

Digitaaliset metadata-arkistot ja mesoanalyysi kulttuurisen vaihdon kartoittamisessa

—

*Viola Parente-Čapková, Kati Launis ja
Jasmine Westerlund*

K

— irjallisuudentutkijoille sana ‘arkisto’ assosioituu tavallisimmin fyysiseen paikkaan, jossa säilytetään kirjailijoiden jäämistöjä. Paitsi materiaaliin (perinteisesti papereihin, kuviin, äänitteisiin tai tiedostoihin) arkisto viittaa myös instituu-tioon, arkistolaitokseen, missä materiaalisia arkistoja säilytetään, luokitellaan ja hallinnoidaan. Abstraktimmassa merkityksessä arkisto on “lausumien muotoutumisen ja muodonmuutoksen yleinen järjestelmä” (Foucault 1969/2005, 172) ja kirjoittamisen mahdollisuuksia määrittävä systeemi” (Hyttinen & Kivilaakso 2010, 8). Kuten Hyttinen ja Kivilaakso (mt.) tuovat esiin, tämä abstrakti merkitys pätee tarvittaessa koko kirjallisuuteen, kun se määritellään “todellisuutta ja historiaa koskevien väitteiden ja kuvittelun” arkistoksi. Metatasoja riittää heti määrittelyvaiheessa, sillä kirjallisuus on esitellyt runsaasti arkistoja ja kirjastoja myös temaattisella tasolla, Umberto Econ *Ruusun nimen* (1989) kaltaisista bestsellereistä vähemmän tunnettuihin teoksiin. Sekä kirjailijat että tutkijat ovat jo kauan sitten kiinnittäneet huomiota valtaan ja sensuurin eri muotoihin näiden aiheiden yhteydessä (ks. myös Hyttinen & Kivilaakso 2010, 12). Arkistot ovat kulttuurisen muistin paikkoja, niiden on joskus pakko toimia pitkään yksityisissä paikoissa tai “maan alla”, ennen kuin niistä voi tulla instituutioita sanan varsinaisessa mielessä. Ja joskus on pakko palata epäviralliseen toimintaan sen jälkeen, kun on saanut toimia hetken virallisesti, kuten venäläisen Memorialin tapauksessa.¹ Esimerkki osoittaa, miten arkistot elävät ajassa ja reagoivat siihen, mitä ympärillä tapahtuu.

Ajankohtainen kysymys on myös se, mitä uutta digitaalisuus on tuonut arkistojen toimintaan, niiden määritelmiin, niiden käyttöön ja tutkimiseen sekä niitä koskeviin teoreettis-metodologisiin pohdintoihin kirjallisuudentutkimuksen näkökulmasta. Mitä tai mitä kaikkea voisi tarkoittaa “digitaalinen arkisto”? Usein sillä tarkoitetaan yksinkertaisesti arkistoa, jolla on samantyyppinen päämäärä kuin fyysisellä arkistolla, mutta digitaaliseen arkistoon sisältyvät asiakirjat ja menneisyydestä kertovat esineet on digitoitu – esimerkiksi skannaamalla tai valokuvaamalla. Toisaalta arkisto voi olla syntyjään digitaalinen. Joidenkin mielestä koko digitaalinen arkisto -nimitys ei ole mielekäs: valtaosassa arkistolaitoksia aineistojen digitointi on aloitettu jo kauan sitten ja syntyjään digitaalisten aineistojen yhteydessä on itsestään selvää, että ne ovat olemassa digitaalisessa muodossa. *Digital archive* -nimitystä on kuitenkin varsinkin englanninkielisessä maailmassa käytetty ja käytetään edelleenkin melko runsaasti, vaikka sen määritelmät ovat vaihtelevia.

Newton Gresham library -sivuston mukaan useat digitaaliset arkistot voi jakaa kolmeen pääryhmään. Ensimmäiseen ryhmään kuuluvat fyysisten arkistojen, kirjastojen, yliopistojen, museoiden, valtioiden, (paikallisten, valtion tai kansallisten) agentuurien tai muiden kulttuuristen seurojen luomat ja ylläpitämät arkistot, jotka ovat usein ilmaisia. Joskus niiden käyttö on kuitenkin jol-

lain tavalla rajattua tai vaatii sen, että käyttäjä rekisteröityy saadakseen pääsyn aineistoihin. Esimerkiksi Suomen Kansalliskirjaston digitaalisten aineistojen digitoidut lehdet ovat tällä hetkellä vapaasti käytettävissä vuoteen 1939 (ruotsinkieliset lehdet vuoteen 1949), mutta vain rekisteröityneet tutkijat pääsevät lukemaan uudempia lehtiä. Norjalaiseen nettiarkistoon, joka on nimeltään juuri Digital Arkivet ja joka toimii erilaisten kokoelmien alustalla (*platform*), tarvitaan myös tunnistautuminen. Tunnistautumisen edellyttämiseen on yleensä joku – usein tärkeä – syy, esimerkiksi aineistoissa voimassa oleva tekijänoikeus tai muu käyttörajoitus, joka estää aineiston vapaan jakamisen tai vaatii käyttäjältä käyttösitoumuksen allekirjoittamista. Monet digitaaliset arkistot toimivat kuitenkin open access -periaatteella, eli ovat vapaasti käytettävissä Internetissä. Toiseen ryhmään kuuluvat digitaaliset arkistot ovat yksityisten ihmisten, pienten ryhmien tai voittoa tavoittelemattomien järjestöjen luomia, ja usein myös vapaasti käytettäviä. Kolmanteen ryhmään kuuluvat voittoa tavoittelevien yritysten kuten kustantamojen luomat arkistot, jotka ovat usein maksullisia.

