös tutkimuksen toistettavuutta. Nykytilanteessa tutkija saattaa tehdä suljettujen käyttöliittymien avulla monimutkaisiakin tietokantahakuja ja muodostaa niiden perusteella erilaisia päättelyketjuja tiettyjen termien käytön historiallisesti kehityksestä, mutta hauista, niiden sisällöstä ja päät-

telyketjusta ei välttämättä jää edes tutkijalle itselleen selkeää kirjanpitoa. Analyysien automatisointi tarjoaa ongelmaan ratkaisun. Samalla muutkin tutkijat voivat hyödyntää avointa lähdekoodia omassa työssään ja kehittää sitä halutessaan edelleen omiin tarpeisiinsa ja jakaa työnsä tulokset takaisin yhteisön käyttöön. Globaalin tiedeyhteisön näkökulmasta tällainen malli voi vähentää merkittävästi päällekkäistä työtä ja tarjoaa mahdollisuuksia verkostomaiseen, perinteiset institutionaaliset ja maantieteelliset rajat ylittävään yhteistyöhön joka hyödyttää kaikkia alan toimijoita ja kiihdyttää alan tutkimusta. Hajautettu versionhallinta mahdollistaa laajemman tiedeyhteisön osallistumisen välineiden kehittämiseen ja soveltamiseen. Vastaavaa lähestymistapaa on sovellettu muun muassa bioinformatiikan tutkimuksessa menestyksellä jo pitkään. Murros on jo käynnistynyt myös humanistis-yhteiskuntatieteellisellä alalla, mutta uusien käytäntöjen vakiintuminen vie aikaa. On silti erittäin tärkeää, että aatehistorian kohdalla laskennallisten menetelmien suhteen alaa suunnataan kohti parhaita tieteellisiä toimintamalleja. Nopeat omaksujat voivat saada etulyöntiaseman ja tehdä merkittävää pioneerityötä muokaten alan kehityssuuntaa.

Viime aikoina esillä ollut avoimen tieteen käsite korostaakin läpinäkyvien ja yhteisöllisten toimintamallien merkitystä. Perinteisesti avoimen tieteen käsitteen katsotaan viittaavaan mm. tutkimuksen lähdeaineistojen, tutkimusmenetelmien ja tieteellisen julkaisemisen avoimuuteen. Avoimuus sisältää vapauden tallentaa, muokata ja jakaa edelleen tutkimusmateriaalia tavanomaisten tieteellisten viittauskäytäntöjen mukaisesti ilman muita rajoituksia. Tutkimusaineistojen koon kasvaessa ja niiden käsittelyyn tarvittavien informaatiotekniikan menetelmien monimutkaistuessa tutkimuksen avoimuus on kohdannut uudenlaisia haasteita. Näihin luonnontieteissä vakavasti otettuihin kysymyksiin ei ole kuitenkaan samalla tavalla vielä kiinnitetty huomiota ihmistieteiden tutkimustradition piirissä joka on avointen käytäntöjen osalta jäänyt kansainvälisesti jälkeen luonnontieteistä. Kun luonnontieteen alalla tutkimuslaitokset jopa kilpailevat avointen tietokantojen laadulla^[39], humanistis-yhteiskuntatieteellisen tutkimuksen kannalta merkittäviä aineistoja ylläpitävät laitokset vaikuttavat toisinaan välinpitämättömiltä aineistojen avoimen tutkimuskäytön edistämisen suhteen. Myös kannustimet ovat perinteisesti olleet heikkoja, mutta viime aikoina monet rahoittajat ja julkaisijat ovat alkaneet edellyttää avointa julkaisemista tukevaa aineistohallintaa. Tutkimustoiminnassakin voi vielä törmätä kirjaviin näkemyksiin ja jopa virheellisiin käsityksiin tutkimusaineistojen ja julkaisemisen avoimuuteen liittyen. Tämä on helmasynti humanistis-yhteiskunnallisessa tutkimuksessa pitäen sisällään aatehistorian. Jos tutkijat eivät itse osaa vaatia aineistojen ja tutkimustulosten vapaata saatavuutta, sitä ei voitane olettaa aineistojen ylläpitäjiltäkään. Aineistoja hallinnoivat tahot, kuten kirjasto- ja muistiorganisaatiot, tulisikin sitoa lähdeaineistojen ja menetelmien avaamisen osalta tiiviimmin mukaan tutkimukseen. Lisäksi tulisi tutkia mah-

