**Digitaaliset ihmistieteet (*Digital Humanities*) ja historiantutkimus[[1]](#footnote-2)**

“Tutkija astelee arvokkaasti arkistoon ja jättää asiaankuuluvasti berberinsä naulakkoon. Mutta pienen agenttisalkkunsa hän pitää mukanaan … Yllätys yllätys! Agenttisalkku ei olekaan salkku! Sen kannen alta paljastuu näppäimistö, jonka takaa kääntyy ylös pieni näyttöruutu ... Olisiko kuvattu ‘agenttisalkku’ mahdollinen? Olisiko se tarpeellinen vai pelkkää hienostelua? Ja ennen kaikkea: olisiko siitä hyötyä esimerkiksi neulakortteihin verrattuna?”[[2]](#footnote-3) [Historiantutkimuksen tulevaisuuden visiointia vuonna 1983]

Digitaalisten aineistojen käytöstä historiantutkimuksessa on keskustelu viime aikoina paljon. Itseään ”digihistorioitsijoiksi” tituleeraavien henkilöiden määrä kasvaa maailmalla tasaisesti ja historiantutkimusta tehdään myös Suomessa ”ilman papereita” niin kuin 1980-luvulla visioitiin. Tästä huolimatta digitaalisten menetelmien käyttö on vielä vain harvoissa tapauksissa vaikuttanut historiantutkimuksen traditioiden ytimeen tai muokannut historian alan tutkijakoulutusta.[[3]](#footnote-4) Digitaalisten menetelmien hyödyntämisessä historiantutkimuksessa ei kuitenkaan itsessään ole mitään uutta. Jo Viljo Rasila 1960-luvulla kirjoitti tietokoneiden käytöstä historiantutkijoiden apuvälineenä *Historiallisessa Aikakauskirjassa*.[[4]](#footnote-5) Kannettavasta tietokoneesta ”agenttisalkkuna” puhunut Jussi T. Lappalainen jatkoi 1970- ja 1980-luvuilla tästä keskustelua Suomessa.[[5]](#footnote-6) Todellinen muutos syntyy kuitenkin vasta silloin, kun uusia datatieteen menetelmiä aletaan soveltaa aiempien jatkumoiden ja tutkimustraditioitten varaan. Datatiede on yleistermi, joka viittaa modernien tietojenkäsittelytieteen menetelmien kehittämiseen ja sovelluksiin laajojen digitaalisten tietoaineistojen hallinnoinnissa ja tilastollisessa analyysissä. Datatieteen menetelmiä ja nopeasti karttuvia digitaalisia tietoaineistoja käytettäessä myös lähdekritiikki saa uusia muotoja ja merkityksiä. Tulkintaongelmat ja virhelähteet muuttuvat yhä lukuisemmiksi ja monisyisemmiksi. Historiantutkimuksen perusmetodit eivät silti tulevaisuudessakaan katoa, vaan lähdekritiikin hallinnasta tulee yhä merkittävämpää menetelmäosaamisen rinnalla. Historiantutkijan koulutus alkaa toden teolla muuttua digitaalisten ihmistieteiden vaikutuksesta kun saadaan aikaan tutkimustuloksia joihin ”perinteisin” keinoin ei olisi päästy – tämä paradigman muutoksen aika ei vielä ole kokonaisvaltaisesti koittanut, mutta etenemme jo vauhdilla sitä kohti.

Tämä luku tarkastelee digitaalisten ihmistieteiden (*digital humanities*) merkitystä historiantutkimuksessa ja pohtii tämän suhdetta erilaisiin teoreettisiin kysymyksenasetteluihin.[[6]](#footnote-7) Digitaalisilla ihmistieteillä tarkoitetaan erityisesti modernin datatieteen ja laajojen digitaalisten tietoaineistojen hyödyntämistä humanistis-yhteiskuntatieteellisessä tutkimuksessa.[[7]](#footnote-8) Aineistot voivat sisältää esimerkiksi tekstiä, kuvaa, ääntä tai liikettä. Suomen Akatemia päätyi keväällä 2015 käännökseen *digitaaliset ihmistieteet*, joka kattaa sekä humanistiset että yhteiskuntatieteet.[[8]](#footnote-9)

Tarkastelemme mahdollisuutta muodostaa teoriaa digitaalisista ihmistieteistä historiantutkimuksessa työskentelymenetelmien, tutkimusaineistojen ja monitieteisen yhteistyön näkökulmasta. Tämä sisältää nuoren tutkimusalan luonteen arvioinnin suhteessa muuhun tutkimukseen. Lisäksi pohdimme uudenlaisten menetelmien ja tutkimuksen piiriin saatujen lähdeaineistojen merkitystä. Heiluriliike näyttäisi vievän taas kohti aikaa jossa historiantutkimuksen linjat uudistuvat mikrohistoriallisesta otteesta kohti ”Ison Historian” näkökulmaa. Massadata (eng. *big data*) –tutkimus vaatii teknisiä erityistaitoja ja välineitä, mutta samalla myös tutkimustavat voivat yhtenäistyä.[[9]](#footnote-10) Punnitsemme tämän kehityksen hyviä ja huonoja puolia.

Tekstissä pohditaan myös kulttuurinmuutosta, jonka uudenlaisten lähestymistapojen käyttöönotto synnyttää. Tietotekniikan käytöllä ei humanistisessa tutkimustraditiossa ole itseisarvoa eikä esimerkiksi pelkän aineistojen digitoinnin sinänsä katsota vielä kuuluvan alan varsinaisen tutkimuksen piiriin. Digitaalisen tutkimuksen määrittely on haastavaa, sillä tutkimusaineistojen keruu ja järjestäminen on keskeinen – välttämätön, mutta ei riittävä – osa tieteellistä tutkimusprosessia. Monitieteisyys, jossa yhdistellään tutkimuskysymyksiä, menetelmiä ja lähdeaineistoja useiden tutkimusalojen piiristä, määrittää digitaalisten ihmistieteiden tutkimusta.[[10]](#footnote-11) Historiantutkimuksen kannalta mielenkiintoista on erityisesti se, minkälaista uutta tutkimusta aiempien tutkimustraditioitten ydinkysymyksiin liittyen voidaan tehdä, kun käyttöön avautuu uusia aineistoja ja menetelmiä.

Kärjistäen voidaan todeta, että verrattaessa esimerkiksi biotieteisiin historiasta on vuosisatojen aikana kehittynyt vahvemmin yksittäisen tutkijan henkilön ympärille rakentuva tieteenala. Tämä näkyy selvästi esimerkiksi verrattaessa eri alojen julkaisumalleja. Historiankirjoituksessa yhteisjulkaisuja ei perinteisesti arvosteta ja ihanteena on yksittäisen kirjoittajan monografia. Tämä voi johtaa erilaisiin lieveilmiöihin kuten epäsuhtiin rahoitusmalleissa.[[11]](#footnote-12) Samalla, kun moderni luonnontieteellinen tutkimus rakentuu monialaisen yhteistyön varaan, on historiantutkimuksessa tyypillistä, että yksittäisten tutkijoiden ja koulukuntien näkemykset voidaan kokea jopa tärkeämpinä kuin objektiivisuus ja näkökantojen tueksi esitetty tausta-aineisto. Luonnontieteitä tavanomaisesti painottavissa sovelluksissa, vaikkapa luonnonsuojelun alalla, voitaisiin toki hyödyntää runsaammin humanistien osaamista. Humanistisessa tutkimuksessa tärkeää roolia näyttelee hermeneuttisen otteen merkitys. Siinä, missä luonnontieteet kuvaavat universaaleja lainalaisuuksia, humanistisessa tutkimuksessa ilmiöiden omalaatuisuus ja tulkinnallinen monimielisyys näyttelee keskeistä roolia. Miten säilyttää tämä humanismin ydin lisäten kuitenkin samalla keskustelun koheesiota ja laajuutta? Tämä on yksi historiantutkimuksen tulevaisuuden keskeisistä haasteista.

**Digitaalisuus historiantutkimuksessa**

Miksi tulisi puhua digitaalisista ihmistieteistä tai digitaalisesta historiantutkimuksesta ja pyrkiä määrittelemään, mihin näillä termeillä viitataan?[[12]](#footnote-13) Tilastollisten menetelmien hallinta ei voi korvata sovellusalaan liittyvää asiantuntemusta, mutta voi täydentää sitä ja tuoda tutkimukseen uusia vaikutteita. Huomiota tulisi kiinnittää erityisesti tutkimustraditioihin, joilla on paras uudistumispotentiaali. Samalla digitaalisten ihmistieteiden tulee pystyä asemoimaan itsensä perinteiseen humanistisen tutkimuksen kentällä. Digitaalisuuden painottaminen muuttaa paitsi tutkimusta myös tutkimuskohdetta: esimerkiksi digitaalinen sanomalehti lähteenä ei ole sama asia kuin painettu sanomalehti.[[13]](#footnote-14)

Siinä oleva tieto saattaa osittain säilyä uudessa muodossa, mutta se ei koskaan voi korvata materiaalista lähdettä. Aikaisemmasta alkuperäinen/kopio-keskustelusta onkin nykyään edetty sekä ”perinteisten” että digitaalisten kulttuuriaineistojen ymmärtämiseen omanlaisina objekteinaan.[[14]](#footnote-16) Digitaalinen tieto sisältää elementtejä, joita alkuperäisessä kulttuuriperintöaineistossa ei ole. Cambridgen yliopiston piirissä toimiva *Concept Lab* pyrkii muuttamaan käsitteiden tutkimusta korostamalla sitä, että digitaalista ja painettua ei tutkita aatehistorian kannalta samalla tavoin.[[15]](#footnote-17) Aiemman tutkimuksen hypoteeseja ja tuloksia voidaan testata ja todentaa uudella tavalla. Datatieteen menetelmien soveltaminen mahdollistaa uudenlaiset kysymyksenasettelut, mutta edellyttää sujuvaa vuorovaikutusta eri tieteenalojen asiantuntijoiden välillä. Tutkimuksen uudet menetelmät ja muodot muokkaavat tutkimuskysymyksiä ja luovat myös uusia. Esimerkiksi käy Lundin yliopiston projekti, jossa pohditaan filosofisesti tiedon luonnetta digitaalisessa maailmassa.[[16]](#footnote-18)

Aineellisten ja aineettomien tutkimuskohteiden maailmassa tapahtuvat muutokset ovat oleellisia myös tutkimuksen teoreettisesta näkökulmasta. Aiempi keskustelu määrittää voimakkaasti uusien tutkimustulosten merkitystä. Jos tutkii vaikkapa David Humen teoksia aihemallintamisen avulla, tulee pystyä perustelemaan kytkös myös muuhun Hume-tutkimukseen. Toisin sanoen, jos tutkii, miten David Hume käyttää vapauden käsitettä, ei riitä, että tutkimuksen kohteena ovat Humen tekstit yleisesti. Uusien tulkintojen tulee pystyä asemoimaan itsensä Humen vapauden käsitettä aikaisemmin tutkineen kirjallisuuden kenttään. Tämä aatehistorioitsijoille itsestään selvä näkökulma ei aina ole sitä digitaalisissa ihmistieteissä, joiden tutkijat saattavat tulla toisenlaisten tutkimustraditioitten piiristä. Osa tutkijoista tuntuu ajattelevan, että uusia menetelmiä käytettäessä tutkimus voitaisiin irrottaa aiemmista viitekehyksistä. Digitaalisten ihmistieteiden tutkimus ei voi kuitenkaan täysin irrota aikaisemmasta, samankaltaista kulttuuriperintöaineistoa koskevasta tutkimustraditiosta. Uusien kysymystenasettelujen tulee kyetä operoimaan luontevasti historiantutkimuksen kentässä.[[17]](#footnote-19)

Digitaalisten ihmistieteiden määritelmä on tätä taustaa vasten suoraviivainen: kun tutkimusta tehdään esimerkiksi digitoiduilla sanomalehdillä, datatieteen menetelmien hyödyntäminen tekee tutkimuksesta digitaalisten ihmistieteiden piiriin kuuluvaa. Pelkkä digitaalisten sanomalehtien katselu ei kuitenkaan vielä kuulu digitaalisten ihmistieteiden piiriin eikä puhtaan teoreettinenkaan menetelmäkehitys itsessään vielä ole historiantutkimusta, vaan lukeutuu esimerkiksi tietojenkäsittelytieteen tai kieliteknologian piiriin. Soveltava menetelmäkehitys voidaan sen sijaan katsoa digitaalisten ihmistieteiden piiriin kuuluvaksi tutkimukseksi.