Monien tutkijoiden mukaan analogisen ja digitaalisen vastakkainasettelu ei tunnu enää hedelmälliseltä (Tolonen 2017; Roine & Piippo 2021, 22). Perinteisten, analogisten aineistojen parissa työskennelleen kirjallisuudentutkijan näkökulmasta on kuitenkin edelleenkin aiheellista pohtia nykyajan arkistotutkimuksen keskeistä asiaa: miten digitaalinen muoto vaikuttaa tutkijan kokemukseen, arkistomateriaalille asetettuihin kysymyksiin ja tutkimustyön tuloksiin? Otetaan ajatusleikki: sanotaan, että tutkimme SDP:n kansanedustajana, Työläisnaisliiton puheenjohtajana ja runoilijana toimineen Hilja Pärssisen perinteisessä analogisessa muodossa olevaa aineistoa Työväenarkistossa. Hämeenlinnan naisvankilassa tuomiotaan istunut Pärssinen kirjoitti runoja paperin puutteessa lakanankappaleisiin, ompelutuvan tilkkuihin ja paperilappuihin. Näitä säilytetään arkistossa. Runot on kirjoitettu vieri viereen, kankaan jokainen kohta hyödyntäen. Runoja täynnä olevien kankaankappaleiden pitelemine, kankaan tuntu ja tuoksu, materia, kuljettaa tutkijan kirjoitushetkeen. (Launis 2010, 28.) Jos sama aineisto olisi digitaalisessa muodossa, kuvina kangaslapuista, tuntu olisi aivan toinen – vaikka sisältö eli runoteksti olisikin tutkijan edessä. Samoin L. Onervan postikorttialbumien fyysinen selaaminen on täysin erilaista arkistossa kuin näytön ääressä: jotkut postikortit pysyvät edelleenkin hienosti albumin sivuilla, joidenkin kiinnikkeet ovat menneet rikki, joitakin on lisätty sivujen väliin ikään kuin ylimääräisinä. Näitä seikkoja on korostettu monesti, mutta uskomme, että on aiheellista tuoda ne esiin aina uudestaan tutkijoiden konkreettisten kokemusten yhteydessä. Analogisessa arkistossa materiaalin määrä ja laatu konkretisoituvat eri tavalla kuin digitaalisessa. Koteloiden määrä on nähtävissä ja liput, laput, muistivihot ja erilaiset kirjepaperit tuntuvat erityisiltä. Tekstin vieressä voi olla kuivattuja kukkia, joissa on vielä jäljellä tuoksua, tai paperissa rosoinen, nopeasti repäisty reuna,

kun teksti on kirjoitettu kiireessä – näin on ollut Olly Donnerin arkistojen tutkimisen tapauksessa.²

Joka tapauksessa arkisto-sana herättää kirjallisuudentutkijoissa odotuksia siitä, että on mentävä toisaalle, jotta pääsisi aineistoihin käsiksi: siirtyminen tähän toiseen paikkaan aineistojen pariin on merkittävä vaihe. Arkisto ympäristönä on ainutlaatuinen tila: siinä on tutkimusrauhaa ja samalla arkistoon liittyy monenlaista portinvartijuutta ja rituaaleja. Toisaalta verkkoympäristö tekee helpoksi aikaisemmin mahdottoman, koska se mahdollistaa tietojen ja sivustojen linkittämisen, erilaisten kokonaisuuksien yhdistämisen ja tiedon sekä aineiston rikastamisen. On mahdollista linkittää vaikkapa kirjallisia teoksia koskevaa metadataa³ varsinaisiin kirjallisiin teksteihin. Metadatan ja sen visualisoinnin avulla päästään tutkimaan erilaisia – esim. kirjailijoiden tai muiden kulttuuristen toimijoiden – verkostoja ja kysymään uusia tutkimuskysymyksiä. Näin tehdään parhaillaan esimerkiksi Fennicaa eli Suomen kansallisbibliografiaa hyödyntävässä projektissa, jossa tutkitaan Suomen kirjallisuushistoriaa bibliografisen metadatan avulla.⁴ Näin esimerkiksi erilaisista metadataa sisältävistä digitaalisista tietokannoista voi tulla laajempi virtuaalinen ympäristö eli omanlaisensa portaali, tietäntyyppinen portti digitoituihin tai syntyjään digitaalisiin aineistoihin, varsinaisiin digitaalisiin arkistoihin. On kuitenkin muistettava, että käytettävissämme oleva metadata on erilaisten ristiriitaisten, usein pitkän ajan kuluessa kerääntyneiden (mm. ideologisten ja poliittisten) intressien välisten keskustelujen osa ja tulosta (Baca 2016; Dahlgren et al. 2020).

