

Joonas Korhonen

Taistelu historiallisesta tiedosta – saksalaiset digitointiprojektit Googlen haastajana

Historiantutkimus demokratisoituu. Internetin välityksellä yhä useammat lähteet maailman laidoilta ovat kaikkien ihmisten saatavilla. Vielä muutama vuosi sitten vain ulkomaisista kirjastoista saatavissa ollut lähdekirja saattaa hyvällä onnella löytyä jo nyt internetistä digitaalisesti valokuvattuna faksimileversiona. Internetiin sijoitetut valokuvatut lähteet haastavat esimerkiksi mikrofilmit saatavuudessaan mutta usein myös kuvanlaadussaan. Kun hakukonejätti Google aloitti taannoin mittavan faksimiledigitointiprojektinsa, monet ei-englanninkieliset maat säikähtivät historiallisen tiedon angloamerikkalaistumisesta. Syntyi kansallisille intresseille pohjautuvaa kilpailua historiallisesta tiedosta.

Onko [Google Book Searchille](http://books.google.com/) (GBS) < <http://books.google.com/> > osoitettu kritiikki aiheellista? Jäävätkö esimerkiksi saksalaiset lähteet saatavuutensa puolesta angloamerikkalaisen kulttuuri-imperialismin varjoon? Tässä katsauksessani teen ekskursion internetissä oleviin lähteisiin ja tutkin esimerkin valossa saksankielisten lähteiden nykyistä ja tulevaa edustavuutta internetissä. Rajaan tarkasteluni 1700-luvun lopun saksalaisalueilla tuotettuihin painotuotteisiin, koska ne ovat omien tutkimusintressieni takia minulle kaikkein tutuimpia. 1700-luvun lopun aineistoa on suomalaisissa kirjastoissa verrattain vähän, joten rajaus on sikälikin perusteltu. Pohdin myös internetissä olevan aineiston käyttökelpoisuutta suomalaiselle tutkijalle. Tarkastelen työssäni lähdekriittisistä syistä vain faksimileaineistoa. Esimerkiksi Project Guttenberg, jossa on vanhojen kirjojen tekstejä elektronisessa muodossa, jää tarkasteluni ulkopuolelle.

GBS vrs. DFG?

Hakukone Google julisti vuoden 2004 loppupuolella aloittavansa mahtipontisen digitalisointiprojektin, jonka tarkoituksena on digitoida jopa kymmeniä miljoonia tekijänoikeuksien ulkopuolella olevia vanhoja kirjoja ja tuoda ne internetiin kaikkien ihmisten ulottuville. Koska Google tekee yhteistyötä suurien kirjastojen kanssa Yhdysvalloissa ja Isossa-Britanniassa, on projektiin latautunut suuria odotuksia historioitsijoiden keskuudessa. Pianhan suuret määrät vanhoja tekstejä olisi luettavissa tietokoneen näyttöpäätteeltä, ilman että tarvitsisi jäljittää alkuperäistekstiä jopa toisessa maassa sijaitsevaan kirjastoon.¹

Googlen historialliseen digitointiprojektiin liittyy monien mielestä kuitenkin paljon ongelmia. Erityisesti GBS:n eli Google Book Searchin lähteiden edustavuus ja dokumentointi tuntuu saavan paljon kritiikkiä osakseen.² Vaikka yhdysvaltalainen hakukone digitoi vanhoja kirjoja muun muassa laajoista Harvardin ja Oxfordin yliopistojen kirjastoista, tulee digitaalisesta kokoelmasta selvästi englanninkieliseen materiaaliin painottunut. Koska hakukone pystyy tarjoamaan tutkijoille helppoa materiaalia, monet eurooppalaiset pelkäävät, että tulevaisuudessa Euroopan ja muun maailman historia nähdään yhä enemmän angloamerikkalaisten silmälasien läpi.³

Euroopassa on syntynyt vastarintaliikkeitä Googlen ylivaltaa vastaan. Vaikka asetelmaa on mediassa verrattu Daavidin ja Goljatin taisteluksi, aikoo eurooppalaisten kirjastojen koalitio tosissaan haastaa Googlen orastavan hegemonian historiassa. Ranskalaiset ovat jopa maan ylintä johtoa myöten nähneet tarpeelliseksi tarjota vastapainoa angloamerikkalaiselle kulttuuri-imperialismille. Itse presidentti Jacques Chirac liputti huhtikuussa 2005 näyttävästi eurooppalaisen digitointiprojektin puolesta Pariisissa EU:n kulttuuriministerien tapaamisen yhteydessä.

