

Ennen ja nyt - Historian tietosanomat

Sähköiset aineistot tarjoavat uusia mahdollisuuksia historiantutkimukselle

2014-01-08 13:01:18 Lauri Keskinen

Marraskuu oli antoisaa aikaa sähköisten arkistojen parissa työskenteleville, aineistoja hyödyntäville opiskelijoille sekä asiasta kiinnostuneille harrastajille kuten sukututkijoille. Kuun puolivälissä järjestettiin Arkistolaitoksen toimesta tutkijayhteisöille suunnattu [Lisätietoja verkossa! -seminaari](#), sekä Historiallisen yhdistyksen Digitaaliarkistoseminaari, joka oli tarkoitettu laajemmalle historiasta ja arkistoista kiinnostuneelle yleisölle.

Kummankin seminaarin varsin rikas anti on tiivistettävissä toteamukseen: digitaalisia aineistoja on jo paljon olemassa, mutta tarve ja kiinnostus kasvavat lähes eksponentiaalista vauhtia. Se, että nuori opiskelija- ja tutkijapolvi tottuu maailmaan, jossa kaikki tieto on saatavilla sähköisessä muodossa, ja usein vielä kotikoneelta käytettynä, tulee jatkossa aiheuttamaan valtavan paineen arkistojen digitoinnille ja saatavuuden varmistamiselle. Siksi ohjeistukset hyviin digitointikäytäntöihin on luotava nyt. Aika on myös sikäli otollinen, että sähköisten aineistojen, joukkoistamisen ja internetin mahdollistaman interaktiivisuuden yhteispelissä on vielä laajoja kartoittamattomia alueita.

Keskustelua aiheista käytiin sattumalta marras-joulukuun vaihteessa myös *Helsingin Sanomien* (6.12.) mielipidepalstalla, jossa muiden muassa tutkijat Tuuli Toivonen ja Joonas Lehtomäki esittelivät avoimen datan tuottamisen niin hyviä kuin huonojakin puolia. Hyviä puolia olivat esimerkiksi yhteistyömahdollisuuksien kasvattaminen, tulosten avoimempi testattavuus ja datan täydentäminen useiden tutkijoiden toimesta. Suurimmat heikkoudet piilivät vaihtelevissa käytännöissä, joilla tiedot kirjataan ylös sekä datan jakelemiseen soveltuvan infrastruktuurin puutteessa. Sama tuli esiin myös seminaareissa, ja ne tarjosivat niin arkistojen tuottajille kuin käyttäjillekin luontevan foorumin pohtia jatkotoimenpiteitä.

Käyttökokemusta parantamassa – sähköisesti

Lisätietoja verkossa! -seminaari alkoi Arkistolaitoksen kehittämispäällikkö Tomi Ahorannan katsauksella laitoksen tunnettuustyöhön ja sen onnistumiseen. Laitoksella oli vuonna 2009 havahduttu huomioon, että sen palveluista, toiminnasta ja aineistoista liikkui internetin keskustelupalstoilla niin vanhentunutta kuin väärääkin tietoa, eikä siellä myöskään aina osattu neuvoa kysyjä. Myöskään sähköisten aineistojen koko kirjoa ei tunnettu. Ahoranta totesikin Arkistolaitoksen tiedottaneen toiminnastaan siinä mielessä väärillä foorumeilla, ettei tieto saavuttanut kyseisiä keskustelupalstoja eli laajempaa yleisöä. Suunnitelmallisen tunnettuustyön, jonka tuli keskittää tiedottaminen laitoksen omille sivuille, lopputuloksena mahdollinen asiakaskunta jaettiin heidän tietojenmaksumistapojensa perusteella lukijoihin, kuuntelijoihin ja katselijoihin. Näille muotoiltiin tiedotuskanavat [Arkistojen Portti](#), [Arkistolaitoksen äänitepankki](#), ja [ViisaanHiiren kanava Youtubessa](#).

Ahorannan alustuksen punaiseksi langaksi osoittautui se, kuinka aktiivista työtä Arkistolaitoksen kaltainen muistiorganisaatio tekee saattaakseen yhteen aineistot ja käyttäjät. Tämä ei aina ole itsestään selvää, sillä sitkeästi elää näkemys, jonka mukaan muistiorganisaatiot tekevät aineistoja vain omaan tai korkeintaan tutkijoiden käyttöön, näin unohtaen muun potentiaalisen asiakaskunnan.