Käsilä olevassa katsauksessamme pohdimme sitä, millaisia mahdollisuuksia ja haasteita kirjallisuuden vastaanottoa dokumentoivat digitaaliset arkistot – tai nimenomaan metadata-aineistot – tarjoavat uudelle datarikasteiselle⁵ tavalle kirjoittaa kirjallisuushistoriaa. Esimerkkitapauksemme tulee parhaillaan käynnissä olevasta, aikaisempiin kansainvälisiin hankkeisiin perustuvasta Emil Aaltosen Säätiön rahoittamasta *Tekstit liikkeessä: Naiskirjailijoiden tuotannon vastaanotto Suomessa ja Venäjällä 1840–2020* -hankkeesta, jossa digitaaliset aineistot, tietokannat ja kokoelmat ovat keskeisiä.⁶

Tekstit liikkeessä -hanke on ensimmäinen suomalais-venäläistä naiskirjallisuuden historiaa laajasti tarkasteleva projekti. Sen tavoitteena on “kartoittaa venäläisten naiskirjailijoiden vastaanottoa Suomessa ja vastaavasti Suomen naiskirjailijoiden vastaanottoa Venäjällä. Projekti tutkii kulttuurista, nimenomaan kirjallista, vastaanottoa (kääntämistä ja muuta tekstien ylijäräistä liikkumista) ja tuo siten tutkimuksen keskiöön erilaisia ja yhteiskunnan eri tasoilla toimivia kirjallisia ja sosiaalisia verkostoja.” Hanke linkittyy vahvasti naiskirjallisuuden tutkimuksen alalla kehitteillä oleviin digitaalisiin tutkimusmenetelmiin ja digitaalisen median tutkimukseen ja luo uudenlaista yhteistyötä ylijäräisen kirjallisuus- ja kulttuurihistorian tieteellisen perustutkimuksen, humanististen alojen sekä digitaalisen tutkimuksen kesken. (*Tekstit*

liikkeessä -sivusto.) Vastaanotto ymmärretään mahdollisimman laajasti kulttuurisena vaihtona, ja hankkeessa tarkastellaan niin käännöksiä, arvosteluja, adaptaatioita, erilaisia mainintoja kuin mainoksiakin – ja tietenkin näiden kaikkien tekijöitä, kulttuurisen kentän vaikuttajia ja toimijoita, jotka usein ovat jääneet piiloon tai taustalle. Myös tekijyys on määritelty laajimmalla mahdollisella tavalla: tekijöitä ovat kirjailijat, kääntäjät, toimittajat ja muut tekstejä tuottaneet henkilöt.

Hankkeessa kartoitettu vastaanotto ja muu tieto kulttuurisesta vaihdosta on kerätty hanketta varten suunniteltuun TeLi-tietokantaan. Kyseessä on nk. *custom build* -eli hankkeelle räätälöity tietokanta, jota ylläpidetään Tampereen yliopistossa. Siihen on koottu metadataa Suomen ja Venäjän naiskirjailijoista ja heidän kirjoittamistaan teoksista, teosten käännöksistä sekä muusta vastaanotosta Venäjällä ja Suomessa.⁷ Tietokanta sisältää tietoja teoksista aina 1830-luvulta 2020-luvulle, ja teksteistä on toimitettu täydet bibliografiset tiedot. (Ks. Westerlund & Mihailova, tulossa.) Metadatan yhtenäisyys on yksi haaste varsinkin sellaisissa hankkeissa, joissa – kuten tässäkin esimerkissä – metadata on koottu monesta eri lähteestä.

Tietokannan tärkeimpänä mallina on toiminut kansainvälinen virtuaalinen tutkimusympäristö NEWW VRE eli Women Writers -tietokanta, jota ylläpidetään Alankomaissa. Tietokanta rakennettiin Huygens instituutissa Haagissa/Amsterdamissa ja sen ylläpito on tänä vuonna siirtymässä Radboudin yliopistoon Nijmegeniin. Tämän virtuaalisen tutkimusympäristön sisällöstä ja konseptista vastaa DARIAH-EU:n *Women Writers in History* -työryhmä. Tietokanta on kaikille käyttäjille avoin ja sisältää tietoa teksteistä ja kirjallisista toimijoista 505 eri maasta. Tällä hetkellä kehitteillä on tietokannan kolmas, paranneltu versio. Tietokanta sisältää reilut 8000 kirjailijaa Sappofasta alkaen. Ajallinen takaraja on ensimmäisen maailmansodan paikkeilla. Tietokannan painopiste on nimenomaan naisten välisten kirjallisten verkostojen esiintuomisessa; kuka käänsi kenenkin teoksen, kuka oli kirjeenvaihdossa kenen kanssa ja kuka kommentoi ketäkin. Näin esiintuodut verkostot kulkevat läpi historian, vuosikymmenten ja lukuisien eri maiden. (Westerlund & Mihailova, tulossa.) Tarkoituksena on linkittää tulevaisuudessa tietokannat keskenään, jotta Suomen ja Venäjän välillä tapahtuva kulttuurivaihto asettuisi laajempaan kansainväliseen kontekstiin.

