
Digitaalinen historia



Seppo Eskola ja Lauri Leinonen

Suomen keskiajan asiakirjalähteistä uusi verkkopalvelu

Kansallisarkisto käynnisti vuonna 2015 Koneen Säätiön tuella *Diplomatarium Fennicum* -hankkeen, jossa samanniminen Suomen keskiaikaisten asiakirjojen tietokanta uudistetaan täysin ja päivitetään modernin keskiajan tutkimuksen vaatimukset täyttäväksi. Uudistunut *Diplomatarium Fennicum* -tietokanta muodostaa monitieteisen, suomalaista keskiajan tutkimusta palvelevan verkkopalvelun ja luo samalla infrastruktuurin keskiaikaisten asiakirjalähteiden pitkäjänteiselle editointityölle ja tutkimukselle. Uuden verkkopalvelun betaversio julkaistaan 1.10.2016 Turun kirjamesseilla.

Asiakirjat ovat yksi keskiajan tutkimuksen tärkeimmistä lähderyhmistä, ja niitä koskevien tieteellisten editioiden laatiminen on keskeinen tapa tuoda asiakirjalähteitä laajamittaisesti tutkijoiden käyttöön. Pohjoismaissa keskiaikaisten asiakirjojen editiosarjoja, joita usein nimitetään *diplomatariumeiksi*, on laadittu systemaattisesti 1800-luvun alkupuolelta lähtien. Suomessa keskiaikaisen asiakirja-aineiston tutkimus perustuu

Reinhold Hausenin (1850–1942) massiiviselle editointityölle. Hausen kokosi, editoi ja julkaisi kaikki löytämänsä Suomen keskiaikaa koskevat asiakirjat teossarjassaan *Finlands Medeltidsurkunder* (FMU 1910–1935) sekä niin sanotun Turun Tuomiokirkon Mustan kirjan editiossa (*Registrum Ecclesiae Aboensis eller Åbo Domkyrkas Svartbok*, REA 1890). Hausenin yli 6700 asiakirjaa sisältävät julkaisut ovat olleet suomalaisen keskiajantutkimuksen kulmakivi jo yli sadan vuoden ajan. Erityisen arvokkaana hänen työssään voidaan pitää koko Suomen keskiajan kattavan asiakirjakorpuksen muodostamista. Hausenin työ muodostaa *Diplomatarium Fennicum*in (DF) perustan.

Keskiajan tutkimus on viime vuosikymmeninä jatkuvasti monipuolistunut ja vahvistanut asemaansa Suomessa, ja samalla vaatimukset tutkimuksen apuvälineille ovat kasvaneet. Uusimman tutkimuksen näkökulmasta monet Hausenin työssään tekemät valinnat ovatkin jo vanhentuneet. Editointityön periaatteet ovat kehittyneet, historialliset tulkinnat muuttuneet ja monet Hausenin editioiden tiedot, kuten viittaukset arkistolähteiden sijainteihin, ovat yksinkertaisesti käyneet vuosikymmenien kuluessa paikansa pitämättömiksi. Nykytutkimus korostaa myös monitieteisyyttä, kun taas Hausenin editiot ovat pitkälti soveltumattomia esimerkiksi historiallisen kielitieteen tarpeisiin. Tarve tilanteen parantamiselle on jo pitkään ollut ilmeinen.

Kansallisarkistossa luotiin jo 1990-luvulla ensimmäinen versio *Diplomatarium Fennicu-*

FM **Seppo Eskola** työskentelee projektipäällikkönä ja FM **Lauri Leinonen** tutkijana Kansallisarkiston *Diplomatarium Fennicum* -hankkeessa. Sähköpostit: seppo.eskola@arkisto.fi ja lauri.leinonen@arkisto.fi.

mista. Tuolloin tavoite oli saattaa FMU:n ja REA:n sisältämät tekstit sähköiseen muotoon ja asettaa ne verkkotietokantana tutkijoiden ja suuren yleisön käyttöön. Hanke julkaisikin parannellun version FMU:sta sekä osan REA:a (ensimmäiset 150 asiakirjaa) ja saavutti näin pääosin tärkeimmän tavoitteensa: Hausenin painetut julkaisut saatiin digitaaliseen muotoon ja mahdollistettiin tiedonhaku laajasta tekstiaineistosta. Ensimmäinen DF-tietokanta onkin ollut kohta kaksikymmentä vuotta tutkijoiden ahkerassa käytössä. Samalla teknologian nopea kehitys on kuitenkin jättänyt 90-luvun tietokannan jälkeensä, ja nykynäkökulmasta sen puutteet ovat yhtä ilmeiset kuin Hausenin editioidenkin. Tutkijat ovat jo pitkään toivoneet tietokannan niin sisällöllistä kuin teknistäkin päivittämistä.