Suomen historian dosentti Anu Lahtinen jatkoi samasta teemasta käsitellessään Arkistolaitoksen toista projektia, jolla pyritään yhtenäistämään sähköisten aineistojen

viittauskäytäntöjä. Siinä missä Ahoranta käsitteli asiakkaan alustavaa tiedonhakuja ja aineistojen ääreen saapumista, Lahtinen keskittyi asiakkaan toimintaan lähteiden parissa sekä myöhempään julkaisuutoimintaan. Lahtinen alusti PID-hankkeen (*Persistent Identifier*) pilottivaiheesta, jonka aikana hän oli kirjoittanut PID-viitetyökalua hyödyntävän [artikkelin](#) *Ennen ja nyt* -verkkolehteen. PID on työkalu, jolla käyttäjä pystyy luomaan pysyvän ja muuttumattoman viitteen sähköiselle arkistoinnille. Tunnetustihan ongelmia on ilmennyt siinä, että sähköisiin aineistoihin viittaavat linkit lakkaavat toimimasta tai ne viittaavat väärään tiedostoon. PID:in avulla näistä ongelmista pitäisi siis päästä eroon, ja PID mahdollistaa myös tiedostojen välittömän jakamisen esimerkiksi Facebookissa. Käytännössä työkalu toimii niin, että se luo aineistolle kuvailutietojen lisäksi pysyvän linkin käyttäjän hyödynnettäväksi vaikka julkaisun lähdeluettelossa. Linkkiä klikkaamalla lukija puolestaan pääsee käsiksi kirjoittajan hyödyntämään aineistoon.

Lahtisen mukaan PID-työkalun tärkein hyöty on sen tavassa tuoda näkyviin aineistojen arkistointi- ja kokoelman merkintälogiikka. Samalla viitetyökalu voi auttaa yhtenäistämään viitteiden merkintätapoja, eikä myöskään tule unohtaa PID:in tärkeintä ominaisuutta eli pysyvyyttä. Lopuksi Lahtinen kertoi vielä PID-kokeilun jatkosta, jossa *Ennen ja nyt* tulee näyttämään merkittävää roolia; lehdessä tullaan ensi vuoden aikana julkaisemaan joukko PID-työkalua hyödyntäviä artikkeleita. Kirjoituskutsu julkistetaan vielä tämän vuoden puolella.

Kansallisarkiston sektorijohtaja István Kecskeméti nappasi Lahtisen heittämän PID-pallon todetakseen, että PID:ien luominen näyttää helpolta samoin kuin karttojen digitointi. Todellisuudessa molemmat ovat todella mutkikkaita prosesseja, jotka vaativat useita rinnakkaisia ja perättäisiä työvaiheita. Alustuksessaan Kecskeméti keskittyi karttojen digitointiin sekä siihen, miten georeferointi parantaa niiden käytettävyyttä. Historiallisille kartoille sijoittuvia paikkoja voi olla varsin hankala löytää, sillä paikannimet voivat muuttua melkoisesti aikojen kuluessa. Yksi keino karttojen käytettävyyden parantamiseksi on sijoittaa ne georeferoinnin avulla Googlen karttapohjalle. Näin ollen karttoja eri ajoilta voidaan katsella samalla tavalla kuin mitä tahansa moderniakin karttaa – ja vielä kätevästi selaimessa!

Tavallinen kartta on yksiulotteinen, mutta georeferoituun historialliseen karttaan on ikään kuin luotu uusi ulottuvuus, aika. Kecskeméti esittelemällä [karttapohjaisella aineistohaualla](#), historiallisia karttoja voi etsiä klikkaamalla jotain yleissilmäyslehden neljästä sadasta ruudusta, joista kustakin aukeaa kartta. Testimielessä Arkistolaitos oli georeferoinut viisi karttaa ja upottanut ne Google Earth -karttapalveluun – tämä voisi siis jatkossa olla yksi tavoista, joilla etsiä karttoja. Liukusäätimellä käyttäjä voi tarkastella maanmuotojen ja rakennuskannan muutoksia siirtymällä historiallisesta kartasta nykyiseen, todennäköisesti vain muutaman vuoden vanhaan karttaan. Näidenkin palveluiden käyttöliittymät ja toimintaperiaatteet paljastivat, kuinka ansiokasta työtä Arkistolaitos tekee palvelujensa ja aineistojensa saavutettavuuden lisäämiseksi.

