

Digitaaliset ihmistieteet tutkimuskartalle

■ Kaius Sinnemäki ja Mikko Tolonen

Humanistisille ja yhteiskuntatieteille (so. ihmistieteet) tyypillinen jaottelu kvalitatiivisen ja kvantitatiivisen tutkimusotteen välille on nopeassa murroksessa. Digitaalisten tietoaaineistojen määrä on kasvanut kiihtyvällä tahdilla ja laskennallisia menetelmiä on yhä enemmän tarjolla tutkijoille. Esimerkkinä kehityksestä on FIN-CLARIN-konsortio, joka on osa Suomen Akatemian tutkimusinfrastruktuurien tiekarttaa vuosille 2014–20.

Nopeasta kehityksestä huolimatta digitaalisten aineistojen ja laskennallisten menetelmien hyödyntäminen humanistisilla ja yhteiskuntatieteellisillä aloilla on huomattavasti vähäisempää kuin luonnontieteissä niin Suomessa kuin kansainvälisestikin. Digitaalisuuden voimakas lisääntyminen on kuitenkin väijäämätön tulevaisuuden trendi, joka vaatii herkkää korvaa toimijoilta. Tässä kehityksessä selvä uhka on humanistis-yhteiskuntatieteellisten alojen lisääntyvä pirstaleisuus, jossa niin sanotut analogiset ihmistieteet ja digitaaliset ihmistieteet eriytyvät omiksi riiteleviksi tieteenaloikseen, yksi painottaen syvällistä ymmärtämistä, toinen laskennallisia menetelmiä ja digitaalisten aineistojen määrää.

Samalla voi toteutua se pahin painajainen, ettei humanistien ja yhteiskuntatieteiden piirissä pian ymmärretä asioita enää muuta kuin pintapuolisesti ja että laskennallisia menetelmiä kehitellään ilman, että niillä olisi konkreettista vaikutusta tieteenalan varsinaisiin tutkimuskysymyksiin (toisin sanoen ”perinteinen” ei kykene uusiutumaan ja ”uusi” keskittyy matkimaan menetelmiä muualta).

Tätä uhkakuvaa torjumaan suunnittelimme Helsingin yliopistossa 1.–3. joulukuuta 2014 kaksiosaisen tapahtuman koskien big dataa ja

digitaalisia ihmistieteitä (#HDHD). Tapahtuman ensimmäisenä osana järjestimme big datan käyttöä historian ja kielen tutkimuksessa käsittelevän symposiumin Helsingin yliopiston tutkijakollegiumissa 1.–2.12.2014. Sen tavoitteena oli muodostaa käsitys Suomessa ja muualla tehtävästä big data -tutkimuksesta näillä aloilla.

Tapahtuman toisena osana järjestimme yhteisen työpajapäivän 3. joulukuuta 2014 Helsingin yliopistossa. Sen tarkoituksena oli tutkijoiden ja palveluntarjoajien – eli muistiorganisaatioiden ja tutkimusyksiköiden – kesken kehittää digitaalisten ihmistieteiden tutkimusta Suomessa ja oppia eri puolilla maailmaa tehtävästä tutkimuksesta tällä alalla.

Tapahtumiin lähti mukaan laaja rintama toimijoita, jotka ovat kiinnostuneet big datasta ja ihmistieteistä Suomessa: mm. Helsingin yliopiston humanistinen tiedekunta, Kansalliskirjasto, FIN-CLARIN-konsortio, Kuluttajatutkimuskeskus, Yhteiskuntatieteellinen tietoarkisto sekä VARIENG- ja BAULT-tutkijaverkostot.