dollisuuksia hyödyntää viranomaisten ja muiden tahojen kasvavassa määrin tarjoamien avoimen datan voimavarojen hyödyntämistä tutkimuksen tukena. Argumentti siitä miten tähän parhaiten päästään, on tiivistää merkityksellistä yhteistyötä tutkimuksen ja sitä tukevien organisaatioiden välillä.

Avoimet käytännöt voivat uudistaa tieteenalan toimintakulttuuria ja suuntaa. Kansainvälisessä kilpailussa on pystyttävä tekemään hajautettua yhteistyötä, jossa yhteisten pelisääntöjen kautta pyritään avaamaan koko tutkimusprosessi niin että samoista tai samankaltaisista kysymyksistä kiinnostuneiden tutkijoiden ei tarvitse keksiä pyörää uudestaan, vaan voidaan yhteisvoimin kehittää rikkaampia ja laadukkaampia kokonaisuuksia. Esimerkiksi tekstilouhinnan osalta kotimaisten tutkimusryhmien kannattaisi tehdä yhteistyötä suomalaisten sanomalehtiaineistojen jalostamisessa ja analysoinnissa, jotta niukat tutkimusresurssit eivät valuisi hukkaan tutkimusryhmien tehdessä päällekkäistä työtä. Humanistilla aloilla puhutaan usein avoimesta julkaisemisesta, mutta vähemmän tutkimusaineistojen ja tutkimusprosessien avoimuudesta, vaikka nämä saattavat olla tutkimuksen kehittymisen kannalta jopa vielä keskeisempiä. Paradoksaalisesti tätä merkitystä ei aina havaita, kun ajatellaan, että tiedeyhteisön kannalta hyödyllisin julkaisuuyksikkö on lopullinen, useaan kertaan puntaroitu ja viilattu tutkimustulos. Aatehistoria on tässä asioiden viimeistelyssä esittänyt ajoittain jopa humanistisessa kontekstissa ylilyövää pikkutarkkuutta. Kuitenkin myös tutkimusmenetelmät ja tutkimusprosessin aikana käytävä ajatustenvaihto ovat keskeisiä tutkimuksessa. Näiden avoin jakaminen tiedeyhteisön kesken tarjoaa laadullisten tulosten lisäksi myös arvokasta materiaalia ja välineitä joiden pohjalta voidaan tehokkaasti rakentaa uutta tutkimusta. Avoimia käytäntöjä omaksuneilla tieteenaloilla on hyödynnetty linux-maailmasta tuttua ”julkaise aikaisin ja usein”-periaatetta, jonka mukaan koko kvantitatiivinen tutkimusprosessi ja sen eteneminen avataan yleisölle.^[40] Avoimuuden merkitystä tutkimusetiikan ja metodiikan ytimessä tuleekin korostaa. Näin tutkimusprosessi ja -tulokset liikkuvat sujuvammin tutkijoiden ja tutkimusalojen välillä ja niitä voidaan paremmin hyödyntää innovaatiotoiminnassa, opetuksessa ja kansalaisyhteiskunnassa.