Suomen Akatemiaa edustaneen professori Matti Sintosen puheenvuoro keskittyi Akatemian rooliin tutkimusinfrastruktuurien kehittämisessä ja rahoituksessa. Hieman irralliselta vaikuttanut puheenvuoro avautui ainakin itselleni paremmin vasta Youtubesta katsottuna ja kaikki muut alustukset nähtyäni. Alustus käsitteli kotimaista FIRI-asiantuntijaryhmää, joka on yksi Akatemian rahoitusta myöntävistä jaostoista. Ryhmän tavoitteina on muun muassa laatia kokonaiskuva tutkimuksen infrastruktuureista, toimia alan kansallisena asiantuntijaorganisaationa, järjestää rahoitushaku infra-hankkeille sekä tehdä ehdotuksia valtion talousarvioon hankkeiden rahoittamiseksi. Varsinaista infrastruktuuri-käsitettä avatessaan Sintonen korosti tutkimuksellisten yhteistyömahdollisuuksien hyötyvän eniten toimivasta infrasta. Samaten infrastruktuurit houkuttelevat ulkomaista huippuosaamista sekä tukevat tutkimusta ja tiedon avoimuutta.

Alustuksen merkittävyys historiantutkijoille ja arkistotoiminnalle liittyi siihen parhaillaan

meneillään olevaan keskusteluun, jossa hahmotellaan erityisesti humanistisille tieteille infrastruktuuria tutkimusaineistojen keräämiseen, pitkäaikaiseen säilytykseen ja jakeluun. Etenkin Jessica Parland-von Essen, joka esitellään kattavasti myöhemmin, on käyttänyt voimakkaita puheenvuoroja kyseisen infrastruktuurin aikaansaamiseksi. Hänen mukaansa suuret määrät aineistoa jäävät pienen piirin tai pahimmassa tapauksessa vain yhden tutkijan hyödynnettäväksi, kun jakamisen mahdollistavaa järjestelmää ei ole. Usein datan avoimuus onkin jäänyt juhlapuheiden ja kannanottojen tasolle.

Kaatuneista ja keskiaikaisista koodekseista

Iltapäivä jatkui Helsingin sotaturmat 1918 -hankkeen johtaja Jarmo Niemisen alustuksella siitä, minkälainen voisi parhaimmillaan olla sotaoloissa menehtyneiden kansallinen tietokanta. Nieminen avasi näkemystään käyttöliittymän, paikkatiedon ja verkostoitumisen käsitteiden avulla. Niemisen mukaan käyttöliittymän tulee olla selkeä, ja siinä on oltava kattavat hakutoiminnot. Lisäksi sen pitää tarjota linkit – esimerkiksi edellä mainittua PID-työkäluä hyödyntäen – suoraan lähteisiin, tiedot on saatava kopioitua omalle koneelle, ja omat muistiinpanot sekä haut tulee saada taltioitua.

Paikkatiedoista puhuessaan Nieminen sivusi Kecskemétin georeferointi-alustusta, sillä sotaoloissa menehtyneiden kansallisen tietokannan tulisi pystyä sijoittamaan tietonsa myös kartoille. Dataa – esimerkiksi sodissa menehtyneiden kaatumis- ja hautauspaikat – visualisoimalla tutkija voisi nopeasti saada yleiskuvan tilanteesta, tai se voisi toimia herätteenä aivan uusille tutkimuskysymyksille.

Verkostoitumisella Nieminen viittasi siihen lisäarvoon, jota harrastajat ja ylipäätään muistiorganisaatioiden ulkopuoliset käyttäjät pystyvät tiedoillaan luomaan tietokantaan. Niemisen mukaan käyttäjällä pitäisi olla kattavat mahdollisuudet lisätä tietoa muista lähteistä, muokata olemassa olevia tietoja ja vaihtaa tietoja muiden käyttäjien kanssa. Nieminen siis ylipäätään kaipasi arkistokokoelmiin avoimuutta sekä interaktiivisuutta.