Käynnissä oleva toinen *Diplomatarium Fennicum* -hanke on syntynyt Kansallisarkiston ja tutkijayhteisön dialogin pohjalta vastaamaan näihin toiveisiin. Hanke nostaa Suomen keskiaikaisten asiakirjojen tutkimusinfrastruktuurin nykypäivän vaatimustasolle.

Tietovaranto ja tutkimusväline

*Diplomatarium Fennicum*in uusi verkkopalvelu on teknisesti täysin uudistettu, ja tietokannan sisältö on laajentunut merkittävästi. Tavoitteena on luoda yhtäältä tietovaranto, jonne voidaan tallentaa laajasti tietoa Suomen keskiaikaa koskevista asiakirjalähteistä, ja toisaalta työväline, jolla tätä tietoa voi hallita ja tutkia. Käytännössä sekä tietokannan sisältö että toiminnallisuudet on rakennettu uudestaan: Hausenin julkaisujen rinnalle on lisätty muita editiosarjoja, asiakirjaeditiot on yhdistetty muihin niitä koskeviin verkkotietokantoihin, alkuperäislähteiden arkistosijainnit on pyritty päivittämään ajantasaisiksi, ja tietokannan käytettävyydelle olennaista metatietoa on yhdenmukaistettu ja monipuolistettu. Erityisen tärkeänä on pidetty alkuperäislähteiden digitaalisten kuvien liittämistä asiakirjaeditioihin aina kun mahdollista. Teknisiä ominaisuuksia suunniteltaessa on panostettu käytettävyyteen: hakuominaisuudet ovat kattavat, kuvien ja editioiden käsittelemiseen on laadittu monipuoliset työkalut, ja parhaillaan kehitetään toiminnallisuuksia esimerkiksi karttojen ja tilastojen hyödyntämiseen.

Asiakirjojen tutkimushistoria on uudessa tietokannassa haluttu tehdä läpinäkyväksi, ja tietokantaan on liitetty FMU:n ja REA:n ohella myös tärkeimmät Hausenin lähteenään käyttämät editiosarjat, kuten A. I. Arwidssonin *Handlingar till upplysning af Finlands häfder* (1846–1858) sekä tietyt hansakaupunkien asiakirjojen editioita sisältävät julkaisusarjat. Aiemmin tietokannasta pääosiltaan puuttunut Turun tuomiokirkon Mustan kirjan editio – yli seitsemänsataa asiakirjaa – on kokonaisuudessaan digitoitu, oikoluettu ja lisätty tietokantaan. Myös muuta tutkimusta tukevaa kirjallisuutta on digitoitu, esimerkkinä mainittakoon Hausenin *Finlands Medeltidsigill* (1900), jonka sisältämät sadat sinettijäljennökset on yhdistetty asi-aankuuluviin asiakirjoihin. Tärkeimmät ulkomaiset tietokannat on myös huomioitu, ja kukin asiakirja on yhdistetty *Svenskt Diplomatariums huvudkartotek* (SDHK) tietokantaan Ruotsin Valtionarkiston sivuille.

Korkealaatuiset digitaaliset kuvat ovat editioiden ohella nykyaikaisen asiakirjatutkimuksen perusedellytys. *Diplomatarium Fennicum*in on tarkoitus aikaa myöten sisällyttää kuvat kaikista aineistoon sisältyvistä alkuperäisasiakirjoista tai – milloin tämä ei ole mahdollista – linkittää asiakirjaeditiot alkuperäislähdettä säilyttävän instituution palveluihin. Tehtävä ei ole vaille haasteita, sillä alkuperäislähteitä säilytetään kymmenissä eri arkistoissa ja kirjastoissa ympäri Eurooppaa, vaikka suuri enemmistö löytyykin Tukholman ja Tallinnan arkistoista tai Suomesta Kansallisarkistosta ja Kansalliskirjastosta. Tällä hetkellä lähes kaikki suomalaisissa instituutioissa säilytettävät keskiaikaiset asiakirjat on digitoitu, ja kuvat löytyvät joko suoraan tietokannasta tai linkin takaa. Ulkomaisten instituutioiden asiakirjojen kuvien hankkiminen on aloitettu, ja niitä löytyy tällä hetkellä tietokannasta toista sataa.

