

➤ Kirjallisuudentutkijan näkökulma
syntysähköisten käsikirjoitusten
arkistointiin^{1,2}

—

Veijo Pulkkinen

V
—

altaosa nykykirjailijoista kirjoittaa tietokoneilla, etsii tietoa internetistä, viestii sähköpostin ja erilaisten pikaviestipalveluiden avulla sekä julkaisee tekstejä blogeissa ja erilaisissa verkko-yhteisöpalveluissa, mikä luo haasteita niin kirjallisuudentutkimukselle kuin kirjailijoiden henkilöarkistoja tallentaville arkistoille. Yksi keskeinen huoli on se, että "[u]sujen medioiden myötä olemme tilanteessa, jossa perinteisiä kirjeaineistoja, teosten käsikirjoituksia korjausmerkintöineen tai muita kirjailijan työtä ja kirjoittamista valaisevia arkistodokumentteja ei synny", kuten asia ilmaistiin vuonna 2012–2015 toteutetussa *Puhuva kirjailija – kertova arkisto* -hankkeessa, jossa haastateltiin kirjailijoita koskien heidän työtään 1900-luvun lopulla ja 2000-luvun alun Suomessa (Puhuva kirjailija).

Erityisen huolissaan tilanteesta on oltu geneettiseksi kritiikiksi kutsutun kirjallisuudentutkimuksen suuntauksen piirissä. Sen tutkimuskohteena on kirjallisen teoksen syntyprosessi, jota tutkitaan nimenomaan kirjoitusprosessista säilyneiden arkistodokumenttien ja niissä näkyvien erilaisten muutosmerkintöjen, kuten poistojen, lisäysten ja korvausten pohjalta. Kirjoitusprosessin digitalisoituminen on uhannut tehdä uudemman kirjallisuuden geneettisestä tutkimuksesta mahdotonta. (Biasi 2010, 171; Grésillon 2001, 13–14; Mathijsen 2009, 234; Vauthier 2016, 164–166.)

Tilanne on kuitenkin muuttunut sen jälkeen, kun Matthew Kirschenbaum esitteli tutkimuksessaan *Mechanisms: New Media and the Forensic Imagination* (2008), miten digitaalisen forensiikan eli rikostutkimuksen menetelmiä ja välineitä voidaan soveltaa tekstuaalitieteissä. Forensisilla tietokoneohjelmilla voidaan tallennusmedioilta, kuten levykkeiltä ja kiintolevyiltä, palauttaa poistettuja tekstitiedostoja ja tekstifragmentteja ja niihin liittyvää metadataa sekä löytää yksittäisistä tekstitiedostoista geneettisiä kerrostumia, jotka eivät näy tavanomaisissa tekstinkäsittelyohjelmissa.

Kansainvälisellä tasolla digitaalisen forensiikan menetelmiä ja välineitä on viimeisen viidentoista vuoden aikana kehitelty ja sovellettu etenkin syntysähköisten aineistojen arkistoinnissa, mutta Suomessa tähän ollaan vasta heräämässä. Digitaalisen forensiikan soveltaminen syntysähköisten käsikirjoitusten geneettiseen tutkimukseen on toistaiseksi ollut paljon vähäisempää. Edelläkävijöitä tällä saralla ovat etenkin saksalaisten runoilijoiden Thomas Klingin ja Michael Speierin syntysähköisiä käsikirjoituksia tutkinut Thorsten Ries (2010; 2017; 2018; 2023) ja Aurèle Crassonin johtaman monitieteisen tutkimusryhmän hanke *Derrida Hexadecimal*, jossa tutkitaan filosofi Jacques Derridan digitaalisia kirjoitusstrategioita (Crasson 2019; Crasson, Lebrave & Pedrazzi 2019).

Tässä artikkelissa käsitellään syntysähköisten aineistojen arkistointiin liittyviä yleisempiä kysymyksiä, jotka ovat nousseet esille tähänastisessa tutkimuksessani. Syntysähköisten käsikirjoitusten ja kirjoitusprosessien geneettinen tutkimus asettaa aineistolle tiettyjä vaatimuksia, jotka kytkeytyvät niin aineiston vastaanottoon, hallintaan kuin asiakaskäyttöön saattamiseen. Tämän

artikkelin tarkoituksena on käydä läpi näitä seikkoja tutkijan näkökulmasta ja tarjota virikkeitä syntysähköisten aineistojen arkistointikäytäntöjen kehittämiseen sekä herättää yleisempää keskustelua syntysähköisten aineistojen luonteesta ja merkityksestä sekä niiden arkistoinnista ja tutkimuksesta.

Syntysähköiset käsikirjoitukset geneettisen kritiikin tutkimusaineistona

Geneettisen kritiikin tutkimuskohteena ovat luomisprosessit, kuten kirjallisen teoksen kirjoitusprosessi. Merkittävintä geneettisessä kritiikissä on sen prosuaaliseen tekstikäsitukseen perustuva näkökulma kirjallisuuteen. Toisin kuin perinteinen kirjallisuudentutkimus, joka keskittyy valmiiseen ja julkaistuu tekstiin, tutkii geneettinen kritiikki sitä prosessia, joka johti tuon tekstin syntymiseen. Tästä näkökulmasta julkaistu teksti näyttyy vain yhtenä vaiheena monien joukossa eikä se ole välttämättä edes viimeinen, sillä monet kirjailijat revisioivat teoksiaan, kun siihen tarjoutuu mahdollisuus uusien laitojen julkaisemisen myötä. Geneettisen kritiikin näkökulmasta ensisijaisen tutkimusaineiston muodostavat kirjailijan tekemät muistiinpanot, suunnitelmat, luonnokset ja erilaiset teoksen tekstin käsikirjoitusversiot. Tutkimussuunta onkin olennaisella tavalla riippuvainen arkistoista, mutta toisaalta se on myös tuonut arkistoaineistot uudella tapaa osaksi kirjallisuudentutkimusta. Vaikka käsikirjoitusaineisto voi rikastuttaa julkaistusta tekstistä tehtyjä tulkintoja, tämä ei ole geneettisen kritiikin varsinainen päämäärä. Pikemminkin tarkoituksena on horjuttaa käsitystä yhdestä valmiista teoksen tekstistä. (Biasi 2004, 43; Grésillon 1994, 7; Hay 2002, 103; Lernout 2002, 58.) Hyvä esimerkki on Joel Lehtonen, joka hioi lakkaamatta tekstiään eikä ollut siihen tyytyväinen sen ilmestymisenkään jälkeen (Pulkinen 2019). Muun muassa *Mataleenan* (1905) hän kirjoitti suureksi osaksi uusiksi *Kootuissa teoksissaan* (1931–1935).

Syntysähköisten käsikirjoitusten geneettisessä tutkimuksessa keskeisin kiinnostuksen aihe ovat tekstiin tehtyjen muutosten jäljet, joita voidaan löytää digitaalisen forensiikan keinoin. Tutkimuksen tavoitteena voi olla esimerkiksi selvittää julkaisemattomien syntysähköisten käsikirjoitusten ja fragmenttien suhdetta kirjailijan julkaistuu tuotantoon. Tämän makrogeneettisen tutkimuksen lisäksi voidaan tehdä mikrogeneettistä tutkimusta, jossa analysoidaan yksittäisten syntysähköisten käsikirjoitusten syntyprosesseja. Lisäksi voidaan tutkia digitaalisen kirjoitusvälineen vaikutusta kirjailijan kirjoittamiseen ja ajatteluun. Esimerkiksi *Derrida Hexadecimal* -projektissa on löydetty yhteyksiä Derridan tietokoneen käytön ja hänen filosofisissa kirjoituksissaan esiintyvän arkiston käsitteen välillä. Derridan tapa nimetä tiedostoja muistuttaa myös hänen teksteissään esiintyvää kielellistä leikittelyä (Vrt. Crasson 2019; Crasson, Lebrave & Pedrazzi 2019, 192–196.)

