

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Kokko, Marja

**Title:** Kirjastojen arkisto-aarteista kaiken kansan luettavaksi

**Year:** 2014

**Version:**

**Please cite the original version:**

Kokko, M. (2014). Kirjastojen arkisto-aarteista kaiken kansan luettavaksi. Tietue :  
Jyväskylän yliopiston kirjaston lehti, 2014(2).  
<https://jyx.jyu.fi/dspace/handle/123456789/43717>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Tietue 2 / 2014

Tietue on Jyväskylän yliopiston kirjaston lehti, joka ilmestyy verkossa neljä kertaa vuodessa.

ISSN-L 1798-4890

ISSN 1798-4890

## Kirjastojen arkisto-aarteista kaiken kansan luettavaksi

21.5.2014 Tekijä [Jyväskylän yliopiston kirjasto](#)

[Avoimen tiede ja tutkimus –hankkeen](#) onnistuminen edellyttää, että tiedetään, miten ja missä tutkimuksen dataa säilytetään ja julkaistaan, kuka sen omistaa sekä miten datasta ja sen olemassaolosta kerrotaan muille. Humanistisen tutkimuksen kohdalla sama kysymyksenasettelu koskee digitoituja aineistoja, joiden hyödyntäminen aikaisemmin on voinut olla käytännössä mahdotonta esimerkiksi aineistojen haurauden ja muuten heikon kunnan vuoksi.

Humanistisen tutkimuksen hyödyntämiä aineistoja on digitoitu jo vuosikymmenten ajan, mutta keskustelu tutkimuksen avoimuutta korostavasta tiedepolitiikasta on nostanut aineistojen saavutettavuuden (tutkimustyön tehostaminen) rinnalle lisäksi keskustelun tutkimustulosten arvioitavuudesta. Kun julkaisuun voidaan liittää linkki alkuperäislähteeseen, voivat lukijat arvioida tutkimuksen tasokkuutta myös tutkimusaineistojen tulkinnan näkökulmasta.

Arkistolaitoksen [Digitaaliarkistosta](#) saa vaivattomasti kattavan kuvan siitä aineistosta, joka on jo digitoitu ja mitä tullaan digitoimaan.

Humanistisessa tutkimuksessa tieteen avoimuuden haasteena ovatkin erilaiset toisistaan irrallaan olevat projektit, jotka tuottavat arvokkaita aineistoja tutkimuksen käyttöön, mutta ei ole Digitaaliarkiston tuottaman tiedon tavoin saatavissa koottua tietoa.

## Kirjaston kokoelmat tutkimusaineistona

Kirjasto on yliopistonyhteisön keskeinen tutkimuksen tietovaranto. On hyvä muistaa, että kirjaston kokoelmat toimivat myös tutkimusaineistoina. Esimerkiksi historian- ja kirjallisuudentutkimukselle kirjasto edustaa samanaikaisesti sekä tutkimuskirjallisuuskokoelmaa että tutkimusaineistoa. Kirjallisuudentutkija ja etenkin kirjahistoriantutkija on voinut käyttää digitoituja kirjoja astumatta lainkaan kirjastoon esimerkiksi [Project Gutenberg](#) ja [Project Runeberg](#) –tietokantojen kautta. Mainitut tietokannat ovat kaikille avoimia, joten kuka tahansa pääsee tarkastelemaan tekijänoikeuksista vapautuneita suomalaisia kirjallisuuden klassikkoteoksia Kalevalasta lähtien.

Jyväskylän yliopiston kirjasto tarjoaa yliopistoyhteisön jäsenille käyttöön lukuisia niin kutsuttuja lisensoituja sekä kotimaisia että kansainvälisiä tietokantoja, jotka palvelevat erinomaisesti kirjallisuudentutkimusta tai kirjahistoriantutkimusta.

Vapaakappalekirjastona meillä on tarjota runsaasti Kansalliskirjaston ylläpitämään Doria-julkaisuarkistoon [digitoituja aineistoja](#), joista lisää kirjastonhoitaja Marjut Katajan [artikkelissa](#).