Datatieteen menetelmien kirjo on laaja samoin kuin käyttäjien tarpeet. Esimerkiksi aineistojen haku ja selailu tähän suunniteltujen käyttöliittymien avulla vaatii erilaisia teknisiä ratkaisuja ja palvelee erilaisia käyttäjiä kuin samoihin tietoaineistoihin kohdistuva laskennallinen analyysi, jossa tilastollisten ohjelmointikielten avulla tutkitaan aineistossa esiintyviä laajoja säännönmukaisuuksia ja niiden tilastollista merkitsevyyttä, kuten vaikkapa kirjapainotoiminnan leviämisvauhdin vertailua eri kielialueilla ja tieteenaloilla. Perinteisempi sanomalehtitutkimus, jossa digitoitua sanomalehteä voidaan käyttää esimerkiksi tavanomaisen sanahaun kautta ja uusien laskennallisten menetelmien käyttö voivat kuitenkin jakaa yhteisen tutkimusperinteen. Aikaisemmasta fyysisestä lähteestä muodostettu digitaalinen ”kuva” ei silti vielä itsessään edusta digitaalista tutkimusta.

Digitaalisen historiallisen aineiston ja muun kulttuuriperinnön ja digitaalisen historiallisen aineiston erosta yysisestä lähteestä on nyt digitaalinen " on nähty edustavan ainoastaan sitä,ön erottelu on keskeistä digitaalisten ihmistieteiden määrittelylle. Kun digitaalisten menetelmien käyttöönotto historiantutkimuksessa on ollut toistaiseksi vaatimatonta, tulisi huomio kiinnittää erityisesti tutkimuksen uudistamiseen.[[18]](#footnote-20)

Tutkimusaineistot voidaan jaotella digitaalisena syntyneisiin (kuten sosiaalinen media) tai muuten syntyneisiin (esimerkiksi digitoidut sanomalehdet) aineistoihin. Vaikka tämän tekstin painopiste on varhaisemman ajan tutkimuksessa ja digitoiduissa historiallisissa aineistoissa, on huomioitava, että myös historiantutkimusta voidaan tehdä digitaalisina syntyvistä aineistoista. Yksi esimerkki tästä on Kansalaisten mielenliikkeet -hanke, jossa tutkitaan Suomi24-keskustelupalstan aineistoa. Historiantutkimusta hankkeessa edustavat Jaakko Suominen ja Anna Sivula, jotka tutkivat keskustelupalvelua muuttuvana alustana – siis historiallisena ilmiönä. Tämä esimerkki osoittaa, miten historiantutkimusta voidaan uudistaa uusilla menetelmillä, samalla siirtäen historiantutkimuksen traditiota kohti modernin maailman ilmiöitä. Kyseisessä hankkeessa uusi ja vanha kohtaavat.[[19]](#footnote-21)

**Massadata ja ilmiöpohjainen tutkimus**

Käsi kädessä digitaalisuuden kanssa kulkee käsite nimeltä massadata.[[20]](#footnote-26) Tällä tarkoitetaan erityisen laajoja tietoaineistoja, joiden käsittelyyn tarvitaan uudenlaista tietojenkäsittelyinfrastruktuuria ja tilastollisia menetelmiä. Yksi määrittely massadatalle on, ettei aineisto mahdu yksittäiseen tietokoneeseen, vaan tarvitaan järeitä infrastruktuuriratkaisuja ja skaalautuvia menetelmiä aineiston hallintaan, käsittelyyn ja analysointiin. Suomessa näitä tarjoaa etenkin Tieteen Tietotekniikan Keskus (CSC). Tällaisessa tutkimuksessa ja pienemmänkin mittakaavan data-analyysissä aineistoista pyritään tyypillisesti etsimään yleisiä tilastollisia trendejä. Lähestymistavan etuna on, että havaintoja voidaan tehdä mittakaavassa, joka ei ole aiemmin ollut mahdollinen; karkeakin aineisto voi osoittautua arvokkaaksi, kun etsitään ja kuvataan laajoja tilastollisia säännönmukaisuuksia. Laajaan kvantitatiiviseen analyysiin perustuvassa tutkimuksessa datasta voidaan tehdä myös odottamattomia havaintoja. Nämä voivat suunnata tutkimusta ennakoimattomalla tavalla. Molemmilla lähestymistavoilla – huolella laadituilla täsmä-analyyseillä ja laajempien karkeiden aineistojen tutkimuksella – on paikkansa. Lähestymistavasta riippumatta aineiston siistiminen käyttökelpoiseen muotoon on kuitenkin tutkimuksessa usein yksi työläimmistä ja eniten aikaa vievistä vaiheista.

**Digitaalinen historiantutkimus käytännössä**

Digitaalisten ihmistieteiden piirissä on käyty mittavaa keskustelua metodeista ja niiden merkityksestä.[[21]](#footnote-29) Mitä ne voivat todella tarjota historiantutkimukselle? Massadataan liittyen on muodikasta alleviivata, että aineiston kuvaaminen uudella tavalla voi tarjota uusia näkökulmia tutkittavaan aiheeseen. Datan visualisoinnista on muodostunut oma metodinen suuntauksensa, joka ulottuu perinteisten tieteenalojen yli. Visualisointi on sekä tiedeviestinnän väline että kokeileva tutkimusmetodi, jolla aineistosta voidaan seuloa havaintoja, jotka voivat parhaimmillaan johtaa uudenlaisiin tutkimuskysymyksiin.[[22]](#footnote-30) Värikkäät visualisoinnit eivät voi kuitenkaan korvata syvällisempää tilastollista mallinnusosaamista.

Aineiston kokoon liittyvien ongelmien ratketessa tutkimuksen painopiste sirtyy kohti sovelluksia, kuten esimerkiksi tiedontuotantoa sanomalehdissä.[[23]](#footnote-31) Näitä tarkastellaan monimuotoisen digitaalisen aineiston kautta erilaisista näkökulmista usein osana laajempaa tutkimuskokonaisuutta, jossa yhdistyy esimerkiksi sosiolingvistinen tai sosiologinen tutkimus uusiin datatieteen menetelmiin. Sanomalehdet tarjoavat oivan esimerkin digitaalisen humanismin monimuotoisesta tutkimuskohteesta, jossa sanomalehtien materiaalisuus ja fyysiset ominaisuudet muodostuvat mielenkiintoiseksi tutkimuskohteeksi perinteisen sisältöanalyysin ohella. Niiden tarkastelu voi olennaisesti auttaa laadullisissa tulkinnoissa Suomessa tapahtuneesta julkisesta keskustelusta, kuten myöhemmin osoitamme.

Toinen esimerkki ilmiöpohjaisesta lähestymisestä on kaupunkitutkimus. Voidaan tutkia, miten kaupunkien toimintaan liittyvää avointa ja suljettua lähes reaaliaikaistakin dataa voidaan ymmärtää ja hyödyntää kaupunkien kehityksen kontekstissa. Kaupunkeihin ja yhteiskuntaan liittyvää dataa ja laskennallisia menetelmiä on saatavilla kasvavassa määrin vakiintuneiden laadullisen tutkimuksen tueksi. Humanistien panos voi tuottaa näkökulmia tällaisten ilmiöiden kehittymisestä ja niiden yhteyksistä ihmisten käyttäytymiseen, hyvinvointiin, vuorovaikutukseen ja maailmankuvaan. Tähän liittyy ilmiön historiallinen analysointi, mutta tutkimuksen tavoitteena voi olla löytää myös sellaisia näkökulmia, joilla on merkitystä esimerkiksi poliittisessa päätöksenteossa.[[24]](#footnote-32)

Eri tieteenalojen kohdatessa ilmiöitä voidaan tutkia monista täydentävistä näkökulmista. Tämä pakottaa tutkijat ulos asiantuntijarooleistaan ja perinteiseltä mukavuusalueeltaan. Kun tarkoituksena on ymmärtää laajoja ja monimutkaisia kokonaisuuksia, kasvaa tarve yhteistyölle eri alojen asiantuntijoiden kesken. Tästä humanisteilla on viimeisten vuosikymmenien eriytymisen aikana vähän kokemusta luonnontieteiden tutkijoihin verrattuna. Digitaalisissa ihmistieteissä sen sijaan korostuu monitieteinen tutkimusote, yhteistyö ja pragmaattisuus. Historiantutkimuksen uudistamisen ohella keskeistä on myös viedä alan osaamista rikastuttamaan muita tutkimusaloja; oma oppiala tulee asettaa osaksi laajempaa humanististen ja sosiaalitieteiden perinnettä, jossa ihmisen toiminnan mallintaminen on kulkenut jo jonkin aikaa eri reittejä suhteessa historiantutkimukseen. Digitaalisten menetelmien mahdollisuudet pitää pystyä hyödyntämään menettämättä humanistisen tutkimuksen hermeneuttista erityispiirrettä. Lisäksi tulisi kiinnittää huomiota tutkimusaineistojen edustavuuteen, esikäsittelyyn, tilastollisten tulkintojen merkitsevyyteen ja toistettavuuteen. Digitaalisten ihmistieteiden tutkimusta onkin perusteltua tehdä hankkeissa, joihin kootaan eri tieteenalojen osaamista.

Digitaalinen historiantutkimus keskittyy usein tekstiaineistoihin. Tärkeimpiä näistä ovat aineiston kuvailutiedot (*metadata*) sekä kokoteksit (Kuva 1).[[25]](#footnote-33) Näitä hyödyntää muun muassa COMHIS-yhteistyöhanke (2016-2019)[[26]](#footnote-34), jossa tutkitaan suomalaista julkisen keskustelun historiallista kehitystä eurooppalaisessa kontekstissa (Kuva 2). Hankkeessa kiinnitetään erityistä huomiota aineistonhallintaan. Tätä tukemaan on kehitetty Octavo-niminen avoimen tieteen alusta, jolla tutkimusaineistoja voidaan yhdistää ja analysoida tehokkaalla ja läpinäkyvällä tavalla.[[27]](#footnote-35) Automatisoidut ratkaisut mahdollistavat sujuvasti uusien lähdeaineistojen mukaan tuonnin, aineistoista löytyvien virheiden korjauksen ja tilastollisten yhteenvetojen täydentämisen.

Kirjastoluetteloiden sisältämien kuvailutietojen avulla pystytään kartoittamaan tiedontuotannon yleinen kehitys huomattavasti aikaisempaa kattavammin ja objektiivisemmin. Voimme 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 menetelmillä kuin laajempia kokotekstitietokantoja, koska kuvailuaineistojen koko on rajatumpi ja rakenne selkeä. Tämä täydentää merkittävästi tekstipohjaisen tutkimuksen mahdollisuuksia. Koneluettavien kuvailutietokantojen hyödyntämistä alan tutkimuksessa on merkittävästi aliarvioitu.

Olemme tutkineet kuvailutietojen käyttöä muun muassa vertailemalla ruotsalaisen Kungliga-kirjaston luettelointitietoja Suomen kansallisbibliografia Fennicaan. Tämän pohjalta pystymme tekemään suoraviivaista analyysiä kirjojen eri julkaisumuotojen vaikutuksesta tiedontuotantoon varhaismodernina aikana (Kuva 3). Kotimaisen historiantutkimuksen kannalta kiintoisan esimerkin tarjoaa tieteellisten laskentakirjastojen laadinta suomalaisen kansallisbibliografian Fennican tutkimiseen, vaikka samalla tuleekin tiedostaa ja korjata Fennican alkuperäisaineistosta vielä löytyvät puutteet. Alustava analyysi osoittaa esimerkiksi suoraan historiallisiin tapahtumiin kytkeytyvät tiedontuotannon huiput ja aallonpohjat Vaasassa, Helsingissä, Turussa ja Suomessa vuosina 1764–1917 (Kuvat 4–8). Kartoittamalla vaikkapa ruotsalaisiin tai venäläisiin paikannimiin tehtävien viittausten määrää ja laatua suomalaisissa sanomalehdissä voimme saada uutta tietoa suomalaisen julkisen keskustelun rakenteesta ja kehityksestä historian eri vaiheissa. Samaa voidaan tehdä paikallisemmassa mittakaavassa tutkimalla viittauksia naapurikuntien ja suurempien kaupunkien välillä. Kansallisbiografian historiallisten henkilöiden nimien paikantaminen ja kartoitus voi myös osaltaan kehittää suomalaisen julkisen sfäärin tutkimusta. Näin uudet menetelmät voivat antaa historian tutkimuksen käyttöön uudenlaista materiaalia ja näkökulmia.

