

Digitoitu arkistoaineisto kulttuurintutkimuksessa -projekti Digitised Archive Material in Cultural Studies

Pasi Klemettinen (toim.)

”Ei se synny synnyttämättä”

Selvitys digitointiprojektin vaiheista ja työprosesseista

Projektin johtaja: professori Pirkko Nuolijärvi (Kotus)
Koordinaattori: Pasi Klemettinen (SKS)
Projektin rahoittaja: Suomen Akatemia (2001-2004)

Kirjoittajat:

Niina Hämäläinen (Turun yliopisto)
Kirsi Keravuori (SKS / Biografiakeskus)
Pasi Klemettinen (SKS / Kansanrunousarkisto)
Raija Majamaa (SKS / Kirjallisuusarkisto)
Lasse Nirhamo (Suomen Tietojenkäsittelymuseoyhdistys ry)
Jukka Saarinen (SKS / Kansanrunousarkisto)
Pekka Tolonen (Turun yliopisto / Kalevala-instituutti)
Matti Uusivirta (Kotimaisten kielten tutkimuskeskus)

Suomen Akatemian projekti nro 52990
© Digitoitu arkistoaineisto kulttuurintutkimuksessa -projekti

Suomalaisen Kirjallisuuden Seuran raportti 1/2006
ISBN 951-746-850-4 (PDF)



Seuraavassa hankeraportissa keskitytään käsikirjoitusten digitointiprosessiin liittyvien teknisluonteisten työvaiheiden kuvaamiseen. Selvityksen tavoitteena on tiedon välittäminen eteenpäin digitointia harjoittavassa organisaatiossa (SKS) sekä tätä kautta mm. hankkeessa tuotettujen digitaalisten tietovarantojen säilyvyyden, käytettävyyden ja hallittavuuden turvaaminen. Toisena keskeisenä pyrkimyksenä on käytännön kokemusten jakaminen muiden muistiorganisaatioiden ja yliopistollisten sidosryhmien välillä. Yksityiskohtaisesta selvityksestä toivotaan olevan hyötyä tulevien digitointihankkeiden suunnittelussa ja esimerkiksi digitoinnista aiheutuvien kustannuksien arvioinnissa (työntekijät, laitteet, työtilat ja hankkeiden aikataulut). Aineiston pitkäaikaistallennuksen osalta kustannuksien arviointia vaikeuttaa se tosiseikka, että teknologian nopea kehitys laskee jatkuvasti siirto- ja tallennusvälineiden hankintahintoja suhteessa niille tallennettavissa olevaan tietomäärään. Mainittakoon, ettei selvityksessä oteta kantaa siihen, miten aineistot tulisi priorisoida digitoitavaksi tai miten resursseja voimakkaasti sitovat digitointihankkeet tulisi mahdollisesti rahoittaa.

Sisällys

Johdanto	3
SKS:n taustahankkeita	4
Digitoinnin perusteita	6
Digitointilinjan käynnistäminen	8
Käsikirjoitusaineiston valinta	11
Käytännön työprosessit	12
Aineistojen skannaaminen	14
Tiedostojen nimeäminen	16
Metatiedot	18
Digitointiprosessissa syntyviä tietoja	19
Aineistoluettelot	21
Digitoitujen aineistojen käytettävyydestä	22
Tekstintunnistaminen (OCR)	23
Aineistojen suhteelliset digitointinopeudet	26
Projektin rinnakkaishankkeita	29
I Elias-kirjetietokanta	31
Korpuksesta tietokannaksi	31
Kuvaus Trip-tietokannoista Elias ja Elias2	33
Hakusivujen toiminta	35
Huomioita, kokemuksia ja ajatuksia	35
II SKVR-tietokanta	36
Johdatus	36
Suomen Kansan Vanhojen Runojen digitointi	36
1. Tekstin skannaus ja merkkimuotoon muuttaminen	37
2. Tekstin rakenteistaminen (xml)	38
3. Alustava SKVR-tietokanta	39
4. SKVR-tietokanta – lopullinen versio	41
Digitaalisen tekstin (SKVR) ja arkistomuistiinpanojen suhde	41
Lopuksi	42
III Valistus-digitointihanke	43
Yhteenvetoja	43
Lähteitä	47

Liitteet

Liite 1	Digitoitujen aineistojen luettelo	49
Liite 2	Konferenssit ja työseminaarit	58
Liite 3	Lönnrotiana ja Elias Lönnrotin kirjeenvaihto tutkimuskohteeksi digitaaliarkistona	60
Liite 4	CPT-tiedostojen pelastusoperaatio	73
Liite 5	Digitoinnin teknisiä lisätietoja	75
Liite 6	SKVR-korpuksen rakenne	78
Liite 7	SKVR-korpuksen datatyyppin määritelmä (DTD)	79
Liite 8	Kuvamuotoisena ja merkkipohjaisesti digitoitujen aineistojen yhteiskäyttö – selvitys mahdollisuuksista yhdistää alkuperäisaineistot kopioihinsa	82
Liite 9	Joitakin huomioita SKVR-tietokannasta käyttäjän näkökulmasta	88
Liite 10	Valistus-digitointihanke	90
Liite 11	Sanakirjojen skannaus	91

JOHDANTO

Suomen Akatemian rahoittama yhteishanke *Digitoitu arkistoaineisto kulttuurintutkimuksessa* käynnistyi syksyllä 2001. Professori Pirkko Nuolijärven johtamassa 3-vuotisessa hankkeessa olivat mukana Suomalaisen Kirjallisuuden Seura (SKS), Kotimaisten kielten tutkimuskeskus (Kotus) sekä Turun yliopiston Kalevala-instituutti ja kulttuurien tutkimuksen laitos (uskontotiede). Tässä selvityksessä painopiste on SKS:n digitointiprosessien ja niihin suoraan tai välillisesti liittyvien toimintojen kuvauksessa. Projektin laaja-alaisena tavoitteena oli laadullisesti korkeatasoisten mutta samalla taloudellisesti mielekkäiden ratkaisujen löytäminen seuraaviin digitaalisia tietovarantoja koskeviin osa-alueisiin:

- Aineiston muuttaminen digitaaliseen muotoon (skannaaminen, OCR)
- Aineiston säilyvyys (tallennusmuodot ja -välineet)
- Aineiston hallittavuus (hakujärjestelmät, tietokannat, XML-dokumentit)
- Aineiston sovellettavuus (julkaisut, tutkimukset)
- Aineiston saatavuus ja käytettävyys (etäkäyttö, käyttökopiot)

Digitointiin ja yleisemmin tietoyhteiskuntaan liittyvät suuret EU-hankkeet ovat viime vuosina tuottaneet lukuisan joukon erilaisia ohjeistuksia ja suosituksia, joiden pohjalta erilaisten muistiorganisaatioiden on mielekästä suunnitella ja kehittää omia toimintojaan. Projektin selvityksessä ei katsottu tarpeelliseksi referoida metatiedoista, standardeista ja ns. parhaista käytänteistä käytyä laajaa ja vilkasta keskustelua. Vuoden 2004 eräänä tärkeänä julkaisuna voidaan kuitenkin mainita Minerva-projektin työryhmä 6:n toimittama käsikirja, joka kattaa tiiviissä muodossa koko digitointi-projektin elinkaaren: hankkeen suunnittelun, aineiston valinnan, valmistelutyöt, alkuperäisaineistojen käsittelyn, digitointiprosessit, metatiedot, aineiston säilyttämisen ja julkaisemisen, tekijänoikeudet sekä projektin koordinoinnin¹. Minerva-projektin web-sivuilta löytyy käsikirjaa täydentävää materiaalia sekä linkkejä muihin vastaaviin lähteisiin.

Pitkällä aikavälillä tarkasteltuna yhtä oikeaa tapaa digitoida tai käyttää digitaalisia aineistoja ei ole olemassa. Tähän on syynä teknologian nopea ja ennustamaton kehitys sekä toisaalta tutkijoiden yms. aineistojen käyttäjäryhmien muuttuvat tiedonintressit. Erilaisten IT-kehityshankkeiden viidakossa tulee muistaa, että muistiorganisaatioiden ensisijaisena tehtävänä on arkistoitujen analogisten ja digitaalisten tietovarantojen säilyvyyden turvaaminen tuleville sukupolville. Tällöin yksinkertaiset, käyttäjäystävälliset ja helposti lähestyttävät tietotekniset ratkaisut ovat usein toimivimpia. Tietotekniikan ei tulisi olla rajoittavana vaan helpottavana tekijänä aineiston aktiivisen hallinnan ja käytön suhteen. Mottona voitaneen pitää mm. jääkiekkovalmennuksesta tuttua sanontaa "keep it simple".