TeLi-tietokantaan on kerätty metadataa monesta eri lähteestä: vanhoista metadatan-arkistoista kuten kansallisbibliografioista (Suomessa Fennicasta), käännösbibliografioista, kirjastoluetteloista ja arkistoista (Suomessa Finnasta), ja ennen kaikkea lehdistöstä. Lehtihaut on tehty Kansalliskirjaston digitoiduissa sanomalehdissä erilaisilla hakusanoilla ja -tavoilla: on haettu erilaisilla sanayhdistelmillä ja kirjailijoiden nimillä sekä luettu läpi lehtien vuosikertoja manuaalisesti. Metadatan yhtenäisyys on aina ongelma ja hankkeemme tapauksessa tämä ongelma korostuu. Lähteet ovat olleet sekä digitoituja että

analogisia; uutta aikaa on kartoitettu myös syntyjään digitaalisten aineistojen, kuten blogien ja kirjallisuussivustojen, avulla. Koska projektin aikana tuli ensin pandemia ja sitten Venäjän hyökkäyssota Ukrainaan, tutkimuksen on ollut pakko keskittyä Venäjä-Suomi -suuntaan, eli Venäjän kirjailijoiden vastaanottoa Suomessa on ollut mahdollista tutkia paljon perinpohjaisemmin kuin päinvastoin. Samoista syistä datan keräämisestä ja tietokantaan viedystä datasta on tullut paljon digipainotteisempaa (eli valtaosa tiedoista on kerätty jo valmiiksi digitaalisista lähteistä) kuin mitä aiemmin suunniteltiin: pandemian aikana ei päästy edes Suomen arkistoihin, ja Venäjän matkoja on ollut mahdollista tehdä vain muutama projektin alussa.

TeLi-tietokanta on siis metadatan digitaalinen säilytyspaikka (*digital repository*), jota voi kokonaisuutena pitää tietyn tyyppisenä arkistona ja kulttuuriperinnön osana, metadata-arkistona (ks. Enqvist & Pikkanen, tulossa).⁸ TeLi-tietokannan tekeminen ja siinä työskenteleminen on saanut projektin tutkijat kysymään uusia tutkittavaan aineistoon liittyviä sekä yleisemmin kirjallisuushistoriallisia tutkimuskysymyksiä ja kehittämään hankkeen teoreettista ja metodologista taustaa. Hankkeessa on lisäksi jouduttu pohtimaan, kuinka löytää kirjallisia toimijoita, jotka eivät koskaan ole olleet korostetusti esillä, kuinka kaivautua nimimerkkien taakse, kuinka jäljittää käännosten ja adaptaatioiden ketjuja. Erilaiset hakumahdollisuudet paljastavat uusia kirjailijoiden sekä teosten välisiä suhteita ja mahdollisuus laskennalliseen tutkimukseen korjaa aikaisempia Venäjän ja Suomen kulttuurisen vaihdon ja kirjallisuuden vastaanoton tutkimuksen tuloksia, joista nousee esiin vain muutaman naiskirjailijan läsnäolo kirjallisella kentällä.

Aineiston laajuuden takia metadatan keräämisessä on keskitytty tiettyihin, hankkeessa ”aikaikkunoiksi” kutsuttuihin, ajanjaksoihin. Lisäksi on tehty erilaisia tapaustutkimuksia, joissa on hyödynnetty sekä kirjallisuudentutkijoille tyypillistä ”lähilukemista” (*close reading*) sanan laajassa mielessä eli perinteistä kirjallista analyysia ja tulkintaa. Laajojen aineistojen yhteydessä on ainakin osittain päästy hyödyntämään Franco Morettin (2013) kehittämää ”kauko- tai etälukemisen” metodia (*distant reading*). ”Kaukolukeminen” tarkoittaa laajojen aineistojen tarkastelua laskennallisin menetelmin. Tällä tavalla on mahdollista kartoittaa esim. jonkun kirjallisen lajin levikkiä jonakin tiettyinä aikakautena (Moretti 2005) ja saada paljon luotettavampaa tietoa kuin manuaalisten laskelmien tai vanhojen tilastotietojen avulla. *Tekstit liikkeessä* -hankkeessa on hyödynnetty sekä kauko- että lähilukemista, sekä usein jotain niiden välimuotoa, jota voi kutsua mesoanalyysiksi (*mesoanalysis*). Kyseessä on Paul Saint-Amourin tunnetuksi tekemä käsite, jolla hän tarkoittaa eräänlaista keskitietä lähi- ja kaukoluvun tai mikro- ja makrotason välillä, näiden menetelmien kompromissia.⁹ Saint-Amour ei ole ainut keskitien metodologian peräänkuuluttaja. Moni digitaalisia menetelmiä käyttävä kirjallisuudentutkija on peräänkuuluttanut juuri tällaista keskitien metodologiaa, joka sallii työskentelemisen laajojen

digitoitujen aineistojen parissa ja samalla korostaa kirjallisuudentutkimuksen asiantuntemusta sekä laajasti ymmärrettyä ja kontekstualisoivaa lähilukua (ks. esim. Bode 2018). Kontekstoivat, historiallistavat kirjallisuudentutkijat ovat aina tehneet tutkimusta, jossa tuodaan yhteen yksittäisten teosten lähilukemista, tapaustutkimuksia ja kirjallisen kentän laajempia kehyksiä ja ilmiöitä. Kontekstoivan tulkittamisen ja hermeneuttisen kehän perusedellytyksenä on ollut lukemisen, suhteuttamisen ja uudelleen suhteuttamisen vuorovaikutus.