Myöskään saksalaiset kirjastot eivät ole jääneet makaamaan kuoppiinsa taistelussa historiallisesta tiedosta. Kuvaavaa on, että yhden Saksan suurimman digitaalista lähdemateriaalia tuottavan laitoksen [Göttinger Digitalisierungs-Zentrumin](http://gdz.sub.uni-goettingen.de/de/index.html) eli GDZ:n www-etusivulla < <http://gdz.sub.uni-goettingen.de/de/index.html> > on linkki Spiegel Onlineen 19.4.2006 julkaisemaan kirjojen digitalisointia käsittelevään artikkeliin "Was Google falsch macht" ("Mitä Google tekee väärin"), jossa verrataan Googlen

digitointihanketta Göttingenin yliopiston huolelliseen ja mittavaan digitointityöhön. Artikkeleihin haastateltu Göttingenin digitointikeskuksen johtaja Ralf Stockman kertoo, että Googlen digitointisysteemi aiheuttaa ongelmia digitaalisen valokuvan laadussa. Göttingenissä vanhoja kirjoja digitoivat ihmiset, jotka digitoinnin yhteydessä pystyvät tarkkailemaan kuvanlaatua. Google Book Searchissa digitoinnin hoitavat puolestaan robotit, jotka eivät pysty valvomaan digitointijälkeä.⁴ Kuvanlaatu Googlen digitoimissa kirjoissa onkin monesti erittäin keho ja paikoin jopa lukukelvoton.

Saksalaisten on helppo puolustaa moninaisia ja alati laajenevia digitointiprojektejaan. Jo vuodesta 1997 Saksan valtion tutkimusyhteisö Deutsche Forschungsgemeinschaft on tukenut saksalaisten aineistojen digitointia. Se rahoittaa myös Göttinger Digitalisierungs-Zentrumin toimintaa. GDZ:in sivuilta löytyy [listaus](http://gdz.sub.uni-goettingen.de/de/index.html) < <http://gdz.sub.uni-goettingen.de/de/index.html> >, jossa on yli 70 Deutsche Forschungsgemeinschaftin rahoittamaa saksalaista digitointiprojektia. Saksalaisen digitaalisen aineiston haku kannattaa aloittaa täältä. Myös [NetbibWiki](http://wiki.netbib.de/coma/StartSeite) < <http://wiki.netbib.de/coma/StartSeite> > sisältää suuren määrän linkkejä internetistä löytyvään saksalaiseen aineistoon. Sivusto on wiki-pohjainen eli sen käyttäjät vastaavat sen päivityksestä. Periaate on siis sama kuin Wikipediassa. Sivustolta löytyvä Alte Drucke Digital (vanhoja painotuotteita digitoidusti) sisältää laajan linkkikokoelman saksalaisten instituutioiden digitoimia faksimile- eli näköiskopioita ennen 1800-lukua painetuista painotuotteista. Linkkejä kannattaa katsoa myös [Heidelbergin yliopiston kirjaston sivuilta](http://www.ub.uni-heidelberg.de/) < <http://www.ub.uni-heidelberg.de/> >, jonne linkit on järjestetty teemoittain.

Digitoidut 1700-luvun lopulla ilmestyneet saksalaiset kirjat

Google Book Searchista ei ole ainakaan vielä tullut merkittävää 1700-luvun lopulla painettujen kirjojen kirjastoa. Kirjahakukone löytää 1700-luvun kolmelta viimeiseltä vuosikymmeneltä kaikkiaan vain yli 300 digitoitua teosta, joista vain viitisenkymmentä on saksankielistä. GBS:än suurin etu on, että siinä pystyy tekemään hakuja kirjojen leipäteksteihin. Kuitenkin useat tekstit ovat joiltakin osiltaan hyvinkin epätarkkoja, eikä haku leipätekstiin toimi tällaisissa valokuvissa. Valitettavasti haku ei toimi usein silloinkaan

jos teksti on painettu fraktuuralla, kuten saksalaisten 1700-luvun tekstien kohdalla usein on.