Diplomatarium Fennicum on relaatiotietokanta, jonka toiminta perustuu siihen tallennetun metatiedon yhdistelylle. Metatietojen kehittäminen on hankkeen keskeisiä haasteita ja edellytys tietokannan tehokkaalle käytölle, tiedon systematisoinnille ja laajoille haku- ja muille toiminnallisuuksille. Pitkällä aikavälillä tavoitteena on tuottaa tietokantaan muun muassa yhdenmukaistetut nimi- ja paikkatiedot

koko asiakirjakorpuksesta, kategorisoida ja asiansanoittaa asiakirjat historiallisen kontekstinsa ja sisältönsä mukaan sekä kerätä alkuperäisasiakirjoista kodikologiset tiedot. Tämä on luonnollisesti erittäin suuri urakka ja voidaan toteuttaa nykyisen hankkeen puitteissa vain osin. Työ koostuu sekä Hausenin editioiden tietojen yhdenmukaistamisesta ja siirtämisestä koneluettavaan muotoon että täysin uuden tutkimustiedon tuottamisesta. Tulevaisuudessa joukkoistaminen tarjoaa kuitenkin lupaavia mahdollisuuksia metatietojen laajentamiseen – koskien esimerkiksi paikatietojen keräämistä – mikä on huomioitu verkkopalvelun kehittämisessä.

Hyvän esimerkin metatiedon kehittämiseen liittyvistä haasteista muodostavat tiedot asiakirjojen arkistosijainneista. Hausenin käyttämät lähdeviitteet ovat kuuluisan epätarkkoja (*”Original på papper i Finlands Riksarkiv”*) tai aikaa sitten vanhentuneita eivätkä siksi enää palvele tutkimusta. Näin alkuperäislähteiden käyttäminen onkin tutkijoille usein hyvin työlästä. Tietokannassa viitataan yli tuhanteen Suomessa sijaitsevaan arkistolähteeseen, jotka DF-hanke on valtaosiltaan jäljittänyt. Lisäksi lähteiden arkistosignumit on lisätty tietokantaan. Samalla on mahdollistettu tärkeimpien aineistojen digitoiminen ja digitaalisten kuvien yhdistäminen tietokantaan. Keskiakaisia asiakirjalähteitä on Suomesta paikallistettu noin 230, joista yli 140 on Kansalliskirjastossa ja lähes 90 Kansallisarkistossa (tai muutamassa yksittäistapauksessa eri maakunta-arkistoissa). On myös hyvä muistaa, että satoja keskiajan Suomeen liittyviä asiakirjoja on säilynyt vain myöhempiä kopioina esimerkiksi 1600- ja 1700-lukujen tuomiokirjoissa, joiden kuvat löytyvät nyt tietokannasta. Ulkomailta sijaitsevien lähteiden paikallistaminen on kesken, mutta suuri osa Ruotsissa säilytettävien asiakirjojen arkistotunnuksista on jo löydetty. Kuten yllä todettiin, aineistoja löytyy kymmenistä ulkomaisista arkistoista, joista merkittävimpiä ovat Ruotsin valtionarkisto, Tukholman kaupunginarkisto, Tallinnan kaupunginarkisto, Vatikaanin arkisto sekä lukuisat vanhojen hansakaupunkien arkistot.

Kaikkiaan DF-tietokanta sisältää tällä hetkellä jo yli 23 000 viitettä lähde-editioihin, arkistolähteisiin, tutkimuskirjallisuuteen ja muihin tietokantoihin. Julkaisemalla tietokannasta betaversion Kansallisarkisto haluaa jo kehitys-

vaiheessa tarjota tutkijoille mahdollisuuden jakaa lähdetietouttaan DF-hankkeen kanssa sekä osallistua muilla tavoin tietokannan kehittämiseen.