Päivän toiseksi viimeisestä alustuksesta vastasi SLS:n tutkijakoordinaattori Jesse Keskiäho, joka kertoi meneillään olevasta [Codices Fennici -hankkeesta](#). Hankkeen tämänhetkisenä tavoitteena on tutkimuksellisesti kartoittaa kaikki Suomen alueeseen liittyvät keskiaikaiset ja 1500-luvun käsikirjoitukset. Myöhemmässä vaiheessa toiveena on digitoida aineistot. Kiinnostavaksi hankkeen tekee se, ettei se rajoitu vain tiettyyn muistiorganisaatioon tai edes johonkin nykyiseen kansallisvaltioon, vaan aineistoja kerätään mistä tahansa säilytyspaikasta. Hankkeen kartoittama aineisto koostuu koodeksimuotoisista käsikirjoituksista, joita on kaikkiaan 99. Yli puolet aineistosta on luonteeltaan liturgista, mutta on mukana myös teologiaa, historiaa ja lakeja.

Niemisen tavoin Keskiähokin puhui avoimen datan puolesta. Codices Fennici -hankkeessa datan avoimuus tarkoittaa kuvien ja luettelointitietojen vapaata käyttöä. Kysymysosiossa Keskiäholta tiedusteltiin, miksi aineiston määrä vaikuttaa niin pieneltä. Tämä kieltämättä kummastutti itseänikin. Keskiäho tyytyi toteamaan, että naapurimaiden aineistot ovat kooltaan kymmen- tai satakertaisia. Varsinaista syytä aineiston pienelle määrälle alustaja ei maininnut, mutta sivuissa laskettuna aineistoa kertyy ”kohtuullisesti”.

Lähdekritiikki ja louhinta puhuttivat kummassakin seminaarissa

Päivän viimeinen puhuja, Brages Pressarkivin arkistopäällikkö Jessica Parland-von Essen käsitteli valtaosin samoja teemoja myös viikkoa myöhemmin Digitaaliarkistoseminaarissa, joten tässä käsitellään kahden alustuksen yhteisiä sisältöjä. Parland-von Essen muistutti alustuksessaan lähdekritiikin haasteista

sähköisten aineistojen osalta: Vaikka digitaalinen lähde koostuu ykkösistä ja nolista, ja näin ollen sen pitäisi pysyä siirroista ja kopiointikerroista huolimatta muuttumattomana, ei näin kuitenkaan todellisuudessa ole. Uusintadigitoinnit, formaatin vaihdokset ja digitiedostojen muokkaaminen paremmin ajan vaatimuksia vastaavaksi ovat kaikki askelia, joissa sisältö mahdollisesti muuttuu, oli kyseessä sitten kartta, kirkonkirja tai äänitetiedosto.

Parland-von Essen liputti myös joukkoistamisen, visualisoinnin ja aineiston louhimisen hyötyjä. Hyvä esimerkki joukkoistamisesta on [The Great War Archive](#), jota suuri yleisö on mukana rakentamassa luovuttamalla ensimmäiseen maailmansotaan liittyviä aineistoja ja muistitietoa. Visualisointi, joka saattaa vanhakantaisimmissa historiantutkijoissa herättää sekä ihmetystä että kysymyksen sen hyödyllisyydestä, viittaa aineistojen muuntamiseen visuaaliseen muotoon lisätiedon ja yleiskuvan saavuttamisen toivossa. Parland-von Essen näytti esimerkkinä Euroopan karttaa, johon oli merkitty eräiden historiallisten kirjeiden kulkureittejä tietyinä aikoina. Suurilla aineistomäärillä tätä voisi esimerkiksi hyödyntää valtasuhteiden tarkastelussa tai aatteiden leviämisen tutkimuksessa.

Louhiminen puolestaan viittaa suurten digitaalisten aineistojen tarkasteluun ja muokkaamiseen tiettyjen parametrien mukaisesti. Esimerkiksi Google Books -palvelu voisi jatkossa mahdollistaa tiettyjen sanojen etsinnän sekä tulosten jaottelun sanan esiintymistiheyden, ajan ja asiayhteyden mukaan. Lopuksi Parland-von Essen esitti, että Suomeen olisi saatava mitä pikimmiten infrastruktuuri, joka mahdollistaisi historian tutkijoiden ja harrastajien tuottaman aineiston tallentamisen ja jakelun avoimen datan periaatteita noudattaen.