Digitaalisen kirjoitusprosessin tapauksessa geneettisen kritiikin tutkimusaineisto koostuu paitsi syntysähköisiä tekstejä ja tekstiversioita sisältävistä näkyvistä tekstitiedostoista, joita kirjailija on tallentanut tallennusmedioille,

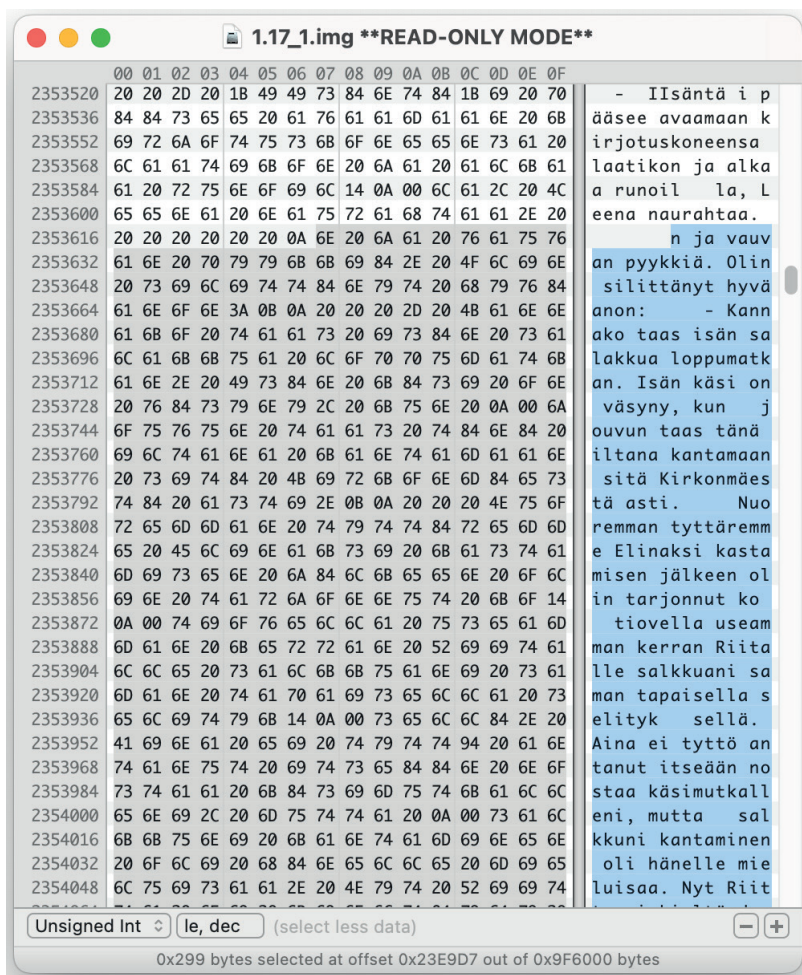
kuten levykkeille, cd-levyille, muistitikuille, tietokoneen, tabletin ja älypuheli-
men kiintolevyille, ulkoisille kiintolevyille ja pilvipalveluille, myös poistetuista
tai muutoin näkymättömissä olevista tekstitiedostoista ja -fragmenteista, joita
voidaan palauttaa digitaalisen forensiikan menetelmin.

Keskityn tässä artikkelissa varsinaisen kirjoitusprosessin tuloksena synty-
neisiin niin kutsuttuihin endogeneettisiin aineistoihin. Sivuhuomioina totean
kuitenkin, että näiden lisäksi kirjailijan käyttämiltä digitaalisilta kirjoitusväli-
neiltä voi löytyä eksogeneettisiä jälkiä. Eksogeeniksellä tarkoitetaan tausta-
tutkimuksen ja intertekstuaalisuuden kaltaista vuorovaikutusta ulkoisten
aineistojen kanssa. (Biasi 1996, 45–46 Debray-Genette 1977, 24; Van Hulle 2019,
8–9.) Syntysähköisen kirjoitusprosessin kontekstissa erityisesti verkkoselaimen
selainhistoria voi tarjota merkittävää tietoa hänen hyödyntämistään ulko-
puolisista lähteistä. Kirjailijan selaimen käyttöä on tutkittu syntysähköisten
kirjoitusprosessien geneettisessä tutkimuksessa, joka perustuu kirjailijan tieto-
koneelle asennettavalle näppäilytallentimelle. Tällaiset tutkimukset tehdään
yhteistyössä kirjailijan kanssa, jolloin myös sovitaan tietokoneen käytön seura-
misesta. (Buschenhenke 2016, 2, 8; Van Hulle 2022, 178.) Arkistoidun aineiston
kohdalla selainhistorian tutkiminen ja poistettujen tiedostojen palauttamisen
kaltaisten forensisten tutkimusmenetelmien käyttö nostaa esille yksityisyyttä
koskevia eettisiä ja juridisia kysymyksiä, joita käsittelem tarkemmin seuraa-
vassa luvussa.

Tiedostojen kaivertamiseksi (engl. *file carving*) kutsuttu poistettujen tie-
dostojen palauttamisen menetelmä perustuu siihen, ettei tiedoston data tosi-
asiassa poistu tallennusmedialta, kun se deletoidaan ja vielä tyhjennetään niin
kutsutusta roskakorikansioista, vaan ainoastaan siihen liittyvä metadata häviää
tiedostojärjestelmästä. Tiedostojen data on usein jakaantuneena eri paikoissa
fyysistä tallennusmediaa ja tiedostojärjestelmä pitää yllä tietoa tiedostofrag-
menttien sijainneista. Kun tiedosto poistetaan, tämä sijaintitieto katoaa osaksi
tai kokonaan ja tiedoston datan sijainneista tulee niin kutsuttua allokoima-
tonta eli vapaata tilaa, jonka päälle voidaan kirjoittaa uutta dataa. Tiedostoja
tai niiden rippettä voidaan yrittää koota uudelleen käsin heksaeditorin avulla
tai tiedostokaivertimiksi tai -karvereiksi (engl. *file carver*) kutsuttujen ohjelmien
avulla analysoimalla mahdollisia sijaintitietojen rippettä ja raakadataa, joka
sijaitsee allokoimattomassa tilassa. Poistettujen tiedostojen dataa voi löytyä
myös niin kutsutusta *slack*-tilasta. Sillä tarkoitetaan tiedoston tallentamisesta
ylijäänyttä ”väljää” tilaa, jonka käyttöjärjestelmä on varannut tiedostolle, mutta
jota tiedosto ei tarvitse (engl. *slack space*). (Pal & Memon 2009, 60–61; Ries 2018,
392–393.)

Esimerkiksi Kalle Päätalon tietokoneen kiintolevyllä oleva tiedosto
TXTKAP.056 sisältää *Pölvökanto Iijoen törmässä* (1998) -romaanin toisen luvun
käsi­kirjoituksen (Päätalo 1997b). Tiedoston looginen koko on 102871 tavua. Kiin-
tolevy on alustettu FAT16-tiedostojärjestelmällä, joka tallentaa tiedostot 2048
tavun lohkoina. TXTKAP.056:n tallentamiseen tarvitaan näin 51 lohkoa, joiden
yhteenlaskettu koko on 104448 tavua. Tiedoston tallentamiseen tarvittava fyysi-

nen koko on siis suurempi kuin sen looginen koko, jolloin käyttämättä jää 1577 tavua. Tämä ylijäämä muodostaa tiedoston slackin. (Carrier 2005, 139–140; Sammes & Jenkinson 2007, 94–94.) Tarkastelemalla heksaeditorin avulla kiintolevyn levykuvaa bittitasolla saadaan selville, että tiedoston perässä oleva ylijäämä sisältää 193 sanaa pitkän tekstifragmentin (Päätaalo 1997a), joka on suurimmalta osin³ todennäköisesti peräisin romaanin kolmannen luvun varhaisemmasta versiosta (vrt. Päätaalo 1998b, 91) (Kuva 1). TXTKAP.056 on tallennettu tuon varhaisen version päälle, mutta osa siitä on jäänyt jäljelle tiedoston slack-tilaan. Vertaamalla fragmenttia kolmannen luvun käsikirjoitukseen (Päätaalo 1998a) ja julkaistuun tekstiin saadaan tietoa esimerkiksi siitä, miten romaanin tekstuaalisuuden eli tekstimassan tuottamisen aikana tehdyt muutokset eroavat niistä muutoksista, joita on tehty tekstin revisioiden aikana eli silloin, kun tekstiä on kirjoitettu uudestaan teoksen kokonaisuuden ollessa selkeämmin tiedossa.