Terve skeptisismi

Aatehistorian piirissä on oltu pidättyväisiä digitaalisten aineistojen käyttöön ja tiedonlouhintaan koneavusteisin menetelmin samalla kun kieliteknologia ja siihen liittyvät laskentamenetelmät ovat kehittyneet jo ainakin kolmenkymmenen vuoden ajan humanistisenkin tutkimustradition puitteissa.^[41] Perinteisessä aatehistorian tutkimuksessa yksittäinen sankaritutkija lukee ja kirjoittaa, eikä automatisoinnilla tai tieteellisellä laskennalla nähdä juurikaan lisä-

arvoa. Toisaalta aatehistorian tutkijat eivät ole välttämättä olleet tietoisia uusien informaatiotieteiden menetelmien mahdollisuuksista. Itsenäistä laajaan asiantuntemukseen pohjautuvaa luovaa ajattelua mikään kone ei tee tutkijan puolesta. Tutkimusta voitaisiin kuitenkin monissa tapauksissa oleellisesti tehostaa ottamalla käyttöön tilastollinen kehys laajempien aineistojen tarkasteluun. Tiedonlouhinnalla ja digitaalisten aineistojen käytöllä voidaan myös laajentaa tutkimuksen aikajännettä sujuvasti vuosikymmenistä vuosisatojen aikana tapahtuviin muutoksiin.^[42] Tämän suuntainen ajattelu ei ole uutta ja ainakin käsitehistorian piirissä tätä mahdollisuutta on Suomessakin pohdittu pidempään.^[43] Samalla digitaalisten aineistojen mahdollisuuksien realisoituminen aatehistorian tutkimuksessa vaatii muutoksia tutkimuskulttuurissa monitieteisen yhteistyön osalta.

Humanistisessa ja historian tutkimustraditiossa monitieteisten ja ennakoluulottomasti uusia informaatiotieteen menetelmiä perinteisten kysymysten ratkaisemisessa hyödyntävät tutkijaryhmät ovat yhä harvinaisia.^[44] Konkreettinen ongelma on myös, että useat digitaalisen ihmistieteiden hankkeet keskittyvät perifeerisiin aiheisiin eivätkä tutkimustradition keskiössä olevat henkilöt kykene ymmärtämään digitaalisen komponentin lisäarvoa ennen kuin tutkimuksella on vaikuttavuutta tieteenalan perusongelmien ratkaisemisessa. Lisäksi monet visualisointiin liittyvät ratkaisut eivät välttämättä tuota varsinaisia tuloksia, vaan keskittyvät tapaan miten tietoa esitetään. Joskus tämä saattaa valitettavasti olla myös puhdasta sirkusakrobatiaa. Samalla on selvää, että informaatiotieteen ja tieteelliseen laskentaan liittyvän menetelmäosaamisen puutetta ihmistieteiden ja erityisesti humanismin tutkimusperinteessä on paikattava. Tämä ei missään muodossa ole helppo tehtävä, koska täysverisistä laskennallisen tieteen ammattilaisista on pulaa kaikilla sovellusaloilla. Aatehistoriaan liittyvien tieteellisen laskennan sovellusten viive saattaa selittyä sillä, että aatehistorian ydinosajat eivät ole lähteneet menetelmiä kehittämään ja näin ei ole muodostunut aatehistorian tarpeisiin vastaavia tietojenkäsittelyä hyödyntäviä menetelmiä. Uusi tutkimusala tarjoaakin lupaavia mahdollisuuksia ja tuoreita tutkimuskohteita myös menetelmätutkijoille. Haasteena on, että laskennallisten tieteiden edustajat eivät välttämättä osaa ennakoida aatehistorioitsijoiden tarpeita. Sovellettavia menetelmiä löytyy, mutta tutkimuskulttuurit eivät ole vielä kohdanneet merkittävällä tavalla.^[45] Uusien lähestymistapojen ilmestyessä tutkimuskenttään korostuu samalla tarve korostaa humanistisen tutkimustradition omien tutkimuskysymysten merkitystä ja ylipäänsä hahmottaa mikä on aatehistoriallinen perustutkimuksen ydin. Mutkikkaat laskentamenetelmät ja hienot visualisoinnit eivät yksinään riitä vastaamaan sovellusalan keskeisiin kysymyksiin. Toisaalta vaikka teoreettiset keskustelut ovat tärkeitä, käytännön tulokset ja ymmärrys sanelevat tutkimuksen lopullisen arvon. Mikään tutkimusmenetelmä tai aineisto maailmassa ei korvaa hyviä tutkimuskysymyksiä ja -hypoteeseja. Tutkimuskysymyksen mää-