Alkuperäiset kuvailutietueet eivät sellaisenaan sovellu tilastolliseen analyysiin, vaan ne pitää ensin jalostaa sopivaan muotoon. Tämä sisältää muiden muassa maantieteellisten ja henkilönimien erilaisten kirjoitusmuotojen yhtenäistämistä, samannimisten kirjoittajien erottelua, kirjoitusvirheiden käsittelyä ja kirjojen sivumäärien arviointia bibliografisten standardimerkintöjen pohjalta. Tietoja voidaan lisäksi täydentää ja rikastaa muiden lähdeaineistojen avulla. Kirjoittajista voidaan selvittää puuttuvat elinvuodet ja sukupuoli sekä muuta taustatietoa kuten esimerkiksi painopaikkojen perusteella painomaa. Valtaosa työstä voidaan ja kannattaa 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 räätälöidyt laskentakirjastot tarjoavat erinomaisia mahdollisuuksia alan tutkimukselle verrattuna valmiisiin ohjelmistoihin, jotka olettavat että aineisto on valmiiksi saatettu kulloinkin käytössä olevan ohjelmiston vaatimaan muotoon ja siihen sisältyvät virheet korjattu. Tämä on merkittävä puute, sillä käytännön työssä on arvioitu 80 % työajasta kuluvan aineiston esikäsittelyyn.[[28]](#footnote-36) Tieteellisen ohjelmoinnin hyödyntäminen auttaa automatisoimaan kaiken mikä automatisoitavissa on, jättäen tutkijalle samalla vapaat kädet lopullisiin joustaviin, aineistokohtaisesti räätälöityihin ratkaisuihin. Laskentakirjastot tarjoavat tutkijalle pääsyn sekä aineistojen jalostamisen että tilastollisen analyysin ja visualisointien kannalta keskeisiin työkaluihin, ja koko työvirta voidaan toteuttaa läpinäkyvällä ja toistettavalla tavalla, tieteellisen avoimuuden periaatteita kunnioittaen.

Tekstiaineistojen tutkimukseen liittyy kaksi toisiaan täydentävää lähestymistapaa, joita ovat rakenteettoman tekstin louhinta sekä erilaisten aineistojen kuvailutietojen tutkimuskäyttö joka pitää sisällään niin kirjasto-, arkisto- kuin myös museoluetteloinnit. Näiden hyödyntäminen tutkimuksessa ei poista perinteisen lähiluennan merkitystä, vaan täydentää sitä. Tekstilouhinnan tutkimuspotentiaali on valtava. Jos historioitsija on esimerkiksi kiinnostunut tutkimaan oikeudenmukaisuuden käsitteen muuttumista brittiläisessä julkisessa keskustelussa, sitä voidaan tutkia esimerkiksi louhimalla kansainvälisiä aineistoja. Näitä ovat muiden muassa 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. Esimerkkejä muista vastaavista aineistoista on ranskalainen kansallinen Gallica-projekti.[[29]](#footnote-37)

Samalla louhintaan voidaan yhdistellä käsitehistorian metodeja. Verrattuna perinteiseen lähiluentaan käsitteen muutosta voidaan ryhtyä tutkimaan aivan eri tavoin kun käytössä on automatisoituja, laajojen tietoaineistojen käsittelyyn skaalautuvia mallintamismenetelmiä. Kotimaisena esimerkkinä on kaikkien historiallisten suomalaisten sanomalehtien tekstilouhinta, jossa aineistoista haetaan paikannimiä ja henkilöitä.[[30]](#footnote-38) OCR-menetelmä, jonka avulla kuvattu aineisto käännetään koneluettavaan muotoon, ei ole täydellinen; parhaimmassakin tapauksessa sanojen tunnistustarkkuus on vain noin 80 %. Menetelmiä pyritäänkin kehittämään siihen suuntaan, että ne voivat toimia vakaasti aineiston luonnollisista puutteista huolimatta.[[31]](#footnote-39) Uusien menetelmien käyttö kehittyy ja yleistyy nopeasti, muokaten alan tutkimuskäytäntöjä.

**Avoimen tieteen merkitys**

Digitaalisten aineistojen merkitys on nähty ensisijaisesti aineistojen parantuneen saatavuuden kautta. Silti aineistojen avoimuteen ja käsittelyyn liittyy merkittäviä periaatteellisiakin ongelmia. Viime aikoina avoimen tiedon (*open knowledge*) tai avoimen tieteen (*open science*) periaatteet on alettu nähdä laajempana kokonaisuutena, joka tarjoaa runsaasti uusia välineitä, lähestymistapoja ja mahdollisuuksia historian tutkimukseen.

Avoin tiede korostaa läpinäkyvien ja yhteisöllisten toimintamallien merkitystä. Käsitteen katsotaan viittaavaan esimerkiksi lähdeaineistojen, tutkimusmenetelmien ja julkaisemisen avoimuuteen. Avoimuus sisältää vapauden tallentaa, muokata ja jakaa edelleen tutkimusmateriaalia tavanomaisten tieteellisten viittauskäytäntöjen mukaisesti muitta rajoituksitta. Tutkimusaineistojen koon kasvaessa ja niiden käsittelyyn tarvittavien informaatiotekniikan menetelmien monimutkaistuessa tutkimuksen avoimuus on kohdannut uudenlaisia haasteita. Pelkkä tutkimusprosessin yleinen kuvaus ei vielä takaa, että tutkimus on toistettavaa, tai että muu tutkimusyhteisö pystyisi rakentamaan sen varaan tehokkaasti uutta tutkimusta. Verrattuna luonnontieteisiin kuten fysiikkaan tai biolääketieteeseen, joissa avoimuuden puutteeseen liittyvät ongelmat ovat tulleet esiin, kvantitatiivisen tutkimuksen haasteisiin ei ole vielä kiinnitetty samanlaista huomiota ihmistieteissä.[[32]](#footnote-40) Kun luonnontieteissä tutkimuslaitokset jopa kilpailevat avointen tietokantojen laadulla[[33]](#footnote-41), humanistis-yhteiskuntatieteellisen tutkimuksen kannalta merkittäviä aineistoja ylläpitävät laitokset vaikuttavat jopa välinpitämättömiltä aineistojen avoimen käytön edistämisen suhteen. Tieteenala on avointen käytäntöjen osalta jäänytkin kansainvälisesti jälkeen luonnontieteistä. Muiden muassa historiassa voi törmätä kirjaviin ja jopa virheellisiin käsityksiin tutkimusaineistojen ja julkaisemisen avoimuudesta. Tärkeä muutos tähän on mahdollisesti tulossa Opetus- ja kulttuuriministeriön Avoin Tiede ja Tutkimus (ATT)–hankkeen myötä.[[34]](#footnote-42) Yksi keskeinen hankkeen aikaansaannos on huomion kiinnittäminen tekijänoikeuksien ja yksilösuojan merkitykseen ja haasteisiin tiedonlouhinnan osalta.[[35]](#footnote-43) Eurooppalainen lainsäädäntö muuttuu tiedonlouhinnan osalta vuoteen 2018 mennessä.[[36]](#footnote-44)

Avoin data[[37]](#footnote-45) viittaa siihen, että tutkimuksen alkuperäiset tietoaineistot ovat saatavilla 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 määritelmän ja vakiintuneiden viittauskäytäntöjen kanssa.

Luonnontieteellinen tutkimus on edelläkävijä aineistojen avoimessa jakelussa. Luonnontieteissä on tiedostettu alkuperäisaineistojen saatavuuden merkitys tutkimuksen läpinäkyvyydelle ja avoimuudelle, joka on yksi tieteellisen toiminnan kulmakivistä. Aineistojen jakeluun esimerkiksi fysiikassa ja biotieteissä on perustettu keskitettyjä julkisia tietokantoja, joiden pitkän tähtäimen rahoitus on turvattu. Humanistis-yhteiskuntatieteellisen tutkimuksen lähdeaineistot on koottu kaupallisen mallin puitteissa tai esimerkiksi kirjastojen toimesta sillä ajatuksella, että niiden käyttöoikeuksia voidaan myöhemmin myydä. Avoimen julkaisemisen ei ole perinteisesti nähty tuovan merkittävää lisäarvoa. Tulevaisuudessa vaikkapa perustavanlaatuiset aatehistorian metodologiset kysymykset kuten skinneriläinen kontekstualismi ja koselleckilainen käsitehistoria voivat kuitenkin näyttäytyä uudenlaisessa perspektiivissä digitoitujen aineistojen tutkimuksen myötä. Tämä edellyttää, että substanssi- ja menetelmäosaaminen löytävät toisensa ja historian tutkimuskysymysten ehdoilla etenevä tutkimus voi ammentaa kieliteknologiasta, koneoppimisesta ja muista informaatiotieteiden traditioista. Samalla olisi tärkeää, että yhteistyötä tutkimuksen tietoaineistoja hallinnoivien kirjastojen ja muiden muistiorganisaatioiden kesken syvennetään huomattavasti.

Käytännössä pääsy historian tutkimuksessa keskeisiin alkuperäisaineistoihin on osoittautunut haastavaksi jopa luottamuksellisen tutkimusyhteistyön kautta puhumattakaan samojen aineistojen avoimesta julkaisemisesta verkossa.[[38]](#footnote-46) Nykytilanteessa on hämmentävää, ettei historiallisia aineistoja useissa tapauksissa pystytä saamaan tutkimuksen käyttöön ilman monimutkaisia sopimusneuvotteluja samalla, kun tekijänoikeuksiin tai yksilönsuojaan liittyviä ongelmia aineistojen käyttöön ei välttämättä liity lainkaan. Tämä on esimerkki siitä, kuinka aineistojen käyttö ja hallinta eivät ole tutkijakeskeisiä, vaan toimintaa säätelee aggressiivinen tekijänoikeuden haltijoiden edunvalvonta.[[39]](#footnote-47) Keinotekoisten kaupallisten ja teknisten rajoitteiden yleisyys digitoitujen aineistojen saatavuudessa poikkeaa merkittävästi luonnontieteiden tutkimuksessa yleisestä käytännöstä, jossa alkuperäiset mittausaineistot tallennetaan julkaisun yhteydessä avoimiin tietokantoihin. Selittävänä tekijänä ovat erot tutkimuskulttuurissa sekä aineistojen jakelun järjestämisessä ja rahoitusmalleissa. Toivottavasti esimerkiksi Kansalliskirjaston strategisten painopisteiden muutoksen myötä nykyinen käytäntö astuu kohti nykypäivää, jotta digitaalisten aineistojen käyttö ja sitä varten rakennettavien työkalujen hyödyntäminen muuttuu nykyistä ketterämmäksi. On tärkeää, että tutkimuksen mahdollistaminen etenee myös Suomessa mahdollisimman suoraviivaisesti.

Avoimuuden puutteet muodostavatkin tällä hetkellä merkittävän pullonkaulan, johon liittyy humanistis-yhteiskuntatieteellisessä tutkimuksessa 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 lähdeaineistoihin, joita tarvittaisiin kattavan kokonaiskuvan muodostamiseen. Yksityiskohtaisempi ja alaa uudistava tekstilouhinta edellyttää pääsyä koko alkuperäisaineistoon ja mahdollisuuksia sen vapaaseen jatkokäsittelyyn. 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.