Kuva 1. Osa *Pöyhökanto lijojen törmässä* (1998) -romaanin toisen luvun käsikirjoituksen tiedostoa Hex Fiend -heksaeditorilla katsottuna. Maalattuna on tiedoston slack-tilassa olevaa dataa.

Näkyvien ja poistettujen tekstitiedostojen ja -fragmenttien lisäksi syntysähköisten käsikirjoitusten geneettisessä tutkimuksessa on otettava huomioon tiedostoihin liittyvä metadata esimerkiksi niiden luonti- ja muutosajankohdasta sekä väliaikaisten tiedostojen (engl. *temporary files*) ja tiedostojen varmuuskopioiden (engl. *backup files*) kaltainen data, jota syntyy automaattisesti ja tallentuu tietojärjestelmän eri paikkoihin, kun kirjailija käyttää tietokonetta. Tämä tuo esille syntysähköisen geneettisen tutkimusaineiston riippuvuuden siitä tietojärjestelmän kontekstista (laitteistokokonaisuus, käyttöjärjestelmä, ohjelmisto, käyttötiedot), jossa se on luotu. Tarvitsemme sitä, jotta tiedämme, mistä etsiä tutkimuksen kannalta relevanttia dataa, tiedostoja ja tiedosto-fragmentteja, sekä ymmärtääksemme, miten ja miksi ne ovat päätyneet sinne, missä ne ovat. (Ries 2018, 394.)

Käytännössä paras lähtökohta syntysähköisten käsikirjoitusten geneettiselle tutkimukselle on, jos saatavilla on kirjailijan tietokoneen kiintolevystä ja mahdollisista ulkoisista tallennusmedioista, kuten levykkeistä ja ulkoisista kiintolevystä, tehdyt forensiset levykuvakopiot. Forensinen levykuva (engl. *forensic image, forensic copy*) on täsmällinen bittikohtainen kopio koko tallennusmedian sisällöstä ja rakenteesta mukaan lukien sen allokoidun tila, jolta on mahdollista palauttaa poistettuja tiedostoja ja tiedostofragmentteja. Levykuva tarjoaa tietokoneen sisäisen kiintolevyn tapauksessa näkymän tietokoneeseen kokonaisuutena mukaan lukien muun muassa järjestelmäasetukset, hakemistorakenteet, tiedostojen sisällöt ja niiden metadatan. (Kirschenbaum, Ovenden & Redwine 2010, 9–12, 25–28; Ries 2018, 392.)

Luovutus

Digitaalisen forensiikan välineillä voidaan muun muassa tarkastella selainhistoriaa sekä palauttaa poistettuja tai muutoin näkymättömissä olevia tiedostoja, jotka mahdollisesti sisältävät esimerkiksi arkaluontoista ja yksityistä tietoa, jota tekijä ei ole tarkoittanut kenenkään nähtäväksi. Tämä nostaa esille eettisiä ja juridisia kysymyksiä, jotka olisi otettava huomioon jo ennen syntysähköisen aineiston luovuttamista, sillä ne saattavat vaikuttaa siihen, miten ja missä määrin aineistoa voidaan jatkossa käsitellä ja tutkia. Forensisten menetelmien mahdollinen käyttö on tuotava selkeästi esille aineiston luovuttajalle ja sovittava hänen kanssaan luovutussopimuksessa mahdollisista käyttörajoi- tuksista. (Carroll et al., 2011, 67–68; Kirschenbaum, Ovenden & Redwine 2010, 46–47, 51, 56.)

Kenenkään etu ei kuitenkaan ole, että luovuttajaa peloteltaisiin kieltämään kaikki digitaalista forensiikkaa hyödyntävä tutkimus. Näitä samoja välineitä voidaan nimittäin myös käyttää yksityisyyden suojaamisen välineenä etsittäessä ja poistettaessa tietoja, joiden käytön luovuttaja on estänyt tai jota hän on halunnut rajoittaa luovutussopimuksella (Kirschenbaum, Ovenden &

Geneettisen kritiikin näkökulmasta ensisijaisen tutkimusaineiston muodostavat kirjailijan tekemät muistiinpanot, suunnitelmat, luonnokset ja erilaiset teoksen tekstin käsikirjoitusversiot. Tutkimussuunta onkin olennaisella tavalla riippuvainen arkistoista, mutta toisaalta se on myös tuonut arkistoaineistot uudella tapaa osaksi kirjallisuudentutkimusta.

Redwine 2010, 32, 55–56, 58). Esimerkiksi syntysähköisten aineistojen arkistointiin kehitetty ilmainen BitCurator-ohjelmistoympäristö sisältää ohjelmia, joilla voidaan levykuvista identifioida, raportoida ja poistaa henkilötietoja, kuten nimiä, osoitteita, potilastietoja, puhelin-, rekisteri- ja tilinumeroita, sekä muuta sensitiivistä informaatiota (Lee et al. 2013, 20–22; Woods & Lee 2015, 2–3). Forensisten välineiden käyttö ei siis suinkaan rajoitu geneettiseen tutkimukseen, vaan siitä on tullut olennainen osa syntysähköisten aineistojen arkistoinnin nykykäytäntöjä useissa arkistoissa, kirjastoissa ja museoissa.

Käyttörajoitusten sisällöstä voidaan ottaa esimerkiksi Emoryn yliopiston Salman Rushdie -kokoelma. Suojatakseen omaa ja perheensä sekä kolmansien osapuolten yksityisyyttä pääsy hänen juridisiin ja taloudellisiin asiakirjoihinsa on evätty hänen elinaikanaan sekä perheeseen liittyviin asiakirjoihin 70 vuoden päähän aineiston luovutuksesta. Jos perheenjäsen kuitenkin kuolee tätä ennen, tulee häntä koskeva aineisto käytettäväksi. Rushdie on halunnut myös rajoittaa päiväkirjojensa käyttöä tekeillä olevan elämäkerran vuoksi. Kirjeenvaihto ja sähköpostiviestit ovat käytettävissä, mutta niistä on täytynt poistaa tiettyjen henkilöiden puhelin- ja faksinumerot sekä kotiosoitteet. Viimeksi mainittu osoittautui kuitenkin liian työlääksi, joten arkisto päätti toistaiseksi rajoittaa pääsyn lähes koko tähän aineistoon. Rushdie ei myöskään halunnut syntysähköisiä aineistojaan käytettävän vapaasti internetissä, joten ne ovat luettavissa

paikan päällä olevalta tietokoneelta. Ehkä kiinnostavinta on, että arkisto päätti itse jättää palauttamatta tallennusmedioilta mahdollisesti löytyvät poistetut tiedostot eettisistä ja ammatillisista syistä. Rushdien päätös tämä ei ymmärtääkseni ollut, vaan se liittyi kokoelman erityisluonteeseen. Päätös ei myöskään koske kaikkia arkiston kokoelmia, vaan tiedostojen palautuksesta neuvotellaan tapauskohtaisesti kunkin luovutuksen kohdalla. (Carroll et al. 2011, 68.)

Arkistoissa on myös syntysähköisiä aineistoja jo kuolleilta kirjailijoilta, joilta ei ole kysytty heidän suhtautumistaan digitaalisen forensiikan käyttöön, sillä luovutussopimuksia laadittaessa ei aikanaan ole otettu huomioon syntysähköisten aineistojen ja niiden tutkimuksen erityispiirteitä, vaan ne ovat perustuneet perinteisiin paperisiin arkistoaineistoihin. Ongelmana on, ettemme voi jälkeenpäin tietää, mitä tekijä olisi ajatellut asiasta. Suomalaisen Kirjallisuuden Seuran (SKS) arkisto on toistaiseksi ratkaissut ongelman kieltämällä digitaalista forensiikkaa soveltavan tutkimuksen vedoten luovutussopimukseen, jossa tällaista käyttöä ei ole mainittu. Ratkaisu ei ole mitenkään poikkeuksellinen. Tuoreessa kyselyssä amerikkalaisille, irlantilaisille ja brittiläisille arkisto- ja kirjastoalan ammattilaisille tuli esille, että useimmiten syntysähköisten aineistojen asiakaskäytön esteenä ovat tekijänoikeudet, tietosuojaa koskeva lainsäädäntö, aineiston suuri määrä sekä osaamisen puute. Epävarmoissa tilanteissa muistiorganisaatioilla on myös taipumus tulkita lakia pääsyy rajoittavalla tavalla suojellakseen mainettaan (Jaillant 2022, 419, 430).