Digitaalisista sekä analogisista aineistoista ja olemassa olevista metadata-aineistoista rakennetut digitaaliset tietokannat eli uudet metatieto-arkistot mahdollistavat kriittisen katseen vanhoihin metadata-aineistoihin sekä arkistoihin yleensä, niiden vinoutumiin ja "hiljaisuuksiin" (ks. Enqvist & Pikkanen tulossa). Digitaalisia sekä perinteisiä arkistoja ja kulttuuriperinnön instituutioita on jo kauan arvosteltu (muun muassa) "valkoisuudesta" (esim. McPherson 2012). Arvostelun kohteena on ollut myös se, miten kirjastojen, arkistojen ja muisti-instituutioiden digitointistrategiat ovat marginalisoineet ei-valkoisia ja vähemmistöjen aineistoja sekä monia muita aineistoja, jotka eivät kuulu valtavirtaan esimerkiksi tekijän tai kohteen sukupuolen perusteella (esim. Sanz 2013; Wernimont 2013; Bergenmar & Golub 2020; Parente-Čapková 2021). Tässä kirjallisuudentutkimuksen vastaanottoa ja kulttuurista vaihtoa sisällyttävää metadata-arkistoa koskevat pohdinnat liittyvät myös laajempiin, digitaalisia ihmistieteitä sekä arkistollista kulttuuriperintöä koskeviin keskusteluihin (ks. Hupaniittu & Peltonen 2021; Enqvist & Pikkanen, tulossa).

Kun lähdetään liikkeelle hyvin laajasta vastaanoton määritelmästä, jää kartoitettu aineisto väistämättä vain jäävuoren huipuksi. Jopa joidenkin rajattujen ajanjaksojen tai lähteiden systemaattisempi läpikäyminen jättää aukkoja, puhumatta siitä, että itse aineistot ja vanhempi metadata ovat aukkoisia eikä tärkeää aineistoa välttämättä ole kovin paljon säilynyt kuvailtavaksi. Siitä huolimatta tutkimuksen tulokset ja uusi metadata voivat näyttää suuntaa ja inspiroida kysymään uusia kysymyksiä. TeLi-hankkeen tähänastiset tulokset vahvistavat jo olemassa olevan tiedon Suomen ja Venäjän välisestä kulttuurisesta vaihdosta siinä mielessä, että miesten kirjoittamien teosten vastaanottoa (samoin kuin mieskirjailijoiden vastaanottoa eli niitä koskevia mainintoja tai jopa kokonaisia teoksia) on paljon enemmän kuin naisten kirjoittamien teosten ja itse naiskirjailijoiden vastaanottoa. Hankkeessa on kuitenkin löydetty hyvin paljon todistusaineistoa siitä, että venäläisiä naiskirjailijoita on otettu vastaan ja tunnettu Suomessa sekä Venäjällä ja että heidän ja heidän tekstiensä vastaanotto tuo uutta tietoa ja uutta näkökulmaa Suomen ja Venäjän välisiin suhteisiin sekä varsinkin siihen, miten Venäjän kuva on Suomessa rakennettu 1800-luvun puolivälistä lähtien: kansallisesta heräämisestä ja sortovuosista Venäjän vallankumousten kautta suomettumiseen, perestroikaan, jälkineuvostoliittolaiseen aikaan ja nykypäivään.