Big data: aineistot ja lähestymistavat historian ja kielen tutkimuksessa

Big data -symposiumiin kutsuimme historioitsijoita, kielen- ja kirjallisuuden tutkijoita, yhteiskuntatieteilijöitä, tietojenkäsittelijöitä ja kirjastoalan asiantuntijoita. Päälimmäisin huomio useista puheenvuoroista oli se, että ihmistieteissä big data (isot aineistot) ei ole eikä ehkä koskaan voikaan olla samankaltaista kuin luonnontieteissä. Siinä missä FIN-CLARINin kieliaineistot kattavat tällä hetkellä viisi miljardia sanaa suomalaista sanoma- ja aikakauslehtitekstiä ja kaukokirjallisuutta vuosia kestäneen digitoinnin tuloksena, saman tallennuskapasiteetin vaativa datamäärä voidaan kerätä luonnontieteellisessä

hankkeessa muutamissa minuuteissa (vrt. Large Synoptic Survey Telescope Project).

Se, mistä humanistis-yhteiskuntatieteellisessä tutkimuksessa voidaan kuitenkin puhua, on big data -lähestymistavasta tutkimukseen. Tämä tarkoittaa uusien menetelmien soveltamista digitaalisiin aineistoihin yhdessä data-analysoidijien kanssa sekä toimintakulttuurin muutosta, joka mahdollistaisi todellisen tieteidenvälisen tutkimuksen ihmistieteissä.

Toinen tärkeä huomio symposiumista oli, että merkittävä osa siitä humanistis-yhteiskuntatieteellisestä big data -tutkimuksesta, mitä Suomessa ja maailmalla tehdään, pohjaa lopulta kuitenkin visioihin siitä, mitä se ehkä voisi jonakin päivänä olla. Maailmallakin on vielä yllättävän vähän kunnan tutkimustuloksia tällä alueella, erityisesti siinä, miten historiallista muutosta mallinnetaan. Huomattava ongelma tässä on, että luonnollisen kielen prosessointi soveltuu jopa yllättävän huonosti historiallisiin aineistoihin tällä hetkellä.

Suomalaisen tutkimuksen kannalta tämä on myönteinen havainto, jos se osataan ottaa oikein vastaan. Vaikka erityisesti angloamerikkalaisessa maailmassa digitaalisissa ihmistieteissä ollaan Suomea tietyssä mielessä edellä, resurssien järjestyksellä sijoittamisella on mahdollista kuroa tuota etumatkaa umpeen. Verrattain vähäisiä panoksiamme ei kuitenkaan kannata tuhata sellaiseen tutkimukseen, jossa maailmalla ollaan kaukana edellä, vaan sinne, missä globaalistikin ollaan vielä alkutekijöissä. Digitoidut historialliset aineistot (toisin sanoen ”ei-digitaalisina syntyneet”) vaikuttivat symposiumin perusteella olevan yksi tällainen alue, jonka tutkimuksellisesti hyödyntämisessä ja kehittämisessä Suomessa voitaisiin luoda kansainvälisesti korkeinta tasoa olevaa tutkimusta jo olemassa olevien suomalaisten aineistojen avulla.

Kolmas tärkeä huomio liittyi tutkimuksen moniäänisyyteen. Vaikka aineistot voivat olla tutkijoilla samantyyppisiä, kielentutkijat ja historioitsijat lähestyvät tutkimusta kuitenkin erilaisista tarpeista ja lähestymistavoista käsin. Monet kielentutkimuksen lähtökohdista syntyneet organisaatiot, joista suurin on FINCLA-

DIGITAALINEN HUMANISTINEN TUTKIMUS SUOMESSA ULKOPUOLISEN SILMIN

■ Kaius Sinnemäki ja Mikko Tolonen

Joulukuussa 2014 järjestetyn Helsinki Digital Humanities Dayn (#HDHD) yksi ulkomaisista pääpuhujista oli Sussexin yliopiston media- ja viestintätieteiden professori Caroline Bassett. Lukuvuoden 2014–15 hän on viettänyt Helsingin yliopiston tutkijakollegiumissa Helsingin Sanomain Säätiön vierailevana professorina. Haastattelimme professori Bassettia liittyen digitaaliseen humanistiseen tutkimukseen Suomessa.