Projektiin liittyvissä osatutkimuksissa evaluoitiin ja hyödynnettiin digitaalisten aineistojen erilaisia käyttömahdollisuuksia. Tutkimusaiheita olivat mm. suomalainen mytologia ja Lönnrotin työ Kalevalan kokoajana. SKS:ssa tuotettuja digitaalisia aineistoja ovat projektin tutkimussuunnitelmassa nimetyistä tutkijoista käyttäneet Niina Hämäläinen, Pasi Klemettinen, Raija Majamaa, Jyrki Pöysä ja Jukka Saarinen. Ranskassa nykyisin asuvan Hämäläisen folkloristisen väitöskirjan aiheena on Kalevalan Kullervo-runojen tekstualisaatiokehitys. Vuodesta 2002 demokäytössä ollut SKVR-tietokanta on mahdollistanut Hämäläiselle kalevalamittaisten vertailu- ja rinnakkaisaineistojen joustavan etäkäytön (ks. [liite 9](#)).

¹ Good Practices Handbook 2004.

Kansainvälisestä yhteistyöstä mainittakoon konferenssit, työseminaarit ja asiantuntija-tapaamiset Irlannin, Latvian, Liettuan, Ruotsin, Unkarin ja Viron sidosorganisaatioiden edustajien kanssa vuosina 2001–2004 (ks. liite 2). Kansanrunousarkistossa konkreettisinta ja pitkäkestoisinta arkistollinen vuorovaikutus on ymmärrettävistä syistä ollut virolaisten kanssa (SKVR-hanke vuodesta 1998). Digitoinnin suunnittelua, toteuttamista ja kehittämistä laajemmin tarkasteltuna tulee muistaa, että sisarlaitosten ohella tärkeitä yhteistyötahoja ovat atk- ja tietokantapalveluja tarjoavat yritykset ja organisaatiot sekä laitevalmistajat (esim. CSC). Muistiorganisaation kannalta on usein tarkoituksenmukaista keskittyä hallittavissa oleviin osaamisalueisiin ja turvautua tarvittaessa ulkopuolisiin IT-osaajiin, vaikka se toisikin mukanaan tietoteknisen riippuvuuden kolmansista osapuolista (mm. laitteet, ohjelmistot ja tekninen tuki). Henkilöstö- ja talousresurssien säästäminen kriittisessä vaiheessa voi osoittautua virheeksi ja johtaa puutteellisiksi osoittautuvien järjestelmien myöhempään päivittämiskiarteeseen tai niistä luopumiseen.

Projektin alkuvaiheen jälkeen kävi ilmeiseksi, että Kansanrunousarkiston tutkijoiden osalta voimavarat oli tarkoituksenmukaista kohdistaa SKVR-tietokannan kehittämiseen ja aktiiviseen testikäyttöön (Klemettinen ja Saarinen). Tietokantojen osalta mainittakoon, että Elias-digitoinnissa (SKS) ja Kalevala-instituutissa suunniteltiin ja kehitettiin vuosina 2002–2004 eri jaksoissa Elias-kirjetietokantaa, josta on tällä hetkellä olemassa teknisesti hyvin toimiva pilottiversio (ks. Tietokantaratkaisut).

Kotimaisten kielten tutkimuskeskuksen osalta hanke keskittyi lingvististen tekstiaineistojen rakenteen, metatietokuvauksen ja näiden kanssa yhteensopivien käyttöliittymien tutkimiseen. Tarkoituksena oli kehittää kansainvälisiin standardeihin perustuva rakenteinen tietomalli Kotuksen omia korpuksia varten sekä kartoittaa tähän malliin soveltuvia käyttöliittymämahdollisuuksia. Rakenteen ja metatietokuvauksen tutkimisesta vastasi pääosin Mikko Lounela ja käyttöliittymien tutkimisesta Outi Lehtinen. Tutkimustuloksiin perustuen rakennettiin Kotuksen nykyisin käytettävissä oleva korpusmalli². Tätä mallia noudattaa suurelta osin myös Kotuksessa valmisteilla oleva (osittain) julkinen, WWW-pohjainen korpuspalvelu, jonka arvioitu julkaisuaika on vuoden 2006 loppupuolella.

SKS:N TAUSTAHANKKEITA

Digitoitu arkistoaineisto kulttuurintutkimuksessa -projektiin keskeisesti kytkeytyviä taustahankkeita ovat olleet SKVR-projekti ja julkaisemattoman kalevalamittaisen kopiokortiston digitointi (Kansanrunousarkisto). SKS:n rahoittama ja koordinoima SKVR-hanke toteutettiin vuosina 1999–2001 akateemikko Arvo Krikmannin johdolla Tartossa Eesti Kirjandusmuuseumissa³. Hankkeessa digitoitiin vuosina 1908–1948 ja 1997 julkaistu SKVR-julkaisusarja, joka sisältää yhteensä 34 nidettä, 24 000 sivua ja 86 000 muistiinpanoa. Julkaisut sisältävät merkittävän osan Kansanrunousarkistoon tallennetusta kalevalamittaisesta epiikasta, lyriikasta ja loitsustosta. Digitaalisen tekstikorpuksen jatkokäsittely tehtiin Tartossa sekä myöhemmin Kansanrunousarkistossa vuosina 2001–2004. Hankkeen erillisistä työvaiheista mainittakoon tekstintunnistaminen (OCR), tekstien oikolukeminen, muokkaaminen ja ns. normalisoiminen sekä XML-rakenteistaminen CSC:n⁴ ylläpitämää tietokantaa varten (2002–2004). Jälkikäteen arvioiden SKVR-hankkeen työllistävän alkuvaiheen ulkoistaminen oli erittäin onnistunut ja kustannusedullinen ratkaisu SKS:lta. On vaikea kuvitella, kuinka SKS:n

² Lounela 2002; Lehtinen ja Lounela 2004.

³ <http://www.kirmus.ee/>

⁴ <http://www.csc.fi/>

olisi ollut mahdollista palkata viisi kokopäiväistä työntekijää useaksi vuodeksi suhteellisen kapea-alaiseen erillisprojektiin Suomessa. Luonnollisesti on varsin helppoa osoittaa potentiaalisia ongelmakohtia kyseisen järjestelyn ja menettelyn osalta; käytännössä ulkomailla toteutettavan projektin aktiivinen ohjaus ja seuranta voi olla hyvin hankalaa. Tässä tapauksessa akateemikko Arvo Krikmannin ja erikoistutkija Jukka Saarisen dynaaminen yhteistyö oli avain projektin osatavoitteiden saavuttamiseen, eikä suurempia ongelmia tai epäselvyyksiä päässyt syntymään.

Kalevalamittaisesta aineistosta puhuttaessa tulee muistaa, että KRA:n käsikirjoituskokoelmissa on runsaasti muistiinpanoja, joita ei ole julkaistu SKVR-julkaisusarjassa. Suurin osa tästä materiaalista on kerätty eri perinnealueellisten osien julkaisujankohdan jälkeen. Kopio-kortteille konekirjoitettu julkaisematon aineisto sisältää niin ikään toimituksellisin ja laadullisin perustein julkaisuista pois jätettyjä muistiinpanoja (mm. kirjallisperäiset runot ja ns. itse-sepityt rahvaanrunot). Julkaisemattoman kopiokortiston skannaaminen kuvamuotoon aloitettiin syyskuussa 2000 ja saatettiin päätökseen toukokuussa 2001. Hankkeeseen osallistuivat digitoijina Atro Niiniluoto ja Pekka Allonen sekä projektin koordinaattorina Pasi Klemettinen. Kortiston digitointi kuvatiedostoiksi toteutettiin kahdella edullisella tasoskannerilla (Epson Perfection 1200) ja A5-kokoisia kortteja skannattiin päivätasolla 500–1000 kpl. Hankkeessa tuotettiin yhteensä lähes 93 000 PNG-tiedostoa, jotka toimitettiin Lönnrotin aineistoja lukuun ottamatta Tarton työryhmälle jatkokäsitteltäväksi (tekstintunnistus ja oikoluku). Tulevaisuudessa digitaalinen SKVR-korpus tulee käsittämään yhteensä noin 150 000 erillistä perinnetekstiä. Edellä mainittujen kalevalamittaisten aineistojen liittäminen samaan tietokantaan tulee kuitenkin olemaan ongelmallista, sillä julkaisematon materiaali ei ole läpäissyt KRA:ssa samaa huolellista tarkastus- ja indeksointiprosessia kuin julkaistu runoaineisto.