rittely ja siihen vastaaminen saatavilla olevien aineistojen ja tutkimusmenetelmien perusteella ratkaisee viime kädessä onnistumisen aatehistorian saralla. Jos tutkimusta tehdään liian aineisto- tai menetelmälähtöisesti, sen vaikuttavuus voi helposti jäädä näennäiseksi.

Lopuksi

Kuvailutietojen, tekstinlouhinnan ja informaatiotieteiden yhdistely tarjoaa uusia tapoja ymmärtää esimerkiksi painotuotteita fyysisinä objekteina, niiden tuottamiseen liittyviä näkökulmia, sekä laajempia käsitteiden muutoksia.

Digitaalisten aineistojen osalta aatehistorian tutkimuksen kannalta käyttökelpoisia aineistoja on jo hyvin tarjolla. Esimerkiksi Suomen kansalliskirjaston kokoelmat ovat tällaisia. Kotimaiset aineistot tarjoavat kiintoisan ja luontevan tutkimuskohteen aatehistorian tutkimukselle. Lisähaasteena tutkimuksessa on kehittää menetelmiä joilla tekstiaineistoita yhdistäviä ja erottavia piirteitä voitaisiin sujuvasti analysoida yli kielirajojen. Suomen kaksikielisyys tarjoaa myös erinomaisia mahdollisuuksia pilotoida rajatummassa mittakaavassa eri kielten samanaikaista käyttöä ja yhdistelyä tekstilouhinnassa. Suomalaisten sanomalehtien tutkimuksessa voidaan yhdistää monia kotimaisen aatehistorian osaamisen parhaita puolia liittyen aineiston sisältöön, käsitehistoriaan ja lingvistiikkaan. Lisäksi tämä tarjoaa erinomaisia mahdollisuuksia kehittää yhteistyötä tieteellisen laskennan asiantuntijoiden kanssa tavalla joka mahdollistaa uusien tutkimusideoiden syntymisen sekä ihmistieteiden tutkimukseen että tieteellisen laskennan teorian ja menetelmien kehittämiseen. Ihmistieteiden saralla seuraavaksi tavoitteeksi voitaisiin ottaa pyrkimys kvantitatiivisten tutkimusmenetelmien yhtenäistämiseen ja yhteistyön edistämiseen luonnontieteissä jo vakiintuneen avoimen toimintamallin keinoin. Samalla tulee kehittää alan opetusta, jotta pystymme jatkossa auttamaan uusia sukupolvia haastamaan skinnereitä ja koselleckeja aatehistorian omista lähtökohdista, mutta mahdollisesti osaten ammentaa tähän lisäarvoa monitieteisessä yhteistyössä. Olisikin hyvä kehittää aatehistorian ja digitaalisten ihmistieteiden yhteisiä linjauksia alan kotimaisessa opetuksessa. Humanistisen tutkimustradition sosiaaliseen muutokseen liittyy monitieteisyyden, uusien tutkimusmenetelmien ja aineistojen käytettävyyden edistäminen. Tarvitaan konkreettisia ja pitkäjänteisiä toimia uusien avointen toimintamallien tuomiseksi alan toimintakulttuuriin. Samalla keskeistä on, että laskennan asiantuntijoita saadaan mukaan ihmistieteiden piiriin tavalla joka palvelee ihmistieteiden perinteisten ydinteemojen tutkimusta.

Mikko Tolonen on digitaalisten aineistojen tutkimuksen (vt.) professori ja aatehistorian tutkija Helsingin yliopistolla.