Devil's bible: Kuva: Wikimedia Commons

Juuri äskettäin avattiin [Henrik-tietokanta](#), joka avaa tutkijoiden käyttöön kirjanomistamiseen (vuoteen 1809) liittyvän tietovarannon kaikkiaan yhdeksän suomalaisen kaupungin näkökulmasta. Jo aikaisemmin Kansalliskirjasto avasi [Fragmenta membranea –kokoelman](#), joka sisältää puolestaan keskiaikaisen Ruotsin valtakunnan itäosissa, Turun hiippakunnassa, laadittujen ja käytettyjen kirjojen katkelmia (ks. lisää [Doria](#)).

Vastaavia aineistoja löytyy ympäri maailmaa ja uutuuksien seuraaminen alkaakin olla jo totista työtä. Hiljattain [kerrottiin](#) julkisuudessa, että [Ruotsin kansalliskirjaston](#) LIBRIS-tietokanta on liitetty WorldCat-luetteloon, joka on puolestaan maailmanlaajuinen kirjastojen yhteistietokanta ja joka sisältää runsaasti avoimia aineistoja. LIBRIKsen sisällyttäminen WorldCatiin on huomiotu myös siksi, että nyt tutkija pääsee katalogin kautta tutustumaan yhteen keskiajantutkimuksen merkittävimpään kokoelmaan: [Devil's Bibleen](#) (kokoelma tuotiin sotasaaliina Ruotsiin 1600-luvulla nk. kolmekymmenvuotisen sodan melkskeissä). WorldCatiin liittymisestä hyötyvät myös ruotsalainen tutkimus ja kustantaminen, joiden kansainvälinen näkyvyys lisääntyy.

## Tutkimusaineistojen tuntemus osana kirjasto-osaamista

Kaikki edellä mainitut esimerkit kuvastavat sitä, että kirjasto-osaamiseen sisältyy tänä päivänä hyvin vahvasti myös tutkimusaineistojen tuntemus ja kaiken siihen liittyvän informaation suodattaminen opiskelijoille ja tutkijoille. Uutisia uusista digitoiduista aineistoista voi lukea liki päivittäin, mutta missään ei ole paikkaa, josta kaikki olemassa oleva aineisto olisi vaivattomasti löydettävissä. Linkkilistat eivät auta kirjastoasiantuntijaa eivätkä vie tutkimusta eteenpäin, vaan aineistoja pitää pystyä kuvailemaan jo yksinomaan niiden löydettävyyden vuoksi, mutta myös informaatioähkyn ehkäisemiseksi. Jotta tässä saadaan aikaiseksi jotakin merkityksellistä, kirjastoasiantuntijan pitää seurata tutkijoiden keskuudessa käytävää keskustelua (ks. esim. [Digitaalinen vallankumous uudistaa kirjahistorian tutkimusta](#)).

Historiantutkijoille ja esimerkiksi sukututkimuksen harrastajille on jo tullut tutuksi kirjaston vapaakappalekokoelmaan kuuluva Kansalliskirjaston [Historiallinen sanomalehtikirjasto](#) (1771-1910) sekä yksinomaan kirjastoon sijoitetun arkistotyökoneen kautta luettavat sanomalehdet (ks. lisää [Lehdet](#)). Kansalliskirjasto on digitoinut myös Ruotsissa julkaistuja suomenkielisiä sanomalehtiä (ks. lisää [Ruotsinsuomalaiset lehdet digitoidaan](#)). Sanomalehtien digitointityötä on tehty myös yksityisesti, joista esimerkkinä mainittakoon ruotsinkielisten lehtien [Brages Pressarkiv](#) –tietokanta (ks. lyhyt [suomenkielinen esite](#)).