Laajemmassa mittakaavassa tarkasteltuna kalevalamittaisen runoaineiston keruuhistoriallinen ja arkistollinen linkaari ulottuu pääpiirteissään jo lähes 200 vuoden ajalle. On selvää, että lukuisten käsittely- ja työvaiheiden jälkeen tutkijat joutuvat yhä uudelleen palaamaan vanhojen originaalikäsikirjoitusten ja jatkossa toivottavasti yhä useammin niistä digitoitujen korkealaatuisten kuvatiedostojen äärelle (aineistojen lähdekriittinen tarkastelu). Ainoastaan alkuperäismuistiinpanoista käy seikkaperäisesti ilmi yksittäisten perinnetekstien tekstuaalinen konteksti sekä esimerkiksi 1800-luvun kerääjien tyyli ja tapa koostaa tekstejään erilaisin lisäyksin, poistoin ja lyhentein. Oman lukunsa perinneaineistojen tuottamisen historiassa ja problematiikassa muodostavat luonnollisesti kerääjien ja tutkijoiden vaihtelevat kenttätyömetodit sekä heidän henkilökohtaiset vaikuttimensa, valintansa ja ratkaisunsa. Vanhojen kenttämuistiinpanojen yms. käsikirjoitusten väridigitoinnin merkitys tuleekin varmasti jatkossa korostumaan uusien tuotettavien asiakaspalvelujen ja tietokantaratkaisujen osalta.

Seuraava pelkistetty taulukko havainnollistaa Kansanrunousarkistossa säilytettävän kalevalamittaisen käsikirjoitusaineiston pitkäkestoista ja monivaiheista prosessointia. Kaikki perinnetekstit eivät luonnollisesti ole käyneet läpi jokaista henkilöresursseja sitovaa työvaihetta. Huomattakoon, että perusdigitointiin liittyvät työvaiheet eivät liity suoranaisesti toisiinsa, sillä kortistojen ja alkuperäiskäsikirjoitusten digitoinnit on mielekästä suunnitella ja toteuttaa teknisesti erillisinä mutta sisällöllisesti toisiaan tukevinä hankkeina. Viime kädessä arkistolliseksi tavoitteeksi tulee asettaa kaikkien aineistotyyppien ja formaattien keskitetty hallinnoiminen saman dynaamisen tietojärjestelmän puitteissa.

Taulukko: Kentältä tietokantaan

LÄHTÖKOHTA	TYÖVAIHE	TULOS
informantti (esittäjä)	(pika) kirjoitus	kenttämuistiinpano
kenttämuistiinpano	puhtaaksikirjoitus	arkistomuistiinpano
arkistomuistiinpano (sidossarjat)	konekirjoitus	kopiokortisto
arkistomuistiinpano (sidossarjat)	digitointi (skannaus tai kuvaus)	faksimile-laatuinen värikuvatiedosto
manuaalinen kopiokortisto	digitointi	digitaalinen näköiskortisto
kuvatiedosto	tekstintunnistus (OCR)	merkkimuotoinen tekstiedosto
tekstiedosto (ohjelmistosidonnainen)	XML-koodaus (datatyyppin määrittely)	rakenteistettu teksti (ohjelmistoriippumaton)
rakenteistettu teksti	tietokantamääritykset	digitaalinen tekstikorpus (erillistietokannat)
kuvat ja tekstit (äänitteet, videot etc.)	metatietojen ja eri aineistotyyppien linkittäminen	aineistojen holistinen haku ja hallinnointi (arkistotietojärjestelmä)

DIGITOINNIN PERUSTEITA

Yleisellä tasolla tarkasteltuna muistiorganisaatioiden tulee parantaa aineistojen käytettävyyttä ja näkyvyyttä tietoyhteiskunnassa. Käytännössä tämä edellyttää digitaalisten tietovarantojen aktiivista kartuttamista sekä uusien dynaamisten ja innovatiivisten haku- ja tietojärjestelmien kehittämistä verkkoympäristöön. Tulee muistaa, ettei palveluita luoda ainoastaan asiakkaille vaan myös organisaatioiden työntekijöille, joiden tulisi sitoutua uusien sähköisten työvälineiden ja -menetelmien omaksumiseen ja aktiiviseen käyttöön. Seuraava toteamus voi vaikuttaa triviaalilta idealismilta, mutta koko yhteiskunnan tavalla tai toisella läpäisevä digitalisaatio tulisi nähdä investointina tulevaisuuteen eikä hallitsemattomana uhkakuvana.

Perinteiseen analogiseen maailmaan tukeutuvia tai siihen suhteutettavia tulevaisuuden uhkakuvia on helppo luoda, mutta ne tulisi asettaa nykyisessä toimintaympäristössämme aina oikeisiin mittasuhteisiin. Sama suhteellisuuden ja terveen järjen vaatimus koskee ymmärrettävästi myös tietoyhteiskuntahankkeisiin liittyviä tulevaisuusvisioita, jotka perustuvat pahimmassa tapauksessa katteettomalle kasvu- ja kehitysideologialle. Kärjistäen voidaankin todeta, että tasapainottelemme teknologisessa mielessä kahden maailman rajalla, jossa kohtaavat tai risteilevät analogiset konventiot ja toisaalta digitaaliset innovaatiot. Muistiorganisaatioiden vaativana ja haasteellisena tehtävänä on käydä kyseisellä rajaseudulla päättymätöntä viivytystaistelua "kaiken maallisen katoavaisuutta" vastaan.

Tietoyhteiskunnan asettamista haasteista ja tavoitteista toimii esimerkkinä opetusministeriön kulttuuriperinnön digitointityöryhmän KULDIn vuonna 2003 esittämä visio: "Vuoteen 2010 mennessä keskeinen museoissa, arkistoissa ja kirjastoissa oleva kulttuuriperintö on digitoitu priorisointisuunnitelmien mukaisesti. Aineistot ovat tietoverkkojen kautta kaikkien ulottuvilla, ja ne edistävät kansakunnan sivistysperinnön saatavuutta ja tunnettuutta kansallisesti ja kansainvälisesti. Kulttuuriperintöaineistot muodostavat merkittävän voimavaran tutkimukselle, opetukselle ja sisältötuotannolle."⁵ Perinnearkiston näkökulmasta vision tekee ongelmalliseksi

⁵ Kulttuuriperintö tietoyhteiskunnassa 2004.

sen rajoittuminen suomalaisen kulttuuriperinnön kannalta keskeisiin asiakirjoihin, julkaisuihin, teoksiin ja museokokoelmiin. Mitä ovat nämä arvokkaat kulmakivet, joita kansalaisille ja koko maailmalle halutaan sähköisessä muodossa esitellä? Missä määrin valinnat ja priorisoinnit edustavat suomalaisen kulttuuriperinnön koko kirjoa eri aikakausilta? Onko tavoitteena ainoastaan kulttuuristen ja kansallisten reliikkien sähköistäminen virtuaalisiin näyteikkunoihin? Digitointityöryhmän loppuraportissa ei oteta kantaa esimerkiksi laajojen kansanperinne- ja muistitietoaineistojen digitointiin, joiden priorisointi ja arvottaminen on vaikeaa. Niin ikään tekijänoikeussäädökset ja aineistokohtaiset käyttöehdot estävät uudempien perinneaineistojen julkaisemisen ja vapaakäytön verkkoympäristössä, joten ko. aineistot eivät edes lähtökohtaisesti sovellu tietoyhteiskuntahankkeiden digitointiohjelmiin.