Digitaalisia menetelmiä ja perinteistä editointityötä

Diplomatarium Fennicum on osa Suomessakin koko ajan kasvavaa digitaalisen humanismin kenttää. Suomessa on tällä hetkellä käynnissä tai hiljan valmistuneina useita keskiaikaa koskevia tutkimushankkeita, jotka julkaisevat aineistonsa verkossa. Suomalaisen Kirjallisuuden Seuran *Codices Fennici* -hankkeessa luetteloidaan kaikki suomalaiset ja Suomeen liittyvät 1600-lukua edeltävät koodeksit ja luodaan kansallinen normisto käsikirjoitusten kuvailulle sekä kodikologiselle tutkimukselle. Hanke pyrkii myös digitoimaan koodeksit ja julkaisemaan ne avoimessa virtuaalisessa kirjastossa. Kansalliskirjaston *Fragmenta Membranea* -kokoelma puolestaan sisältää fragmentteja 1500:sta keskiajan Suomeen liittyvästä käsikirjoituksesta ja on näin merkittävä tietovaranto Suomen keskiaikaisesta kirjallisuudesta. Koko aineisto on Kansalliskirjaston digitoimana käytettävissä verkossa. Myös kotimaisten kielten keskuksen vanhojen asiakirjanimien kokoelma ja muut nimistökoelmat tarjoavat merkittävän aineiston keskiajantutkimukselle, ja niiden saavutettavuus verkossa paranee jatkuvasti. Kaikkien näiden tietovarantojen yhteiskäytön mahdollisuuksien parantaminen ja metatietojen jakaminen niiden välillä on tällä hetkellä keskustelun alla, ja Kansallisarkistossa järjestettiin keväällä 2016 hankkeiden yhteinen keskustelutilaisuus yhteistoiminnan tiivistämiseksi. Kaiken kaikkiaan tutkimusaineistojen saavutettavuus on tällä hetkellä parempi kuin koskaan ennen, ja digitaalinen humanismi näyttää lunastavan sille asetetut odotukset ainakin tältä osin. Muistiorganisaatioiden hankkeet luovatkin edellytyksiä harppauksille suomalaisessa keskiajan tutkimuksessa.

Digitaaliseen humanismiin liitetään odotuksia paitsi aineistojen uudenaikaisesta saavutettavuudesta myös tutkimusmenetelmien kehittymisestä ja jopa uusien kysymysten asettamisesta. *Diplomatarium Fennicum* vastaa alkuvaiheessaan tutkimusmenetelmien osalta odotuk-

siin luomalla tutkimusalustan keskiaikaisten asiakirjojen tutkimukselle. Tärkein ja käytetyin toiminnallisuus *Diplomatarium Fennicum* kaltaisessa tietokannassa on tiedon hakeminen. Verkkopalvelun hakuominaisuudet mahdollistavat muun muassa täyden tekstihaun, ja metatietojen osalta myös ajankohdan, kielen, henkilöiden ja paikkojen, arkistosijainnin ja julkaisusarjan mukaisen tiedonhaun. Metatietojen parantuaessa ja betavaiheen edetessä hakuominaisuudet parantuvat entisestään, ja samalla kehitetään tekniikkaa sallimaan muun muassa niin sanotut sumeat haut. Asiakirjojen tutkimiseen on kehitetty kuvankatselutyökalu, jonka rinnalle kehitetään tilastollisia työkaluja, tiedon visualisoinnin apuvälineitä sekä karttasovellusta. Toiminnallisuuksia tullaan kehittämään käyttökokemusten perusteella, ja hanke ottaa mielellään vastaan palautetta ja kehitysideoita.

Teknisesti tietokanta toteutetaan siten, että sitä voidaan tarvittaessa uudistaa pala palalta teknologisten mahdollisuuksien lisääntyessä. Teknologinen kehitys on nopeaa, ja Kansallisarkisto on mukana esimerkiksi laajassa käsinkirjoitetun tekstin tunnistushankkeessa (READ, *Recognition and Enrichment of Archival Documents*), jonka tuloksia voidaan mahdollisesti soveltaa tulevaisuudessa keskiaikaisiin aineistoihin. Laajojen tekstimassojen käsittelytyökalut tulevat joka tapauksessa kehittymään, ja tähän varautuakseen hanke tukee tiedon mahdollisimman avointa käyttöä. Valmiin tietokannan sisältö tulee olemaan ladattavissa tutkijoiden omille tietokoneille, jotta käyttäjät voivat vapaasti hyödyntää kaikkia *Diplomatarium Fennicum* tutkimusmahdollisuuksia.

Kaikista mahdollisuuksistaan huolimatta digitaalinen humanismi ei korvaa asiakirjatutkimuksen ydintä, lähteiden editointityötä. *Diplomatarium Fennicum* rakentuu 1800-luvulta alkavan tutkimusperinteen varaan, ja keskiajan tutkijoiden työ alkuperäislähteiden editoinnin parissa jatkuu yhä. Suomalaisen keskiajan tutkimuksen näkökulmasta Hausenin työn vahvuutena on koko keskiajan kattavan asiakirjakorpuksen laatiminen, mikä on antanut hyvän pohjan tutkimukselle. Tilanne ei myöskään ole täysin tavanomainen: esimerkiksi Ruotsissa *Diplomatarium Suecanum* editiotyö on edennyt vasta 1300-luvun lopulle. Hausenin julkaisuihin sisältyy lisäksi asiakirjojen ohella myös

