

## **Kansalliskirjasto ja digitaaliset lähdeaineistot**

**Kansalliskirjasto on rakentanut digitaalisiin aineistoihin liittyviä palveluita jo puolentoista vuosikymmenen ajan. Kirjaston kokoelmat sisältävät sekä digitaalisena syntyneitä että painetusta formaatista digitoituja aineistoja. Osa digitaalisista kokoelmista on vapaasti käytettävissä verkon kautta, osa rajoitetusti vapaakappalekirjastojen tiloissa.**

Suomalaisen historian tutkimuksen piirissä digitaalisiin aineistoihin liittyvistä kysymyksistä on toistaiseksi keskusteltu melko vähän. Vaikka suomalaisten muistiorganisaatiot eli kirjastot, arkistot ja museot ovat rajallisten resurssien vuoksi voineet rakentaa digitaalisia palveluita toivottua hitaammin, työ on kuitenkin ollut käynnissä jo pitkään. Etenkin digitaalisen aikakauden lähdeaineistojen tallentamisella alkaa olla jo kiire.

Parin edellisen vuosikymmenen kuluessa Internetin käyttö on arkipäiväistynyt nopeasti, ja yhä suurempi osa tiedonvälityksestä ja julkisesta keskustelusta tapahtuu verkossa. Tämä muutos näkyy monien muiden elämänalueiden lisäksi myös suomalaisessa politiikassa: vaalikoneiden myötä ehdokkaiden etsiminen on suurelta osin siirtynyt verkkoon, ja erilaisista verkon keskustelufoorumeista on tullut merkittäviä poliittisen mielipiteenmuodostuksen ja poliittisten toimijoiden rekrytoinnin kanavia.

Tämä merkitsee uutta haastetta niin historian tutkijoille kuin tutkimuksen lähdeaineistoja tallentaville muistiorganisaatioillekin. Onko esimerkiksi perussuomalaisten viimekeväisen vaalivoiton taustaa ja puolueen sisälle syntyneen monikulttuurisuuskriittisen liikkeen historiaa mahdollista rekonstruoida tyydyttävästi ilman Hommaforumin kaltaisten keskustelupalstojen ja liikkeen aktiivien blogien sisältämiä tekstejä? Ovatko nämä aineistot tulevaisuudessa tutkijoiden käytettävissä?

Muutama vuosi sitten voimaan tullut *Laki kulttuuriaineistojen tallettamisesta ja säilyttämisestä* (28.12.2007/1433) määrittä kotimaisen verkkoaineiston tallentamisen Kansalliskirjaston tehtäväksi. Kansalliskirjaston näkökulmasta verkkoaineiston keruu on jatkoa perinteikkäälle vapaakappaletoiminnalle, joka oli aiemmin keskittynyt painettuihin julkaisuihin ja audiovisuaalisiin tallenteisiin.

### **Digitaalisen aikakauden lähdemateriaalit**

Verkkoaineistoa kerätään talteen kahdella eri tavalla, toisaalta luovutuksina, toisaalta automaattisen verkkoharavoinnin avulla. Verkkoaineiston keruulle on tyypillistä, että kaikkea verkossa julkaistua aineistoa ei ole mahdollista saada talteen, vaan keräystulos on väistämättä jonkinlainen otos kaikesta verkon kautta tarjolla olleesta materiaalista. Suomalaisen ja suomalaiselle yleisölle suunnatun verkkoaineiston keruussa lisähaasteena on myös verkon kansainvälinen luonne – suomalaiset ja kansainväliset aineistot liittyvät usein saumattomasti toisiinsa, ja suomalaiset aineistot saattavat sijaita jossain kansainvälisessä verkkopalvelussa. Lisäksi monet suurellekin yleisölle suunnatut verkkopalvelut ovat käyttöoikeudeltaan rajattuja, eli niiden saaminen talteen edellyttää yhteistyötä palvelun tuottajan kanssa.

Käytännössä keruukohteita on pakko priorisoida jollain perusteella. Aiemman vapaakappaletoiminnan peruna verkkoaineiston keruussa on painotettu mm. verkossa julkaistua

musiikkia ja painettujen julkaisujen kaltaisia verkkojulkaisuja, esim. elektronisia kirjoja, joilla on ISBN-numero. Toisaalta etenkin tutkijoiden suunnalta on esitetty argumentteja sen puolesta, että keräystoiminnassa pitäisi panostaa enemmän aineistoihin, jotka ovat nimenomaan verkolle tyypillisiä – esim. keskustelupalstoihin, blogeihin ja sosiaalisiin verkkopalveluihin. Valinnan tekeminen edellyttääkin valistunutta näkemystä siitä, millaisia aineistoja tutkijat ja muut verkkoarkiston käyttäjät tulevaisuudessa eniten tarvitsevat. Tältä kannalta olisi hyödyllistä, jos aiheesta käytäisiin nykyistä vilkkaampaa julkista keskustelua tutkijoiden ja muiden keskeisten viiteryhmien kesken.