KULDIn raportissa todetaan Kansallisarkiston osalta seuraavasti: "Aineistojen laajuuden vuoksi vain osa asiakirjallisesta kulttuuriperinnöstä on mahdollista digitoida, ja huomattavaan osaan aineistoa kohdistuu niin vähäinen kysyntä, että digitointi ei ole perusteltua. Lähtökohtana on, että kukin laitos pystyy parhaiten arvioimaan omien aineistojensa merkityksen ja valitsemaan digitoitavat kohteensa." Periaatteessa muistiorganisaation digitointipolitiikan ei tulisi suunnata aineistojen tutkimuskäyttöä, mutta kulttuuriperinnön valikoivassa digitoinnissa on edellä esitetyistä käytännön syistä juuri tästä kysymys. On luonnollisesti helppo priorisoida digitointisuunnitelmissa vanhat, hauraat, harvinaiset ja vaikeasti saavutettavat aineistot, mutta asiat eivät aina ole hyvin uudempienkaan materiaalien osalta. Kansanrunousarkistossa säilytetään esimerkiksi runsaasti heikkolaatuisella musteella puuhiokkeelliselle paperille konekirjoitettuja ja tulostettuja tekstiaineistoja, joiden pitkäaikaistallennus tulee varmasti muodostumaan haasteelliseksi. Lähitulevaisuudessa jouduttaneen kaikissa arkistoissa tekemään radikaaleja päätöksiä sen suhteen mitä hitaasti mutta varmasti tuhoutuvista aineistoista säilytetään ja museoidaan alkuperäisessä muodossaan, ja milloin on tarkoituksenmukaista luopua konservoinnista ja tallentaa aineistojen tietosisältö ainoastaan mikrofilmeille ja/tai digitaaliseen bittiavaruuteen. Rajallisten resurssien puitteissa mikrofilmauksen, konservoinnin ja digitoinnin mielekäs toteuttaminen tulee jatkossa edellyttämään selkeitä linjauksia ja painopisteiden määrittelyä.

Miksi kulttuuriperintöä tulee digitoida? Seuraavat yleispätevät periaatteet on esitetty KULDIn loppuraportissa:

1. Aineistojen digitoiminen edistää informaation hallintaa ja saatavuutta, kokoelmien käytettävyyttä ja säilyvyyttä sekä tallentavien laitosten yhteistyötä ja työnjaon toteuttamista.
2. Aineistojen digitointi mahdollistaa aineistojen joustavan ja poikkitieteellisen tutkimuskäytön, aineistojen alueellisesti tasa-arvoisen saatavuuden sekä kulttuuristen sisältötuotteiden toteuttamisen.
3. Digitoidun materiaalin käyttö edellyttää yhteensopivien menetelmien käyttöä, toimijoiden määrittelemää hierarkiaa ja priorisointia, digitointiprosessin osaamista ja sen hallintaa, materiaalin käyttöön liittyvien tekijänoikeudellisten kysymysten ratkaisua sekä pitkäaikaissäilytyksen varmistamista.

Edellä esitettyihin argumentteihin tukeutuen SKS:n eräänä keskeisenä digitointiperusteena on originaaliaineistojen suojelemisen ja säilyttämisen edistäminen. Tämä peruste koskee erityisesti Kirjallisuusarkiston ja Kansanrunousarkiston vanhimpia käsikirjoitusaineistoja, joiden luettavuus mikrofilmiltä on vaikeaa ja jotka sisältävät eri värisiä ja sävyisiä muste- ja värikynämerkintöjä. Aineistojen systemaattinen digitointi ja tallentaminen värikuvatiedostoiksi mahdollistaisi osaltaan puutteellisissa olosuhteissa säilytettävien arvokkaiden kokoelmien siirron pysyvästi lepoarkistoon. Kansanrunousarkistoon ennen vuotta 1935 kerättyjen ja tallennettujen aineistojen systemaattinen digitointi vaatisi nykyistä digitointilinjaa (tasoskanneria) käyttäen arviolta kolme henkilötyövuotta. Kyseiset käsikirjoitusaineistot

koostuvat ns. tähtikokoelmasta (aineisto 1600-luvulta 1870-luvulle), vähäisiä keräelmiä -sarjasta (VK-sarja 1880–1930), kerääjäsidoksista (S-sarja) sekä kansatieteelliset kuvaukset -sarjasta (E-sarja vuodesta 1889). Nyt annetussa esimerkissä E- ja S-sarjat on katkaistu “keinotekoisesti” vuoteen 1935 ja näin kaikkien edellä mainittujen sarjojen sisältämä kokonaissivumäärä jäänee selvästi alle 150 000 sivun. Mainittakoon, että vuonna 1935 järjestetyn Kalevalan riemuvuoden kilpakeruun myötä arkiston aineistomäärä kaksinkertaistui.⁶

SKS:n erityisperusteena digitoinnin välttämättömyydelle tulee mainita kustannus- ja julkaisu-toiminnan vaihtelevat tarpeet, jolloin kysymys on sekä organisaation sisäisestä että ulkoisesta palvelutoiminnasta (tilaustyöt). Julkaisujen kuvitukseksi digitoidaan valokuvia ja piirroksia sekä jossain määrin myös käsikirjoituksia Kansanrunousarkiston ja Kirjallisuusarkiston kokoelmista. Tarkasti valittujen kohteiden kuvadigitointi poikkeaa, tai ainakin tulisi poiketa, laadullisesti ja määrällisesti käsikirjoitusten massadigitoinnista. Massadigitoinnilla tarkoitetaan tässä yhteydessä laajojen aineistokokonaisuuksien eräskannausta, jolla pyritään digitaalisten tallenteiden ja tutkimusaineistojen suunnitelmalliseen tuottamiseen. Massadigitoinnin perustavoitteena on tekstiyksiköiden (sivu, aukeama tai liuska) hyvä tulostettavuus ja luettavuus sekä edelliseen liittyen kaikkien mahdollisten informatiivisten yksityiskohtien zoomattavuus eli lähennettävyyys (tekstirajaukset). Toistaiseksi A4-kokoisten tai sitä suurempien dokumenttien skannausresoluution minimivaatimuksena on 300 pikseliä tuumalla (ppi). Pienempien digitoitavien kohteiden osalta resoluutio eli tarkkuus tulee arvioida tapauskohtaisesti. Käytännössä massadigitoinnin laatuvaatimuksia ei ole mielekäästä asettaa tarpeettoman korkealle yksittäistapausten ja erityistoiveiden perusteella. Tulee muistaa, ettei esimerkiksi Kansanrunousarkiston kenttätöihin, kyselyihin ja kilpakeruisiin perustuvia käsikirjoitus-kokoelmia ole oletusarvoisesti tarkoitettu julkaistavaksi monivärikuvina julisteissa tai kirjoissa.

SKS:n lähitulevaisuuden tavoitteena on digitoinnin keskittäminen valokuva- ja käsikirjoitus-kokoelmien osalta digitointikeskukseen, jonka toiminnassa pyritään ottamaan huomioon kaikki edellä viitatu tarpeet. Käytännössä pysyvän digitointipisteen perustaminen tulee edellyttämään digitointiin ja kuvankäsittelyyn kouluttautuneen ammattilaisen palkkaamista. Digitoinnin keskittämisellä saavutetaan synergiaetua mm. laiteympäristön, pitkäaikais-tallennuksen ja käytännön työprosessien yhdenmukaistamisen ja hallittavuuden kannalta. Arvioitavaksi ja suunniteltavaksi jää, missä määrin massadigitointeja voidaan toteuttaa ulkoistettuina palveluina suurten muistiorganisaatioiden tms. laitojen tehokkaissa tuotanto-yksiköissä (esim. mikrofilmiskannaus).

DIGITOINTILINJAN KÄYNNISTÄMINEN

”Ei se synny synnyttämättä” (Paavo Haavikko: Rauta-aika)

Hankkeen ensimmäisessä vaiheessa keskityttiin Suomalaisen Kirjallisuuden Seuran toimintoihin sijoitetun digitointilinjan käynnistämiseen (Mariankatu 19 D 46). Työtilasta käytetään erotukseksi muista SKS:n digitointitiloista nimitystä Elias-digitointi. Käytännön tasolla digitointilinjan perustamiseen liittyy keskeisesti tila-, laite- ja henkilöresurssien suunnittelu. Laiteympäristön näkökulmasta käynnistysvaihe sisälsi 1) yksittäisten ATK-laitteiden hankkimisen, kokoamisen ja asentamisen, 2) lisälaitteiden ja ohjelmistojen integroinnin

⁶ http://www.finlit.fi/kra/kra4_2_kk.htm

digitointilinjaan sekä 3) digitointi- ja tallennusvälineiden koekäytön. Laitteiden hankinnasta ja asennuksesta vastasi pääosin ATK-päällikkö Matti Anttila. Eri työvaiheissa olivat mukana myös Pekka Allonen, Heikki Hakkarainen, Hannu Hautala, Pasi Klemettinen ja Jukka Saarinen.