Uutta todistusaineistoa sisältävä TeLi-metadata-arkisto on toiminut tärkeänä alustana ja motivoijana pohdittaessa laajasti ymmärrettyyn kirjallisen vastaanottoon, kulttuurisen vaihtoon sekä kirjallisuushistoriaan liittyviä teoreettisia ja metodologisia lähtökohtia. Datan vieminen tietokantaan edellyttää jo itsessään tulkintaa: siitä on keskusteltu kansainvälisen Women Writers -tietokannan kehittämisen yhteydessä jo yli kymmenen vuoden ajan. Erilaisen aineistojen läpikäynnin tulokset on saatava tietokantaan kvantitatiiviseen muotoon, vaikka moni asia ei tunnukaan taipuvan tietokannassa oleviin kategorioihin, joista sitten jatkuvasti keskustellaankin. TeLi:n kaltaiseen hankkeen yhteydessä ongelmia tulee heti alussa: mikä tai kuka on naiskirjailija? Feministisinä tutkijoina hankkeen jäsenet tietenkään vastaustavat binaarista sukupuolta koskevaa ajattelua sekä essentialisoivaa jakoa eri sukupuoliin. Toisaalta on selvää, että tarvitaan tietoa esimerkiksi siitä, kuinka moni ”naiskirjailija” -kategoriassa toiminut kirjailija näkyy kirjallisella kentällä 1800-luvun ”aikaikkunoissa”. Projekti on lähtökohtaisesti ylijäräinen myös transnationaalissa mielessä, mutta lähtee kuitenkin liikkeelle ”Suomi” ja ”Venäjä” -käsitteistä ja pakottaa tutkijoita pohtimaan uudestaan ja uudestaan näiden käsitteiden määrittelyä eri aikoina – onko kyseessä esimerkiksi Suomi suurruutinaskuntana tai Venäjä osana Neuvostoliittoa. Entä mikä rooli kielellä on suhteessa kansallisuuteen? Kategorioiden muodostamisen yritys herättää monenlaisia kysymyksiä rajatapauksista ja rajaamisen vaikeudesta, karjalaisesta ja saamelaisesta kirjallisuudesta, kirjailijoista ja kulttuurisista toimijoista Venäjältä ja Neuvostoliitosta maanpakoon lähteneisiin kirjailijoihin ja heidän – usein eri kielisiin – teoksiin. Jo hankkeen alkuvaiheessa huomattiin seikka, joka on jatkuvasti noussut esiin myös laajemmassa yhteiskunnallisessa ja kulttuurisessa keskustelussa viimeisen vuoden aikana: moni itsensä ukrainalaiseksi määrittävä kirjailija on otettu vastaan tai kuvailtu venäläiseksi. Näitä samoja kysymyksiä on pohdittu – ja pohditaan edelleen – myös Women Writers -tietokannan yhteydessä laajemmassa eurooppalaisessa kontekstissa.

Arkistoista ja erilaisista metadata-aineistoista saadun tiedon paikantaminen muun muassa sukupuolen ja (yli)kansallisuuden kentällä (ks. Aarnio tulossa) sekä tietojen taivuttaminen ja ”pakottaminen” yksittäisiin kategorioihin vaatii usein yksinkertaistamista ja eri luokkien välillä luovimista. Laskennallisten syiden lisäksi tämä on tärkeää myös visualisointimahdollisuuksia ajatellen. Visualisoinnit auttavat ”havaitsemaan aineistosta trendejä, joita datasta oli aiemmin ollut vaikeampi havaita” (Aarnio tulossa). Visualisointeja tehtiin *Tekstit liikkeessä* -tutkimuksen aikana ja julkaisuja varten niihin laadittiin niitä kommentoivat ja selittävät tekstit. Hankkeen visualisointien tekijän Pyry Aarnion (tulossa) mukaan koko prosessia aineiston läpikäynnistä ja tietojen viemisestä TeLi:n metadata-arkistoon eli tietokantaan voi kuvata eräänlaiseksi sykliksi: ”tekstistä numeroiksi, numeroista kuvaksi ja kuvasta sanoiksi”. Kaikki

nämä yksittäiset askeleet edellyttävät jatkuvaa analyysia ja tulkintaa sekä hankkeen tuottamien metatietojen ja erilaisten metatasojen reflektointia.

Jenny Bergenmar ja Katariina Leppänen (2017) ovat tuoneet esiin, miten angloamerikkalaisten tutkijoiden julistama transnationaalinen ote sivuuttaa sekä sukupuolen että maailman kirjallisuuksien monikielisyys. TeLi-hanke ja sen tuottama metadata-arkisto haluaa kontribuoida marginaalissa olleiden ilmiöiden perintöyttämiseen tai perintöytymiseen (*heritagization*), mikä tarkoittaa “toisiinsa limittyvien perintöprosessien ja perintödiskurssien muodostamaa kokonaisuutta ja dynaamista kehystä”, “joka mahdollistaa ja määrittää tiettyjen menneisyyden jälkien, ilmiöiden ja ilmentymien kehkeytymistä perinnöksi” (Tieteen termipankki: *heritagization*; Enqvist & Pikkanen, tulossa). Toisaalta tuodaan esiin hankkeen tuloksia avoimen metadata-arkiston, visualisointien sekä julkaisujen ja esitelmien, hyvin usein erilaisten mesoanalyysien muodossa. Toisaalta korostetaan sekä teknisiä että teoreettis-metodologisia haasteita, jotka digitaalisen metadata-arkiston toteuttamisprosessi on tuonut esiin. Lisäksi tutkijan täytyy tottua siihen, että arkistojen tavoin (vrt. Hyttinen & Kivilaakso 2010, 8) myös metadata-arkistot ovat aina keskeneräisiä.