Sussexissa Bassett johtaa nelivuotista hanketta Sussex Humanities Lab. Hanke tutkii, missä määrin laskennalliset teknologiat määrittelevät humanistisia tieteitä uudella tavalla. Siinä tarkastellaan digitaalisen käänteän tuomia muutoksia tieteen tekemisen kulttuuriin ja käytänteisiin humanistisissa tieteissä käyttäen apuna kriittistä analyysia. Suomalaisen humanistin mittapuulla lähes kolmen miljoonan punnan hanke on varsin mittava erityisesti, koska se itsessään on vain siemenraha.

Miten näet big data -tutkimuksen?

Big data on sekä hypeä että todellista. Ilmassa on monia väitteitä siitä, mitä big data aiheuttaa humanistisille tieteille. Yksi niistä kuuluu, että se olisi loppu tietyille humanistisille aloille ja perinteisesti ymmärretylle hermeneuttiselle tutkimukselle. Toisaalta big data mahdollistaa uusien tietolajien yhdistämisen. Tällöin voi syntyä uutta tietoa ja tiedon tuottamista, jossa ovat mukana sekä ihmiset että koneet. Tällainen tutkimus ei tuo mukanaan pelkästään uusia menetelmiä vaan myös muuttaa tiedon tuottamisen kulttuuria ja tapoja, joilla tutkimuskysymyksiä muodostetaan. Ehkä tämä uusiutumisen ja moniäänisyyden näkökulma on juuri se tärkein. Big datalla voi siis olla merkittävä humanistisen tutkimuksen perinteitä uudistava vaikutus, jolloin sen rooli ei rajoitu ainoastaan menetelmien kehittelyyn.

Miten näet digitaalisen humanistisen tutkimuksen Suomessa?

Joulukuun tapahtuman kommentaattorit puhuivat paljon siitä, että se oli ensimmäinen laajempi digitaalisen humanismin tapahtuma Suomessa. Sitä se

ei kuitenkaan ollut. Suomessa on tehty jo paljon relevanttia tutkimusta tälläkin alalla ainakin 15–20 vuoden ajan. Kyse on ehkä pikemminkin siitä, missä määrin alan tutkimus on keskustellut itsensä kanssa Suomessa. Mielenkiintoista on jatkossa nähdä, miten nouseva digitaalinen humanistinen tutkimus suhteutuu Suomessa jo olemassa oleviin vahvuuksiin, kuten kielitieteelliseen tutkimukseen.

Suomen etuna on se, että maa on sekä verrattain pieni että kiinnostavalla ja samalla paradoksisella tavalla yhtenäinen. Tällöin on mahdollista saada samasta asiasta kiinnostuneet relevantit toimijat, kirjastot ja muut tahot, saman pöydän ääreen, kuten näytti tapahtuvan joulukuun tapahtuman yhteydessä. Vierailuni aikana olen havainnut, miten tiiviitä yhteyksiä eri alojen toimijoiden välillä Suomessa on mm. viestintäalan, median ja taiteiden, muotoilun ja tietojenkäsittelyn kesken.

Mitä ehdotuksia sinulla on liittyen digitaaliseen humanismiin Suomessa?

Pahin tulevaisuudenkuva digitaaliselle humanistiselle tutkimukselle Suomessa on se, että alan tutkimus aiheuttaisi vain lisää uusia hyödyttömiä jakolinjoja. Pikemminkin alalla tarvitaan lisää tieteidenvälisyyttä ja vanhojen tieteenalajojen ylittämistä. Tässä kehityksessä avoin tiede on tärkeä: digitaaliset humanistiset hankkeet ovat taloudellisesti katastrofeja ilman avoimuutta.

Alalla on toki tietty jännite jo olemassa olevien vahvuuksien ja sen välillä, millaisille uusille avauksille ovet täytyy pitää auki. Tässä on haasteita, sillä esimerkiksi Sussexin yliopisto vaatii meidän hankkeeltamme pysyvyyttä ja halua nähdä meidät muutosvoimana. Tällöin kysymykseksi nousee, miten itse näemme digitaalisen humanistisen tutkimuksen roolin ryhmittelemässä vanhoja tutkimusalueita uudelleen.