Arkistot digitoivat, käyttäjät haluaisivat auttaa

Digitaaliarkistoseminaarissa toisessa alustuksessa Kansalliskirjastoa edustanut tietojärjestelmäasiantuntija Jyrki Ilva esitteli, minkälaista työtä Kansalliskirjasto on tehnyt digitaalisten aineistojen tuottamiseksi. Kansalliskirjaston toimintaa sähköisten aineistojen saralla ohjaavat sen tärkeimmät tavoitteet: kattavien kokoelmien ja digisisältöjen kriittisen massan saavuttaminen sekä autenttisten ja luotettavien lähteiden luominen. Tunnetuin Kansalliskirjaston kokoelmista lienee Historiallinen sanomalehtikirjasto, jossa on noin 2 miljoonaa lehtisivua. Vähemmän tunnettu mutta laaja kokoelma puolestaan koostuu [Turun Akatemian kaikista väitöskirjoista](#), joista nyt on digitoitu jo lähes puolet – loputkin tullaan digitoimaan seuraavan parin vuoden aikana. Ainakin itselleni aivan uutena tietona tuli se, että amerikkalaiselle [The Wayback Machine](#) -sivustolle on myös suomalainen vastineensa, Kansalliskirjaston ylläpitämä [Suomalainen verkkoarkisto](#). Esikuvansa tavoin se on hakupalvelu, jolla voi tarkastella ”historiallisia” verkkosivuja eli sivustojen aiempia versioita. Pikainen testi päättyi pettymykseen: hakusanalla ”YLE” löytyi yli tuhat tulosta, mutta yhtäkään ei päässyt katsomaan kotikoneelta, ainoastaan vapaakappalekirjastoista. Juuri tämänkaltaisista vastoinkäymisistä tavallinen käyttäjä saattaa vetää kielteiset johtopäätökset arkistoaineistojen todellisesta saatavuudesta.

Alkusyöksystä väitellyt historiantutkija, Kansallisarkiston kehittämispäällikkö Outi Hupaniitti ei tällä kertaa edustanut työnantajaansa, vaan kertoi SLS:n julkaiseman [Tutkijoiden ääni ja sähköiset aineistot -tutkimuksen](#) tuloksista. Tutkimuksessa selvitettiin muistiorganisaatioiden ja yksittäisten käyttäjien näkemyksiä sähköisten aineistojen käytettävyydestä sekä organisaatioiden ja käyttäjien yhteistyöstä. Kävi ilmi, että suurin käyttäjien joukko ovat akateemiset käyttäjät, ja nämä eivät juuri eronneet käytettyjen aineistojen perusteella ei-akateemisista käyttäjistä. Kahden ryhmän välillä ei esiintynyt eroja käyttäjäkokemuksessaan; yhdessä tuumin käyttäjät moittivat esimerkiksi ohjelmien heikkoa käytettävyyttä ja huonoja käyttöliittymiä.

Hupaniitti toi Parland-von Essenin tavoin esille joukkoistamisen edut ja käyttäjien

halun osallistua aineistonkeruuseen. Valitettavasti muistiorganisaatiot olivat nähneet asiassa pikemminkin huonoja puolia. Ne pelkäsivät muiden kuin arkistoalan ammattilaisten luomien aineistojen olevan laadultaan heikkoja tai vähintäänkin epätasaisia. Pelkoa herätti myös se ylimääräinen työmäärä, joka syntyisi tällaisten aineistojen jatkokäsittelystä.

Tämänkaltaiset pelokkaat asenteet kertovat paljon nykytilanteesta muistiorganisaatioiden ja käyttäjien rajapinnalla: sähköiset aineistot ovat rikkaus ja kätevästi käytettävissä, mutta toisinaan ne voivat olla liiankin helposti muokattavissa. Suurella yleisöllä ei nimittäin ole käytettävissään niin laitteistoa kuin tietotaitoakaan aineistojen luomiseksi tasolle, joka vastaisi virallisten muistiorganisaatioiden vaatimuksia. Tähän ongelmaan viitattiin myös alussa mainitussa Helsingin Sanomien mielipidekirjoituksessa ”Tutkimusaineistojen avaaminen on taitolaji”.