Leo Lahti on laskennallisen tieteen sovelluksiin ja avoimeen tieteseen erikoistunut Suomen Akatemian tutkijatohtori, joka toimii tällä hetkellä päätoimisesti Helsingin yliopistolla sekä Alankomaissa Wageningenin yliopistossa.

1. Kiitämme Jani Marjasta, Niko Ilomäkeä, Kaius Sinnemäkeä, Sirkka Havua, Eeva Murtomaata sekä kahta referee-lausunnon kirjoittajaa. Tekstin runko perustuu vuoden 2014 lopulla Oulussa pidettyyn esitelmään. Kiitokset Jouni-Matti Kuukkaselle sekä Oulun yliopiston historian filosofian tutkimuskeskuksen yleisölle stimuloivasta keskustelusta digitaalisten aineistojen mahdollisuuksista. [Takaisin]
2. Kielen ja kuvan yhdistämisestä suomalaisessa tutkimuksessa, ks. Tuomo Hiippalan väitöskirja, *Modelling the structure of a multimodal artefact* (2013): <https://helda.helsinki.fi/handle/10138/41736>. [Takaisin]
3. Laajempi katsaus aiheeseen löytyy tämän lehden sivuilta: <http://www.ennenjanyt.net/2015/02/koodaamisen-ja-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/> [Takaisin]
4. Esimerkki uuden profiilin tutkimuskeskuksesta on Sussex Humanities Lab, jossa painopiste on muualla kuin perinteisessä tekstien ja kielen tutkimuksessa. <https://humslab.wordpress.com> [Takaisin]
5. Ihmistieteiden alaa sivuavia ajankohtaisia trendejä ovat esimerkiksi "biohakkerointi", jossa henkilö monitoroi elintoimintojaan kirjanpidolla ja mittalaittein ja voi saadun tiedon perusteella pyrkiä optimoimaan päiväryhtiään, ruokavaliotaan ja liikuntatottumuksiaan. "Oma data"-liikkeessä edistetään henkilökohtaisten ja toisi-naan luottamuksellistenkin rekisteritietojen saatavuutta omaan henkilökohtaiseen käyttöön. [Takaisin]
6. Esim. <http://organisaatioviestinta.com/2015/01/31/onko-laskennallinen-yhteiskuntatiede-viimein-saapu-massa-suomeen/> [Takaisin]
7. David Lazer et al. 'Computational Social Science.' *Science* 323(5915), pp. 721-723, 2009 [Takaisin]
8. Hans-Georg Gadamer, *Reason in the Age of Science*, trans. Frederick G. Lawrence. Cambridge, MA: MIT Press, 1981. [Takaisin]
9. Lähtökohtaisesti esim. digitaalisia ihmistieteitä edustavan Dariahin (<https://www.dariah.eu/about/mision.html>) ja kieliteknologian perinteestä kasvavan Clarin-verkoston (<http://www.clarin.eu/content/mission>) välillä nähtiin jännitteitä, mutta kansainvälinen suuntaus on tällä hetkellä näiden verkostojen luonnollinen yhdentyminen. [Takaisin]
10. ks. myös <http://widerscreen.fi/numerot/2013-2-3/kieltaydyn-maaritlemasta-digitaalista-kulttuuria/> [Takaisin]
11. Franco Moretti, 'Conjectures of world literature', *New Left Review*, 1, 2000, 54-67 and Moretti, 'More conjectures', *New Left Review*, 20, 2003, pp. 73-81. Morettin tekstejä on kerätty kokoelmaan: *Distant Reading*, Verso, 2013. [Takaisin]
12. Erinomaisena osoituksena laajojen aineistojen etäluennan käyttömahdollisuuksista: Sara Klingenstein, Hitchcock, Tim and DeDeo, Simon, 'The civilizing process in London's Old Bailey', *Proceedings of the National Academy of Sciences*, 111, 2014, pp. 9419-8242. Ks. myös <http://www.oldbaileyonline.org/> [Takaisin]
13. Esimerkki tällaisesta menestyksekkäästä korpus-tutkimuksesta on Helsingin yliopiston englanninkielen variaation tutkimusyksikkö, VARIENG, <http://www.helsinki.fi/varieng/> [Takaisin]