Oman tutkimukseni tapauksessa olemme soveltaneet olemassa olevia tutkimuslupasopimuksia ja rajanneet pois sellaisen aineistojen käytön, jotka eivät liity kirjallisten tekstien syntyprosessiin. Tällaisessa tilanteessa korostuu tutkijan vastuullisuus aineiston käytössä. Tutkijana olen sitoutunut Tutkimuseettisen neuvottelukunnan (TENK) laatimiin eettisiin ohjeisiin hakiessani tutkimusrahoitusta Suomen Akatemialta. Niihin sisältyy muun muassa tutkittavien henkilöiden yksityisyyden sekä aineellisen ja aineettoman kulttuuriperinnön kunnioittaminen ja siitä huolehtiminen, ettei tutkimuksesta aiheudu tutkittaville ihmisille tai tutkimuskohteille merkittäviä riskejä, vahinkoja tai haittoja (Kohonen, Kuula-Luumi & Spoofo 2019).

Myös arkistojen tilauslomakkeiden yhteydessä asiakkaat allekirjoittavat sitoumuksen, jossa he sitoutuvat tekijänoikeuksien ja tietosuojalainsäädännön sekä muita aineiston käyttöä koskevien säädöksiin ja ohjeiden lisäksi siihen, ettei aineistoa käytetä aineistoa koskevien henkilöiden maineen vahingoittamiseen. Väärinkäytökset voivat johtaa vakaviin seurauksiin, kuten tulevan tutkimusrahoituksen epäämiseen, arkistojen luottamuksen kadottamiseen ja aineistojen tutkimuslupien menettämiseen sekä viimekädessä tutkijanuran katkeamiseen.

Silti väärinkäytöksiä voi esiintyä. Ikävä esimerkki on tapaus, jossa kaksi amerikkalaista tohtorikoulutettavaa, joilla oli pääsy Susan Sontagin syntysähköisiin aineistoihin, julkaisi tietoja hänen ja Annie Leibovitzin suhteesta, jonka Sontag oli kiistänyt elinaikanaan. Lisäksi he toivat esille, että Sontagin

aineistossa on arkaluontoisia tietoja Leibovitzin sijaisraskaudesta. Vaikka tapauksessa ei rikottu lakia, se on nähty eettisesti hyvin kyseenalaisena ja se on vaikuttanut myös joidenkin arkistoalan toimijoiden näkemyksiin aineistojen pääsyn rajoittamisesta. (Jaillant 2022, 428.) Onkin syytä pohtia, voitaisiinko esimerkiksi syntysähköisten aineistojen tutkimuslupakäytäntöjä kehittää siten, että niissä huomioitaisiin paremmin väärinkäytön uhka. Pitäisikö esimerkiksi mahdollisesti arkaluontoisiin aineistoihin edellyttää tutkijalta tavallista tarkempia perusteluita tai voisiko tutkimuslupasopimuksessa velvoittaa jonkinlaisen julkaisujen eettiseen esitarkistukseen ennen niiden julkaisemista.

Tilannetta helpottaisi, jos digitaalisen forensiikan menetelmät ymmärrettäisiin myös syntysähköisten aineistojen arkistoinnin työkaluiksi eikä vain aineiston käytön yhteydessä sovellettaviksi tutkimusmenetelmiksi. Tällöin arkiston henkilökunta voisi tarkistaa (engl. *sensitivity review*) myös kuolleen luovuttajan syntysähköisen aineiston arkaluontoisten tietojen ja henkilötietojen varalta (engl. *personally identifying information*) ja määritellä sallitaanko pääsy aineistoon, poistetaanko se, vai rajoitetaanko sitä jollakin tavoin. (Sloyan 2016, 29–30.)

Talteenotto

Maamme keskeisimpiin kirjailijoiden käsikirjoituksia säilyttäviin arkistoihin, kuten Suomalaisen Kirjallisuuden Seuran sekä Svenska Litteratursällskapet i Finlandin arkistoihin, on parin viime vuosikymmenen aikana jo kertynyt jonkin verran syntysähköisiä aineistoja ja odotettavissa on, että niiden tarjonta moninkertaistuu tulevaisuudessa. SKS ja SLS ovat pyrkineet varautumaan tilanteeseen vuonna 2015 ja 2020 Suomen Kirjailijaliiton ja Suomenruotsalaisen kirjailijaliiton (FSF) jäsenille suunnatulla kyselyllä, jossa kartoitettiin, millaisia digitaalisia välineitä kirjailijat käyttävät, millaisia aineistoja heidän työstään syntyy ja miten he organisoivat ja arkistoivat näitä aineistoja. Tarkoituksena on ollut käyttää tätä tietoa hyväksi suunniteltaessa tulevien digitaalisten aineistojen vastaanottoa arkistoissa. (Kivilaakso 2015; Miinalainen 2016; 2020.) Kyselyn pohjalta SLS on voinut esimerkiksi laatia ohjeita, jotka helpottavat paitsi yksityishenkilöitä syntysähköisten aineistojen säilyttämisessä myös arkistoja, joille ne aikanaan mahdollisesti luovutetaan (Miinalainen 2019).

Syntysähköistä aineistoa on luovutettu arkistoille alkuperäisillä tallennusmedioilla, kuten levykkeinä, kiintolevyinä ja kokonaisina tietokoneina. Aineistoa on myös saatu muun muassa cd-levyille ja muistitikuille kopioituina tiedostoina. Arkistoista on myös käyty luovuttajien luona kopioimassa tiedostoja talteen esimerkiksi tilanteissa, joissa luovuttaja on halunnut pitää tietokoneen vielä omassa käytössään. Forensisia levykuvia ei näistä aineistoista ole pyritty tekemään, vaan levykkeiden ja kiintolevyjen kaltaisista tallennemedioista on

kopioitu yksittäiset tiedostot talteen, luetteloitavaksi ja asiakkaiden saatavaksi. SKS:llä esimerkiksi ei ole menetelmää, jolla varmistaa, että tiedostot pysyvät muuttumattomina, kun ne avataan toisilla laitteilla.

Näkyvät tekstitiedostot ovat tärkeitä myös geneettiselle tutkimukselle. SLS:n ja SKS:n kyselystä esimerkiksi selviää, että useat kirjailijat tallentavat varsin järjestelmällisesti eri tekstiversiot erillisinä tiedostoina. Lisäksi esimerkiksi Word-dokumenttien binaarisesta rakenteesta on mahdollista löytää heksaeditorin avulla pikatallennustoiminnon tuottamia geneettisiä kerrostumia (Ries 2018, 403). Irralliset tekstitiedot ovat silti riittämättömiä laajemmalle geneettiselle tutkimukselle. Toisin kuin paperisten käsikirjoitusten kohdalla, jossa muutosmerkinnät näkyvät useimmiten suoraan käsikirjoitusivulle tehtyinä poistoina, lisäyksinä ja korvauksina, ovat syntysähköisen kirjoitusprosessin jäljet levittyneenä eri puolilla tietojärjestelmää. Tämän vuoksi on olennaista, että yksittäisiä tiedostoja voidaan tutkia alkuperäisessä tietojärjestelmän kontekstissa ja että poistettuja tiedostoja voidaan palauttaa allokoimattomasta levytilasta.

Etenkin angloamerikkalaisissa kirjastoissa, arkistoissa ja museoissa forensisten levykuvien tekemisestä on tullut standardinomaisen toimenpide synty-sähköisten aineistojen arkistoinnissa. Forensisen levykuvan katsotaan olevan linjassa monien arkistoinnin periaatteiden ja arvojen kanssa säilyttäessään tallennusmedialle tallennetun datan alkuperäisen järjestyksen siinä digitaalisessa kontekstissa, jossa se on luotu (sisäinen provinienssi) sekä säilyttäessään datan virheettömästi ja muokkaamattomana siinä muodossa, johon se on suunniteltu (eheys). Myös datan alkuperän, oikeellisuuden ja väärentämättömyyden tunnistaminen (aitous) voidaan tehdä luotettavalla tavalla laskemalla tarkistussumma (engl. *checksum*), jolla voidaan varmistaa, että tallennusmedialta kopioitu data on säilynyt muuttumattomana. (Chassanoff, Woods & Lee 2016, 99; Carroll et al., 2011, 71; Lassere & Whyte 2021, 6–8, 11; Lee 2012, 133–135.)