Viitteet

- 1 Arkistot ja kirjastot, jotka syntyivät “laittomasti” totalitaristisissa järjestelmissä, keskittyivät yleensä valtion terrorin dokumentoimiseen, kuten tekstissä mainittu Memorial-ihmisoikeusjärjestö. Myös Prahassa sijaitseva, alun perin yksityisissä tiloissa ylläpidetyt *Libri prohibiti* -arkisto ja -kirjasto, sisältävät paljon “kiellettyä kauno-kirjallisuutta” ja siihen liittyvää materiaalia. Sama koskee erilaisia vähemmistöjen perintöä ja kirjallisuutta koskevia arkistoja.
- 2 Nämä konkreettiset tapaukset liittyvät meidän, tämän katsauksen kirjoittajien konkreettisiin kokemuksiin. Olemme monesti istuneet arkistoissa ja tutkineet naiskirjailijoiden jäämistöjä.
- 3 Metadata tarkoittaa tietovarannosta tai sisältöyksiköstä muodostettua tietoa, joka on kuvailevaa ja määrittävää.
- 4 Konsortio *Digital History for Literature in Finland* (Suomen Akatemia, 2022–2026; konsortion johtaja Leo Lahti, kirjallisuushankkeen PI Kati Launis, Kansalliskirjaston hankkeen PI Osma Suominen) <https://www.kansalliskirjasto.fi/fi/projektit/digitaaliset-menetelmat-kotimaisen-kirjallisuushistorian-uudistajana> (2.3.2023)
- 5 Datarikasteisella kirjallisuushistorialla (engl. *data rich literary history*) tarkoitetaan laskennallisia menetelmiä käyttävää kirjallisuushistoriallista tutkimusta. On kuitenkin tärkeää pyrkiä määrällisen ja laadullisen

analyysin yhdistämiseen (ks. Parente-Čapková 2017; Bode 2018; Karkulehto et al. 2022).

6 Hanke on toiminut Turun yliopistossa ja Tampereen yliopistossa. Hankkeessa on työskennellyt myös muiden Suomen yliopistojen (Helsingin yliopisto, Itä-Suomen yliopisto) tutkijoita ja harjoittelijoita. Projektin johtajana on toiminut Viola Parente-Čapková ja tutkijatiimiin ovat kuuluneet Arja Rosenholm, Kati Launis, Jasmine Westerlund, Marja Sorvari, Saara Ratilainen, Irina Savkina, Tintti Klapuri, Jenniliisa Salminen ja Natalia Mihailova. TeLi-tietokannan eli metadata-arkiston rakensi Juho Härme, tiedot on siihen vienyt Natalia Mihailova ja visualisoinnit on tehnyt Pyry Aarnio. Jasmine Westerlund on toiminut NEWW Women Writers ja TeLi-tietokannan välillä.

7 TeLi-tietokanta sisältää tällä hetkellä yli 1000 nimeä suomalaisia ja venäläisiä naiskirjailijoita, kääntäjiä, tutkijoita, toimittajia ja muita kulttuurialan tekijöitä. Tietokannassa on yli 6000 tietuetta. (Ks. <https://puolukka.rd.tuni.fi/tekstiliikkeessa/#/>) (2.3.2023).

8 Digitaalista kuratointia ja tutkimusaineistojen hallintaa koskevissa asiakirjoissa puhutaan data-arkistoista, joista jotkut ovat metadata-arkistoja, ks. esim. <https://libguides.tuni.fi/tutkimusaineistojen-hallinta/metadata>.

9 Morettin kollega Matthew Jockers (2013) kehitti kaukolukemista yleisemmän makroanalyysi-käsitteen (*macroanalysis*) vastakohtana lähilukemista vastaavalle mikroanalyysille. Ks. lisää Parente-Čapková & Launis 2021; Parente-Čapková 2022.

Lähteet

- Aarnio, Pyry tulossa. Sanoista numeroiksi, numeroista kuvaksi, kuvasta sanoiksi. *Ajan kohina*. Baca, Murtha 1998/2016. *Introduction to Metadata*: Third Edition. Getty Research Institute 15 June, 2016.
- Bergenmar, Jenny & Katariina Leppänen 2017. Gender and Vernaculars in Digital Humanities and World Literature. *NORA – Nordic Journal of Feminist and Gender Research* 25(4), 232–246. DOI: 10.1080/08038740.2017.1378256.
- Bergenmar, Jenny & Koraljka Golub 2020. Subject Indexing: The Challenge of LGBTQI Literature. *DHN*, 203–210.
- Bode, Katherine 2018. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press.
- Dahlgren, Anna, Karin Hansson, Ramón Reichert & Amanda Wasielewski 2021. Introduction: The Politics of Metadata. Eds Anna Dahlgren, Karin Hansson, Ramón Reichert, & Amanda Wasielewski. *The Politics of Metadata. Digital Culture and Society* 6(2), 5–16. DOI: 10.14361/dcs-2020-0202
- Digital Archives 2022. Some Definitions and Explanations. *Archive Discovery: A How-To Guide*. Newton Gresham Library: Sam Houston State University. October 19, 2022. <https://www.libpro.cz/en/archive/> (2.3.2023)
- Digital Arkivet 2023. <https://www.digitalarkivet.no/content/about-da> (2.3.2023)