Göttingen Digitalisierung-Zentrum sivulta löytyy tällä hetkellä (13.7.2007) yhteensä jopa 6 451 digitoitua teosta, mikä on yhteensä yli 2300000 digitaalista kuvaa. Suuri osa sivuilta löytyvästä materiaalista on juuri 1700-luvun lopusta. Tieteellisten julkaisut, kartat mutta erityisesti matkakertomukset ovat GDZ:issä hyvin edustettuina. Sadat saksalaiset matkakertomukset 1700- ja 1800-luvuilta tarjoavat mielenkiintoisen kurkistuksen vanhaan maailmaan. Sivuilta löytyy myös paljon sellaista materiaalia, jota ei ole saatavilla suomalaisista kirjastoista. Toisin kuin GBS:ta GDZ:ista voi tehdä hakuja pelkästään otsikon ja tekijän mukaan.

Nyt voi myös itse vaikuttaa siihen, mitä saksalaista aineistoa digitoidaan - se tosin maksaa. GDZ:iin on tullut uusi palvelu nimeltään DigiWunschBuch (digitoivekirja), jonka kautta kuka tahansa voi pyytää Göttingenin kirjastossa olevasta ennen vuotta 1900 painetusta kirjasta digitoinnin ryhtymällä kirjakummiksi (Buchpate). Kirjakummi maksaa noin 25 senttiä jokaista digitoitua sivua kohden. Kun kirja on digitoitu ja palvelumaksu maksettu, digitoitu kirja julkaistaan Digi Wunsch Buch -sivustolla kirjakummin nimen kera. Tällä hetkellä sivuilta löytyy 286 toivekirjaa. Tällä hetkellä matkakirjat näyttäisivät olevan halutuimpia digitalisoinnin kohteita.

Matkakertomuksien kohdalla tulee ajatelleeksi, miten muun maalaiset ovat nähneet saksalaiset alueet 1700-luvun lopulla. Bibliothèque nationale de France [Gallica-sivustolta](http://gallica.bnf.fr) < <http://gallica.bnf.fr> > löytyy tällä hetkellä jo yli 1200 digitoitua teosta, mikä on siis vielä pientä Googlen ja DFG:n projekteihin verrattuna. Saksaa käsittelevää matkakirjallisuutta on sivuille digitoitu muutamia niteitä. Lähteitä 1700-luvun lopun saksasta etsivän kannattaa katsoa myös ECCO eli Eighteen Century Collections Online, josta löytyy jopa 150 000 Englannissa 1700-luvulla julkaistua kirjaa. Kirjojen teksteihin voi tehdä myös hakuja. Sivusto sivuaa Saksan 1700-luvun lopun historiaa siinä, että sieltä löytyy monia matkakertomuksia ja muita kuvauksia, kuten historioita, 1700-luvun lopun Saksasta. Econ käyttö on maksullista, mutta sivuja voi käyttää ilmaiseksi useiden suomalaisten yliopistojen kautta

Verkossa edustava määrä saksalaisia sanoma- ja aikakauslehtiä

Ainakaan toistaiseksi Google ei ole digitoinut yhtään kokonaista lehteä verkkoon. Yleisemminkin tällä hetkellä verkossa on luettavissa vain pieni määrä vanhoja sanomalehtiä. Suomessa Historiallinen sanomalehtikirjasto 1771-1890 on digitoinut kaikki Suomessa kyseisinä vuosina ilmestyneet sanomalehdet. Yhdysvaltalaisten kirjastojen ylläpitämä Making of Amerika -sivusto tarjoaa puolestaan mahdollisuuden tutustua muutamiin 1800-luvulla Yhdysvalloissa ilmestyneisiin aikakauslehtiin. Britannialainen Internet Library of Early Journals tarjoaa muutamia vuosikertoja 1700- ja 1800-luvuilla ilmestyneistä aikakauslehdistä. Englannin osalta valoa näkyy kuitenkin jo putken päässä, kun The London Gazetten digitoimisprojekti valmistuu lähitulevaisuudessa kokonaan. Thomson Gale tarjoaa internetissä mielenkiintoista mutta maksullista The Times -sanomalehden faksimileversiota.

