

Elektronisen julkaisemisen uusia tuulia - sanomalehtiä ja aikakausjulkaisuja taannehtivasti verkossa

AGRICOLA

Vaikka verkkobisneksen ja uusmedia-alan kovasti mainostetulla "sisällöntuotannolla" ei nykymuodossaan näytä olevan paljonkaan yhteistä humanismin kanssa, kaiken kaupallisen turhuuden varjossa on edelleen työn alla sellaisiakin hankkeita, jotka hyödyttävät myös aivan perinteistä humanistista tutkimusta. Useissa ajankohtaisissa koti- ja ulkomaisissa julkaisuprojekteissa ollaan tuomassa verkon kautta saataville sekä sanomalehtien että tieteellisten aikakausjulkaisujen vanhoja vuosikertoja taannehtivasti aina viime vuosisadalta lähtien.

JSTOR - elektroninen lehtivarasto

Tieteellisiin aikakauslehtiin erikoistunut amerikkalainen [JSTOR](#) on yksi historioitsijan kannalta kiinnostavimmista julkaisuhankkeista. Nimi JSTOR on lyhennys sanoista *journal storage*, mikä viittaa hankkeen taustaan: aikakausjulkaisujen säilyttäminen vaatii kirjastoilta runsaasti hyllytilaa, ja yksi JSTOR:in perustamisen taustalla olleista motiiveista oli nimenomaan pyrkimys pienentää yksittäisten kirjastojen varastointikustannuksia tarjoamalla vanhat numerot keskitetysti saataville verkon kautta. Lisäksi tavoitteena oli luoda keskeisistä julkaisuista koottu peruskokoelma, joka olisi tasaveroisesti kaikkien (amerikkalais)yliopistojen käytettävissä. Hanke pääsi vauhtiin vuonna 1995, ja nyt se on edennyt jo seuraavaan vaiheeseen, jossa nykyistä kokoelmaa laajennetaan luonnontieteiden suuntaan.

Tällä hetkellä JSTOR sisältää yhteensä 117 amerikkalaisen, etupäässä humanistisen ja yhteiskuntatieteellisen aikakausjulkaisun täydelliset vuosikerrat elektronisessa muodossa. Useissa tapauksissa

lehtien varhaisimmat numerot ovat peräisin 1800-luvun puolelta. Joukkoon mahtuu useita historioitsijan kannalta keskeisiä julkaisuja, mm. *American Historical Review*, *Journal of Modern History*, *Journal of Economic History* ja *Journal of the History of Ideas*, joista on saatavilla kaikki yli viisi vuotta vanhat numerot. Lehtien uusimpia vuosikertoja palvelusta on turha etsiä, sillä kyseessä on nimenomaan varasto, johon uudet lehdet siirretään vasta myöhemmin. Tämä on luonnollisesti myös lehtien kustantajien etujen mukaista.

JSTOR:in tekninen toteutus on mielenkiintoinen. Yhdistämällä samaan tietokantaan sekä kuvatiedoston että tekstintunnistusohjelmalla luodun tekstitiedoston JSTOR ratkaisee näppärästi kokotekstijulkaisuihin tähän asti liittyneet ongelmat. Nykyiset OCR-ohjelmat tekevät näet väistämättä ainakin jonkin verran virheitä, joiden korjaaminen ihmisvoimin olisi erittäin hankalaa ja kallista. JSTOR-projektissa tekstintunnistuksen tarkkuudeksi on saatu 99.95%, eli OCR-ohjelman läpi ajatusaineistossa on keskimäärin yksi virhe 2000 merkkiä kohden. Tämä tarkkuus riittää kyllä hyvin hakuohjelmien tarpeisiin, mutta prosessin tuloksena syntynyt tekstitiedosto ei kuitenkaan sinällään kelpaa alkuperäisen paperille painetun julkaisun korvikkeeksi.

Tekstitiedostosta poiketen elektroninen faksimile taas säilyttää muuttumattomana sekä alkuperäisen informaation että julkaisun ulkoasun, mutta toisaalta pelkistä kuvatiedostoista ei voi tehdä minkäänlaisia tekstihakuja. JSTOR yhdistää nämä kaksi formaattia siten, että hakuohjelmalla on käytössään OCR-ohjelman läpi ajettu tekstitiedosto, kun taas lukijalle annetaan samat sivut kuvatiedostoina. Näin voidaan kätevästi hyödyntää kummankin formaatin hyvät puolet.

JSTOR:iin tallennetut artikkelit voi halutessaan myös tulostaa, eli niitä ei ole pakko lukea pelkästään näyttöruudulta. Aineistoa voi käyttää joko selailemalla lehtien vuosikertoja alkuperäisessä järjestyksessä numero kerrallaan tai sitten tekemällä aineistosta hakuja, joiden avulla on mahdollista löytää yksittäisiä artikkeleita tai jopa jonkin erisnimen tai sanan esiintymiä jossakin aineistoon sisältyvässä lehdessä jonakin tiettyinä aikana. Niinpä palvelusta on jo nyt ollut hyötyä esim. kielentutkijoille. JSTOR:in sisältämästä aineistosta on voitu tutkia esim. tiettyjen sanojen ja sanontojen esiintymistä eri ikäisissä teksteissä, ja joissakin tapauksissa

aikaisimmat löytyneet esiintymät ovat olleet vuosikymmeniä aiemmin tunnettuja esiintymiä varhaisempia. Muutenkin jotain tiettyä aihetta käsittelevät artikkelit löytyvät tietokannasta vaivattomasti, siinä missä saman materiaalin etsiminen bibliografioiden tai indeksien avulla olisi aiemmin saattanut viedä päiväkausia.