Avoimen datan lisäksi keskeistä tutkimuksen läpinäkyvyydelle ja edelleenkäytölle onkin tutkimusmenetelmien avoin julkaiseminen.[[40]](#footnote-49) Avoimesta lähdekoodista keskusteleminen saattaa tuntua historian tutkijoille vieraalta, mutta digitaalisten menetelmien käytön lisääntyessä avointen toimintamallien edistäminen on merkittävää. Käytäntö on nopeasti yleistynyt erityisesti luonnontieteen alalla, jossa vertaisarvioitujen tutkimusten laadintaan liittyviä lähdekoodeja julkaistaan hajautetuissa versionhallintapalveluissa.[[41]](#footnote-50)

Kvantitatiivisen tutkimuksen ohessa syntyy aineiston jalostamiseen ja tilastolliseen analysointiin soveltuvaa lähdekoodia, joka voidaan järjestää tieteelliseksi laskentakirjastoksi ja hyödyntää myöhemmin muiden vastaavien aineistojen tutkimuksessa. Avoin tieteellinen laskentakirjasto (Kuva 9) viittaa toimintaympäristöön, joka tarjoaa yhteisöllisesti kehitettäviä avoimen lähdekoodin välineitä aineistojen tutkimuskäyttöön tilastollisissa ohjelmointiympäristöissä. Jokainen analyysiprosessin yksityiskohta alkuperäisestä raakadatasta lopullisiin yhteenvetoihin on läpinäkyvä. Tämä vähentää mahdollisuuksia vääristymien piilotteluun ja parantaa tutkimuksen laatua.[[42]](#footnote-56) Valmiisiin käyttöliittymiin verrattuna laskentakirjastot tarjoavat joustavampia rakennuspalikoita, joista alan tutkija voi rakentaa omiin tarpeisiinsa sopivan kokonaisuuden. Tämä vähentää rutiininomaisia työvaiheita. Historiallisiin lähteisiin liittyen esimerkiksi kirjastoluetteloiden kuvailutiedot ovat varsin yhtenäisiä ylläpitäjästä riippumatta. Tämä tarjoaa mahdollisuuksia eri maiden kirjastoluetteloiden yhdistämiseen ja tiedontuotannon tutkimiseen kansainvälisellä tasolla.

Avoimuus parantaa myös tutkimuksen toistettavuutta ja läpinäkyvyyttä. Nykytilanteessa tutkija saattaa tehdä suljettujen käyttöliittymien avulla monimutkaisiakin tietokantahakuja ja muodostaa niiden perusteella heuristisia päättelyketjuja tiettyjen termien käytön historiallisesti kehityksestä, mutta tehdyistä hauista, niiden sisällöstä ja päättelyketjusta ei välttämättä jää edes tutkijalle itselleen selkeää kirjanpitoa. Automatisointi tarjoaa tähän ratkaisun. Avoimuuden ansiosta muutkin voivat hyödyntää samoja ratkaisuja omassa työssään, jatkokehittää sitä omiin tarpeisiinsa, ja jakaa tulokset takaisin yhteisön käyttöön. 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 vauhdittaa tutkimusta. Onkin tärkeää, että myös dataintensiivisessä historiantutkimuksessa voidaan ennakkoluulottomasti kokeilla datatieteen parhaita käytäntöjä.

Avoimuus voi 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. Monien luonnontieteen alojen tutkimuksessa tämä on jo arkipäivää avointen tietokantojen ja tutkimusyhteisön kehittämien ohjelmistokirjastojen myötä. Digitaalisten ihmistieteiden alalla kotimaisten tutkimusryhmien kannattaisi tehdä yhteistyötä esimerkiksi suomalaisten sanomalehtiaineistojen analysoinnissa, jotta niukat tutkimusresurssit eivät valuisi hukkaan tutkimusryhmien tehdessä päällekkäistä työtä.

Poikkeamat mistä tahansa avoimuuden osa-alueesta heikentävät koko tutkimusprosessin avoimuutta. Humanistilla aloilla puhutaan usein avoimesta julkaisemisesta, mutta vähemmän tutkimusaineistojen ja tutkimusprosessien avoimuudesta, vaikka nämä saattavat olla tutkimuksen kehittymisen kannalta jopa vielä keskeisempiä. Paradoksaalisesti tämän merkitystä ei aina havaita, kun perinteisesti ajatellaan, että tiedeyhteisön kannalta hyödyllisin julkaisuyksikkö on lopullinen puntaroitu ja viilattu tutkimusjulkaisu. Kuitenkin myös tutkimusmenetelmät ja tutkimusprosessin aikana käytävä ajatustenvaihto ovat keskeisiä tutkimusprosessin osia. Näiden avoin jakaminen tiedeyhteisön kesken tarjoaa laadullisten tulosten lisäksi 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 vertaisarviointeineen avataan yleisölle.[[43]](#footnote-57) Avoimuuden merkitystä tutkimusetiikan ja metodiikan ytimessä tuleekin korostaa. Näin tutkimusprosessi ja -tulokset liikkuvat sujuvammin tutkijoiden ja tutkimusalojen välillä, ja niitä voidaan paremmin hyödyntää muussakin innovaatiotoiminnassa, opetuksessa ja kansalaisyhteiskunnassa. Avoimuuden tuottaman lisäarvon tutkimukselle ja yhteiskunnalle on arvioitu merkittävästikin ylittävän avoimuuden edistämiseen käytetyn taloudellisen panoksen.[[44]](#footnote-58)

**Digitaalisuuden asemasta historian tutkimustraditiossa**

Historiantutkimuksessa on oltu pidättyväisiä digitaalisten aineistojen käytön ja tiedonlouhinnan suhteen, vaikka eimerkiksi kieliteknologian menetelmät ovat kehittyneet jo ainakin kolmenkymmenen vuoden ajan humanistisen tutkimustradition puitteissa.[[45]](#footnote-59) Kärjistäen voidaan todeta, että perinteisessä historiantutkimuksessa yksittäinen sankaritutkija lukee ja kirjoittaa, eikä tilastopuuhastelun nähdä välttämättä tuovan mainittavaa lisäarvoa. Itsenäistä asiantuntemukseen pohjautuvaa luovaa ajattelua mikään kone ei voikaan tutkijan puolesta tehdä. Toisaalta tutkijat eivät ole välttämättä tietoisia datatieteen avaamista uusista tutkimusmahdollisuuksista.

Historian alalla digitaalisen ihmistieteen sovellusten puute selittyy osin sillä, että alan ydinosaajat eivät ole lähteneet menetelmiä kehittämään ja näin ei historian tarpeisiin vastaavia tietojenkäsittelyä hyödyntäviä menetelmiä ole päässyt syntymäänkään. Haasteena on sekin, etteivät laskennallisten tieteiden edustajat välttämättä osaa ennakoida historioitsijoiden tarpeita. Menetelmiä löytyy, mutta tutkimuskulttuurit eivät ole vielä kohdanneet merkittävällä tavalla.[[46]](#footnote-62) Uusi tutkimusala tarjoaakin lupaavia mahdollisuuksia ja tuoreita tutkimuskohteita myös menetelmätutkijoille.

Tutkimusta voitaisiin monissa tapauksissa tehostaa ottamalla käyttöön tilastollinen kehys laajempien aineistojen tarkasteluun. Digitaalisten aineistojen käytöllä voidaan laajentaa tutkimuksen aikajännettä sujuvasti vuosikymmenistä vuosisatojen aikana tapahtuviin muutoksiin.[[47]](#footnote-63) Tämän suuntainen ajattelu ei ole uutta ja ainakin käsitehistorian piirissä tätä mahdollisuutta on Suomessakin jo pohdittu pidempään.[[48]](#footnote-64) Mahdollisuuksien realisoituminen historian tutkimuksessa vaatii kuitenkin monitieteistä yhteistyötä ja muutoksia tutkimuskulttuuriin.

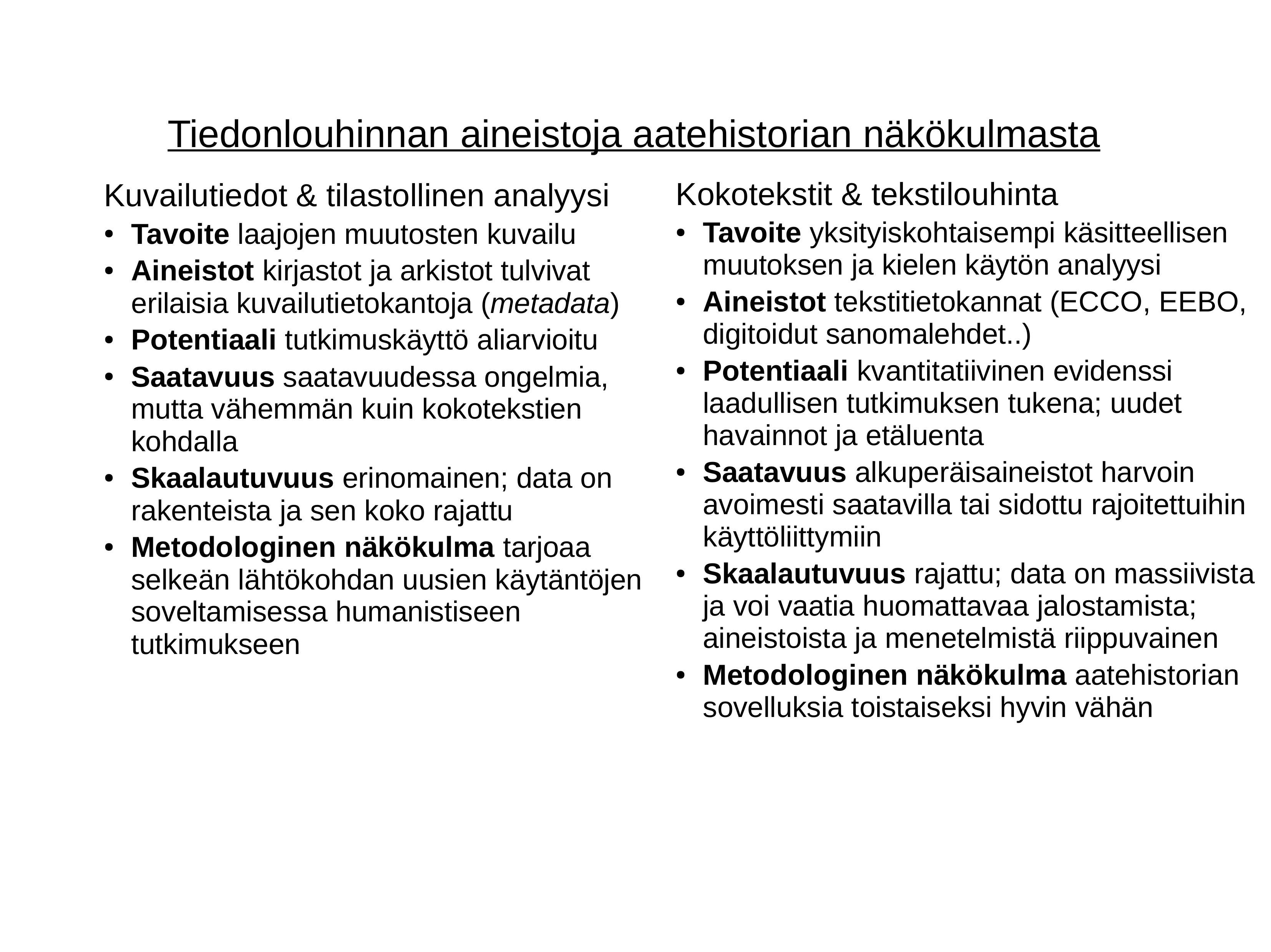
Uusien lähestymistapojen ilmestyessä tutkimuskenttään korostuu sekä tarve korostaa humanistisen tradition omien tutkimuskysymysten merkitystä että hahmottaa alan perustutkimuksen ydin. Käytännön tulokset ja ymmärrys sanelevat tutkimuksen lopullisen arvon. Mikään tutkimusmenetelmä tai -aineisto ei korvaa hyviä kysymyksiä ja hypoteeseja, joiden onnistunut määrittely ja niihin vastaaminen saatavilla olevien aineistojen nojalla ratkaisee viime kädessä onnistumisemme tutkijoina.