Yksittäisten tiedostojen kopioiminen ja avaaminen voi muuttaa tai pahimmassa tapauksessa tuhota alkuperäisen tallennusmedian ja sen sisältämää dataa, jos käytössä ei ole kirjoitussuojaimia (engl. *writeblocker*), jotka voivat olla erillisiä laitteita tai ohjelmia. Alkuperäisten tallennusmedioiden toistuvasta käyttämisestä voi myös seurata fyysistä kulumista ja vaurioita. Lisäksi levykkeet ja kiintolevyt altistuvat ”bittilaholle” (engl. *bit rot*), jolla tarkoitetaan datan lukukelpoisuuden heikkenemistä, kun tallennusmedia rappeutuu fyysisesti sekä tallennusmedioiden ja tiedostoformaattien vanhentumista, jonka seurauksena niiden avaaminen vaikeutuu huomattavasti sopivien laitteiden ja ohjelmien puutteessa. (Chassanoff, Woods & Lee, 2016, 99–100.) Jo tästä syystä olisi tarpeellista tehdä tallennusmedioista levykuvat, vaikka niiden sisältöjen inventointiin ja prosessointiin ei olisikaan vielä resursseja. Lisäksi forensisen levykuvan tekeminen säästää alkuperäistä tallennusmediaa ja sille tallennettua dataa, sillä sen jälkeen alkuperäiseen tallennusmediaan ei tarvitse enää kajota. (Glisson & Maxwell 2010, 16; Woods, Lee & Garfinkel 2011, 58.)

Se, miten syntysähköistä aineistoa luovutetaan ja säilytetään vaikuttaa sen saatavuuteen ja käytettävyyteen. Esimerkiksi SLS:n ja SKS:n nykykäytäntöjä voidaan kuvailla migraation käsitteellä, jolla tarkoitetaan prosessia, jossa pyritään varmistamaan aineiston sisällön aitous, eheys, luotettavuus ja käytettävyys siirtämällä se tiedostomuodosta ja -järjestelmästä toiseen (Carroll et al. 2011, 77). Syntysähköisten aineistojen kohdalla tämä tarkoittaa esimerkiksi 1980-luvulla suomalaisella Teko-tekstinkäsittelyohjelmalla luotujen tekstitiedostojen konvertoimista PDF/A-muotoon, jota suositellaan pitkäaikaissäilytystä varten. Käytäntöä on verrattu paperisiin aineistoihin siten, että arkisto on kiinnostunut käsikirjoituksista eikä niinkään kirjoituspöydästä, jonka laatikossa niitä on säilytetty. Saatavuuden ja käytettävyyden kannalta migraatio tarkoittaa, että asiakas voi tutkia syntysähköistä käsikirjoitusaineistoa esimerkiksi PDF:ksi konvertoituna tiedostona arkiston lukusalissa olevalta päätteellä. Tekijänoikeudenhaltijan luvalla hän voi mahdollisesti saada tiedostosta kopion tai paperitulosteen mukaansa.

Migraatio pitkäaikaissäilytykseen sopiviin tiedostoformaatteihin ja -järjestelmiin on välttämätön osa syntysähköisten aineistojen arkistointia ja suuri osa asiakkaista on pääasiassa kiinnostunut vain aineistojen sisällöistä. Mutta myös syntysähköisen aineiston fyysisellä tallennusmedialla, alkuperäisellä laitekontekstilla, ohjelmistoympäristöllä ja tiedostomuodolla voi olla suuri merkitys. Niin antiikin, keskiajan kuin modernien käsikirjoitusten tutkimuksessa dokumentin fyysisten ominaisuuksien tutkimus on keskeistä. Sama koskee kirjahistorian ja analyttisen bibliografian tutkimusta sekä kirjallisuudentutkimusta ja kriittistä editointia, joissa on jo pitempään oltu kiinnostuneita tekstejä kehystävistä periteksteistä ja bibliografisista koodeista, kuten kirjan sidonnasta, koosta, musteen ja paperin laadusta ja typografiasta. Helposti kopioitavan ja eri laite- ja ohjelmistoympäristöissä luettavan elektronisen tekstin kohdalla tämä tekstin ”materiaalisuudeksi” kutsuttu ulottuvuus on voinut päästä unohtumaan. Mutta kukaan tuskin kieltää, etteikö sillä laite- ja ohjelmistokontekstilla, jolla kirjailija on kirjoittanut käsikirjoituksensa olisi merkitystä. Esimerkiksi Pääatalo päätyi suomalaiseen Teko-tekstinkäsittelyohjelmaan sen helppokäyttöisyyden vuoksi.

Laitteisto- ja ohjelmistokontekstin jäljittelyä tai rekonstruoimista nimitetään emulaatioksi. Siinä syntysähköisiä aineistoja käsitellään pikemminkin uniikkeina artefakteina kuin pelkästään sisältöinä samaan tapaan kuin paperisia käsikirjoituksia pidetään ainutkertaisina. Pyrkimyksenä on myös säilyttää mahdollisimman paljon tietoa siitä kontekstista, jossa aineisto on luotu ja jossa sitä säilytetään. (Carroll et al. 2011, 78.) Tunnettu esimerkki syntysähköisten aineistojen emulaatiosta on Emoryn yliopiston kirjaston replika Rushdien Power Macintosh 5400 -tietokoneesta. Sen avulla asiakkaat pääsevät tutustumaan tietokoneeseen ohjelmistoihin ja tiedostoihin sellaisena kuin se oli Rushdien käyttämänä 1990-luvun puolivälissä. (Mt. 84–85.)

Laitteisto- ja ohjelmistokontekstin jäljittelyä tai rekonstruoimista nimitetään emulaatioksi. Siinä syntysähköisiä aineistoja käsitellään pikemminkin uniikkeina artefakteina kuin pelkästään sisältöinä samaan tapaan kuin paperisia käsikirjoituksia pidetään ainutkertaisina.

Tietokoneiden emulointi on kuitenkin työlästä ja kallista, mihin nähden sen hyödyt voivat jäädä tutkimuksen näkökulmasta vähäisiksi (Gooding et al. 2019, 6; Molloy 2019, 339). Emulointi voisi tulla kyseeseen esimerkiksi Pääatalon kohdalla, sillä Professori Kalle Pääatalon arkisto on keskittynyt häneen liittyviin aineistoihin. Pääatalo käytti koko uransa aikana vain yhtä tietokonetta, joka on arkiston kokoelmassa Taivalkoskella ja edelleen toimintakunnossa. SLS:n ja SKS:n kaltaisten suurempien arkistojen kohdalla vastaava kirjailijoiden tietokoneiden emulointi ei liene mielekäästä, kun jo pelkästään useiden tietokoneiden säilyttäminen kokonaisuuksina voi tuottaa ongelmia.

Pohjois-Carolinan osavaltionyliopiston kirjastossa emulaatiota on lähes-tytty säilyttämällä syntysähköiset aineistot tallennusmedioiden muodostamina arkistoyksikköinä. Esimerkiksi tietokoneen kiintolevyn sisältämiä yksittäisiä kansioita ja tiedostoja ei irroteta tästä yhteydestä ja järjestetä uusiksi esimerkiksi lajityyppien mukaan, vaan ne säilytetään yhtenäisenä kokonaisuutena siinä järjestyksessä, jonka luovuttaja on niille antanut. Tämä säästää aikaa ja rahaa sekä mahdollistaa datan eheyden ja aitouden ylläpitämisen. Lisäksi se tarjoaa tiedostoja, ohjelmisto- ja laiteympäristöä koskevaa metadataa (muun muassa tiedostonimi, hakemistopolku, tiedostotyyppi, tiedostokoko, luonti- ja muutos aika sekä mahdollisesti tekstin otsikko ja lista sen sisältämistä sanoista), jonka avulla voidaan helpottaa aineiston hakemista. (Dietz 2016; Groth 2014.)