Kansainvälisesti vertailtuna Saksassa 1700-luvun lopulla ilmestyneet sanomalehdet ovat internetissä kiitettävästi edustettuina. 1700-luvulla sanomalehdet vakiinnuttivat paikkansa tärkeimpänä mediana. Erään arvion mukaan 1780-luvulla nykyisen Saksan alueella ilmestyi yli 700 erilaista sanomalehteä, mutta jo seuraavana vuosikymmenenä sama luku oli jo yli 1200.⁵ Arvioni mukaan internetiin on tällä hetkellä kuvattu noin 50 saksalaista sanoma- tai aikakauslehteä ajalta 1770-1800. Kaikkiin ilmestyneisiin nimikkeisiin nähden määrä tuntuu pieneltä, mutta maailman laajuudesta verkosta löytyvät sanomalehdet ovat siinä mielessä edustavia, että muutamaa poikkeusta lukuun ottamatta lehtien digitoinnit sisältävät kaikki lehdistä ilmestyneet vuosikerrat. Lisäksi internetistä löytyvän materiaalin edustavuutta parantaa se, että pääosa faksimilelehdistä on pitkään ja laajasti ilmestyneitä lehtiä, kun monet 1700-luvulla ilmestyneistä lehdistä olivat pienilevikkisiä ja vähän aikaa ilmestyneitä.

Yliopistokirjasto Bielefeldin tarjoama kirjallisuuslehtien tietokanta [Retrospektive Digitalisierung wissenschaftlicher Rezensionsorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum](http://www.ub.uni-bielefeld.de/diglib/aufklaerung) < <http://www.ub.uni-bielefeld.de/diglib/aufklaerung> > on erittäin kiinnostava aikakauslehtien

digitalisointiprojekti kirjallisuuden- mutta yhtä hyvin myös historiantutkijoille. Laaja lehtiarkisto sisältää 45 eri lehtien vuosikerrat, jotka sisältävät edelleen 82 000 artikkelia 460 000 sivussa (sama määrä sivuja kuin Historiallisessa sanomalehti kirjastossa). Muun muassa sellaiset saksalaisen valistuksen eli *aufklärungin* keskeiset äänitorvet kuten Der Teutsche Merkur ja Deutsche Monatschrift ovat kokonaan kuvattu www-sivustolle. Lehdet sisältävät muun muassa Goethen, Herderin ja Rousseau'n kirjoituksia. Lehtiin voi tehdä hakuja tekijän ja otsikon mukaan.

Monille saksalaisille kirjastoille on tyypillistä, että niiden sivuille on kuvattu paikallissanomalehtiä. Ausburgin yliopiston kirjasto on digitoinut koko [Ausburgische Ordinari Postzeitung](http://www.bibliothek.uni-augsburg.de/dda/dr/ztg/ordpost/) -nimisen sanomalehden < http://www.bibliothek.uni-augsburg.de/dda/dr/ztg/ordpost/> vuosilta 1770-1806. Myös [Freiburgin yliopistokirjasto](http://www.ub.uni-freiburg.de/dipro/index.htm) < http://www.ub.uni-freiburg.de/dipro/index.htm> on digitoinut Freiburgilaisen sanomalehden nimeltään Freiburgerzeitung vuosilta 1784-1943. Näissä digitoinneissa ei ole valitettavasti indeksejä tai hakukoneita. Jälkimmäinen lehden kuvanlaatukin on epätarkka. Itävallan puolella sijaitsee sanomalehtitietokanta [Austrian Newspaper Online](http://anno.onb.ac.at/anno.htm) < http://anno.onb.ac.at/anno.htm> , jonne on digitoitu suuri määrä itävaltalaisia sanomalehtiä pääosin 1800-luvulta. Löytyypä sivustoilta vuosina 1780-1781 ilmestyneet numerot lehdestä Wiener Zeitung. Mainittakoon, että mitkään näistä sanomalehdistä eivät löydy suomalaisista kirjastoista.

Mielenkiintoisen näköalan saksalaiseen aate- ja koulutushistoriaan tarjoaa [Bibliothek für Bildungsgeschichtliche Forschung](http://www.bbf.dipf.de/retro-digit1.htm) -niminen instituutio < http://www.bbf.dipf.de/retro-digit1.htm> , jonka kotisivuille on koottu pedagogisia sanomalehtiä vuosilta 1760-1942. Sivustoille on kuvattu muun muassa sellaisia lehtiä kuin Allgemeine Bibliothek für das Schul- und Erziehungswesen in Teutschland (1773-1786), Deutsche Zeitung oder Moralische Schilderungen der Menschen, Sitten und Staaten (1788-1795), Beylage zu deutsche Zeitung oder Moralische Schilderungen der Menschen, Sitten und Staaten unserer Zeit (1790-1795).