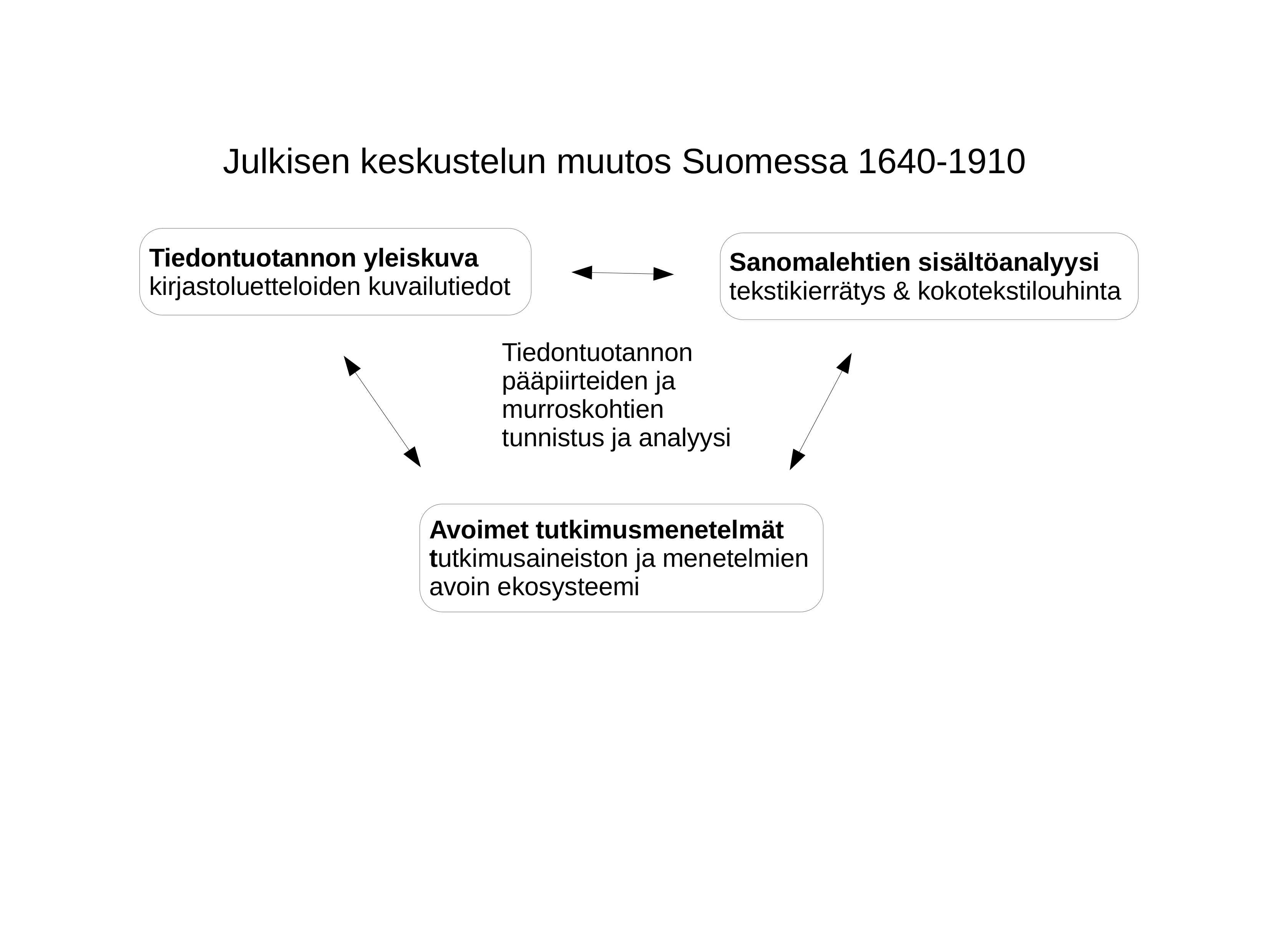
**Lopuksi**

Kuvailutietojen, tekstilouhinnan ja datatieteiden yhdistely tarjoaa uusia tapoja ymmärtää esimerkiksi painotuotteita fyysisinä objekteina, niiden tuottamiseen liittyviä näkökulmia ja laajempia käsitteiden muutoksia. Digitaalisten ihmistieteiden käsite on tarkoitettu väliaikaiseksi luonnehtimaan tietyntyyppistä digitaalisilla aineistoilla tehtävää tutkimusta, jonka varsinaisena päämääränä on uudistaa humanistista tutkimusperinnettä. Todennäköistä on, että digitaalisten aineistojen käyttö leviää suhteellisen nopeasti tutkimuksen valtavirtaan myös historiassa. Tällöin voidaan palata puhumaan historiantutkimuksesta ja humanismista ilman digitaalisuutta tai muita etuliitteitä. Yksittäisten tutkimusperinteiden jopa vuosisataiset jatkumot voivat jatkaa elämäänsä menetelmäkehityksen luomista mahdollisuuksista ammentaen.

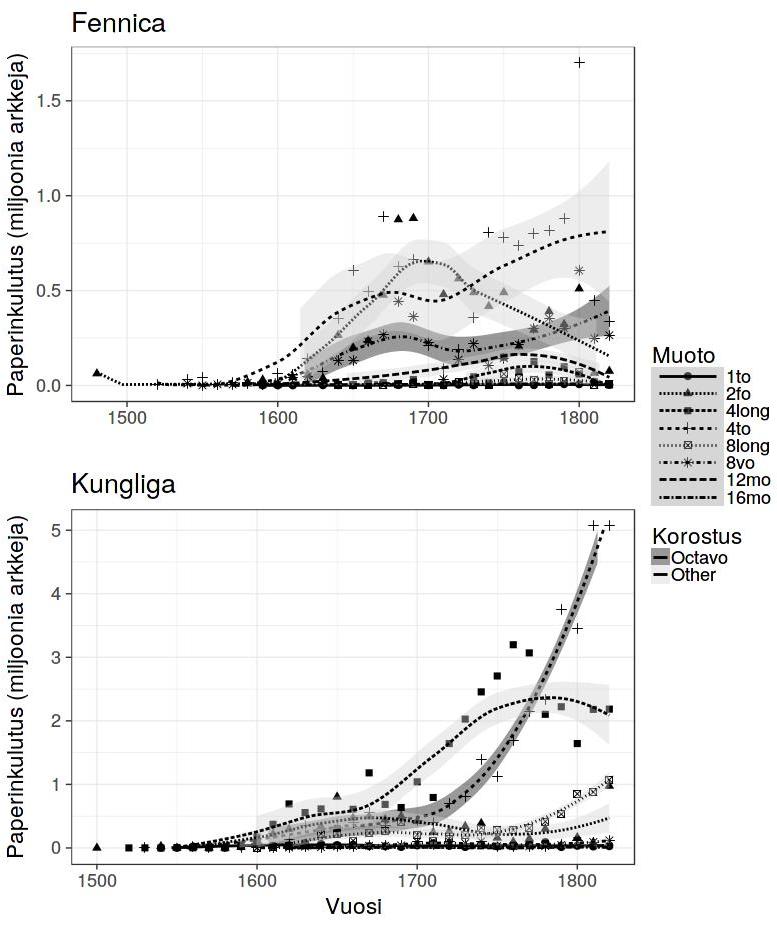
Historian tutkimuksen kannalta käyttökelpoisia digitaalisia aineistoja on jo runsaasti saatavilla. Kotimaiset aineistot, esimerkiksi Suomen kansalliskirjaston kokoelmat, tarjoavat kiintoisan tutkimuskohteen. Lisähaasteena on kehittää menetelmiä, joilla tekstiaineistoja yhdistäviä ja erottavia piirteitä voitaisiin sujuvasti analysoida yli kielirajojen. Suomen kaksikielisyys tarjoaa erinomaisia mahdollisuuksia pilotoida rajatussa mittakaavassa eri kielten käyttöä ja yhdistelyä tekstilouhinnassa. Suomalaisten sanomalehtien tutkimuksessa voidaan yhdistää monia kotimaisen historian osaamisen parhaita puolia liittyen aineiston sisältöön, käsitehistoriaan ja lingvistiikkaan. Lisäksi tämä tarjoaa erinomaisia mahdollisuuksia kehittää yhteistyötä tieteellisen laskennan asiantuntijoiden kanssa tavalla, joka tarjoaa uusia tutkimusideoita sekä ihmistieteiden tutkimukseen että datatieteen menetelmäkehitykseen.

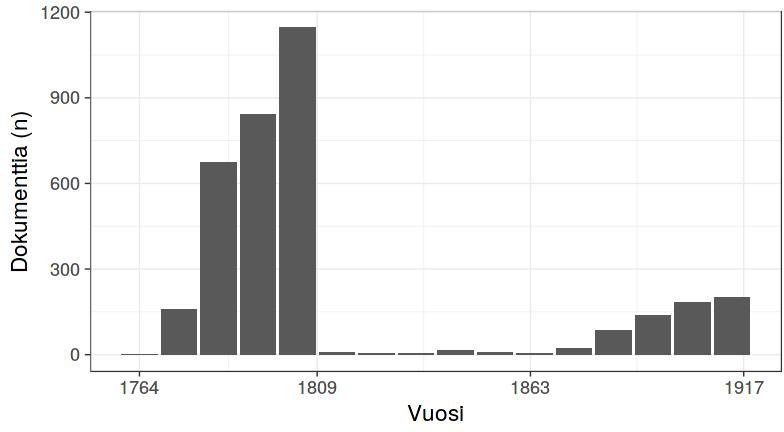
Ihmistieteiden saralla seuraavaksi tavoitteeksi voitaisiin ottaa pyrkimys kvantitatiivisten tutkimusmenetelmien yhtenäistämiseen ja yhteistyön edistämiseen luonnontieteissä jo pidemmälle vakiintuneen avoimen toimintamallin keinoin. Samalla tulee kehittää alan opetusta, jotta pystymme jatkossakin tuottamaan uusia sukupolvia, jotka pystyvät haastamaan aikaisempia tulkintoja tutkimusalan omista lähtökohdista ja osaavat ammentaa lisäarvoa monitieteisestä yhteistyöstä. Olisi hyvä kehittää historian tutkimuksen ja digitaalisten ihmistieteiden yhteisiä linjauksia alan kotimaisesta opetuksesta. Humanistisen tutkimustradition muutokseen liittyy uusien tutkimusmenetelmien ja aineistojen käytettävyyden edistäminen. Tässä avoimen tieteen liikkeen korostamat ajatukset ovat keskeisellä sijalla. Tarvitaan konkreettisia ja pitkäjänteisiä toimia alan toimintakulttuurin uudistamiseksi. Tieteellisen laskennan asiantuntijoita on saatava mukaan ihmistieteiden piiriin tavalla, joka palvelee ihmistieteiden perinteisten ydinteemojen tutkimusta.

Kuva 1. Kuvailutietojen ja kokotekstien sovelluspotentiaali laskennallisessa historiantutkimuksessa.

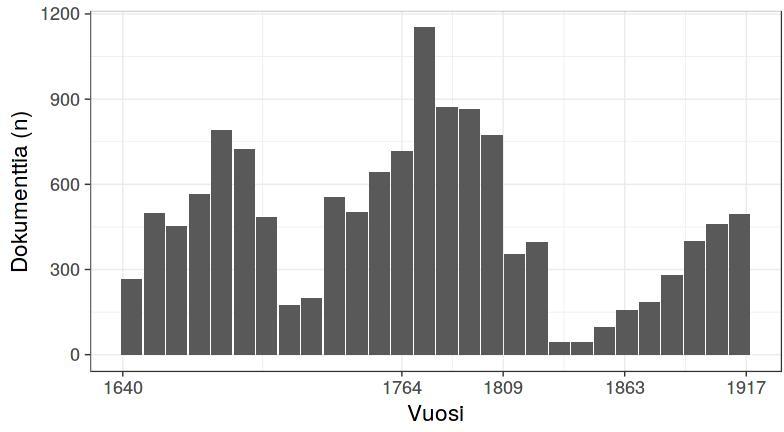


Kuva 2. Kotimaisen tiedontuotannon kehitystä on tutkittu yhdistämällä painetun materiaalin yleisiä kuvailutietoja tekstisisältöihin. Avoimen tieteen menetelmät ovat keskeisiä suurten tietomäärien laadukkaassa tutkimuskäytössä.

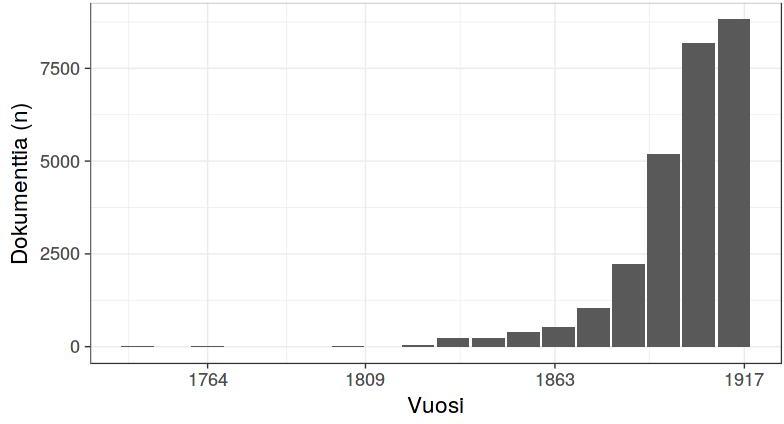
Kuva 3. Paperinkulutus eri kirjaformaattien mukaan 1600-1828 Fennicassa ja Kungligassa. Kuvasta näkyy erityisesti Kungligan ja Ruotsin kohdalla yleiseurooppalainen trendi, jossa octavo-kokoinen kirja (tumma korostus) nousee keskeisimmäksi tiedontuotannon muodoksi 1700-luvun aikana. Tämä kertoo paperinkulutuksen kasvusta, joka kertoo lukemisen leviämisestä laajempien kansanosien keskuuteen valistuksen aikakaudella. Fennicassa havaittavat lievemmät muutokset heijastelevat sitä, että Suomessa tapahtunut painotuotanto ei koskenut niinkään kirjoja, vaan virallisia asiakirjoja. Kirjoja tuotiin Suomeen esimerkiksi Ruotsin alueelta.