Viimeisenä puhujana Kansalliskirjaston projektipäällikkö Jussi-Pekka Hakkarainen esitteli työtään suomalais-ugrilaisten aineistojen [Fenno-Ugrica -kokoelman](#) parissa. Kyseisen projektin aineistolle, joka kattaa noin 17 000 sivua inkeröiden- ja vepsänkielisiä julkaisuja, yli 150 Neuvostoliitossa julkaistua oppi- ja sanakirjaa sekä 25 000 digitoitua sivua marilaisia ja mordvalaisia sanomalehtiä 1920- ja 1930-luvuilta, avattiin kesäkuussa verkkopalvelu. Kyseessä on ensimmäinen kerta, kun entisen Neuvostoliiton alueella julkaistua aineistoa on saatettu vapaaseen kansalaiskäyttöön Kansalliskirjaston tietojärjestelmissä. Hakkarainen kertoi lisäksi Kansalliskirjaston kehittämästä OCR-editorista (*Optical Character Recognition*), jolla koneellisesti tunnistettua tekstiä voidaan muokata kielentutkimusta varten. Lyhyen näytteen perusteella editori vaikutti varsin näppärältä: Hakkarainen näytti miten kuka tahansa käyttäjä voi käden käänteessä korjata OCR:n väärin tai puutteellisesti tulkitseman sanan.

Puhujia riitti – missä kuulijat?

Kumpikin seminaari oli järjestetty ottamaan kantaa mitä ajankohtaisimpiin haasteisiin: avoimen datan määrän kasvattaminen; aineistojen käytettävyyden parantaminen; joukkoistamisen hyödyntäminen aineistonmuodostuksessa; sähköisten aineistojen uudet tulkintatavat historiantutkimuksessa ja niin edelleen. Olisi luullut seminaarien houkutelleen useampiakin alan opiskelijoita tai vastaavasti nuoria tutkijoita – markkinoitiinhan etenkin jälkimmäistä seminaaria ”opinnäytetöiden tekijöille”. Edellä mainitut ryhmät, jotka ovat tottuneet sähköisten aineistojen olemassaoloon ja suureen määräänkin, loistivat kuitenkin poissaolollaan. Sähköisiin aineistoihin tottuneet mitä ilmeisimmin tietävät arkistoaineistojen sijaitsevan myös tulevaisuudessa tietoverkoissa, eivätkä seminaarit näin ollen ylittäneet uutiskynnystä tuolle sukupolvelle. Harmi, sillä itsekin nuoriin tutkijoihin lukeutuvana opin paljon uutta siitä, mitä aineistoilla nykyään voidaan tehdä, ja ennen kaikkea siitä, mitä muistiorganisaatiot ovat tulevaisuutta varten suunnitelleet.

Lisätietoja:

Lisätietoja verkossa! -seminaarin videoidut alustukset:

<http://www.youtube.com/watch?v=NXEHCwfmIZs>

Arkistojen portti: <http://wiki.narc.fi/portti/index.php/Etusivu>

Arkistolaitoksen äänitepankki: <http://extranet.narc.fi/albumi/albumi.php?n=3&a=0>

ViisaanHiiren kanava Youtubessa: <http://www.youtube.com/user/ViisasHiiri>

Karttapohjainen aineistohaku: http://digi.narc.fi/karttapohjainen_aineisto/

FIRI:

<http://www.aka.fi/fi/ATutkijalle/Rahoitusmahdollisuudet/Aakkosjarjestyksessa/Tutkimusinfrastruktuurien-FIRI-rahoitus/>

Codices Fennici -hanke: http://www.finlit.fi/tutkimus/codices_fennici.htm

The Great War Archive: <http://www.oucs.ox.ac.uk/ww1lit/gwa/>

Tutkijoiden ääni ja sähköiset aineistot -tutkimus:

http://www.sls.fi/media/pdf/Forskarnas_rost_slutrapport_28112011_FI.pdf

Turun Akatemian väitöskirjat: <https://www.doria.fi/handle/10024/50699>

Suomalainen verkkoarkisto: <http://webarchive.nationallibrary.fi/>

Fenno-Ugrica -verkkopalvelu: <http://fennougrica.kansalliskirjasto.fi/>

Kirjoittaja on filosofian tohtori, joka työskentelee kulttuurihistorian tutkijana Turun yliopistossa ja tiedekoordinaattorina Ennen ja nyt -lehdessä.