Uudenlaiset tutkimukset vaativat sitä, että niitä tehdään yhteistyössä ja että tässä yhteistyössä nähdään eri toimijoiden vahvuudet. Tästä hyvänä esimerkkinä on Helsingin yliopiston tutkijakollegium, joka on yksikkönä valtavan verkottunut tutkimuskentässä. Kollegium nähdään myös ulkopuolelta ensisijaisesti tutkijayhteisönä eikä pelkästään Helsingin yliopiston yksikkönä. Tämä on hyödyllistä tieteidenvälisen tutkimuksen synnyttämiseen, sillä kollegiumin kaltainen yksikkö voi koota yhteen väkeä eri aloilta ja näin syöttää ajatuksia tieteidenväliseen tutkimukseen, kuten digitaaliseen humanistiseen tutkimukseen.

RIN, joutuvat omaksuma uudenlaisen roolin palvelukseen nimenomaan digitaalisia ihmistieteitä termin laajemmassa merkityksessä. Kun mukaan otetaan vielä laskennallisen tieteen ja muistiorganisaatioiden näkökulmat, on selvää, että digitaaliset ihmistieteet muodostavat monitieteisen tutkimuskentän, jossa eri toimijoiden tarpeista on tärkeä olla selvillä. Tärkeä on tukea myös erityisesti sitä, että aineistot olisivat tutkijoiden saatavissa avoimesti siinä muodossa kuin he sen tarvitsevat. Olisi äärimmäisen tärkeää, että palveluiden tarjoajat ja aineistoja hallinnoivat yksiköt ymmärtäisivät tutkijoiden erilaiset tarpeet eivätkä pyrkisi ajattelemaan tutkijoiden puolesta.

Praktikum-päivä

Tapahtuman toinen osa, Helsinki Digital Humanities Day (#HDHD), keskittyi digitaalisten ihmistieteiden kehittämiseen Suomessa. Kokosimme yhteen tutkijoita, opiskelijoita sekä rahoittajien, kirjastoalan ja hallinnon edustajia oppimaan kansainvälisistä ja kotimaisista digitaalisten ihmistieteiden hankkeista sekä käytännön työpajatyöskentelyn kautta pohtimaan tutkimusalan kehittämistä Suomessa. Tarkoitus oli tutkijalähtöisesti miettiä erilaisten aineistojen käyttöä, niihin liittyviä tekijänoikeuskysymyksiä, tutkimuskysymyksiä ja menetelmiä sekä näin antaa mahdollisuus tutustumiseen sellaisille tutkijoille, joilla ei ole aiempaa kosketuspintaa digitaalisiin ihmistieteisiin.

Ohjelmassa aamupäivän tärkein anti oli saada hyvä ymmärrys digitaalisten ihmistieteiden tutkimuskeskuksista eri puolilla Eurooppaa. Keskuksat vaikuttavat toimivan hyvin erilaisin tavoin profiloituen omanlaisiksi yksiköikseen. Siinä missä hiljattain perustettu Sussex Humanities Lab painottuu kulttuurikritiikkiin, Amsterdamissa toimiva The Huygens Institute for the History of the Netherlands on sateenvarjomainen historian-tutkimukseen painottuva keskus, kun taas UCL Centre for Digital Humanities on pikemminkin monialainen verkosto, joka tekee ja tukee digitaalista humanistista tutkimusta. Huomionarvoista on, että mikään näistä keskuksista ei ole sijoitettu fyysisesti samojen seinien sisään.

Myös Helsingin yliopiston on suunnitteilla digitaalisten ihmistieteiden keskus (HELDIG-hanke) ja sellaisen perustaminen olisikin ehdottoman tärkeää alan kehittämiseksi Suomessa. Oppi maailmalta viittaa siihen, että profiilin löytäminen ja rohkea oman linjan vetäminen luovat toimintaedellytykset tällaiselle keskukselle, mikä edesauttaa Suomen nousua tutkimuksessa ja myös toimintakulttuurin uudistamista paikallisesti.

