

Historia (hyper)textus, historia digitalis

Mikko Kainulainen

Arvosteltu teos: Milligan, Ian 2019. History in the Age of Abundance? How the Web Is Transforming Historical Research. Montreal & Kingston: McGill-Queen's University Press. 310.

Käsitteitä määriteltäessä eräänlaista tukipistettä haetaan usein sanojen etymologisista juurista. Esimerkiksi historian tapauksessa tällainen viitepiste nähdään joskus tarinoissa ja kertomuksissa, toisinaan taas todistamisessa ja tutkimisessa. Ian Milliganin hiljattain ilmestyneessä *History in the Age of Abundance?* -teoksessa historiaa lähestytään ennen kaikkea jälkimmäisestä näkökulmasta. Mutta historian sijaan teosta käsitellessä lähtökohdaksi voisi oikeastaan mielekkäämmin hakea kaksi muuta nykyaikaisen historiantutkimuksen kannalta merkittävää sanaa: *teksti* ja *digitaalinen*.

Historiantutkimuksen aineistoina ja tuotoksina – ja joskus kohteinakin – ovat yleensä erilaiset tekstit. Etymologisesti teksti-sanan juuriksi on esitetty yhtäältä kirjoituksiin ja dokumentteihin, toisaalta punokseen, kudokseen ja verkkoon viittaavia sanoja (vrt. tekstuur, tekstiili). 2000-luvulla historiantutkimus tapahtuu enenevässä määrin digitaalisesti. Tämän sanan etymologia taas viittaa sormiin. Nykyään näppäimistöllä työskentely on useimmille olennainen keino kytkeä yhteen sormet ja digitaalisen informaation käsittely, mutta oikeastaan digitaalisuuden merkityshistoria kumpuaa enemmän sormien ja kymmenjärjestelmän yhteydestä. Lukujärjestelmillä katsotaan yleensä olevan jokin kantaluku: tietokoneiden käyttämässä binäärijärjestelmässä se on kaksi (0–1), desimaalijärjestelmässä kymmenen (0–9). Näistä jälkimmäisen on katsottu perustuvan juuri sormiin, ja käytämme usein vieläkin sormiamme kymmenjärjestelmässä toteutettavaan laskemiseen – ja etenkin sen opetteluun. Tästä näkökulmasta sormet voikin nähdä osana laskemisen teknologiaa.

Kanadalaisessa Waterloon yliopistossa vaikuttavan historian apulaisprofessori Ian Milliganin teoksen keskiössä ovatkin tekstit, verkot, digitaalisuus ja teknologia. Kun Milligan tarkastelee tekstien verkottuneisuutta, kohteena ovat digitaalisen teknologian myötä kehittyneet ilmaisumuodot, erityisesti hypertekstit. Teoksen ytimessä ovat ”digitaalisena syntyneet” sisällöt ja verkkoteknologiat, joihin tällaiset sisällöt kytkeytyvät. Milligan lähestyy nykyistä ja tulevaa historiantutkimusta kahden toisiinsa kietoutuneen kysymyksen kautta. Ylätason kysymyksenä voi nähdä sen, mikä digitaalisuuden merkitys on historiantutkimukselle, ja miten historioitsijoiden tulisi suhtautua digitaalisuuteen ja siihen liittyviin teknisiin ratkaisuihin. Yksityiskohtaisemmalla tasolla Milligan keskittyy digitaalisiin verkkoihin ja verkkojärjestelmiin – etenkin internetiin ja world wide webiin – liittyviin kysymyksiin. Tällä tasolla vastauksia haetaan siihen, miten historioitsijoiden tulisi asennoitua ja asemoi-

tua suhteessa verkkoihin, mitä historioitsijoiden tulisi ymmärtää niistä, ja minkälaisia ratkaisuja digitaalisissa verkoissa muodostuneen ja sijaitsevan aineiston tutkimiseksi on hahmoteltu.

Suurin osa teoksesta käsittelee yksityiskohtaisemman tason kysymyksiä: esittelyjä, tapauskuvauksia ja pohdintoja html-aineistoista sekä www- ja internet-arkistoista. Tämä puoli kirjasta (etenkin luvut 2, 3, 5 ja 6) puhuttelee eniten historian tutkimusta ja arkistotyötä tekeviä. Ylätason kysymykset taas puhuttelevat helpommin myös laajempaa historiasta kiinnostunutta lukijakuntaa. Niitä käsitellään etenkin johdannossa, neljännessä luvussa sekä loppupäätelmissä.