Skannauslaitteiston tilaamista edelsi Hannu Hautalan ja Pasi Klemettisen osallistuminen saksalaisen laitevalmistajan SMA:n⁷ esittely- ja testitilaisuuteen, joka järjestettiin 18.5.2001 Riksarkivetissa Tukholmassa. Tilaisuudessa esiteltiin SMA 32 -laitteistopaketti, johon kuuluu kirjahissin lisäksi Epsonin A3-tasoskanneri. Keskusteluissa kävi ilmi, että laitevalmistajalta oli tulossa markkinoille suurempien sidoksien, lehtien yms. dokumenttien nopeaan skannaukseen soveltuva SMA 21 -laitteisto. Varsin lyhyeksi jääneen harkinta-ajan jälkeen päätettiin elokuussa 2001 käynnistyneen projektin käyttöön hankkia uudempaa ja samalla huomattavasti kalliimpaa teknologiaa edustava laitteisto, johon kuuluu Image Accessin valmistama A2-skanneri⁸. Laitteiston viivästyneen toimittamisen jälkeen skannerin toiminnassa ilmeni lukuisia ongelmia ja puutteita, joita selvitettiin rakentavassa hengessä laitevalmistajan ja -toimittajan kanssa. Projektin toteuttamiseksi suunnitellussa aikataulussa jouduttiin lopulta turvautumaan korvaavaan skannerijärjestelyyn. Käytännössä suunniteltu aineisto skannattiin noin kuudessa kuukaudessa ja näin ensimmäiselle vuodelle asetetut tavoitteet saavutettiin suhteellisen helposti. Tosin kriittisimmässä vaiheessa katsottiin tarpeelliseksi porrastaa tutkimusapulaisten eli skannaajien työaikoja siten, että digitointilinjan nauhavarmistuksineen saatiin tehokkaampaan käyttöön (noin 10 tuntia päivässä). Suurten aineistomassojen nopean digitoimisen ja laitteiden täysipainoisen hyödyntämisen näkökulmasta on digitoijien osalta suositeltavaa harkita jopa pysyviä työvuorojärjestelyjä. Tällöin vähimmäisvaatimuksena on kaksi ammattitaitoista työntekijää yhtä digitointilinjaa kohden.

Jälkikäteen arvioiden digitointilinjan suunnitteluun ja testaamiseen resursoitiin selvästi liian vähän aikaa ja työvoimaa. Lukuisten skanneriteknisten ongelmien seurauksena tehokasta digitointiainaa menetettiin henkilötyövuosina arviolta lähes puolivuotta. Toisaalta tulee muistaa, etteivät tietotekniset ongelmat ole usein ennakoitavissa ja ilmaantuvien ongelmien ratkaiseminen on osa toimivan digitointilinjan ylläpitoa ja kehittämistä. Seuraava lista sisältää joitakin Elias-digitoinnin PC-ympäristössä eri vaiheissa ilmenneitä ongelmia ja puutteita: verkon rajoitukset suurien datamäärien siirrossa, työasemien muistien ja kovalevyjen riittämättömyys sekä cd-aseman ja kovalevyjen rikkoutuminen. Laitehankintojen osalta suunnittelun ja testaamisen merkitystä ei voida tässä yhteydessä liikaa korostaa.

Elias-digitoinnin projektityöntekijät: tutkija Pasi Klemettinen, tutkimusapulainen Pekka Allonen ja tutkimusapulainen Hannu Hautala.

Elias-digitoinnin kaikki projektityöntekijät ovat työskennelleet aiemmin SKS:n arkistoissa, joten originaaliaineiston käsittelyyn ja digitaalisen aineiston luettelointiin ei liittynyt erityisiä ongelmia. Organisaatioissa ja/tai erillisprojekteissa, joissa massadigitointia toteuttaa erillinen tekninen tms. henkilökunta, on syytä konsultoida aktiivisesti originaaliaineistojen säilytyksestä ja konservoinnista vastaavien asiantuntijoiden kanssa. Digitoinnille asetetut määrälliset tavoitteet tuleekin aina realistisesti suhteuttaa vanhojen aineistojen kuntoon ja yleiseen käsiteltävyyteen.

On tärkeää, että muistiorganisaation henkilöstöpolitiikassa pyritään jatkuvuuteen, jottei projekteissa hankittua ja viime kädessä henkilösidonnaista osaamista menetetä yhä uudelleen talon ulkopuolelle. Tilapäis- ja lisätyövoiman käytön edellytyksenä onkin, että organisaatiossa

⁷ <http://www.smadocument.de/english/index.htm>

⁸ http://www.imageaccess.de/eng_index.html

on pysyvää ammattitaitoista työvoimaa itse digitointiin, sen suunnitteluun ja koordinointiin sekä tietojärjestelmien ylläpitoon ja hallintaan (tietotaitojen edelleen välittäminen). Organisaatiossa eri tehtävissä ja mahdollisesti eri osastoilla toimivat avainhenkilöt vaikuttavat olennaisesti myös erillishankkeiden onnistumiseen. Käytännössä digitointiprojektiin liittyvien tietoteknisten toimintojen käynnistäminen täysin alusta ulkopuolisella rahoituksella voi osoittautua mahdottomaksi tehtäväksi (mm. koulutus ja aikataulut). Tämä saattaa muodostua ongelmaksi esimerkiksi EU-hankkeissa, joissa kehitetään ulkopuolisen rahoituksen ja asiantuntijuuden turvin joitakin vuosia teknisesti hyvin toimivia infrastruktuureja, mutta joissa ei sitouduta pitkäjänteiseen digitointityöhön osallistuvan ja kykenevän henkilöstön pysyvään rekrytointiin. "Projektit tulevat ja menevät" ajattelusta voi olla varsin pitkä matka organisaation pysyvään ja suunnitelmalliseen toimintaan.

Lyhyesti todettakoon, etteivät projektille osoitetut tilapäiset työtilat täyttäneet kriittiselle digitointiympäristölle asetettuja laatuvaatimuksia. Ongelmia muodosti mm. suhteellisen avoimessa työtilassa liikkuva pöly, joka tunkeutui skannereiden sisälle ja aiheutti digitaalisiin kuviin häiriöitä. Pölyisyyden seurauksena skannerit jouduttiin avaamaan ja huoltamaan. Lisäongelmien ja työkatkosten välttämiseksi työtilaan järjestettiin viikottainen erikoissiivous ja näin tilanne saatiin hallintaan. Edelliseen liittyen eri tehtäviä hoitavien projektityöntekijöiden (3) sijoittamista samaan huoneeseen ei voida pitää suositeltavana. Erityisesti teknisten ongelmien toistuva selvittäminen projektin alkuvaiheessa vaikutti tavalla tai toisella kaikkien työntekijöiden työrauhaan. Varsinaiset digitointitilat tulisikin erottaa projektin toimistotiloista, joissa esi- ja jälkikäsitellään aineistoja sekä hoidetaan muita päivittäisiä työruutiineja.

Nykyisten kansainvälisten suositusten mukaan työympäristössä tulee kiinnittää erityistä huomiota valaistukseen, seinien ja muiden heijastavien materiaalien väritykseen, ilman- kosteuteen, pölyttömyyteen, originaaliaineistojen käsittelyyn ja säilyttämiseen sekä yleisiin häiriötekijöihin. Näyttöpäätteellä työskennellessä ikkunat tulee tarvittaessa peittää pimennysverhoilla. Seinät ja kattopinnat suositellaan maalattavaksi neutraaliharmaalla Munsell N 8.0 sävyllä. Kyseinen seinä- ja kattopintojen väritys ei vääristä niiltä heijastuvaa valon sävyä (suositus 5000 kelviniä). Tilasuunnittelussa tulee ottaa huomioon myös eri käyttötarkoituksiin hankittavien digitointilaitteistojen erityisvaatimukset. Kamerateknologialle perustuvat laitteistot vaativat enemmän fyysistä työskentelyaluetta, ovat tavallisesti herkempiä pölylle ja hajavalolle sekä tuottavat kohdevaloineen enemmän lämpöä ympäristöön kuin tasoskannerit. Hauraita aineistoja ja itse asiassa mitään ylimääräisiä pölyä sitovia materiaaleja ei tulisi säilyttää kuvadigitointilaitteistojen kanssa samassa tilassa.