JSTOR:in toiminnan rahoitus perustuu sitä käyttäviltä kirjastoilta ja muilta yhteisöiltä kerättäviin vuosittaisiin maksuihin, eli se ei ole vapaata verkkoaineistoa. Suomessa se on kuitenkin Kansallisen elektronisen kirjaston [FinELibin](#) kautta käytettävissä useimmissa suomalaisissa yliopistoissa, ja kuten edellä on toivottavasti käynyt ilmi, kyseessä on historioitsijankin kannalta erittäin hyödyllinen ja suositeltava palvelu.

Entä meillä Suomessa?

Suomessa ei ole käytettävissä miljoonien dollarien budjetteja, mutta mikään ei kuitenkaan estä samanlaisen tekniikan soveltamista hieman pienemmissä ympyröissä. Itse asiassa Helsingin yliopiston kirjaston Mikkelin mikrokuvaus- ja konservointilaitoksen Aurora-projektissa on jo parin vuoden ajan kokeiltu kotimaisen autonomian ajan sanomalehdistön muuntamista elektroniseen muotoon. Tämä projekti on osa laajempaa pohjoismaista hanketta, jossa pyritään saattamaan verkon kautta saataville kaikki pohjoismainen ennen vuotta 1850 julkaistu sanomalehtiaineisto. Suomessa digitoinnin on tämänhetkisten suunnitelmien mukaan tarkoitus edetä aina vuoteen 1860 saakka. Tulevaisuudessa projektia on luonnollisesti mahdollista jatkaa pitemmällekin, joskin intoa vähentää toistaiseksi se, että 1800-luvun lopun aineiston kohdalla alkaa jo tulla vastaan potentiaalisia tekijänoikeusongelmia

Mikkelin projektissa alkuperäiset lehdet kuvataan ensin uudestaan mikrofilmille, josta ne sitten skannataan digitaaliseen muotoon. Aineiston ajaminen tekstintunnistusohjelman läpi vaatii runsaasti ihmistyötä, sillä ohjelma täytyy opettaa jokaisen lehden kohdalla erikseen tunnistamaan kaikki siinä käytettyjen kirjaimien eri variaatiot. Vanhojen lehtien epätasainen painojälki ja monissa lehdissä kirjasinlajina käytetty fraktuura vaikeuttavat prosessia, ja käytännössä samankin lehden kirjaimisto täytyy opettaa ohjelmalle tietyin väliajoin uudestaan sen takia, että lehden painamisessa käytetyt kirjasimet ovat vähitellen kuluneet.

JSTOR:ista poiketen tämä projekti on yhä kesken, ja sen tuloksia voitaneen odottaa yleisesti saataville aikaisintaan parin vuoden päästä. Digitoitu aineisto on tarkoitus koota tietokantaan, joka sitten asetetaan käytettäväksi Internetin kautta. Tietokannan hakuominaisuuksia pyritään parantamaan myös hyödyntämällä 1800-luvulla koottua sanomalehtien artikkelihakemistoa, joka sekin skannataan ja ajetaan tekstintunnistusohjelman läpi. Näin siis myös viime vuosisadan kirjastotätien ja -setien tekemä työ koituisi digitaalisen aikakauden tutkijoiden hyödyksi.

Aurora-projektissa keskitytään vanhaan autonomian ajan materiaaliin, mutta Mikkelissä aloitellaan myös toista digitointi-projektia (Peri+), jossa kokeillaan tieteellisten aikakausjulkaisujen takautuvaa julkaisemista elektronisessa muodossa. Hanke sai Auroran tavoin täksi vuodeksi rahoitusta opetusministeriön tietoyhteiskuntamäärärahoista. Alustavien suunnitelmien mukaan koemateriaalina on tarkoitus käyttää vuodesta 1903 lähtien ilmestynyttä *Historiallista aikakauskirjaa*, jonka vanhoista numeroista ryhdytään siis työstämään elektronista versiota. Samalla täytyy selvittää myös tieteellisissä aikakauslehdissä julkaistun aineiston taannehtivaan digitointiin liittyviä tekijänoikeuskysymyksiä.

Uudempiä tieteellisiä aikakauslehtiä on julkaistu jo aiemmin [Elektra](#)-tietokannassa, joka on kevästä 2000 lähtien tullut FinELibin kautta aiempaa laajemmin saataville kaikissa suomalaisissa yliopistoissa. Elektra on myös siirtymässä uudelle palvelimelle, jolloin se toivottavasti toimii tähänastista juohevammin. Lehtien lisäksi Elektrassa on julkaistu myös [Bibliotheca Historica -sarjassa ilmestyneet väitöskirjat](#), joita on muutamassa vuodessa kertynyt jo kolmisenkymmentä kappaletta. Elektran julkaisut ovat PDF-muodossa, joten palvelun hakuominaisuudet eivät ole samaa luokkaa kuin esim. JSTOR:issa, mutta artikkelit soveltuvat joka tapauksessa hyvin sekä ruudulta luettaviksi että tulostettaviksi.

[Jyrki Ilva](#)

FM, Helsingin yliopisto

[Sisällysluettelo 2/00](#)

[Agricolan Tietosanomien pääsivulle](#)

[Historian äärelle](#) | [Tutkimus, opetus, seurat](#) | [Arkistot, kirjastot, museot](#) | [Ajankohtaista](#)
[Agricolan kartta](#) | [Haku Agricolasta](#) | [Hakemisto](#) | [Uutta!](#)
[Tekijät](#) | [Palaute](#) | [Etusivulle](#)