Tekijänoikeuslainsäädäntö rajaa valitettavasti verkkoarkiston käyttöä radikaalisti. Arkisto on käytettävissä Kansalliskirjaston lisäksi muissa vapaakappalekirjastoissa ja Kansallisessa audiovisuaalisessa arkistossa erillisillä vapaakappaletyöasemilla, joilla aineiston kopioiminen digitaalisessa muodossa on estetty. Arkiston indeksi on kuitenkin käytettävissä verkon kautta (<http://verkkoarkisto.kansalliskirjasto.fi/>).

Digitaalisena syntyneitä aineistoja julkaistaan myös muussa kuin digitaalisessa formaatissa. Useimmat painetussa formaatissa levitettävät aineistot ovat nekin olleet ennen painamista olemassa digitaalisessa muodossa. Kansalliskirjasto on selvittämässä mahdollisuuksia saada niitä talteen sähköisessä muodossa, sillä aineistojen digitointi uudelleen painetusta versiosta on tarpeetonta, jos digitaaliset versiot saadaan talteen ja niiden käyttöoikeudesta sovitaan. Sanomalehtien osalta tätä mahdollisuutta selvitetään vuonna 2011 alkaneessa Comellus-hankkeessa, jossa tehdään yhteistyötä useiden kotimaisten lehtitalojen kanssa ja jossa testataan myös aineistojen tulostamista mikrofilmille pitkäaikaissäilytystä varten.

### **Kulttuuriperintöaineistojen digitointia Kansalliskirjastossa**

Digitaalisena syntyneiden aineistojen tallentamisen lisäksi myös vanhojen painettujen julkaisujen digitointi palvelee tutkimusta monin tavoin. Vaikka tutkijoiden voi olettaa tulevaisuudessakin käyttävän myös painettuja aineistoja, digitoinnilla on kuitenkin tiedepoliittista merkitystä: On todennäköistä, että helposti löydettäviä ja helposti saatavilla aineistoja käytetään enemmän. Niinpä kotimaisten aineistojen digitointi ja saaminen mahdollisimman vapaaseen käyttöön edistää myös niihin kohdistuvaa tutkimusta. Lisäksi suurten yhtenäisten aineistomassojen digitointi avaa mahdollisuuksia uusien automaattiseen tietojenkäsittelyyn ja tiedon louhintaan perustuvien tutkimusmenetelmien hyödyntämiseen.

Nurinkurista kyllä, muutaman vuoden takainen talousromahdus ja sitä seurannut elvytyspolitiikka ovat olleet digitoinnin kannalta siunauksellisia. Elvytyshankkeiden ansiosta digitointiin saatiin oikeastaan ensimmäistä kertaa merkittäviä resursseja sekä kirjasto-, arkisto- että museosektoreille. Vaikka vuosina 2009 ja 2010–2011 toteutetut hankkeet olivat kestoiltaan lyhyitä, niiden puitteissa digitointiin suuria aineistomääriä, joista osa on tulossa julkisesti saataville vasta myöhemmin.

Kansalliskirjaston toiminnan päämääriin on kuulunut digitointiin ja digitaalisiin aineistoihin liittyvien kansallisten infrastruktuurien rakentaminen. Vuonna 2010 valmistuneen Kansalliskirjaston digitointipolitiikan mukaan tavoitteena on tuottaa "autenttisia ja luotettavia" digitaalisia aineistoja, jotka soveltuvat tutkijoiden lähdeaineistoksi. Painettujen aineistojen korvaaminen digitaalisilla käyttökopioilla suojelee hauraita alkuperäiskappaleita kulutukselta ja tarjoaa aineistoille uusia käyttömahdollisuuksia.

Käytännössä toiminnan päälinjana on ollut massadigitointi. Kansalliskirjaston Mikkelissä toimivaan digitointi- ja konservointikeskukseen on rakennettu erityisesti massadigitointiin soveltuva tekninen infrastruktuuri, joka mahdollistaa digitointiprosessin ja digitoitujen aineistojen koko elinkaaren

hallinnan. Automaattiskannereista huolimatta etenkin digitoidun aineiston jälkikäsittely riittävällä tarkkuudella vaatii edelleen paljon ihmistyötä, ja niinpä digitointilinjasto on toistaiseksi voinut toimia täydellä kapasiteetilla ainoastaan elvytyshankkeiden aikana.