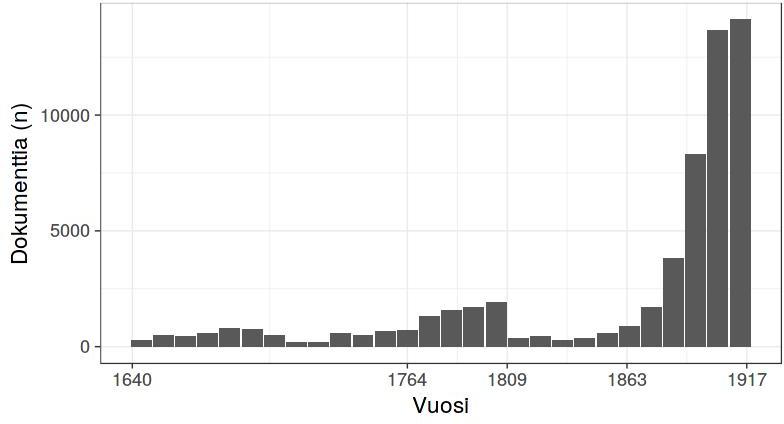
Kuva 4. Julkaisutoiminta Vaasassa vuosikymmenittäin 1760–1917. Aineisto päättyy vuoteen 1917, joten vertailussa on huomioitava, että viimeinen palkki kattaa muihin verrattuna kaksi vuotta lyhyemmän aikajakson (kuvat 4-8). Oheinen kuva viittaa tiedontuotannon romahdukseen Vaasassa välittömästi Venäjän vallan alussa [1]. 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 käänteitä. Suomen kansallisbibliografian Fennican analyysin perusteella julkisuuden kehitys painotuotannon muodossa näyttäisi olevan Venäjän vallan aikana hitaampaa kuin mitä yliopiston siirron ja Helsingin kehittämisen merkitystä korostanut historiantutkimus on aiemmin esittänyt.[[49]](#footnote-65) Mukana on muitakin selittäviä tekijöitä, kuten systemaattisia keisarikunnan aikaiseen painotuotantoon liittyviä puutteita kirjastoluettelossa (on tunnettua, että Fennican luetteloinnista puuttuu mm. 1800-luvun ruotsinkielistä kaunokirjallisuutta). On kuitenkin selvää, että aineistojen kvantitatiivinen analyysi tarjoaa tehokkaan tavan tunnistaa tiedontuotannon historiaan liittyviä trendejä. Lopulliset johtopäätökset tulee varmistaa myös muilla keinoin. Yksi syy siihen, miksi tahdomme tuoda nämä vahvistamattomat tilastolliset näkökulmat esiin on, että emme yksinkertaisesti voi jäädä odottelemaan kansallisbibliografian täydentymistä enää seuraaviksi vuosikymmeniksi, vaan tämä tärkeä luettelointityö on saatava viimeisteltyä mahdollisimman pikaisesti.[[50]](#footnote-66)



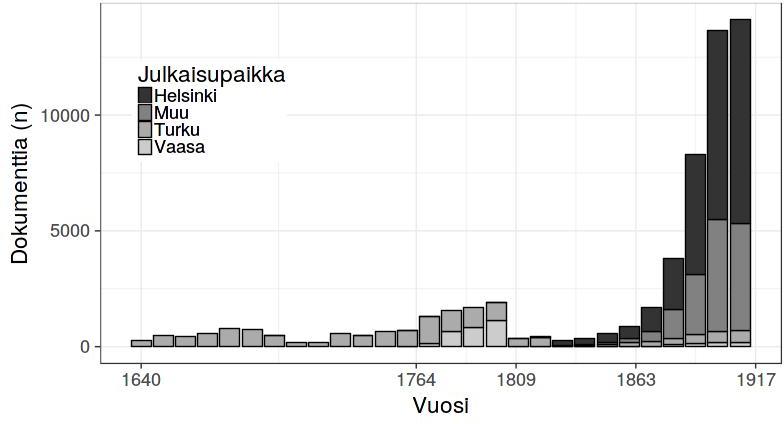
Kuva 5. Julkaisutoiminta Turussa 1640–1917. 1700-luvun kehitys näyttäisi seuraavan eurooppalaista kasvavaa trendiä, kirjapainotoiminnan kehitys 1820-luvulta lähtien on taas nykyisten luettelointitietojen valossa erittäin hidasta.



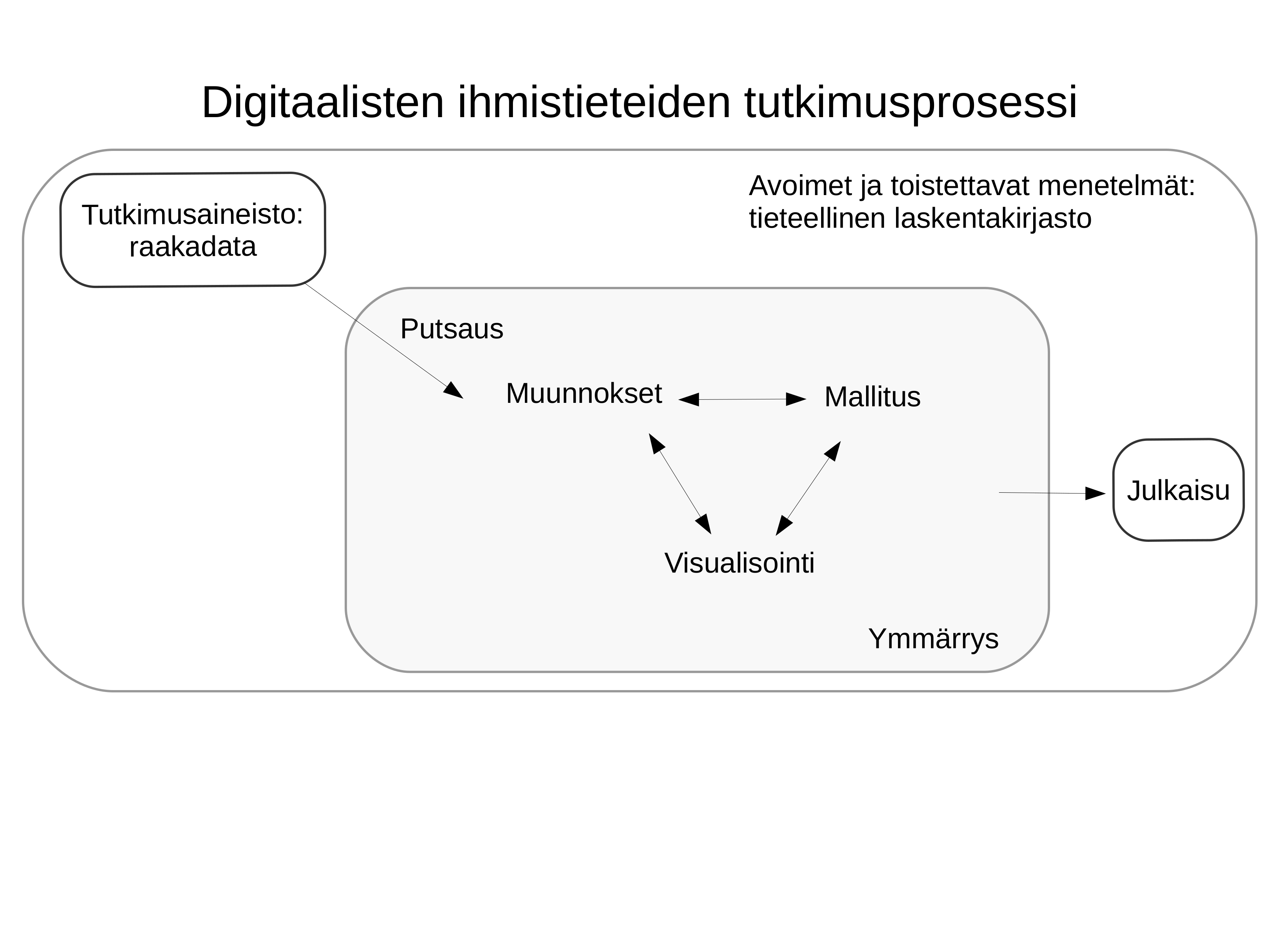
Kuva 6. Julkaisutoiminta Helsingissä 1740–1917. Helsingin kirjapainotoiminnan kehitys Fennican sisältämän aineiston pohjalta viittaa siihen, että ennakkosensuurin osittaisen poistumisen vaikutukset 1850-luvulla ovat yllättävän vähäiset ja varsinainen painotoiminnan kehitys lähtee käyntiin vasta huomattavasti myöhemmin. Yliopiston siirtäminen Helsinkiin näyttää olleen julkaisutoiminnan kannalta, jos ei lamauttava, niin ainakin hidastava toimenpide.



Kuva 7 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 noustaan samalle tasolle kuin Ruotsin vallan aikana. Tämä kuva tulee varmasti tarkentumaan kansallisbibliografian täydentymisen myötä.



Kuva 8 Julkaisutoiminta Suomessa, Helsingissä, Turussa ja Vaasassa 1640-1917. Tässä esitettävä kuvien 4-7 yhteenveto korostaa Suomessa tapahtuneita julkaisutoiminnan muutoksia. Suomalainen painotoiminta painottuu vahvasti Helsinkiin 1800-luvun loppupuolella. Turun painotoiminta on laajassa mittakaavassa stabiilia aina 1800-luvun alkuun asti jolloin myös Vaasan julkaisutoiminta notkahtaa.



Kuva 9. Datasta tietoon. Alkuperäisen tutkimusaineiston tai -aineistojen jalostaminen tiedoksi ja lopulliseksi julkaisuksi on monivaiheinen prosessi, jossa tarvitaan välineitä aineistojen järjestämiseen, yhdistelyyn, ja tilastolliseen analysointiin (Wickham ja Grolemund, 2017).[[51]](#footnote-67) Merkittäviä osia työstä voidaan automatisoida laatimalla tarkoitukseen suunnattuja tieteellisiä laskentakirjastoja. Lähdekoodin avoin julkaisu tekee tutkimusprosessista toistettavan ja läpinäkyvän, mahdollistaa datan ja menetelmien jatkokäytön, ja edistää laajemman tutkijaverkoston osallistumista kehitystyöhön. Lopullinen julkaisu voi sisältää tieteellisen analyysin lisäksi tutkimusaineistoja ja tilastollisia tutkimusmenetelmiä.

**Kirjallisuus**

David Armitage, ‘What’s the Big Idea? Intellectual History and the Longue Durée’, History of European Ideas, 38, 2012, pp. 493-507.

Peter de Bolla, *The Architecture of Concepts: The Historical Formation of Human Rights* (Fordham University Press, 2013).

Fiona Cameron, ‘Beyond the Cult of the Replicant: Museums and Historical Digital Objects—Traditional Concerns, New Discourses’, kirjassa Fiona Cameron ja Sarah Kenderine (toim.), *Theorizing digital cultural heritage* (MIT Press, 2007), pp. 49-75.

Constance Crompton, Richard J. Lane and Raymond Siemens, (eds), *Doing Digital Humanities: Practice, Training, Research* (Routledge, 2017).

Marilyn Deegan and Willard McCarty, *Collaborative Research in the Digital Humanities* (Ashgate, 2012).