- Enqvist, Johanna & Ilona Pikkanen 2023. Kirjeluettelot kulttuuriperintönä ja tutkimusaineistona: metadatan mahdollisuudet digitaalisen käänteän jälkeen. Julkaisematon käsikirjoitus.
- Foucault, Michel 2005: *Tiedon arkeologia (L'archéologie du savoir, 1969)*. Suom. Tapani Kilpeläinen. Tampere: Vastapaino.
- Hupaniittu, Outi & Ulla-Maija Peltonen 2021. Arkistot ja kulttuuriperintö: Johdanto. Toim. Outi Hupaniittu & Ulla-Maija Peltonen. *Arkistot ja kulttuuriperintö*. Tietolipas 268. Helsinki: SKS. DOI:10.21435/tl.268
- Hyttinen, Elsi & Katri Kivilaako. Johdanto. Teoksessa *Lukemattomat sivut. Kirjallisuuden arkistot käytössä*. Toim. Elsi Hyttinen & Katri Kivilaako. Helsinki: SKS, 7–18.
- Jockers, Matthew 2013. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.
- Karkulehto, Sanna, Veronika Laippala, Kati Launis, Jaana Märsynaho, Maija Saviniemi & Minna Sääsikihahti 2022. Selkosten Proust taipuu moneen – Iijoki-korpus ja digitaalisen tekstilouhinnan mahdollisuudet. Teoksessa *Kalle Päätaalo tutkijoiden silmin*. Toim. Maija Saviniemi. Helsinki: Suomalaisen Kirjallisuuden Seura, 170–189.
- Launis, Kati 2010: Vapaus, sisaruus, tasa-arvo. Työläisnaisliikkeen aktiivi Hilja Pärssinen. *Työväen-tutkimus vuosikirja 2010*, Päätoim. Anu-Hanna Anttila, 23–29.
- Libri prohibiti: <https://www.libpro.cz/en/archive/> (2.3.2023)
- Kansalliskirjaston digitaaliset aineistot: https://digi.kansalliskirjasto.fi/etusivu?set_language=fi (2.3.2023)
- McPherson, Tara 2015. Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation. Teoksessa *Debates in the Digital Humanities*. Ed. Matthew K. Gold. Minnesota Scholarship Online. DOI:10.5749/minnesota/9780816677948.003.0017.
- Moretti, Franco 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- 2013. *Distant Reading*. London: Verso.
- NEW Women Writers: Virtual Research environment: <https://womenwriters.resources.huygens.knaw.nl/> (2.3.2023)
- Parente-Čapková Viola & Kati Launis 2021. Kirjallisia kohtaamisia - venäläiset kirjat Turun kaupunginkirjaston vanhassa kokoelmassa. Teoksessa *Kieliä ja kohtaamisia Turun historiassa*. Toim. Leena Kolehmainen, Aino Liira & Kirsi-Maria Nummila. Helsinki: SKS, 239-261. DOI: <https://doi.org/10.21435/skst.1475>
- Parente-Čapková Viola 2021. Digitaaliset ihmistieteet kirjallisuudentutkimuksessa. *Avain – Kirjallisuudentutkimuksen aikakauslehti* 18(4), 86-95. DOI:10.30665/av.111973
- 2022. Matilde Serao in Finland. Teoksessa *Matilde Serao Beyond National Fame: Reframing the Neapolitan Writer's Reception and Network*. Eds Ursula Fanning, Katherine Mitchell & Gabriella Romani. Series Women and Gender in Italy 1500–1900. Paris: Classique Garnier, 131–157.
- Roine & Piippo 2021: Roine, Hanna-Riikka & Laura Piippo 2021. Kirjallisuus ja sen tutkimus digitalisissa ympäristöissä. *Kirjallisuudentutkimuksen aikakauslehti Avain* 18(2), 6–27. DOI: 10.30665/av.102654.
- Sanz, Amelia 2013. Digital Humanities or Hypercolonial Studies? *RICT, Responsible Innovation, Ethical Issues* 2013. <https://core.ac.uk/download/pdf/163094852.pdf> (2.3.2023)
- Tekstit liikkeessä. <https://tekstitliikkeessa.com/tekstit-liikkeessa/> (2.3.2023)
- Teli-tietokanta. <https://puolukka.rd.tuni.fi/tekstitliikkeessa/#/> (2.3.2023)
- Tieteen termipankki: heritagization: <https://tieteentermipankki.fi/wiki/Nimitys:heritagization> (2.3.2023)

- Tolonen, Mikko 2017: Digitaalinen humanismi ja digitalia. Teoksessa *Digitaalinen tieto haltuun. Ratkaisuja digitaalisten aineistojen hallintaan ja käyttöön*. Toim. Miia Kosonen. Mikkeli: Kaakkois-suomen ammattikorkeakoulu. <http://urn.fi/URN:ISBN:978-952-344-020-3> (2.3.2023)
- Wernimont, Jennifer 2013. Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives. *Digital Humanities Quarterly* 7(1). <http://www.digitalhumanities.org/dhq/vol/7/1/000156/000156.html> (2.3.2023)
- Westerlund, Jasmine & Natalia Mihailova, tulossa. Venäläiset naiskirjailijat Suomessa: tietokirjasta tietokantaan. Julkaisematon käsikirjoitus.
- Women Writers in History Working Group: DARIAH-EU. <https://www.dariah.eu/activities/working-groups/women-writers-in-history/> (2.3.2023)