Historiallinen sanomalehtikirjasto (<http://digi.kansalliskirjasto.fi/>) on epäilemättä tunnetuin Kansalliskirjaston tuottama digitaalinen aineistokokonaisuus. Toistakymmentä vuotta sitten alkanut työ on edennyt vähitellen, ja tällä hetkellä palvelu kattaa suomalaiset sanomalehdet 1700-luvun lopulta aina vuoteen 1910 saakka. Vuosien 2009 ja 2010–2011 elvytyshankkeissa keskityttiin sanomalehtien sijasta aikakauslehtiaineistoon, josta suurin osa on nyt digitoitu 1800-luvulta aina vuoteen 1944 asti.

Lisäksi Kansalliskirjasto on digitoinut joitakin pienempiä mutta yhtä lailla merkittäviä aineistokokonaisuuksia. Ilkka ja Ulla Paateron rahaston varoilla digitoitiin vuosina 2008–2009 yli 1 700 Turun akatemian väitöskirjaa (noin 40 prosenttia kaikista ennen vuotta 1828 julkaistuista väitöskirjoista), jotka ovat vapaasti saatavilla Kansalliskirjaston Doria-palvelussa (<http://www.doria.fi>). Lähitulevaisuudessa on tarkoitus digitoida loputkin väitöskirjat. Dos. Tuomas Heikkilän johtaman "Suomen keskiajan kirjallinen kulttuuri"-hankkeen yhteydessä on puolestaan digitoitu Kansalliskirjaston kokoelmiin sisältyvät yli 10 000 pergamenttifragmenttia, jotka nekin ovat tulossa verkkoon lähiaikoina.

Kansalliskirjasto on jo aiemmin digitoinut laajan kokoelman vanhoja teollisuushinnastoja. Lisäksi Kansalliskirjasto on digitoinut runsaasti monografioita, joihin sisältyy sekä kaunokirjallisuutta että matkakirjoja ja muutakin tietokirjallisuutta. Osa aineistosta on digitoitu yritysten, yhteisöjen ja yksittäisten kansalaisten lahjoitusten turvin (Pelasta kirja -hanke). Kansalliskirjaston musiikkikirjaston yhteydessä toimivan Kansallisen äänitearkiston kokoelmista on digitoitu yli 10 000 äänitettä, joista tosin tekijänoikeusyistä johtuen vain pieni osa on voitu julkaista vapaasti verkossa (Dorian Raita-kokoelma).

Kansalliskirjaston digitointipolitiikassa tavoitteeksi on määritelty aineistojen mahdollisimman laaja käyttö ja uudelleenkäyttö. Tekijänoikeuslainsäädäntö kuitenkin rajaa mahdollisuuksia monilta osin. Esim. sanoma- ja aikakauslehdet ovat tällä hetkellä vapaassa verkkokäytössä suhteellisen turvallisena pidettyyn rajavuoteen 1910 asti. Tätä uudempia lehtiä voi käyttää Suomalaisen verkkoarkiston tapaan ainoastaan vapaakappaletyöasemilla, joilta voi lukea aineistoa, tehdä kokotekstihakuja ja tulostaa paperikopioita.

Nykyinen tekijänoikeuslainsäädäntö ei anna tutkimuskäytölle erivapauksia, eli Kansalliskirjaston kokoelmista digitoidun tekijänoikeuden alaisen materiaalin aineistojen tuominen saataville vapaakappalekirjastojen ulkopuolella edellyttää sopimusta oikeudenhaltijoiden kanssa. Tekijänoikeuslakiin sisältyvät sopimuslissenssisäädökset vähentävät kuitenkin oleellisesti laajojen aineistokokonaisuuksien lissensiointiin liittyviä riskejä. Kansalliskirjasto selvittää yhdessä Kopioston kanssa mahdollisuuksia tuoda myös tekijänoikeuden alaisia aineistoja saataville esim. korkeakoulujen käyttöön lissensioitavina maksullisina aineistokokonaisuuksina, joiden tuottamat tulot käytettäisiin oikeudenhaltijoille maksettaviin korvauksiin.