**Kimmo Elo (toim.):** Digitaalinen humanismi ja historiatieteet **(Turku: Turun historiallinen yhdistys ry, 2016).**

Margaret Gardiner-Garden and Timothy Littlejohn, ‘A comparison of microarray databases’, Brief Bioinform, 2001, s.143-158. doi: 10.1093/bib/2.2.143<http://bib.oxfordjournals.org/content/2/2/143.short>

**Matthew K. Gold (toim.) *Debates in the Digital Humanities* (University of Minnesota Press, 2012 2016).** <http://dhdebates.gc.cuny.edu/>

Shawn Graham, et al., *Big Digital History: Exploring Big Data Through a Historian’s Macroscope* (London, 2014).

Jo Guldi and David Armitage, [The History Manifesto](http://scholar.harvard.edu/armitage/publications/history-manifesto) (Cambridge: Cambridge University Press, 2014).

Jyrki Hakapää. *Kirjan tie lukijalle. Kirjakauppojen vakiintuminen Suomessa 1740–1960* (Helsinki: SKS, 2008).

Megan L. Head, Luke Holman, Rob Lanfear, Andrew T. Kahn and Michael D. Jennions, ‘The Extent and Consequences of P-Hacking in Science’, *PLOS*, <http://dx.doi.org/10.1371/journal.pbio.1002106> 2015.

Susan Hockey, ‘The History of Humanities Computing’, in S. Schreibman et al. (eds.), *A Companion to Digital Humanities*, 2004, pp. 3–19. <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=default>

Ari Häyrinen, *Open sourcing digital heritage: digital surrogates, museums and knowledge management in the age of open networks* (<https://jyx.jyu.fi/dspace/handle/123456789/40157>), 2012.

John PA Ioannidis, ‘Why Most Published Research Findings Are False.’.PLoS Med 2(8): e124. doi: 10.1371/journal.pmed.0020124, 2005. <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>

John P. A. Ioannidis, ‘How to Make More Published Research True’, *PLOS*, 2014, <http://dx.doi.org/10.1371/journal.pmed.1001747>

Kimmo Kettunen, Tuula Pääkkönen ja Mika Koistinen, ‘Kansalliskirjaston digitoitu historiallinen lehtiaineisto 1771–1910: sanatason laatu, kokoelmien käyttö ja laadun parantaminen’, *Informaatiotutkimus*, 35, 2016, s. 3-14.

Matti Klinge, *Keisarillinen Aleksanterin Yliopisto, 1808-1917* (Helsinki: Otava, 1989)

Matti Klinge, *Pääkaupunki. Helsinki ja Suomen valtio 1808-1863* (Helsinki: Otava, 2012).

Leo Lahti,Niko Ilomäki ja Mikko Tolonen, ‘A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470-1800’, *LIBER Quarterly*. 25, 2015, pp. 87–116.

Lahti L, da Silva F, Laine M, Lähteenoja V, Tolonen M (2017) Alchemy & algorithms: perspectives on the philosophy and history of open science. Research Ideas and Outcomes 3: e13593. <https://doi.org/10.3897/rio.3.e13593>

Jussi T. Lappalainen ja Vesa Lappalainen, ‘Historiantutkimusta ilman papereita’, *Historiallinen aikakauskirja,* 81, 1983, 75-78.

Jani Marjanen, *Den ekonomiska patriotismens uppgång och fall: Finska hushållningssällskapet i europeisk, svensk och finsk kontext 1720–1840*, 2013.

Willard McCarty, ‘What is Humanities Computing? Towards a Definition of the Field’ [http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty,%20What%20is%20humanities%20computing.pdf](http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty, What is humanities computing.pdf) (1998)

Willard McCarty, *Humanities Computing*, Basingstoke and New York: Palgrave Macmillan, 2005 ja Harold Short, ‘The Role of Humanities Computing: Experiences and Challenges’, *Literary and Linguistic Computing*, 21 (1), 2006, pp. 15–27. <http://llc.oxfordjournals.org/content/21/1/15.full>.

Jenny Molloy, ‘The Open Knowledge Foundation: Open Data Means Better Science’, PLoS Biol 9(12): e1001195. doi:10.1371/journal.pbio.1001195, 2011, <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195>

Andrew Morin, Jennifer M. Urban, Paul D. Adams, Ian Foster, Andrej Sali, David Baker, and Piotr Sliz, ‘Shining Light into Black Boxes’, *Science*, 336, pp. 159-160, DOI: 10.1126/science.1218263, 2012 <http://www.sciencemag.org/content/336/6078/159.summary>

Tito Orlandi, ‘Is Humanities Computing a Discipline?’ in G. Braungart, K. Eibl and F. Jannidis (eds), *Jahrbuch für Computerphilologie*, 4, Paderborn: Mentis Verlag, pp. 51–8, 2002. <http://computerphilologie.uni-muenchen.de/jg02/orlandi.html>

Semi Purhonen ja Arho Toikka, ‘”Big datan” haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät. Esimerkkitapauksena aihemallianalyysi tasavallan presidenttien uudenvuodenpuheista 1935–2015’, *Sosiologia*, 53, 2016.

Viljo Rasila, ”Tietokone historiantutkimuksessa”, *Historiallinen aikakauskirja*, 65, 1967, 140–146.

Viljo Rasila, *Tilastolliset menetelmät* historiantutkimuksessa (Helsinki: Otava, 1977).

Jaakko Suominen ja Anna Sivula, ‘Digisyntyisten ilmiöiden historiantutkimus’, teoksessa Elo, Kimmo (toim.). Digitaalinen humanismi ja historiatieteet. Turku: Turun historiallinen yhdistys, 2016, s. 94-127.

Patrik Svensson, ‘The Landscape of Digital Humanities’, *DHQ*, 4 (1), 2010. <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>

Patrik Svensson, ‘Envisioning the Digital Humanities’, *DHQ*, 6 (1), 2012 <http://digitalhumanities.org/dhq/vol/6/1/000112/000112.html>

Patrik Svensson, *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital* (University of Michigan Press, 2016).

Viveca Still ja Oikeuksien hallinta -työryhmä, ‘Suurimmat juridiset esteet avoimelle tieteelle ja tutkimukselle Suomessa’, Avoin tiede ja tutkimus –hanke, 2016 <http://avointiede.fi/documents/10864/12232/Suurimmat+juridiset+esteet+avoimelle+tieteelle+ja+tutkimukselle+Suomessa/45e6ded5-192b-4c16-8ac6-f0828e146a19?version=1.0>

Mikko Tolonen ja Leo Lahti, ‘Aatehistoria ja digitaalisten aineistojen mahdollisuudet’, *Ennen & Nyt*, 2015

<http://www.ennenjanyt.net/2015/02/koodaamisen-ja-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/>

Mikko Tolonen, Niko Ilomäki, Hege Roivainen ja Leo Lahti, ‘Printing in a Periphery: a Quantitative Study of Finnish Knowledge Production, 1640–1828’, In Digital Humanities 2016: Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków, 2016, pp. 383-385.

John Unsworth, ‘Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?’ Part of a symposium on ‘Humanities Computing: Formal methods, experimental practice’, King’s College, London, 13 May 2000. [http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html](http://digitalhumanities.org/dhq/vol/6/1/000112/000112.html)

Wickham Hadley ja Grolemund Garrick. *R for data science: import, tidy, transform, visualize, and model data* (O’Reilly Media Inc., Canada, 2017).