Käyttäjät voi selata aineistoa kahdella tavalla. Ensinnäkin käytössä on tiedostojen metadatatista koostettu CSV-formaatissa oleva taulukko, jota voidaan esimerkiksi Excel-taulukkolaskentaohjelmaa käyttäen järjestellä eri tavoin, kuten luontipäivämäärän tai tiedostomuodon mukaan. Toiseksi tallennusmediaa voidaan selata verkkoselaimessa toimivalla virtuaalisella tiedostonhallintaohjelmalla (engl. *virtual filesystem browser*), joka muistuttaa esimerkiksi Windowsin Explorer- tai Macin Finder-ohjelmia, joilla voidaan hallita tietokoneen tiedostoja. Näin käyttäjä voi ikään kuin navigoida digitaalisessa ympäristössä, jossa kirjailija työskenteli ja nähdä esimerkiksi, miten hän nimesi ja järjesteli kansioita ja tiedostoja. Molempia voidaan käyttää verkkoyhteyden välityksellä. Niiden kautta voi tarkastella tiedostojen metadatat, mutta itse tiedostoihin ei pääse käsiksi. Ne helpottavat kuitenkin huomattavasti aineistopyyntöjen tekemisessä. (Dietz 2016; Groth 2014.)

Pohjois-Carolinan yliopiston kirjaston emulaatiomallia voi verrata esimerkiksi runoilijan käyttämään vihkoon, joka sisältää paitsi runojen versioita myös muita tekstejä. Arkistossa runoja sisältäviä sivuja ei irroteta vihkosta ja siirretä runoja sisältäviin säilytysyksiköihin. Myöskään jokaista vihkon sisältämää tekstiä tuskin luetteloidaan erikseen, vaan vihkon sisältöä kuvaillaan yleisellä tasolla ja sitä käsitellään kokonaisuutena. Tämä on tärkeää, sillä vihko tarjoaa kontekstuaalista tietoa, joka voi olla olennaista sen sisältämien yksittäisten tekstien ymmärtämiselle. Emulaation periaate, jonka mukaan myös tiedostojen ohjelmisto- ja laitteistoympäristöllä on merkitystä, on hyvä pitää mielessä kehitettäessä syntysähköisten aineistojen arkistointia alkaen luovutustilanteesta, jossa olisi pyrittävä saamaan mahdollisimman paljon aineiston syntyä ja säilytystä koskevaa kontekstuaalista tietoa.

Pohjois-Carolinan yliopiston kirjaston mallia voisi kehitellä järjestelmällä, jonka avulla käyttäjät voisivat halutessaan lisätä tiedostoja ja niiden välisiä suhteita koskevaa tietoa, kuten tekstin otsikon, lajityypin tai aiheen, jos sitä ei ole voitu selvittää digitaalisin välinein metadatan keruun yhteydessä. Huomion arvoista on myös, ettei tiedostoja ilmeisesti ole migroitu, vaan ne ovat saatavilla ainoastaan alkuperäisessä tiedostomuodossa, jolloin käyttäjän on itse huolehdittava niiden lukemiseen tarvittavien ohjelmien hankkimisesta (ks. esim. Alexander Isley Papers). Jos tarjolla olisi myös migroidut versiot tiedostoista, tarjoaisi malli varsin hyvän lähtökohdan tutkimukselle, joskin se merkitsisi lisää kustannuksia. Lisäksi voisi kunkin tallennusmedian yhteyteen liittää tiedot palautetuista ja poistetuista tiedostoista niissä tapauksissa, joissa se on mahdollista. Tällainen järjestelmä palvelisi jo suurinta osaa käyttäjistä.

Syntysähköisiä aineistojen pääsyä ja käytettävyyttä koskevassa keskustelussa näkee harvemmin, että myös koko levykuva olisi saatavissa tutkimuskäyttöön. Tämä johtuu luultavasti siitä, että tavoitteena on tehdä käytettävyydestä mahdollisimman sujuvaa. Geneettisen kritiikin kaltaisen tutkimuksen kannalta olisi kuitenkin suotavaa, että syntysähköisten aineistojen arkistointikäytäntöjä kehitettäessä otettaisiin myös huomioon mahdollisuus tutkia levykuvia koko-

naisuudessaan eikä vain niistä irrotettuja yksittäisiä tiedostoja, vaikka kysyntä tälle jäisikin pieneksi.

Lopuksi

Olen tässä artikkelissa tarkastellut syntysähköisten aineistojen arkistointikäytäntöjen kehittämistä oman tutkimusalani kieltämättä hyvin rajallisesta näkökulmasta, jolloin käsittelyn ulkopuolelle ovat jääneet muun muassa sähköpostit. Näkökulma on myös ollut varsin yksipuolisesti tutkijan näkökulma. Tarkoituksena on kuitenkin ollut herättää yleisempää keskustelua syntysähköisten aineistojen arkistoinnin ja tutkimuksen kehittämisestä, jota ei voi tehdä ilman eri alojen asiantuntijoiden yhteistyötä ja keskinäistä vuorovaikutusta. Syntysähköisten aineistojen arkistointia koskevassa kansainvälisessä keskustelussa pääpaino oli kauan siinä, miten tällaiset aineistot saadaan talteen ja miten ne voidaan arkistoida (Jaillant 2019, 297; Shein 2014, 2–3). Tätä keskustelua käytiin pitkälti digitaaliseen forensikkaan perehtyneiden tietotekniikan asiantuntijoiden ja arkiston työntekijöiden välillä. Viime aikoina keskustelu on laajentunut aineiston asiakaskäyttöön asettamisen kehittämisen kysymyksiin ja siinä on painotettu arkistoinnin ammattilaisten ja aineiston käyttäjien välisen yhteistyön merkitystä (Jaillant 2019, 299; 2022, 418, 423). Kuten Jaillant (mt. 21) toteaa – ja olen tässä artikkelissa yrittänyt tuoda esille – voi arkistokäytännöillä olla huomattava vaikutus tutkimuksen tekemiseen. Silti käyttäjät harvoin osallistuvat niiden suunnitteluun.

Suomessa ollaan vasta heräämässä syntysähköisten aineistojen arkistoinnin haasteisiin, jolloin kehitystyötä on mahdollista tehdä alusta asti vuorovaikutuksessa eri viiteryhmien kanssa. Arkistoissa ollaan ongelmista ja kansainvälisestä keskustelustakin tietoisia. Toimenpiteisiin ei kuitenkaan ole toistaiseksi ryhdytty osaksi siitä syystä, ettei syntysähköisten aineistojen kysyntä ole ollut kovin suurta, mikä on mahdollistanut asian lykkäämisen. Kyse on myös priorisoinnista, johon vaikuttavat muun muassa taloudelliset ja osaamisen resurssit. Uusien järjestelmien rakentaminen on kallista ja hidasta. Syntysähköisiä aineistoja ei voi myöskään esimerkiksi mitata samalla tapaa kuin paperiaineistoja, kun luodaan tilastoja siitä montako hyllymetriä on saatu vuodessa otettua talteen ja inventoitua. On siis tarjottava enemmän tietoa syntysähköisten aineistojen luonteesta, kulttuurihistoriallisesta merkityksestä ja siitä miten niitä voidaan tutkia. Tämä ei tarkoita vain arkiston henkilökunnan kouluttamista, vaan myös esimerkiksi kirjallisuuden opetuksessa olisi ryhdyttävä opettamaan erilaisten syntysähköisten aineistojen arkistointi- ja tutkimusmenetelmiä. Erityisen tärkeää on välittää tietoa arkistoaineiston tuleville luovuttajille, missä onkin jo tehty arvokasta työtä yllä mainitussa SLS:n ja SKS:n toteuttamassa kyselyssä. Ääni olisi myös saatava kuuluville niille tahoille, jotka päättävät, mihin resursseja kohdennetaan. (Vrt. Jaillant 2019, 299–300). Paperiaineistojen digitointi

on ymmärtääkseni Suomessa jo varsin hyvässä vauhdissa. Nyt olisi jo korkea aika ryhtyä kehittämän syntysähköisten aineistojen vastaanottoa, hallintaa ja asiakaskäyttöön tarjoamista.