Onko internet oikeasti merkityksellinen historiallisten merkitysten tuottaja?

Sähköiset arkistot internetissä pystyvät tarjoamaan tällä hetkellä vain pienen murto-osan kirjastoihin ja arkistoihin säilytystä materiaalista. Tämä huomioon otettuna saksalaiset 1700-luvun lopun painotuotteet ovat suhteellisesti hyvin edustettuina internetissä. Eikä eurooppalaisten tarvitse olla huolissaan vielä edes yhdysvaltalaisen Googlen hegemoniasta historiallisen tiedon tuottajana, sillä Google Book Search laahaa monessa tapauksessa perässä myös englanninkielisten faksimilekirjojen tuottajana määrällisesti ja varsinkin laadullisesti. Jos Google onnistuu massiivisessa projektissaan ja pystyy tulevaisuudessa tarjoamaan hyvälaatuisen miljoonien kirjojen internetarkiston, toivotan sen ilolla tervetulleeksi. Tuohan se mukanaan - kuten jo nyt on nähty - toisia historiallisen tiedon vallasta kilpailevia internetarkistoja. Se on varmaa, että historiallisen tiedon kilpavarustelusta hyötyvät loppujen lopuksi eniten historioitsijat ympäri maailmaa, joille kertyy koko ajan enemmän lähteitä tietokoneensa näyttöpäätteelle.

Se, että jonkin tietyn kulttuuripiirin tiettyä aineistoa on helposti saatavilla, ei vaikuta kriittisesti orientoituneen tutkijan valintoihin. Toisin voi olla kylläkin populaarin historioitsijan laita. Koko internetlähteiden kilpavarustelu selittyykin ensisijaisesti populaarin tutkimuksen lähtökohdista käsin. Jos lähtee rajaamaan tutkimuskysymystään helposti löydettäviin lähteisiin, tietysti lähteitä tarjoavilla instituutioilla on silloin valtaa vaikuttaa ihmisten historian kuvaan. Nyt tulee kuitenkin muistaa se, että kaikkien kirjastojen kokoelmat ovat myös karttuneet enempi tai vähempi sellaisissa tiedonetsintäprosesseissa, jossa valta on vaikuttanut.

Lopuksi on muistettava, ettei Google tai mikään muukaan tiedontuottaja saa olla historioitsijalle mikään orwellilainen *telescreen*, jota ei voi vaientaa. Jos ei ole tyytyväinen internetissä oleviin palveluihin, niitä ei tarvitse käyttää. Internetissä olevat lähteet pitää ajatella tutkijan ekstraresursseiksi. Kannattaa ottaa muutama digitaalisia lähteitä hakeva hakukone kirjastohakukoneiden rinnalle materiaalia etsiessä. Voi säästyä paljolta vaivalta.

Kirjoittaja on filosofian maisteri Joensuun yliopistosta, joka on aloittanut tohtorikoulun European University Institutessa syksyllä 2007. Hänen väitöskirjansa käsittelee 1700-luvun lopun saksalaisen valssi-tanssin kulttuuri- ja sosiaalishistoriaa.

Kirjallisuus

Dames, K. Matthew. "Beyond Google." *Searcher* Sep2006, Vol. 14 Issue 8 (2006), p.21-24

Fischer, Heinz-Dietrich 1973. *Deutsche Zeitschriften des 17. bis. 20. Jahrhunderts*. Pullach, München.

Nygård, Antti-Jussi. Googlen kunnianhimoinen kirjahaku herättää kritiikkiä. *Ennen ja nyt* 2005/2, <http://agricola.utu.fi/tietosanomat/numero2-05/google.html> [30.11.2006]

Turvey, Tom. "A Perspective on Google Book Search." *Information Today* Jan2006 Vol. 23 Issue 1, p. 1-25.

O'Leary, Mick. "Google Book Search Has Far to Go." *Information Today* Nov2006, Vol. 23. Issue 10, p. 41-47.

¹ Dames, 2006; Nygård, 2005

² Mm. *Information Today* -nimisessä sanomalehdessä on käyty näyttävää keskustelua Google Book Search -projektista. Kts. Turvey 2006; O'Leary, 2006.

³ Kts. esim. Nygård, 2005.

⁴ Spiegel Online. 19.4. 2006.

⁵ Fischer 1973, 18-19.