Digitointihankkeita suunniteltaessa ja raportoitaessa ylikorostuvat helposti teknologiset kysymykset ja erilaiset tuotannolliset laskelmat. Elimelliseksi osaksi erityisesti pitkäkestoisten toimintojen seuranta ja arvioimista tulisi liittää myös työntekijöiden viihtyvyyden, työssä jaksamisen sekä työn palkitsevuuden kriittinen arvioiminen. Kyseisiin inhimillisiin tekijöihin voidaan vaikuttaa mm. työolosuhteilla, palkkauksella sekä työtehtävien mielekkäällä jakamisella ja aikataulutamisella.

KÄSIKIRJOITUSAINEISTON VALINTA

SKS:ssa hankkeen suunnitteluvaiheeseen liittyi olennaisesti digitointiin soveltuvan vanhan käsikirjoitusaineiston kartoitus ja valinta. Lönnrotiana-kokoelman valikoivaan digitointiin päädyttiin seuraavista syistä:

1. Elias Lönnrotin elämäntyöhön perustuva kokoelma on osa kansallisesti ja kansainvälisesti merkittävää kulttuuriperintöä.
2. Kokoelman kysyntä ja asiakaskäyttö on ollut suhteellisen vilkasta.
3. Kokoelman laajuus ja heikko kunto on vaikeuttanut sen tutkimuskäyttöä.
4. Niin ikään käytettävyyttä on heikentänyt originaalitekstien vaikea luettavuus sekä aineistosta kuvattujen mikrofilmien heikohko laatu (ongelmana tarkkuus, terävyys ja erityisesti harmaasävyisyys).
5. Digitoinnin myötä voidaan historiallisista syistä fyysisesti eri paikoissa säilytettävät aineistot yhdistää virtuaalisesti.
6. Arkistojen tietojärjestelmiin tehtävien uudistusten jälkeen on vanhoja käsikirjoitusaineistoja mahdollista hakea ja tarkastella digitaalisina faksimile-laatusina väritallenteina (näköissidokset).

Avainasemassa käytännön valintaprosessissa oli Kirjallisuusarkiston erikoistutkija Raija Majamaa, tosin myös muilla projektin tutkijoilla oli mahdollisuus esittää omia toivomuksiaan digitoitavaksi aineistoksi (ks. liite 3). Kesään 2002 mennessä Kansanrunousarkiston ja Kirjallisuusarkiston Lönnrotiana-kokoelmasta digitoitiin arviolta 25 000 sivua ja aineistosta tuotettiin eri käyttötarkoituksiin yli 100 000 kuvatiedostoa. KRA:n osalta Lönnrotin aineistot eli sidokset 1–30 skannattiin kokonaisuudessaan (n. 3300 sivua). Kaikkiaan kokoelma sisältää arviolta 80 000 sivua ja/tai liuskaa. Digitoitu aineisto käsittää mm. Kalevalan ja Kantelettaren käsikirjoituksia, Lönnrotin keräämää tai hänen hallussaan ollutta kalevalamittaista lyriikkaa, epiikkaa ja loitsurunoutta, satuja, arvoituksia ja sananparsia sekä matkakertomuksia ja kirjeenvaihtoa.

Erityistä kiinnostusta aineistojen kriittiseen valikointiin ja tulevaan tutkimuskäyttöön osoitti yhteistyökumppaneista Kalevala-instituutti⁹ ja professori Lauri Honko, jonka pyynnöstä digitoidusta aineistosta toimitettiin systemaattisesti käyttökopiot myös ko. instituutille. Käytännössä käsikirjoitusten skannaus, teknisten metatietojen luettelointi sekä käyttökopioiden luominen aloitettiin lokakuussa 2001. Raportin liitteenä on digitoitujen aineistojen luettelo, joka sisältää tietoja juoksevasti numeroitujen sidosten keskeisisällöstä sekä laajuudesta perinneyksikköinä, sivuina/liuskoina ja tiedostoina (ks. liite 1). Aineistoluettelo digitoitiin ja osittain toimitettiin Aarne Anttilan teoksessa julkaistusta luettelosta, joka sisältää jossain määrin virheellisiä ja puutteellisia tietoja¹⁰. Todettakoon, ettei digitoitiprojektiin tulisi sisällyttää arkistollisten luettelointi-, sisältö-, viite- yms. perustietojen tarkistusprosesseja. Projektin suunnitteluvaiheessa onkin syytä arvioida vanhojen ja mahdollisesti päivittämättömien manuaalisten arkistotietojen sovellettavuus ja käytettävyyttä mm. organisaation tietojärjestelmissä ja sähköisissä asiakaspalvelussa. Itse asiassa manuaalisten hakemistojen, kortistojen ja luetteloiden muuttaminen sähköiseen muotoon voi olla suhteellisen helppo ja nopea työvaihe verrattuna tietojen systemaattiseen uudelleen tarkistamiseen originaaliaineistosta.

⁹ <http://kalevala.utu.fi/>

¹⁰ Anttila 1985, 594–612.

Kalevala-insituutin kanssa käydyissä keskusteluissa koettiin tarpeelliseksi temaattisesti toisiaan tukevien erilaisten aineistotyyppien saattaminen tietoverkkojen kautta virtuaalisesti yhteen. Tämä ei tarkoita sitä, että eri tavoin muodostetut digitaaliset aineistot olisivat käytettävissä yhden ja saman tietokannan kautta. Sateenvarjo-ajattelun mukaisesti spesifisiin käyttö-tarkoituksiin räätälöidyt tietokannat sekä laajemmat digitaaliset arkistot on mahdollista saattaa yhteen erilaisin viittauksin ja linkityksin. Näin tiedonhakijalla on käytettävissään suhteellisen selkeät kartat ja opasteet tietyn aihepiirin kartoittamiseksi tai jo olemassa olevan tiedon syventämiseksi. Seuraava KULDI-digitointityöryhmän loppuraportin sitaatti tukee edellä kuvattua ajattelua: "Aineistot voidaan linkittää erikseen sovittavien teemojen ympärille suuremmiksi kokonaisuuksiksi, sillä digitaaliset aineistot ovat toden teolla hyödynnettävissä vasta kun kokoelmien massa on riittävän laaja ja kattava."¹¹ Mainittakoon, että suuria lupauksia herättävän semanttisen webin kautta on mahdollista löytää toisiinsa liittyvää mielekästä informaatiota heterogeenisistä aineistomassoista, joilla ei ole edellä kuvatun kaltaista valmista yhteistä nimittäjää. Raportissa käsiteltävän projektin näkökulmasta yhteisenä nimittäjänä toimii kalevalamittainen folklore sekä siihen liittyvä historiallinen keruu- ja julkaisutoiminta erityisesti Elias Lönnrotin osalta. Tätä kautta luonteva temaattinen kombinaatio on löydettävissä esimerkiksi SKVR-tietokannan, Kalevala-tietokannan ja Elias-kirjetietokannan väliltä. Kalevala-tietokanta on Kalevala-insituutin aiempaa tuotantoa; kirjetietokanta on puolestaan projektissa kehitelty yhteistyöhanke, jonka pilottivaihe toteutettiin osittain Kalevala-insituutin erillis-rahoituksella (Alfred Kordelinin säätiö ja Turun yliopistosäätiö).

Valitettavasti vuoden 2004 loppuun mennessä edellä mainituista tietokannoista ainoastaan SKVR-tietokanta oli rajallisesti tutkijoiden käytettävissä Internetin välityksellä (CSC:n tutkijapalvelut)¹². Alkuvuodesta 2005 SKVR-tietokanta siirrettiin SKS:n palvelimelle ja tietokantaa kehitetään osana seuran tuottamia ja ylläpitämiä avoimia verkkopalveluita. Professori Lauri Hongon äkillinen poismeno kesällä 2002 vaikutti osaltaan projektin toimintaan ja digitoituja aineistoja ei ole toistaiseksi kyetty hyödyntämään täysimittaisesti. Lähitulevaisuudessa yhteistyöhankkeiden edellytyksiä ja mahdollisuuksia tullaan selvittämään mm. Elias-kirjetietokannan viimeistelemiseksi ja saattamiseksi avoimeen verkkokäyttöön. Lisättäköön, että tutkijan tms. tiedonhakijan näkökulmasta olennaista on aineistojen luotettavuus, käytettävyys ja saatavuus eikä esimerkiksi sähköisten palvelujen organisatoriset tuottavuuskysymykset.