Iltapäivän työpajatyöskentely oli praktikumpäivän varsinainen helmi ja tuntui täyttävän sekä osallistujien ja järjestäjien odotukset erinomaisesti. Tapahtumassa emme erotelleet humanistisia ja yhteiskuntatieteellisiä aloja omikseen, vaan osallistujat saattoivat valita kuudesta temaattisesta työpajasta mieluisensa. Työpajat käsitelivät menetelmien perusteita käytännönläheisesti, tutkimuksen avoimuutta, saatavilla olevia digitoituja aineistoja (sanoma- ja aikakauslehdet sekä Suomi24-aineisto), eettisiä ja tekijänoikeusasioita sekä metadataa tutkimuksessa. Työpajoja oli vetämässä sekä kotimaisia että ulkomaisia alan asiantuntijoita.

Tällainen hackaton (koodauspäivä) -tyylinen tapahtuma osoitti hyödyllisyytensä ja sellaisia kannattaa jatkossakin järjestää humanistisilla ja yhteiskuntatieteellisillä aloilla.

Tulevaisuuden näkymiä

Digitaaliset ihmistieteet ovat tulleet ryminällä akateemiselle kartalle kuluneen vuoden aikana. Suomen Akatemia lanseerasi teemasta tutkimusohjelman pikavauhdilla huhtikuun 2015 hakuun ja Helsingin yliopisto otti teeman yhdeksi painopisteeksi 375-vuotisjuhlallisuuksiinsa. Helsingin yliopistossa on lisäksi suunnitteilla digitaalisten ihmistieteiden maisteri-sivuainekokonaisuus sekä jatko-opiskelijaseminaari, jotka käynnistyvät syksyllä 2015. Muissa yliopistoissa, kuten Turun suunnalla, alan tutkimusta ja opetusta on myös kehitetty viime aikoina.

Nähdäksemme alan tulevaisuus Suomessa riippuu siitä, missä määrin muilla tieteenaloilla toimiviksi havaitut käytänteet voivat saada sijaa tällaisessa uudenaikaisessa humanistis-yhteiskuntatieteellisessä tutkimuksessa. Näitä käytänte-

tä ovat koko tutkimusprosessin avoimuus sekä hajautettu yhteistyö monitieteisessä tutkimuskehityksessä. Vastaavat toimintamallit vaikuttivat ratkaisevasti läpimurtoihin ihmisen perimän kartoituksessa viime vuosikymmenellä; niihin panostaminen voi keskeisesti kiihdyttää monitieteistä tutkimusyhteistyötä, kun digitaalisuuden mahdollisuuksia aletaan täysmittaisemmin luodata ihmistieteissä.

Tällaisista avoimen tieteen käytänteiden muutoksista on puhuttu jo pitkään humanistis-yhteiskuntatieteissä, mutta niiden hitaaseen toteutumiseen vaikuttavat edelleen kammo-laskennallisia menetelmiä kohtaan sekä eri tahojen ylenpalttinen oman reviirin suojeleminen, jonka syyt esimerkiksi yliopistojen sisällä ovat pitkälti rakenteelliset. Ratkaisevia läpimurtoja monitieteisessä tutkimuksessa ei kuitenkaan saavuteta ilman kulttuurin muutosta, jossa humanistit ja yhteiskuntatieteilijät lähtevät aidosti yhteistyöhön sekä keskenään että yhdessä laskennallisen tieteen ja tutkimusta tukevien keskustusten kanssa (esim. CSC ja Yhteiskuntatieteellinen tietoaristo).

Tutkimusta tukevat keskuksat ovat myös uudistamistarpeen edessä, sillä nykyisellään tärkeät laskennalliset tukitoiminnot ja humanistis-yhteiskuntatieteellisten alojen toimintakulttuurit ovat aivan liian kaukana toisistaan. Uusiin tieteenvälisiin mahdollisuuksiin tarttuminen vaatiikin kaikilta toimijoilta mukautumiskykyä. Palkintona on kuitenkin jännittävän ja uuden tutkimusalan luominen ja kehittäminen sekä monipuolisempien ja tarkempien vastausten saaminen ihmistieteiden keskeisiin tutkimuskysymyksiin.

Kirjoittajat ovat filosofian tohtoreita ja tutkijoita Helsingin yliopiston tutkijakollegiumissa.