1. Luku on muokattu aikaisemmin julkaistusta ”Aatehistoria ja digitaalisten aineistojen mahdollisuudet”. Tekstiä, esimerkkejä ja viitteitä on päivitetty ja sovitettu historian metodiopetuksen tarpeisiin. [↑](#footnote-ref-2)
2. Jussi T. Lappalainen ja Vesa Lappalainen, ‘Historiantutkimusta ilman papereita’, *Historiallinen aikakauskirja,* 81, 1983, 75-78. [↑](#footnote-ref-3)
3. Digitaalisten metodien kirjosta historiantutkijoille, ks. Graham S et al. *Big Digital History: Exploring Big Data Through a Historian’s Macroscope*. London, 2014. Tämän kritiikistä vastaavasti ks. Tim Hitchcock’s blog on 7th November 2014 at<http://historyonics.blogspot.fi/> [↑](#footnote-ref-4)
4. Viljo Rasila, ”Tietokone historiantutkimuksessa”, *Historiallinen aikakauskirja*, 65, 1967, 140–146; merkittävää on myös, että Rasila oli kiinnostunut myös tilastollisten menetelmien käytöstä historiankirjoitukseen (Rasila, *Tilastolliset menetelmät* historiantutkimuksessa, Otava, 1977). Digitaalisesta historiantutkimuksesta tehtiin keväällä 2016 Petri Pajun johdolla kartoitus johon osallistui 17 vastaajaa eli otos on melko suppea ja tuloksista on hankala vetää pidemmälle meneviä johtopäätöksiä (<https://digihistfinlandroadmapblog.wordpress.com/2016/09/01/raportti-kyselyvastauksista/>). Syksyllä 2016 kartoitettiin Inés Matresin johdolla yleisemmin humanistisen alan digitaalisten metodien käyttöä Suomessa, vastaajia oli 239 ja nämä tulokset kannattaa yhdistää edelliseen (<https://www.helsinki.fi/en/news/finnish-survey-on-digital-research-practice-in-the-arts-and-humanities>). [↑](#footnote-ref-5)
5. Kiitämme tästä vinkistä Jussi T. Lappalaisesta anonyymia arvioitsijaa. [↑](#footnote-ref-6)
6. Samasta aiheesta katso myös **Kimmo Elo (toim.):** Digitaalinen humanismi ja historiatieteet**. Turun historiallinen yhdistys ry, Turku 2016.** [↑](#footnote-ref-7)
7. **Yleisiä “companion” esityksiä Digital Humanities aiheesta löytyy useita, yksi parhaista on Matthew K. Gold (toim.) *Debates in the Digital Humanities*, University of Minnesota Press, 2012 (2016) http://dhdebates.gc.cuny.edu/** [↑](#footnote-ref-8)
8. Eräs mielenkiintoinen kehitys on aiemmin suositun “Humanities computing” termin suhde “Digital humanities” termin käyttöön. “Humanities computing” termiä on määritelty (1998-2006) esim. seuraavissa teksteistä: Willard McCarty, ‘What is Humanities Computing? Towards a Definition of the Field’ [http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty,%20What%20is%20humanities%20computing.pdf](http://staff.cch.kcl.ac.uk/~wmccarty/essays/McCarty, What is humanities computing.pdf) (1998); John Unsworth, ‘Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this?’ Part of a symposium on ‘Humanities Computing: Formal methods, experimental practice’, sponsored by King’s College, London, 13 May 2000. [http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html](http://digitalhumanities.org/dhq/vol/6/1/000112/000112.html); Tito Orlandi, ‘Is Humanities Computing a Discipline?’ in G. Braungart, K. Eibl and F. Jannidis (eds), *Jahrbuch für Computerphilologie*, 4, Paderborn: Mentis Verlag, pp. 51–8, 2002. <http://computerphilologie.uni-muenchen.de/jg02/orlandi.html>; Susan Hockey, ‘The History of Humanities Computing’, in S. Schreibman et al. (eds.), *A Companion to Digital Humanities*, 2004, pp. 3–19. <http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=default>; Willard McCarty, *Humanities Computing*, Basingstoke and New York: Palgrave Macmillan, 2005 ja Harold Short, ‘The Role of Humanities Computing: Experiences and Challenges’, *Literary and Linguistic Computing*, 21 (1), 2006, pp. 15–27. <http://llc.oxfordjournals.org/content/21/1/15.full>. Vertaa näitä “Digital Humanities” alaa määritteleviin blogi-kirjoituksiin vuosilta 2014-2016 jotka UCL listannut “Blogs and Online Material” alle seuraavaan linkkiin: [http://blogs.ucl.ac.uk/definingdh/further-reading/#blogs](http://blogs.ucl.ac.uk/definingdh/further-reading/" \l "blogs) Laajempi katsaus aiheeseen löytyy myös Niin & Näin lehdestä:<http://www.ennenjanyt.net/2015/02/koodaamisen-ja-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/> [↑](#footnote-ref-9)
9. Keskustelua tästä laajempien aikajaksojen tutkimuksesta on käyty esim. teoksessa: Guldi, Jo, and David Armitage, [The History Manifesto](http://scholar.harvard.edu/armitage/publications/history-manifesto). Cambridge: Cambridge University Press, 2014. On hyvä huomioida, että kyseinen teksti on saanut osakseen myös ryöpyn kritiikkiä. [↑](#footnote-ref-10)
10. Monitieteisyyden merkitys korostuu merkittävällä tavalla digitaalisissa ihmistieteissä. Esimerkki tästä on digitaalisten ihmistieteiden opintokokonaisuus Helsingin yliopistossa, jossa koulutuksen tähtäimenä on monitieteinen yhteistyö pikemmin kuin yksittäisten tutkijoiden varustelu tutkimusprosessin kaikkien osien hallintaan: <https://www.helsinki.fi/en/researchgroups/helsinki-digital-humanities/teaching>. [↑](#footnote-ref-11)
11. http://www2.uta.fi/ajankohtaista/uutinen/ihmistieteet-haviavat-rahoitusmallissa [↑](#footnote-ref-12)
12. Hyvä näkökulma digitaalisten ihmistieteiden merkitykseen olemassaolevasta kirjallisuudesta löytyy Patrik Svenssonilta. Katso esim. Svensson, ‘The Landscape of Digital Humanities’, *DHQ*, 4 (1), 2010. <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>;Svensson, ‘Envisioning the Digital Humanities’, *DHQ*, 6 (1), 2012 <http://digitalhumanities.org/dhq/vol/6/1/000112/000112.html> ja Svensson, *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. University of Michigan Press, 2016. [↑](#footnote-ref-13)
13. Tästä aiheesta, katso Ari Häyrisen väitöskirja, Open sourcing digital heritage : digital surrogates, museums and knowledge management in the age of open networks (<https://jyx.jyu.fi/dspace/handle/123456789/40157>), 2012. Kiitämme Jaakko Suomista viitteestä. [↑](#footnote-ref-14)
14. Modaliteettien ja materiaalisuuden merkitys analoginen/digitaalinen debatissa on ollut merkittävä erityisesti museoiden kontekstissa. Tämän jo pitkään jatkuneen debatin hyvä pohdinta on esimerkiksi Fiona Cameron, ‘Beyond the Cult of the Replicant: Museums and Historical Digital Objects—Traditional Concerns, New Discourses’, kirjassa Fiona Cameron ja Sarah Kenderine (toim.), *Theorizing digital cultural heritage*, MIT Press, 2007, pp. 49-75. [↑](#footnote-ref-16)
15. (<https://ccdkconceptlab.wordpress.com/>). Hankkeen teoreettisesta asetelmasta, kts. Peter de Bolla, *The Architecture of Concepts: The Historical Formation of Human Rights,*Fordham University Press, 2013. [↑](#footnote-ref-17)
16. (<http://www.kdw.lu.se/>) [↑](#footnote-ref-18)
17. Vuosittaisen, kansainvälisesti suurimman ADHO-yhteisön organisoiman digital humanities konferenssin abstrakteja tutkimalla vuodelta 2016 (<http://dh2016.adho.org/abstracts/>) käy esimerkiksi ilmi, että digital humanities alalla kytkentä historiallisten aiheiden kohdalta varsinaiseen historialliseen tutkimustraditioon ei ole mitenkään itsestään selvää, vaikka useat aiheet itsessään olisivat historiallisia. [↑](#footnote-ref-19)
18. Kansainvälisinä keihäänkärkinä historiantutkimuksen kannalta digitaalisissa ihmistieteissä kannattaa mainita ainakin Frédéric Kaplanin johtama Digital Humanities Laboratory <http://dhlab.epfl.ch/> ja Stanfordin yliopiston CESTA <https://cesta.stanford.edu/> jossa viety läpi useita historiantutkimusta muokkaavia projekteja. [↑](#footnote-ref-20)
19. Suomi24 aineistosta historiantutkimuksen kannalta ja digisyntyisten ilmiöiden historiantutkimuksesta perusteellisesti katso Jaakko Suominen ja Anna Sivula, ‘Digisyntyisten ilmiöiden historiantutkimus’, teoksessa Elo, Kimmo (toim.). Digitaalinen humanismi ja historiatieteet. Turku: Turun historiallinen yhdistys, 2016, s. 94-127. [↑](#footnote-ref-21)
20. Katso tästä Tolonen ja Lahti, ‘Aatehistoria ja digitaalisten aineistojen mahdollisuudet’, *Ennen & Nyt*, 2015. Katso myös Semi Purhonen ja Arho Toikka, ‘”Big datan” haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät. Esimerkkitapauksena aihemallianalyysi tasavallan presidenttien uudenvuodenpuheista 1935–2015’, *Sosiologia*, 53, 2016. [↑](#footnote-ref-26)
21. Lukuisista aihetta käsittelevistä kokoelmista, katso esim. Constance Crompton, Richard J. Lane and Raymond Siemens, (eds), *Doing Digital Humanities: Practice, Training, Research*. Routledge, 2017 ja Marilyn Deegan and Willard McCarty, *Collaborative Research in the Digital Humanities*. Ashgate, 2012. [↑](#footnote-ref-29)
22. Hyvä yksinkertainen esimerkki tästä on: <http://datascienceplus.com/the-importance-of-data-visualization/> [↑](#footnote-ref-30)
23. Samalla metodit kehittyvät kiihtyvää vauhtia, esimerkki viime kesän DH2016 konferenssista on esim. <http://dh2016.adho.org/abstracts/115>. [↑](#footnote-ref-31)
24. Pohjoismaisista SmartCity-hankkeista, ainakin Aarhusissa humanisteilla on Dariah-verkoston kautta ollut selvä rooli <http://www.smartaarhus.eu/about-smart-aarhus> Hankkeesta enemmän: <http://www.smartculture.eu/sites/default/files/D4.3__PLAN_FOR_THE_INTERNATIONAL_AND_MENTORING_ACTIVITIES.pdf> [↑](#footnote-ref-32)
25. Tästä enemmän täällä: <http://www.ennenjanyt.net/2015/08/aatehistoria-ja-digitaalisten-aineistojen-mahdollisuudet/> [↑](#footnote-ref-33)
26. Suomen akatemian hanke “Digitaalinen historiantutkimus ja julkisuuden muutos Suomessa 1640–1910” (hankenro 293316). [↑](#footnote-ref-34)
27. https://comhis.github.io/posters/octavo/ [↑](#footnote-ref-35)
28. Tämä arvio perustuu COMHIS akatemiahankkeen puitteissa tehtyyn työhön kirjastoluettelojen putsauksessa ennen kuin niillä voidaan tehdä varsinaista tilastollista analyysiä ja tutkimusta. Varoittava esimerkki vastaavien aineistojen käytöstä ilman esikäsittelyä on esimerkiksi <http://www.gdeltproject.org/> jossa dataa on kyllä paljon ja visualisoinnit vastaavia kuin muissakin projekteissa, mutta ongelmat syntyvät siitä että lähdeaineistoon ja näin ollen myöskään johtopäätöksiin ei voida luottaa. Katso myös Leo Lahti, Niko Ilomäki ja Mikko Tolonen, ‘A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470-1800’, *LIBER Quarterly*, 25, 2015, pp. 87–116 ja Mikko Tolonen, Niko Ilomäki, Hege Roivainen ja Leo Lahti, ‘Printing in a Periphery: a Quantitative Study of Finnish Knowledge Production, 1640–1828’, In Digital Humanities 2016: Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków, 2016, pp. 383-385. [↑](#footnote-ref-36)
29. Latinankielisestä aineistosta yleisemmin ks. [www.brepolis.net](http://www.brepolis.net/). [↑](#footnote-ref-37)
30. <http://digi.kansalliskirjasto.fi/sanomalehti> [↑](#footnote-ref-38)
31. Kimmo Kettunen, Tuula Pääkkönen ja Mika Koistinen, ‘Kansalliskirjaston digitoitu historiallinen lehtiaineisto 1771–1910: sanatason laatu, kokoelmien käyttö ja laadun parantaminen’, *Informaatiotutkimus*, 35, 2016, s. 3-14. [↑](#footnote-ref-39)
32. John PA Ioannidis, ‘Why Most Published Research Findings Are False.’.PLoS Med 2(8): e124. doi: 10.1371/journal.pmed.0020124, 2005. <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124> [↑](#footnote-ref-40)
33. Margaret Gardiner-Garden and Timothy Littlejohn, ‘A comparison of microarray databases’, Brief Bioinform, 2001, s.143-158. doi: 10.1093/bib/2.2.143<http://bib.oxfordjournals.org/content/2/2/143.short> [↑](#footnote-ref-41)
34. <http://avointiede.fi/> [↑](#footnote-ref-42)
35. Viveca Still ja Oikeuksien hallinta -työryhmä, ‘Suurimmat juridiset esteet avoimelle tieteelle ja tutkimukselle Suomessa’, Avoin tiede ja tutkimus –hanke, 2016 <http://avointiede.fi/documents/10864/12232/Suurimmat+juridiset+esteet+avoimelle+tieteelle+ja+tutkimukselle+Suomessa/45e6ded5-192b-4c16-8ac6-f0828e146a19?version=1.0> [↑](#footnote-ref-43)
36. <https://ec.europa.eu/priorities/digital-single-market_en> [↑](#footnote-ref-44)
37. Jenny Molloy, ‘The Open Knowledge Foundation: Open Data Means Better Science’, PLoS Biol 9(12): e1001195. doi:10.1371/journal.pbio.1001195, 2011, <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195> [↑](#footnote-ref-45)
38. 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. [↑](#footnote-ref-46)
39. Tästä keskustelua, esim. Hannu Salmi ja Mikko Tolonen, ”Ilman lehtiarkistoja me historiantutkijat emme voi tehdä työtämme”, Helsingin Sanomat, 24.4.2017. <http://www.hs.fi/paivanlehti/24042017/art-2000005182067.html> [↑](#footnote-ref-47)
40. <http://www.sciencemag.org/content/336/6078/159.summary>; John P. A. Ioannidis, ‘How to Make More Published Research True’, *PLOS*, 2014, <http://dx.doi.org/10.1371/journal.pmed.1001747> [↑](#footnote-ref-49)
41. Esim. Github, Gitlab ja Bitbucket [↑](#footnote-ref-50)
42. Megan L. Head, Luke Holman, Rob Lanfear, Andrew T. Kahn and Michael D. Jennions, ‘The Extent and Consequences of P-Hacking in Science’, *PLOS*, <http://dx.doi.org/10.1371/journal.pbio.1002106> 2015 [↑](#footnote-ref-56)
43. John PA Ioannidis, ‘Why Most Published Research Findings Are False.’.PLoS Med 2(8): e124. doi: 10.1371/journal.pmed.0020124, 2005. <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124> [↑](#footnote-ref-57)
44. <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> [↑](#footnote-ref-58)
45. Helsingin yliopistossa esim. Kimmo Koskenniemi. [↑](#footnote-ref-59)
46. Mielenkiintoista kehitystä ja seurattavaa toki riittää, niistä tässä mainittakoon:<http://benschmidt.org/projects/digital-humanities-research/> [↑](#footnote-ref-62)
47. Tästä ks.<http://www.geschichte.uni-frankfurt.de/43010469/jussenb> ja [http://historymanifesto.cambridge.org](http://historymanifesto.cambridge.org/), ks. myös David Armitage, ‘What’s the Big Idea? Intellectual History and the Longue Durée’, History of European Ideas, 38, 2012, pp. 493-507. [↑](#footnote-ref-63)
48. 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. [↑](#footnote-ref-64)
49. Matti Klinge, Keisarillinen Aleksanterin Yliopisto, 1808-1917 (Helsinki: Otava, 1989); Matti Klinge, Pääkaupunki. Helsinki ja Suomen valtio 1808-1863 (Helsinki: Otava, 2012). [↑](#footnote-ref-65)
50. Ks. esim. Jyrki Hakapää. *Kirjan tie lukijalle. Kirjakauppojen vakiintuminen Suomessa 1740–1960.* SKS, 2008, 380) [↑](#footnote-ref-66)
51. Wickham Hadley ja Grolemund Garrick. *R for data science: import, tidy, transform, visualize, and model data* (O’Reilly Media Inc., Canada, 2017). Avoimen tieteen tutkimusprosessin kehityksestä ks. myös Lahti L, da Silva F, Laine M, Lähteenoja V, Tolonen M (2017) Alchemy & algorithms: perspectives on the philosophy and history of open science. Research Ideas and Outcomes 3: e13593. <https://doi.org/10.3897/rio.3.e13593> [↑](#footnote-ref-67)