Viitteet

- 1 Tätä tutkimusta on rahoittanut Suomen Akatemia (projektinnumero: 332487).
 - 2 Haluan kiittää Suomalaisen Kirjallisuuden Seuran arkiston kokoelmapäällikkö Katri Kivilaaksoa ja Svenska Litteratursällskapet i Finlandin arkiston arkistonhoitaja Maria Miinalaista keskusteluista, jotka ovat valottaneet arkiston näkökulmaa syntysähköisten aineistojen arkistoinnissa.
 - 3 Tarkasti ottaen tämä slack-tila koostuu kahdesta erityyppisestä slackista: RAM-slackista ja jäännös-slackista (engl. *residual slack*). FAT16-järjestelmässä lohko koostuu neljästä 512 tavun sektorista. TXTKAP.056:n viimeisestä lohokosta tiedoston tallentami-
- seen on käytetty vain 471 tavua, joten kolme sektoria on jäänyt kokonaan käyttämättä. Ensimmäiseen sektoriin on jäänyt käyttämätöntä tilaa 41 tavua. Vanhemmat käyttöjärjestelmät täyttivät tämän tilan RAM-muistissa olevalla datalla. (Carrier 2005, 139–140.) Esimerkissämme RAM-slack sisältää tekstin: ” n ja vauvan pyykkiä. Olin silittänyt hyvä” (Kuva 1). Tämä fragmentti on *Pölhökanto lijoen törmässä* -romaanin toisen luvun tekstivariantti (vrt. Päätalo 1998b, 79). Jäljelle jääville kokonaisille sektoreille järjestelmä ei kirjoita mitään. Siksi jäännös-slackista on mahdollista löytää tallennusmedialle aiemmin kirjoitettua dataa.

Arkistolähteet

Professori Kalle Päätalon arkisto, Taivalkoski:

- Päätalo, Kalle. 1997a. Kolmannen luvun fragmentti. *Pölhökanto lijoen törmässä*. Syntysähköinen käsikirjoitus. Tiedostokoko: 1577 tavua, tarkistussummat: MD5: 7d578770b03bda-0759ba8f19575c8bfc; SHA256: 67f933558493159f9c14d4f494e3db256feed8d6ce7706e98dc19b-1c25ad9df. Sijainti: kiintolevy: Seagate ST-225, kapasiteetti 20 mb, 5.25”, MFM, FAT16, yksi osio. Sarjanumero: 62292101. Levykuvan tarkistussummat: MD5: 54e752dc91c-514cd4e0a542f76226882; SHA256: 374fae59ee38391aba0bb0ff07ca7686acea1e3a91a88fe-c760eaa65fcc6eea. Tiedostofragmentin offset: 2353616–2355197.
- Päätalo, Kalle. 1997b. Toinen luku. *Pölhökanto lijoen törmässä*. Syntysähköinen käsikirjoitus. Tiedostonimi: TXTKAP.056, tiedoston hakemistopolku: c:/teko, tiedostokoko: 102871 tavua, muokkauspäivä: 3.10.1997, tarkistussummat: MD5: 54fb59678928e834581dbcf37d26ee9c; SHA256: 345c93c4b69450739dcbc6c7d18215270743cf21a0949f99319f290ac206d45e. Sijainti kiintolevy (ks. Päätalo 1997a).
- Päätalo, Kalle. 1998a. Kolmas luku. *Pölhökanto lijoen törmässä*. Syntysähköinen käsikirjoitus. Tiedostonimi: TXTKAP.046; Sijainti: kiintolevy (ks. Päätalo 1997), hakeistopolku: c:/teko/, tiedostokoko: 112032 tavua, muokkauspäivä: 5.1.1998, tarkistussummat: MD5: 56ff709ad34c-5c29a9eae396bcac60f; SHA256: a714539ec45ef39d5a52d27754d51c45fed747058e240d75d-cae6b245a024e6. Sijainti kiintolevy (ks. Päätalo 1997a).

Muut lähteet

- Alexander Isley Papers 1967–2021 2021. North Carolina State University Libraries. <https://www.lib.ncsu.edu/findingaids/mco0534> (28.9.2022).
- Biasi, Pierre-Marc de 1996. What is a Literary Draft? Toward a Functional Typology of Genetic Documentation. *Yale French Studies*, 89, 26–58.
- Biasi, Pierre-Marc de 2004. Toward a Science of Literature: Manuscript Analysis and the Genesis of the Work. Käänt. Jed Deppman. Teoksessa *Genetic Criticism: Texts and Avant-Textes*. Toim. Jed Deppman, Daniel Ferrer & Michael Groden. Philadelphia: University of Pennsylvania Press, 36–67.
- Biasi, Pierre Marc de 2010. Pour une génétique généralisée: L'approche des processus à l'âge numérique. *Genesis* 30, 163–175.
- Buschenke, Floor 2016. Het literaire werk anno 2016. Digitale Schrijfprocessen vastleggen en analyseren. *Vooy's* 34 (4), 8–20. Pre-print: https://pure.knaw.nl/ws/portalfiles/portal/7195159/34.4Buschenhenke_artikelVooy's.pdf (12.12.2022).
- Carrier, Brian 2005. *File System Forensic Analysis*. Crawfordsville: Addison Wesley Professional.
- Carroll, Laura, Erika Farr, Peter Hornsby & Ben Ranker 2011. A Comprehensive Approach to Born-Digital Archives. *Archivaria* 72, 61–92. <https://archivaria.ca/index.php/archivaria/article/view/13360> (28.9.2022).
- Chassanoff, Alexandra, Kam Woods & Christopher Lee 2016. Digital Preservation Metadata Practice for Disk Image Access. Teoksessa *Digital Preservation Metadata for Practitioners*. Eds Angela Dappert, Rebecca Squire Guenther & Sébastien Peyrard. Springer: Cham, 99–109. DOI:10.1007/978-3-319-43763-7_8
- Crasson, Aurèle 2019. Derrida et l'ordinateur: une archive malgré soi. *En attendant Nadeau. Journal de la littérature, des idées et des arts* 91. <https://www.en-attendant-nadeau.fr/2019/11/19/archives-manuscrits-4-derrida/> (12.12.2022).
- Crasson, Aurèle, Jean-Louis Lebrave & Jérémy Pedrazzi 2019. Le "siliscrit" de Jacques Derrida. Exploration d'une archive nativement numérique. *Genesis* 49, 7–12. DOI:10.4000/genesis.4316
- Debray-Genette, Raymonde 1977. Génétique et poétique: Esquisse de méthode. *Littérature* 28, 19–39.
- Dietz, Brian 2016. Let the Bits Describe Themselves. Bloggers! The Blog of SAA's Electronic Records Section. <https://saaers.wordpress.com/2016/01/07/let-the-bits-describe-themselves/> (28.9.2022).
- Drmacich, Jessika, Kate Dundon, Jess Farrell, Christina Velazquez Fidler, Camille Tyndall Watson & Hannah Wang 2022. Legal and Ethical Considerations for Born-Digital Access. Alexandria: Digital Library Federation. DOI 10.17605/OSF.IO/KETR7
- Glisson, Brad & Rob Maxwell 2010. A Digital Forensics Workflow. Teoksessa *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Eds Kirschenbaum, Matthew G., Richard Ovenden & Gabriela Redwine. Council on Library and Information Resources (CLIR) Reports 149, 16. <https://www.clir.org/pubs/reports/pub149/> (28.9.2022).
- Gooding, Paul, Jos Smith, & Justine Mann 2019. The Forensic Imagination: Interdisciplinary Approaches to Tracing Creativity in Writer's Born-Digital Archives. *Archives and Manuscripts*, 47(3), 374–390. DOI:10.1080/01576895.2019.1608837
- Grésillon, Almuth 1994. *Éléments de critique génétique: Lire les manuscrits modernes*. Paris: PUF.
- Grésillon, Almuth 2001. La critique génétique, aujourd'hui et demain. *L'Esprit Créateur* 41 (2), 9–15.
- Groth, Jason Evan 2014. Let the Bits Describe Themselves: Arrangement and Description of Born Digital Objects. Libraries News, Special Collections, NC State, University Libraries. <https://www.lib.ncsu.edu/news/special-collections/let-the-bits-describe-themselves%3A-arrangement-and-description-of-born-digital-objects> (28.9.2022).