Kaiken kaikkiaan historiallisten aineistojen digitointi on jo itsessään kiehtova ala. Kuuntele asiasta erittäin mielenkiintoinen [tallenne](#), jossa haastatellaan Kansalliskirjaston Mikkelin Digitointi- ja konservointikeskuksen johtajaa Majlis Bremer-Laamasta.

## **Kaunista katsottavaa, mutta miten tekstistä saa mitään selvää!**

Vuosisatoja vanhat aineistot ovat kaunista katseltava, mutta aineistojen avoin saavutettavuuskaan ei auta, jos oma tietotaito ei yllä vieraskieliseen vanhaan käsinkirjoitettuun tekstiin. Esimerkiksi edellä mainittua ”pirun raamattua” pitää tyytyä vain ihailemaan, jos taidot eivät riitä keskiaikaisen käsinkirjoitetun latinan tulkintaan. Jo 1800-luvun lopun ja 1900-luvun alun suomenkielisen digitoidun aineiston hyödyntäminen voi olla hankalaa, jos ei huomaa suomen kielen kehitystä ja esimerkiksi hahmota fraktuuralla painettua tekstiä.

Professori Eero Sormunen esitteli *Becoming the Digital Humanist 1.0* –seminaarissa 25.5. 2014 mielenkiintoisia näkökulmia [Kansalliskirjaston Historiallisen sanomalehtikirjaston](#) käyttöön tai pikemminkin siihen, mitä pitää ymmärtää, kun käyttää mainittua aineistoa. Historiallisessa sanomalehtikirjastossa on siis kysymyksessä digitoitujen sanomalehtien kokoelmasta/tietokannasta, jossa on käytössä suhteellisen yksinkertaiset hakuominaisuudet.

Alkuperäismateriaali on saatettu digitaaliseen muotoon käyttämällä automaattista tekstintunnistushjelmaa. Voisi ajatella, että aineistojen digitointi on avain kaiken olemassa olevan tiedon saavutettavuuteen, mutta näin ei kuitenkaan ole. Tietokoneohjelmat eivät ole niin kehittyneitä, että ne pystyisivät tunnistamaan kaikkia sanoja. Oma kokemukseni tukee hyvinkin tätä tietoa, sillä käytin mainittua lehtiarkistoa Jyväskylän yliopiston kirjaston 100-vuotishistoriaprojektissa. Jyväskylän kaupungin nimi tuottaa lähtökohtaisesti vaikeuksia ääkköstensä vuoksi, mutta myös erilaiset kirjoitusmuodot tekivät aluksi niin vaivattomalta tuntuvan työn sangen haasteelliseksi.

Kansalliskirjastossa pistettiin vuoden 2011 alussa pystyyn [Digitalkoot](#) (myös [YouTube-video](#) *Digitalkoista*), jonka aikana vapaaehtoiset avustajat pääsivät osallistumaan tekstin korjaamiseen ja tekstin ”puhdistamiseen”.

Tätä merkkittunnistamisen teknistä problematiikkaa avaa esimerkiksi professori Sormunen ohjauksessa syntynyt Ismo Raitasen [pro gradu –tutkielma](#). Tutkielmassa selvitetään mielenkiintoisella ja hyvin käytännönläheisesti tekstintunnistamisen problematiikkaa juuri Historiallisen sanomalehtikirjaston osalta. Arkiston käyttäjän on jo lähtökohtaisesti huomattava, että kieli on muuttunut ja arkiston aineiston hyödyntäminen edellyttää muun muassa sitä, että käyttäjä tunnistaa keskeiset varhaisnykysuomen sanat.

Historiallisen sanomalehtikirjaston markkinoinnissa ja käyttäjäkoulutuksessa olisikin oivallinen tilaisuus yhdistää kielitieteilijöiden opetus varhaisnykysuomen ominaispiirteistä ja kirjastoasiantuntijan panos käyttöliittymän tekniikan haltuunotossa. Nyt suomen kielen kiemuroihin voi tutustua vaikkapa Kotimaisten kielten keskuksen [Kotuksen verkkojulkaisujen](#) avulla.

*Marja Kokko*