Johdannossa Milligan asemoituu osaksi digitaalista humanismia ja massadatan (big data) tutkimusta, kontekstualisoituu suhteessa yleisempään keskusteluun ”digitaalisesta käänteestä” ja kohdistaa oman fokuksensa verkkoarkistoihin (web archives). Milligan painottaa, että teos kumpuaa nimenomaan hänen kokemuksestaan historioitsijana. Tästä perspektiivistä käsin motivaatio teokselle tiivistyy johdannon virkkeeseen: ”Historioitsijoilla on tapana käyttää pitkälti implisiittisiä metodologioita, mutta siirtyessämme uuden tyyppisten lähteiden pariin, joita koulutuksemme ei ole valmistanut käsittelemään, meidän täytyy olla entistä eksplisiittisempiä.” (s. 6, 60). Historioitsijoita verrataan myös kirjasto- ja arkistolaan, josta käsin Milliganin mukaan on toistaiseksi pidetty keulaa monien historioitsijoille olennaisten digitaalisten kysymysten käsitteilyssä (s. 7).

Johdantoa seuraava ensimmäinen luku on eräänlainen internetin, world wide webin, html:n ja verkkoarkistojen pikakurssi historioitsijoille. Milligan painottaa näiden teknologioiden ymmärtämistä useilla tasoilla: esimerkiksi internetin kohdalla tällaisia ovat fyysinen (laitteiden ja infrastruktuurin) taso, koodin taso ja sisällön taso. Näistä kaksi jälkimmäistä tasoa ovat vahvemmin läsnä teoksessa kokonaisuutena. Olennaisten käsitteiden esittelyn lisäksi ensimmäinen luku on myös pienimuotoinen digitaalisten verkkojen historia. Tältä perustalta lukija johdatellaan joihinkin digitaalisen humanismin kantaviin ajatuksiin ja hankkeisiin, kuten esimerkiksi kirjallisuushistorioitsija Franco Morettin ajatukseen etätai kaukolukemisesta (distant reading). Luvun lopussa Milligan esittää kolme piirrettä, joita hän soisi painotettavan enemmän digitaalisten aineistojen kanssa työskenneltäessä: läpinäkyvyys, nöyryys ja subjektiivisuus. Esimerkiksi erilaiset hakukoneet tulevat problematisoiduksi pitkin kirjaa. Milliganin mielestä historioitsijat ovat hyväksyneet hakukoneet ”mustina laatikkoina” eli välineinä joiden toimintamekanismia ei ymmärretä tai edes tunneta, mutta joita silti käytetään. Myös historioitsija Kati Katajisto (2020, 72) on esittänyt huolen ”hermeneuttisen kehän katkeamisesta” tilanteissa, joissa tutkija soveltaa sellaisia laskennallisen datatieteen menetelmiä, joiden tilastollisia malleja ja algoritmeja hän ei ymmärrä. Milliganin mukaan tällaisten välineiden käytöstä itsestään ei tarvitse luopua, mutta siihen, miten hakukoneiden kaltaisia teknologioita käytetään, tarvitaan läpinäkyvyyttä. Nöyryyttä ja subjektiivisuutta Milligan taas painottaa siitä syystä, että ne ovat vaarassa unohtua käsitellessä algoritmien välityksellä suuria aineistoja – aineistoja, jotka vaikuttavat kaiken katavilta, mutta ovat piilevästi täynnä monenlaisia heikkouksia ja arkistointiin liittyviä valintoja.

Toisessa luvussa Milligan esittelee verkkoarkistojen historiaa ja niiden toimintalogiikkaa. Näistä keskeisin on 1990-luvulla perustettu, sanfranciscolaisessa entisessä kirkossa operoiva Internet Archive ja sen 2000-luvun alussa suurelle yleisölle avattu Wayback Machine -sivusto. Kyseessä on hakurobottien (web crawler) avulla muodostettu tietokanta www-sivuista eri aikoina. Wayback Machinen kaltaisten palveluiden avulla käyttäjä voi tavallaan vierailta joillakin valikoiduilla vanhoilla sivuilla, kuten vaikkapa Helsingin Sanomien vuoden 1996 joulukuun 22:n päivän www-sivulla. Mutta oikeastaan monille vanhoil-