- Hay, Louis 2002. *La Littérature des écrivains: Questions de critique génétique*. Paris: Jose Corti.
- Jaillant, Lise 2019. After the digital revolution: working with emails and born-digital records in literary and publishers' archives. *Archives and Manuscripts* 47 (3), 285–304. DOI: 10.1080/01576895.2019.1640555
- Jaillant, Lise 2022. How Can We Make Born-digital and Digitised Archives More Accessible? Identifying Obstacles and Solutions. *Arch Sci* 22, 417–436. DOI:10.1007/s10502-022-09390-7
- Kirschenbaum, Matthew 2008. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, Massachusetts & London, England: The MIT Press.
- Kirschenbaum, Matthew, Richard Ovenden & Gabriela Redwine 2010. *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Council on Library and Information Resources (CLIR) Reports 149. <https://www.clir.org/pubs/reports/pub149> (28.9.2022).
- Kirschenbaum, Matthew & Doug Reside 2013. Tracking the Changes: Textual Scholarship and the Challenge of the Born Digital. Teoksessa *The Cambridge Companion to Textual Scholarship*. Eds Neil Fraistat & Julia Flanders. Cambridge: Cambridge University Press, 257–273. DOI:10.1017/C09781139044073.013
- Kivilaakso, Katri 2015. Kulttuuriperintö uhkaa kadota bittitaivaan tuuliin. Vähäisiä lisiä: Kirjoituksia kulttuurista, tutkimuksesta ja kulttuuriperinnöstä. Suomalaisen Kirjallisuuden Seura. <http://neba.finlit.fi/blogi/kulttuuriperinto-uhkaa-kadota-bittitaivaan-tuuliin/> (28.9.2022).
- Kohonen, Iina, Arja Kuula-Luumi & Sanna-Kaisa Spoofo 2019. Ihmisen kohdistuvan tutkimuksen eettiset periaatteet ja ihmistieteiden eettinen ennakoarviointi Suomessa. Tutkimuseettisen neuvottelukunnan ohje 2019. Tutkimuseettisen neuvottelukunnan julkaisuja 3/2019. Helsinki: Tutkimuseettinen neuvottelukunta. https://tenk.fi/sites/default/files/2021-01/Ihmistieteiden_eettisen_ennakoarvioinnin_ohje_2020.pdf (28.9.2022).
- Lassere, Monique & Jess M. Whyte 2021. Balancing Care and Authenticity in Digital Collections: A Radical Empathy Approach to Working with Disk Images. *Journal of Critical Library and Information Studies* 3 (2), 1–25. DOI:10.24242/jclis.v3i2.125
- Lee, Christopher 2012. Archival Application of Digital Forensics Methods for Authenticity, Description and Access Provision. *Comma* 2, 133–140. DOI:<https://doi.org/10.3828/comma.2012.2.14>
- Lee, Christopher, Kam Woods, Matthew Kirschenbaum & Alexandra Chassanoff 2013. From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions. Baltimore: Maryland Institute for Technology in the Humanities. <https://bitcuratorconsortium.org/from-bitstreams-to-heritage-putting-digital-forensics-into-practice-in-collecting-institutions/> (28.9.2022).
- Lehtonen, Joel 1905. *Mataleena: laulu synnyinseudulle*. Helsinki: Otava.
- Lehtonen, Joel 1931–1935. *Kootut teokset 1–8*. Helsinki: Otava.
- Lernout, Geert 2002. Genetic Criticism and Philology. *Text* 14, 53–75.
- Mathijssen, Marita 2009. Genetic Textual Editing: The End of an Era. Teoksessa *Was ist Textkritik? Zur Geschichte und Relevanz eines Zentralbegriffs der Editionswissenschaft*. Hrsg. Gertraud Mitterauer, Ulrich Müller, Margarete Springeth & Verena Vitzthum. Tübingen: Niemeyer, 233–240.
- Molloy, Kevin 2019. Literary Archives in the Digital Age: Issues and Encounters with Australian Writers. *Archives and Manuscripts* 47 (3), 327–342. DOI:10.1080/01576895.2019.1631863
- Miinalainen, Maria 2016. "Jag har en gammal dator i garderoben". Det digitala kulturarvet. *Källan* 1, 30–31. <https://www.sls.fi/sv/utgivning/kallan-20161> (28.9.2022).
- Miinalainen, Maria 2019. Ta hand om din historia – arkivvård av digitala filer. Svenska Litteratursällskapet i Finland. <https://www.sls.fi/sv/blogg/ta-hand-om-din-historia-arkivvard-av-digitala-filer> (28.9.2022).
- Miinalainen, Maria 2020. "För närvarande ligger filer i molnet". *Källan* 2, 19–21. <https://www.sls.fi/sv/utgivning/kallan-20202> (28.9.2022).

- Pal, Anandabrata & Nasir Memon 2009. The Evolution of File Carving. *IEEE Signal Processing Magazine* 26 (2), 59–71. DOI:10.1109/MSP.2008.931081
- Puhuva kirjailija – kertova arkisto 2012–2015. <https://www.finlit.fi/fi/puhuva-kirjailija-kertova-arkisto-2012-2015#.Yt5bR8FBzao> (28.9.2022).
- Pulkinen, Veijo 2019. Poistoja. Lisäyksiä. Liimauksia. Kirjoituskoneen asema Joel Lehtosen kirjoitusprosessissa. *Synteesi: taiteidenvälisen tutkimuksen aikakauslehti* 37 (1–2), 66–85.
- Päätaalo, Kalle 1998. *Pölkökanto lijoen törmässä*. Jyväskylä: Gummerus.
- Ries, Thorsten 2010. "die geräte klüger als ihre besizer": Philologische Durchblicke hinter die Schreibszenen des Graphical User Interface. *Editio* 24, 149–199.
- Ries, Thorsten 2017. Philology and the Digital Writing Process. *Cahier voor Literatuurwetenschap* 9, 129–158.
- Ries, Thorsten 2018. The Rationale of the Born-Digital Dossier Génétique: Digital Forensics and the Writing Process: With Examples from the Thomas Kling Archive. *Digital Scholarship in the Humanities* 33 (2). DOI:10.1093/llc/fqx049
- Ries, Thorsten 2023. *Textgenese und digitale Forensik: Exemplarische Studien zu Thomas Kling und Michael Speier*. Beihefte zum Euphorion. Heidelberg: Winter. Tulossa.
- Sammes, Tony & Brian Jenkinson 2007. *Forensic Computing*. Toinen laitos. London: Springer-Verlag.
- Shein, Cyndi 2014. From Accession to Access: A Born-Digital Materials Case Study. *Journal of Western Archives* 5 (1), 1–42. <https://digitalcommons.usu.edu/westernarchives/vol5/iss1/1> (28.9.2022).
- Sloyan, Victoria 2016. Born-digital archives at the Wellcome Library: appraisal and sensitivity review of two hard drives. *Archives and Records* 37 (1), 20–36. DOI:10.1080/23257962.2016.1144504
- Van Hulle, Dirk 2019. Modern Manuscripts. *Oxford Research Encyclopedia, Literature*. Oxford University Press, 1–20. <http://oxfordre.com/literature> (12.12.2022).
- Van Hulle, Dirk 2022. *Genetic Criticism: Tracing Creativity in Literature*. Oxford: Oxford University Press.
- Vauthier, Bénédicte 2016. Genetic Criticism Put to the Test by Digital Technology: Sounding out the (mainly) Digital Genetic File of El Dorado by Robert Juan-Cantavella. *Variants: The Journal of the European Society for Textual Scholarship*. 12–13, 163–186. DOI:10.4000/variants.353
- Woods, Kam, Christopher Lee & Simson Garfinkel 2011. Extending digital repository architectures to support disk image preservation and access. *Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries*. New York: Association for Computing Machinery, 57–66. DOI:10.1145/1998076.1998088
- Woods, Kam & Christopher Lee 2015. Redacting Private and Sensitive Information in Born-Digital Collections. *Archiving Conference* 1, 2–7. Pre-print. https://www.researchgate.net/publication/279736261_Redacting_Private_and_Sensitive_Information_in_Born-Digital_Collections (28.9.2022).