### **Tutkijoiden ja muistiorganisaatioiden yhteistyö**

Kansalliskirjaston digitoidut aineistot ja verkkoaineistojen keruu on tietysti tarkoitettu suurelta osin palvelemaan tutkijoiden tarpeita. Käytännössä tutkijoiden rooli palveluiden suunnittelussa on toistaiseksi ollut vähäinen, ja ne ovat syntyneet suurimmaksi osaksi kirjaston ja digitointityön rahoittajien määrittelemistä lähtökohdista. Tämä on ajoittain herättänyt kritiikkiä tutkijoiden keskuudessa. Suurten kansallisten infrastruktuurien rakentaminen ei välttämättä palvele suoraan

yksittäisen tutkijan ajankohtaisia tiedontarpeita, vaan niiden arvo tulee usein näkyviin vasta vähitellen sitä mukaa kun rakennustyö etenee.

Muistiorganisaatioiden ja tutkijoiden intressien yhteensovittaminen on kuitenkin tärkeää. Ilman tutkijayhteisön tukea muistiorganisaatioilla voi olla vaikeuksia löytää rahoitusta digitoitihankkeilleen, ja lisäksi vaarana on se, että toiminnan prioriteetteja mietitään liiaksi organisaation omien prosessien kannalta ja unohdetaan loppukäyttäjien tarpeet. Etenkin uusiin digitaalisena syntyneisiin aineistoihin liittyy paljon sellaisia uusia kysymyksiä, joista olisi hyödyllistä käydä julkista keskustelua – alkaen siitä, millaisia aineistoja tulevaisuuden tutkijasukupolvet ylipäänsä tarvitsevat.

Käytännössä toimivien yhteistyön mekanismien luominen on osoittautunut haastavaksi. Yhteistyö ei synny itsestään, vaan se edellyttää molemmilta osapuolilta aikaa, panostusta ja yhteisten päämäärien löytämistä. Tutkijoiden kannalta yhteistyö tarjoaisi kuitenkin mahdollisuuksia muistiorganisaatioiden aineistojen hyödyntämiseen ja monissa tapauksissa myös uusien digitaalisiin korpuksiin ja tiedon louhintaan liittyvien tutkimusmetodien hyödyntämiseen. Muistiorganisaatioiden kannalta tutkijayhteistyö taas auttaisi aineistojen saamisessa laajempaan tutkimuskäyttöön ja toisi niiden ulottuville arvokasta sisältöasiantuntemusta.

### **Digitaalisen maailman mahdollisuudet**

Viime aikoina muistiorganisaatioiden keskinäinen yhteistyö ja yhteisten digitaalisten palveluiden luominen on saanut vauhtia opetus- ja kulttuuriministeriön rahoittaman laajan Kansallinen digitaalinen kirjasto -hankkeen (<http://www.kdk.fi>) kautta. Hankkeessa rakennetaan muistiorganisaatioiden aineistoille yhteistä asiakasliittymää, ja sen puitteissa on myös selvitetty mahdollisuuksia yhteisen kansallisen digitaalisten aineistojen pitkäaikaissäilytysjärjestelmän rakentamiseksi.

Kansalliskirjasto on myös käynnistänyt syksyllä 2011 selvitystyön mahdollisuuksista perustaa Helsingin yliopistoon erityinen digitointiin liittyviin kysymyksiin keskittyvä professuuri. Vaikka selvitystyön tulokset valmistuvat vasta myöhemmin, on selvää, että digitoitujen ja digitaalisena syntyneiden aineistojen roolista ja niihin liittyvistä metodisista kysymyksistä pitäisi käydä Suomessakin nykyistä aktiivisempaa keskustelua myös historiantutkijoiden kesken.

Kansainvälinen keskustelu on ollut viime aikoina vilkasta, ja kuvaavaa kyllä esim. American Historical Associationin tammikuun 2012 vuosikokouksen ohjelmassa digitaaliseen historialle (teema *The Future Is Here: Digital Methods in Research and Teaching in History*) on omistettu yli 20 sessiota. Keskustelua on vauhdittanut mm. Googlen digitoimaan kirja-aineistoon perustuva Google Ngram Viewer -palvelu (<http://books.google.com/ngrams/>), joka tarjoaa rajoituksistaan huolimatta vaikuttavan näytön suurten historiallisten tekstimassojen pohjalta tehtävien analyysien mahdollisuuksista.

Innokkaimmissa visioissa on jo ennakoitu humanistisen tutkimuksen kehittyvän tätä kautta luonnontieteiden kaltaiseksi eksaktiksi tieteeksi ("Culturomics", ks. <http://www.culturomics.org>). Vaikkei tähän näkemykseen uskoisikaan, on ilmeistä, että tulevilla historioitsijasukupolvilla on nykyisten välineiden ja tutkimusmetodien lisäksi käytettävissään myös monenlaisia uusia mahdollisuuksia.