le sivuille ei sellaisenaan voi palata. Vain pieni osa verkossa olevasta sisällöstä arkistoidaan. Samalla vajavaiseksi jäävät myös ne sivut, jotka on arkistoitu. Tämä johtuu www-aineiston hypertextisestä luonteesta: useat sisällöt ovat eräänlaisia viitekokoelmia johonkin muualla sijaitsevaan (esim. väriin, kuvioon, ääneen, videoon, kirjoitukseen), jota ei ehkä olekaan arkistoitu. Muun muassa tästä johtuen verkkoarkistot ja niiden data ovat ”sotkuisia”. Tätä sotkua ja sen kanssa työskentelyä Milligan selvittää myös kolmannessa luvussa, jossa hän argumentoi, että perinteisen sisältöjen lähiluvun lisäksi digitaalisten aineistojen – etenkin verkkoaineistojen – parissa työskentelevien historioitsijoiden tulee problematisoida se, mitä pidetään ”sisältönä”, ja tutkia myös metadataa. Karkeasti tiivistäen voisi sanoa, että tällöin tutkimuksen kohteena ei olekaan jokin aineisto, vaan joitain lukuisia aineistoja kuvaileva ja määrittävä informaatio.

Pitkin kirjaa Milligan esittelee lukuisia tapauskuvauksia erilaisista arkistointihankkeista ja historiantutkimuksesta, joista useimmat liittyvät nimenomaan verkkoarkistojen käyttöön. Neljännessä luvussa Milligan reflektoi vierailujaan British Libraryn vapaakappalekokoelmassa keskittyen kiinnostavasti sekä kokoelman yleisiin käyttöehtoihin että digi- ja printtimateriaalien ehtojen eroihin. Kyseisessä tapauksessa digitaalisen aineiston käyttäminen oli huomattavasti jäykempää ja hitaampaa kuin printtiaineiston. Viides luku syventyy GeoCities-palveluun. Luku toteuttaa tavallaan kaksi tehtävää: yhtäältä se esittelee erään verkkoaineiston muutoksineen ja arkistoisprosessineen, toisaalta tarjoaa esimerkin menneeseen verkko yhteisöön kohdistuvasta historiantutkimuksesta. Kuudes luku käy läpi muutamia verkkoaineistojen arkistoisemiseen ja analysoimiseen soveltuvia ohjelmistoja (vapaan ja avoimen lähdekoodin Heritrix, Webrecorder, Gephi ja ImageMagick; avoimen lähdekoodin Archives Unleashed ja Voyant Tools; sekä kaupallinen Archive-It) ja niiden mahdollistamia tekniikoita.

Kirjan päätelmäluvussa tiivistyy vielä sanoma siitä, että historioitsijoiden tulee kehittää ymmärrystään digitaalisuudesta ja siihen liittyvistä teknologioista sekä myös ottaa digitaalisten aineistojen käsittelyyn soveltuvia menetelmiä haltuun. Kun erilaiset aineistot välittyvät monenlaisten digitaalisten palveluiden ja ohjelmistojen kautta, ei tätä välitystä voi ottaa annettuna. Päätelmissään Milligan painottaa digitaalisten aineistojen vaikeuksia ja mahdollisuuksia. Yhtäältä aineistoja on paljon enemmän, uusista ihmisryhmistä ja uudenlaisen toiminnan piiristä. Toisaalta tällaisista aineistoista on helppo luulla liikaa tai väärinkäyttää niitä. Aineistot ovat myös usein uudella tavalla hajanaisia tai katkonaisia, ja niissä on niin paljon informaatiota, että siihen on entistä helpompi hukkua.

Teos puhuttelee ensisijaisesti historioitsijoita, mutta sillä on annettavaa myös historianopetukselle – etenkin sellaiselle historianopetukselle, jossa oppilaiden ja opiskelijoiden on tarkoitus myös *tehdä* historiantutkimusta ja oppia ymmärtämään historiantutkimuksen prosessia eikä pelkästään oppia sen tuotoksista (ks. Rantala ym. 2020). Tällaisissa lähestymistavoissa on painotettu etenkin lähteiden lukemista, kontekstualisoimista, tulkitsemista ja niistä päättelyä. Niiden lisäksi olennaisena voisi nähdä myös ainakin lähteiden etsimisen ja valitsemisen. Digitaalisten tietokantojen myötä monien opettajien mahdollisuudet toteuttaa tai mallintaa arkistoihin liittyvää tutkimusvaihetta opiskelijoiden kanssa ovat helpottuneet huomattavasti. Näihin uusiin mahdollisuuksiin tarttuessa on kuitenkin syytä harjoittaa kriittikkiä myös digitaalisia välineitä ja ympäristöjä kohtaan (ks. Jandrić ym. 2018). Historiaan liittyviä tutkivan oppimisen projekteja suunnitteleville Milliganin kirja tarjoaakin olennais-tietoa verkkoarkistojen rakenteisista ja historiasta, sekä ajatuksia verkkoarkistojen parissa tapahtuvan lähdekritiikin tueksi.