KÄYTÄNNÖN TYÖPROSESSIT

Seuraava esitys sisältää projektissa suunnitellun ja toteutetun digitointiprosessin keskeisimmät työvaiheet (ks. liite 5). Hankkeessa tuotetut digitaaliset tietovarannot mahdollistamaan erilaisia tapoja hyödyntää arkistoaineistoja SKS:n tulevien toimintojen kannalta. Huomattakoon, että 1800-luvulla käsinkirjoitetut aineistot eivät alkuperäisessä muodossaan sovellu merkki-muotoiseen tekstintunnistukseen. Vapaatekstihaut mahdollistavien tietokantaratkaisujen edellytyksenä onkin, että tiettyyn aineistokokonaisuuteen sisältyvät tekstit on kirjoitettu puhtaaksi koneella tai että ne sisältyvät johonkin julkaisuun tai julkaisusarjaan, jolloin ne voidaan tunnistaa ohjelmallisesti merkkimuotoon (OCR). Molemmista tapauksista on projektin osalta myönteisiä kokemuksia, tosin tietyn resurssointivarauksin (ks. Elias-kirjetietokanta ja SKVR-tietokanta).

¹¹ Kulttuuriperintö tietoyhteiskunnassa 2004, 16.

¹² <http://www.csc.fi/suomi/tutkija/index.phtml.fi>

Hankkeen ensimmäisen vaiheen selkeänä ja konkreettisenä tavoitteena oli tuottaa priorisoidusta osasta Lönnrotiana-kokoelmaa pitkäaikaistallennukseen soveltuvat TIFF-tiedostot sekä JPEG-katselukuvat. Henkilöresurssien ja skannauslaitteistojen käyttöä suunniteltiin lähtökohtaisesti ko. tavoitteen saavuttamiseksi. Käytännössä käsikirjoitusten skannaus ja nauhavarmistus suoritettiin samassa pc-työasemassa, josta kuvatiedostot siirrettiin verkon kautta jatkokäsiteltäväksi toiselle työasemalla. Digitointiprosessin sujuvuus edellyttää tässä laajuudessa esitettyinä kahden digitoijan tai tutkimusapulaisen kokopäiväistä työpanosta. Työn mielekkyyden ja motivaation säilyttämiseksi on suositeltavaa jakaa ja jaksottaa työtehtäviä siten, että työntekijät voivat soveltuvien osin osallistua digitointiprosessin eri työvaiheisiin. Tämä varmistaa osaltaan sen, etteivät digitointilinjan päivittäiseen toimivuuteen liittyvät tietotaidot ole liiaksi henkilösidonniaisia.

Originaaliaineistojen skannaaminen

- Käsikirjoitusten valmistelu (fyysisen käsittelyn ja siirron järjestäminen)
- Käsikirjoitusten digitointi (skanneri ja kirjahissi)
- Mastertiedostot (TIFF, RGB 24-bit, resoluutio 300 dpi)

Kuvatiedostojen käsittely

- Tiedostojen nimeäminen sarjoiksi (kokoelma, sidos, signum)
- Laadun tarkistaminen ja mahdolliset kuvakorjaukset

Massakonvertointi

- Konvertoimalla tuotetut käyttökopiot (JPEG)
- Käyttökopioita voidaan hyödyntää tietokannoissa ja Web-sivustoilla

Massatallennus

- Eri tietovälineille tallennettavat JPEG- ja TIFF-tiedostot
- Käyttökopiot tallennetaan lisäksi SKS:n palvelimelle

Nauhavarmistus

- Edelliseen liittyen tiedostojen pitkäaikaissäilytys (mastertiedostot)
- Tandberg SLR 100 (TAR-formaatti)

Metatietojen luettelointi

- Digitointiprosessissa syntyvien metatietojen dokumentointi (ks. Metatiedot)
- Sidoskohtaisten aineistoluetteloiden luominen (sisältö + tiedostonimet)

Tekstintunnistus (optionaalinen)

- Tekstien oikoluku ja editointi
- Mahdollinen XML-koodaus

Mainittakoon, että originaalikäsikirjoitusten lisäksi projektissa skannattiin eräitä Lönnrotin teoksia SKS:n kirjastosta. Ensimmäisen työaseman kovalevyn tuhouduttua menetettiin kuitenkin kyseiset kuvatiedostot lukuunottamatta nuottikirjaa Suomalaisten Runojen ja Laulujen Nuotteja. Julkaisuja ei skannattu uudelleen, sillä kirjojen skannaus eräänlaisena välityönä laiteongelmien keskellä ei ollut projektin kannalta välttämätöntä. Kaikki kirjat oli tallennettu JPEG-formaattiin.

Toimivan digitointilinjan muodostaminen vuosina 2001–2002 oli varsin ongelmallista. Tähän olivat syinä:

1. Riippuvuus laitetoimittajista (aikatauluongelmat).
2. Uusiin laitteisiin liittyvät puutteet ja suoranaiset viat.
3. SLR-nauhavarmistuksen tekninen hitaus (tallennus ja purkaminen).
4. Digitoitavan käsikirjoitusaineiston vaihteleva laatu (konservointitarve ja sidontatekniikka).
5. Projektin työntekijöiden puutteelliset tietotaidot digitointiin liittyvissä lukuisissa teknisissä yksityiskohdissa. Toisaalta merkittävimmät ongelmat eivät liittyneet projektityöntekijöiden varsinaiseen tehtäväkenttään eli aineiston käsittelyyn, skannaukseen ja tallennukseen. Rutinoituneiden työkäytänteiden osalta digitointilinja oli lopulta varsin toimiva.

AINEISTOJEN SKANNAAMINEN

Projektin käyttöön hankittiin syksyllä 2001 saksalaisvalmisteinen skanneri-kirjahissi -yhdistelmä (SMA 21). Kyseinen laitteisto soveltuu sitomattoman sekä kooltaan pienehköjen sidoksien tms. kirjamuotoisen materiaalin digitointiin. SKS:n arkistojen tapauksessa suurimmat sidokset ovat foliokokoa ja ne voidaan suhteellisen turvallisesti digitoida kyseisellä laiteyhdistelmällä. Periaatteena on, että aineiston kunto ja käsiteltävyys tulee arvioida aina tapauskohtaisesti sekä olemassa olevien kuntokartoitusten pohjalta. Mainittakoon, että laitteiston hankintapäätökseen vaikuttivat olennaisesti taloudelliset seikat, sillä suurten dokumenttien laadukkaaseen ja turvalliseen digitointiin käytettävät tarkat digitaalikamerat ovat tasoskannereihin verrattuna selvästi kalliimmassa hintaluokassa. Niin ikään digitaalikameran käyttöönotto, ylläpito ja yleinen laadunvalvonta vaatii digitointiin erikoistuneen reprokuvaajan tietotaitoja. Digitaalikameran hankinta tulee kysymykseen ainoastaan suuremmissa muistiorganisaatioissa, joilla on digitoitikeskuksissaan käytettävissä useita digitointilinjoja erityyppisille materiaaleille (mm. asiakirjat, kartat, valokuvat ja mikrofilmit).

Kirjahissia käytettäessä skannattava analoginen kohde nostetaan mekaanisesti tasoskannerin lasialustaa vasten. Nostoalusta on vaihdettavissa kohteen koon ja paksuuden mukaan yksi- tai kaksiosaiseksi. Kaksiosaisen alustan molemmat puolet liikkuvat keinurakenteen ansiosta vapaasti eri korkeustasoille. Näin eri paksuiset sidokset lepäävät skannauksen aikana alustalla mahdollisimman luonnollisessa asennossa eikä sidoksia repiviä tai murtavia jännitteitä pääse syntymään. Niin ikään kirjahissin puristus-paine on säädettävissä portaattomasti siten, etteivät vanhat kirjat, sidotut aineistot tms. dokumentit vahingoitu. Joissakin tapauksissa puristus-paineen tulee olla varsin kova, jotta sidos saadaan levitettyä lasialustaa vasten mahdollisimman tasaiseksi. Näin vältetään skannerin lukupään tarkennus- ja syväterävyysongelmilta, valoituksen epätasaisuuksilta (varjo- ja katvealueet) sekä tekstin taittumiselta tai katoamiselta tiukan sidoksen keskelle. Arkistoaineistojen kohdalla ei voida suositella sidoksessa keskelle kääntyvien tekstirivien automaattista oikaisemista, joka on optiona kehittyneemmissä skannaus- ja kuvankäsittelyohjelmistoissa. Työntekijöiden digitoitinkoulutuksessa tulee käyttää harjoitus-aineistoja, joilla voidaan vapaasti kokeilla erilaisia työtapoja ja -menetelmiä sekä demonstroida potentiaalisia virhetilanteita.