14. Quentin Skinner, *Visions of Politics Volume 1. Regarding Method*, Cambridge University Press, 2002. [Takaisin]
15. Skinner, *Visions of Politics*, s. 57-102, 159. [Takaisin]
16. Reinhart Koselleck, 'Einleitung', Otto Brunner, Werner Conze & Reinhart Koselleck (hrsg.), *Geschichtliche Grundbegriffe. Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*. Band I, Stuttgart, Klett-Cotta 1972, s. XIII-XXVII. Reinhart Koselleck, *Vergangene Zukunft. Zur Semantik geschichtlicher Zeiten*, Frankfurt am Main, Suhrkamp 1979. Kari Palonen on, aivan oikein, alleviivannut skinneriläisen kontekstualismin yhteyttä käsitehistoriaan, Palonen, 'The History of Concepts as a Style of Political Theorizing. Quentin Skinner's and Reinhart Koselleck's Subversion of Normative Political Theory', *European Journal of Political Theory*, 1, 2002, pp. 91-106. Katso myös, Melvin Richter, 'Toward a Lexicon of European Political and Legal Concepts: A Comparison of Begriffsgeschichte and the 'Cambridge School'', *Critical Review of International Social and Political Philosophy*, 6, 2003, pp. 91-120. Esimerkkinä eteenpäin katsovasta käsitehistoriaprojektista voidaan mainita Bernhard Jussenin johtama keskiaikahanke, <http://www.geschichte.uni-frankfurt.de/43010469/jussenb>; Mehler, Alexander; Schwandt, Silke; Gleim, Rüdiger; Ernst, Alexandra, *Inducing linguistic networks from historical corpora. Towards a new method in historical semantics*, in: Paul Bennett, Martin Durrell, Silke Scheible, Richard J. Whitt (Hrsg.), *New Methods in Historical Corpora*, Tübingen 2013, S. 257-274. Kiitämme tässä erityisesti Jani Marjasta käsitehistoriaprojektien nykytilaan liittyvästä katsauksesta. [Takaisin]
17. Jouni-Matti Kuukkanen on tosin ehtinyt jo ryhtyä tähän jo ennen kuin on alettu pohtia ajatuksia uusien koneavusteisten menetelmien mahdollisuuksista: Jouni-Matti Kuukkanen, 'Making Sense of Conceptual Change', *History and Theory*, 47, 2008, pp. 351-372. [Takaisin]
18. <https://github.com/rOpenGov/fennica>. Tässä julkaisussa Fennicaan liittyvät graafit ovat Niko Ilomäen laatimia. [Takaisin]
19. Mielenkiintoista kehitystä ja seurattavaa toki riittää, niistä tässä mainittakoon: <http://benschmidt.org/projects/digital-humanities-research/> [Takaisin]
20. Suomalaisten sanomalehtien tekstilouhinnassa, suuntaa näyttää tällä hetkellä Tampereen yliopiston Informaatiotieteiden yksikön piirissä tehtävää työ. Myös muita aloitteita mm. Helsingissä ja Turussa on vireillä. [Takaisin]
21. Tämän dynamiikan tutkimuksen idean "isä" on Jani Marjanen tässä meidän välisen keskustelun kontekstissa. [Takaisin]
22. Latinankielisestä aineistosta yleisemmin ks. www.brepolis.net. [Takaisin]
23. <http://digi.kansalliskirjasto.fi/sanomalehti> [Takaisin]
24. Kimmo Kettunen et al. 'Analyzing and Improving the Quality of a Historical News Collection using Language Technology and Statistical Machine Learning Methods.' In: *Proc. of IFLA 2014*, Lyon (2014). [Takaisin]
25. <http://www.textcreationpartnership.org/2014/12/24/eebo-tcp-phase-i-public-release-what-to-expect-on-january-1/> [Takaisin]
26. ropengov.github.io [Takaisin]
27. FINCLARIN on avannut tätä aineistoa toukokuun 2015 lopulla ladattavassa muodossa vuoteen 1875 asti, <https://korp.csc.fi/download/Digilib-pub/> Kansalliskirjasto tulee todennäköisesti jatkamaan tätä vuoteen 1910 asti lähitulevaisuudessa, jakelukanava saattaa hyvin olla kyseinen FINCLARIN ja CSC. [Takaisin]
28. Esimerkiksi Dewey-järjestelmän saaminen tutkimuskäyttöön on osoittautunut erittäin hankalaksi. Samanlaisia kokemuksia liittyy niin ECCO-aineiston raakadatan saamiseen kuin myös eri kirjastoluetteloiden marc-tiedostojen saamisesta tutkimuskäyttöön. [Takaisin]
29. FS Collins et al. The human genome project: lessons from large-scale biology. *Science* 300(5617):286-290, 2003. [Takaisin]
30. Ks. esim. ihmisen perimän kartoitushanke GenBank <http://www.ncbi.nlm.nih.gov/genbank/> ja Euroopan