Verkkoaineistoista kiinnostuneille Milligan tarjoaa oivan johdannon aiheeseen, ja sitä kautta päästään myös kiinni laajempiin digitaalisuuden kysymyksiin historiantutkimuksen

kannalta. Kirja liikkuu voimakkaasti käytännön tasolla, ja sen tarkastelut kiinnittyvät jatkuvasti joihinkin tiettyihin sivustoihin, ohjelmistoihin, arkistoihin ja tapauskuvauksiin. Tämä on ehdottomasti ansio. Samalla toisinaan abstraktimpi, filosofinen – kenties materialistinen tai fenomenologinen – käsittelykin olisi voinut olla paikallaan joidenkin kysymysten kohdalla. Esimerkiksi: mitä ylipäänsä merkitsee se, että, jokin teksti (pinta) on olemassa ja tulee luetuksi muodossa, joka ”pinnan alla” koostuu binäärisistä luvuista elektronisissa laitteissa, eikä vaikkapa musteesta paperilla? Mitä tulisi tietää näiden muotojen tuotannosta ja ylläpidosta? Miten verkkosivujen hypertekstilunne eroaa siitä kielen dialogisuuteen ja eroihin perustuvasta verkottuneisuudesta, joka on kirjattuna jo itse sanaan *teksti*, ja josta vaikkapa Mihail Bahtin ja Jacques Derrida kirjoittivat? Ja pitäisikö juuri historioitsijoiden huomioida jotain erityistä näiden kysymysten suhteen?

Historioitsijoiden herätykseksi – takakannen mukaan – tarkoitettulta kirjalta voinee myös kysyä, missä mielessä sen tarkastelutapa ja argumentti on kriittinen? Ilmiselvää on, että historioitsijana Milligan nojautuu paljon lähdekritiikkiin ja haluaa kirjassaan edistää sen edellytyksiä niin verkkoaineistojen kuin muidenkin digitaalisten aineistojen piirissä. Lähdekritiikin lisäksi Milligan lukee myös kriittisen teorian merkittäväksi resurssiksi verkkoarkistojen käsittelyssä. Tosin hän ei tuo suoraan esille, kuinka toteuttaa tämän kriittisen teorian soveltamisen. Toisenlainen kritiikki teoksessa löytyy siitä, kuinka Milligan itse verkkoaineistoja tutkivana historioitsijana haastaa historioitsijoita yleisemmin: hänen mielestään historioitsijat ovat olleet ”verrattain hitaita mukautumaan tähän mediumin muutokseen” (s. 26).

Samalla teosta voisi tarkastella myös laajemmin osana datafikaation ja massadatan kriittistä tutkimusta (critical data studies; Van Dijck 2014; Iliadis & Russo 2016; Kitchin & Lauriault 2018). Joitain yhteneväisyyksiä Milliganin ja tämän suuntauksen väliltä löytyy esimerkiksi siitä, kuinka esille tuodaan datan – ja sen puutteen – valtaan liittyviä kytköksiä, tai siitä kuinka haastetaan ajatuksia siitä, että massadata ”puhui puolestaan” ja korvaisi muunlaiset lähestymistavat. Milliganin perimmäisenä tarkoituksena näyttäytyy kuitenkin ennen kaikkea verkkoaineistojen ja massadatan käytön edistäminen. Hänen argumenttinsa koskee nimenomaan “verkkoarkistojen merkitystä osana kulttuuriperintöömme ja korvaamattomana historiallisena lähteenä.” (s. 24). Milligan tuntuu kutsuvan historioitsijoita hypäämään datafikaation kelkkaan. Toki hänelle kyse on sellaisesta kelkasta, johon pääsyn edellytyksenä ovat useat eettiset – mm. yksityisyyteen ja saatavuuteen liittyvät – pohdinnat. Milliganin tarkastelu ei kuitenkaan laajene pohtimaan vaikkapa massadataa sellaisesta näkökulmasta, jota Rob Kitchin ja Tracey Lauriault (2018) hahmottelevat datakokoonten (data assemblages) käsitteen kautta. Tällöin mukaan kutsutut historioitsijat nähtäisiin osana niitä asiantuntijoiden, instituutioiden, lakien, infrastruktuurien, ajattelumallien ja muiden yhteenkietoutumia, jotka tuottavat, käyttävät, perustelevat, normalisoivat, kaappavat, louhivat, hyödyntävät ja tulkitsevat dataa.