muuta tekstiaineistoa kronikkakatkelmista piirto- ja karttoihin. Tämä tavallisuudesta poikkeava ratkaisu voidaan nähdä kauaskatseisuutena ja vahvuutena, sillä tällä tavoin laaja tekstiaineisto voidaan koota saman sateenvarjon alle ilman, että kiinnostavia tekstikatkelmia joudutaan suotta rajaamaan korpuksen ulkopuolelle. Laajuudeltaan Hausenin työ ei luonnollisesti ole kaiken kattavaa, ja myöhemmin onkin löytynyt aineistoja, joista Hausen ei ollut tietoinen. Käynnissä olevan hankkeen tarkoituksena on jatkossa yhteistyössä tutkijoiden kanssa julkaista FMU:sta puuttuvien asiakirjojen ja tekstikatkelmien editioita sekä ylläpitää alustaa, jolla Suomen keskiajan asiakirjakorpukseen voidaan tulevaisuudessa liittää yhä uusia editioita.

Tätä tavoitetta varten hankkeen ja keskiajan tutkijoiden yhteistyönä on laadittu editioperiaatteet uusien editioiden tuotannolle. Editioperiaatteet on muodostettu modernista tutkimusnäkökulmasta ottaen huomioon kansainväliset esimerkit (erityisesti *Diplomatarium Suecanum*) sekä eri tieteenalojen tarpeet. Hausen kohdisti editioissaan huomionsa ennen kaikkea tekstien merkityssisältöön, ei niinkään esimerkiksi kieleen tai ulkoasuun. Moderni keskiajan tutkimus analysoi kuitenkin lähteitä monipuolisesti, ja kieli ja kirjoittamiskäytännöt ovat usein tutkimuksen ensisijaisiakin kohteita. *Diplomatarium Fennicum* editioperiaatteet huomioivat nykyisen tutkimuskentän moninaisuuden ja ohjaavat tutkijat laatimaan yhdenmukaisia editioita.

DF-hankkeen puitteissa tuotetaan uusia editioita sekä aiemmin julkaisemattomista että jo julkaistuista asiakirjoista. Asiakirjaeditioiden laatiminen on aikaa vievää ja hankkeen resurssit asettavat lähivuosina tiukat rajat työn laajuudelle. Uusia editioita pyritään ensi alkuun julkaisemaan Viron ja Vatikaanin arkistojen aineistoista yhteistyössä niihin erikoistuneiden tutkijoiden kanssa. *Svenska Kulturfondenin* tuella puolestaan tuotetaan uusia, kielitieteelliset kriteerit täyttäviä transkriptioita Naantalista luostarissa kirjoitetuista asiakirjoista. Kokonaisuutena hanke pyrkii tuottamaan monipuolista tieteellistä aineistoa jo betakehitysvaiheessa ja huomioidaan tietokannan suunnittelutyössä mahdollisimman hyvin eri tieteenalojen tarpeet.

Diplomatarium Fennicum on luonteeltaan jatkuvasti kehittyvä palvelu, jonka ylläpitoon

Kansallisarkisto on sitoutunut. Ensimmäinen teknisesti täysin valmis versio tietokannasta valmistuu Suomen itsenäisyyden juhluvuonna 2017 muistuttaen samalla, että hallinnollisena kokonaisuutena Suomen juuret ovat jo keskiajalla syntyneissä rakenteissa. Sisällönkehittämistyön ja uusien editioiden julkaisemisen on suunniteltu jatkuvan tutkijoiden kanssa pitkälle tulevaisuuteen.

Betaversio avoimeen käyttöön

Diplomatarium Fennicum -tietokannan betaversio julkaistaan 1.10.2016 Turun kirjamessuilla. Päivä on samalla myös Kansallisarkiston vanhimman asiakirjan, niin sanotun Karjalan nais-

rauhan 700-vuotispäivä. Asiakirja on päivätty 1.10.1316, ja siinä kuningas Birger takaa Ruotsin lain mukaisen ruumiin ja omaisuuden suojan Karjalan naisille. Kansallisarkisto järjestää messuilla näyttelyn, jossa Karjalan naisrauha sekä muita keskiaikaisia aineistoja on yleisön nähtävillä. Ohjelmassa on myös paneelikeskusteluja keskiaikaan ja *Diplomatarium Fennicum*iin liittyen. Paikan päällä pääsee tutustumaan tietokantaan ja keskustelemaan sen toteuttajien kanssa.

Lisätietoa hankkeesta ja tietokanta löytyvät osoitteesta: <http://www.arkisto.fi/df>