- hiukkasfysiikan tutkimuskeskuksen CERNin avoimen datan portaali: <http://opendata.cern.ch/> [Takaisin]
31. <http://www.sciencemag.org/content/336/6078/159.summary> [Takaisin]
32. <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001747> [Takaisin]
33. <https://www.authorea.com/>; <https://www.gitbook.com/> [Takaisin]
34. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195> [Takaisin]
35. <http://www.worldbank.org/content/dam/Worldbank/document/Open-Data-for-Economic-Growth.pdf>; http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation [Takaisin]
36. Mark Watson. When will 'open science' become simply 'science'? *Genome Biology* 16:101, 2015. [Takaisin]
37. <http://www.sciencemag.org/content/336/6078/159.summary>; <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001747> [Takaisin]
38. Esim. Github ja Bitbucket [Takaisin]
39. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002106> [Takaisin]
40. <http://bib.oxfordjournals.org/content/2/2/143.short> [Takaisin]
41. Eric S. Raymond (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media. ISBN 1-56592-724-9. [Takaisin]
42. Helsingin yliopistossa esim. Kimmo Koskeniemi. [Takaisin]
43. Tästä ks. <http://www.geschichte.uni-frankfurt.de/43010469/jussenb> ja <http://historymanifesto.cambridge.org>, sekä David Armitage, 'What's the Big Idea? Intellectual History and the Longue Durée', *History of European Ideas*, 38, 2012, s. 493-507. [Takaisin]
44. Ks. esim. Pasi Ihalaisen tuotanto ja Jani Marjanen, *Den ekonomiska patriotismens uppgång och fall: Finska hushållningssällskapet i europeisk, svensk och finsk kontext 1720–1840*, 2013. [Takaisin]
45. Keskiäikaisten käsikirjoitusten tutkimuksessa historian alalla pioneerina Suomessa ovat kuitenkin olleet Teemu Roos ja Tuomas Heikkilä, 'Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets', *Literary and Linguistic Computing*, 24, 2009, s. 417-433. [Takaisin]



[Lataa artikkeli PDF-tiedostona](#)

◀ AATEHISTORIA ▶ DIGITAALISET IHMITIETEET ▶ DIGITALISAATIO ▶ METODOLOGIA

YKSI VASTAUS ARTIKKELIIN "AATEHISTORIA JA DIGITAALISTEN AINEISTOJEN MAHDOLLISUUDET"

Paluuviite: Tankar kring den snabba utvecklingen i Finland år 2015 | Historia i en digital värld