Hakukoneista kirjoittaessaan Milligan nostaa esille, kuinka ne ovat huomattavasti muutaneet historiantutkimuksen käytäntöjä (ks. myös Katajisto 2020). Historioitsija Lara Putnam (2016) on kirjoittanut siitä, kuinka hakukoneilla työskennellessä historioitsijan etsimistä ja löytämistä välittävät varsin erilaiset tekijät. Tällöin paikalliseen asiantuntijuuteen, bibliografioihin ja syvälliseen lukemiseen perustuvan etsimisen sijaan haun tekee algoritmi – sinänsä tehokas, mutta tuntemattomana pysyttelevä keino, joka näennäisessä helppoudessaan aiheuttaa myös hämmennystä. Shoshana Zuboff (2019; ks. myös Melin 2020) on valvontakapitalismin kritiikissään kirjoittanut kahden tekstin ongelmaksi kutsumastaan tilanteesta käyttäen esimerkkinään juuri hakukonetta. Tällöin yksi teksti – se enemmän tai vähemmän julkinen teksti, jonka näemme ja jota luemme ja tuotamme – toimii raaka-ainee-

na toiselle, sitä seuraavalle niin kutsutulle varjotekstile. Meille näkymättömässä toisessa tekstissä dataksi muutettu käyttäytymisemme kumuloituu, ja sitä analysoidaan kaupallisin pyrkimyksin. Tällöin digitaalisten välineiden tuomien tai niiden aiheuttamien muutosten (esimerkiksi sen, miten hakukone välineenä muokkaa historiantutkimusta) lisäksi voisi olla syytä kysyä myös niiden kautta tietämisen ja niiden parissa työskentelemisen performatiivisista eli ilmiöitä tuottavista vaikutuksista. Esimerkiksi: Mitä vaikutuksia hakukäytännöilämme on siihen, millaiset asiat tulevat tulevaisuudessa valikoitumaan ”löydettäväiksi”? Hakujen lisäksi tällainen kysyminen on aiheellista digitaalisille välineille laajemminkin. Vaikka Milligan ei juuri tätä pohdikaan, yksi hänen viesteistään on digitaalisten välineiden problematisointi, ja tässä hän on tärkeällä asialla. Kenties historiantutkimuksessa on tarvetta menetelmällisen kivijalan, lähdekritiikin, rinnalla myös *välinekriittikille* (Koolen ym. 2019) – ja sama pätee varmaankin myös tutkimustyöhön monilla muillakin aloilla.

Kirjallisuus

- Iliadis, Andrew & Russo, Federica 2016. Critical data studies: An introduction. *Big Data & Society* 3 (2), 1–7. <https://doi.org/10.1177/2053951716674238>
- Jandrić, Petar, Knox, Jeremy, Besley, Tina, Ryberg, Thomas, Suoranta, Juha & Hayes, Sarah 2018. Postdigital science and education. *Educational Philosophy and Theory* 50 (10), 893–899. <https://doi.org/10.1080/00131857.2018.1454000>
- Katajisto, Kati 2020. Digitalisaation haasteet historiantutkimuksessa. *Historiallinen Aikakauskirja* 118 (1), 66–75.
- Kitchin, Rob & Laurialt, Tracey P. 2018. Toward critical data studies: Charting and unpacking data assemblages and their work. Teoksessa Thatcher, Jim, Eckert, Josef & Shears, Andrew (toim.), *Thinking big data in geography: New regimes, new research*. Lincoln & London: University of Nebraska Press, 3–20. <https://doi.org/10.2307/j.ctt21h4z6m.6>
- Koolen, Marijn, Van Gorp, Jasmijn, & Van Ossenbruggen, Jacco 2019. Toward a model for digital tool criticism: Reflection as integrative practice. *Digital Scholarship in the Humanities* 34 (2), 368–385. <https://doi.org/10.1093/llc/fqy048>
- Melin, Harri 2020. Kuka tietää, kuka päättää ja kuka päättää, kuka päättää. *Tiede & Edistys* 45 (1), 75–81.
- Putnam, Lara 2016. The transnational and the text-searchable: Digitized sources and the shadows they cast. *The American Historical Review* 121 (2), 377–402. <https://doi.org/10.1093/ahr/121.2.377>
- Rantala, Jukka, Puustinen, Mikko, Khawaja, Amna, van den Berg, Marko & Ouakrim-Soi-vio, Najat 2020. *Näinkö historiaa opitaan?* Helsinki: Gaudeamus.
- Van Dijck, José 2014. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12 (2), 197–208. <https://doi.org/10.24908/ss.v12i2.4776>
- Zuboff, Shoshana 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.

KM Mikko Kainulainen on väitöskirjatutkija Turun yliopiston opettajankoulutuslaitoksella.