Periaatteessa tasoskannerin skannausalueen tulee olla yhtä suuri kuin skannattavan dokumentin, jotta kohde voidaan asettaa skanneria vasten mahdollisimman tasaisesti ja digitoitavaa aluetta ei tarvitse keinotekoisesti rajata. Vaikka sidoksen aukeamalta rajattaisiin ja skannattaisiin molemmat sivut erikseen, on digitointi parempi ja samalla nopeampi tehdä kerran paikoilleen asemoidusta kohteesta. Mainittakoon, että A3-skannerin maksimaalinen

skannausalue on 30 cm x 44 cm, joten sitä ei voida suositella foliokokoisten sidottujen dokumenttien digitointiin (noin 22,5 x 36,5 cm). Joissakin skannausohjelmistoissa on käytettävissä sidoksen aukeaman automaattinen jakaminen ja tallentaminen erillisiksi tiedostoiksi. Projektissa saadun kokemuksen perusteella manuaalinen rajausta on paras vaihtoehto laadukkaaseen digitointiin, jolloin jokainen sivu tms. digitoitava yksikkö muodostaa oman harkitun kokonaisuutensa. SKS:n kirjallisuusarkistossa on esimerkiksi sidoksia, jotka sisältävät vaihtelevan kokoisia itsenäisiä dokumentteja, jotka on rajattavissa erillisiksi skannausalueiksi. Näin kuvatiedostojen laatu, tarkkuus ja yksityiskohtien tarkasteltavuus paranee merkittävästi. Vastaavasti voi olla tarpeen rajata skannausalueeseen mukaan ns. tiukoista sidoksista osa aukeaman toisesta sivusta, jotta digitoitava kohde havainnollistuu paremmin ja tekstillä tms. kuvasisällöllä on tiedostossa riittävästi marginaalia. Kyseinen digitointimenetelmä ei luonnollisesti sovellu käsikirjoitusten pitkälle automatisoituun massadigitointiin, jossa tiedostoja tuotetaan kerran alustalle asemoidun sidoksen sivuja kääntämällä ja pedaalia tms. painiketta painamalla.

Skannerin tekninen kuvanlukuaika on ainoastaan yksi osatekijä koko digitointiprosessin kestoa arvioitaessa, eikä se hidastanut missään vaiheessa projektin suunniteltua aikataulua. Digitoitavan originaaliaineiston pienehkö koko ja skannauksessa käytetty resoluutio (300 dpi/ppi) mahdollistivat yksittäisten kuvanlukuaikojen säilymisen keskimäärin alle 15 sekunnissa. Tulee huomata, että päivittäiset skannausmäärät ja mahdolliset tulostavoitteet tulee suhteuttaa työergonomiaan, lepotaukojen tarpeeseen sekä yleisemmin työssä jaksamiseen (työn mielekkyys). Massadigitoinnin tuottavuutta päivä-, viikko- ja kuukausitasolla tulee arvioida teknisten seikkojen ohella myös monelta osin mekaanisen työprosessin uuvuttavuuden ja kuluttavuuden näkökulmasta. Suositeltavaa olisi, että digitoijan toimenkuvaan sisällytettäisiin soveltuvin osin muita työtehtäviä, jotka eivät suoranaisesti liittyisi skannerin käyttöön ja originaaliaineiston käsittelyyn (esim. digitoidun aineiston luettelointi tietokantaan). Suuremmissa yksiköissä digitointiprosessiin osallistuvat työntekijät voivat periaatteessa vaihtaa työtehtäviään sovitun kiertojärjestelmän puitteissa. Tosin projektissa saatujen kokemusten perusteella työtehtävien vuorottelemisen ja ns. toisen töiden jatkaminen lisää työryhmän sisäisen kontrollin ja tiedonvälittämisen tarvetta (kuka, mitä ja milloin).

Kuvatekniset laatuvaatimukset esimerkiksi resoluutio ja bittisyvyys tulee määrittellä suhteessa projektin tai digitointiohjelman tavoitteisiin ja tallenteiden suunniteltuihin käyttötarkoituksiin. Laadukriteerit tulisi arvioida aina myös tapauskohtaisesti, jolla tarkoitetaan tässä yhteydessä digitoitavaksi valittua aineistokokonaisuutta (mm. aineiston laajuus, värien käyttö, originaalikoko ja yksityiskohtien tarkkuus). Niin ikään teknisluonteisiin valintoihin vaikuttavat keskeisesti digitointiin käytettävissä olevat henkilö- ja laiteresurssit (digitointinopeus) sekä tiedostomassojen jälkikäsitteilyyn ja säilytykseen liittyvät kriittiset kysymykset. Todettakoon, että väriprofiilien käyttämättömyys on toistaiseksi estänyt parhaan mahdollisen laatuasteen saavuttamisen erityisesti tilaustöiden osalta. Toisaalta Elias-digitoinnin laitteistoilla luodut värikuvatiedostot soveltuvat nykyisellään erinomaisesti arkistojen peruskäyttötarkoituksiin.¹³

Värihallinnan onnistumisen perusedellytys on kaikkien prosessin osien toiminnan vakiointi. Laadukkaan lopputuloksen saavuttamiseksi on tärkeää ylläpitää vakaata ja määrääjain säädettyä laiteympäristöä. Kuvalaadun ohjelmalliseksi ja silmämääräiseksi kontrolloimiseksi on olennaista skannerin, näyttöpäätteen ja tarvittaessa myös tulostimen toistuva kalibrointi sekä standardoitujen testikarttojen ja ICC-profiilien käyttö¹⁴. Kontrolloidussa kuvadigitoinnissa ja

¹³ Blatner et al. 2004.

¹⁴ ICC on kansainvälinen standardi (International Colour Consortium), jonka avulla eri laitteiden ja järjestelmien väriavaruudet sovitetaan toisiinsa mahdollisimman tarkasti (<http://www.color.org/index.html>).

kuvankäsittelyssä testikartat eli -targetit toimivat laskennallisina kiintopisteinä epävakaissa ja vaihtelevissa laite- ja ohjelmistoympäristöissä. Eli kaikilla targetin yksittäisille värisävyille on olemassa numeeriset arvot, joihin skannerin samasta kohteesta skannaamia sävyarvoja voidaan verrata. Profiili (tiedosto) kuvaa tietyn laitteen mahdollisuuksia toistaa värejä tunnetuissa ja vakioiduissa olosuhteissa. Painoalalla profiilien avulla pyritään varmistamaan värien toistuminen oikein tuotantoprosessin kaikissa vaiheissa. Käytännössä profileja voidaan luoda skannerille, digitaalikameralle, monitorille, tulostimelle, vedostimelle sekä painokoneelle.¹⁵

Digitointiprosessin yhteydessä ICC-profiileilla voidaan suhteellisen helposti hallita laitteiden tekemiä virheitä ja vääristymiä. Profiilit otetaan käyttöön joko skannerin käyttöliittymässä tai Photoshop-kuvankäsittelyohjelmassa. Tosin ongelmana säilyy yhä edelleen laiteympäristön jatkuva ”eläminen” eli muuttuminen, joka vaatii laitteiden tarkastamista ja säätämistä periaatteessa kuukausittain (monitorit). ”Jokainen samanmallinen ja -merkkinen skanneri on sävyntoistoltaan erilainen. Tähän vaikuttaa mm. lamppujen ikä, lasien likaisuus ja käytettävät asetukset. Skannerin profiloinnilla saadaan yleensä erinomaisia tuloksia varsinkin edullisissa ja keskihintaisissa laitteissa. Sävyt toistuvat originaalin kaltaisina yhdellä skannauskerralla ja samoja asetuksia käyttämällä. Ajansäästö skannauksessa - ja varsinkin kuvien korjailussa - on merkittävä. Skannerin profiloinnin yhteydessä profiloidaan yleensä myös näyttö.”¹⁶ Tässä yhteydessä on lisättävä, että faksimile-laatuun pyrittäessä profiilit tulisi määrittellä erikseen eri aineistotyypeille eli yksi ja sama korjausprofiili ei riitä kaikkiin haastaviin skannaustarkoituksiin.

Nykyisin useat graafisen alan yritykset tarjoavat värinhallintapalveluita ohjelmistojen, mittalaitteiden, kurssien, tukipalveluiden ja konsultaatioiden muodossa. Digitointia harjoittavassa muistiorganisaatiossa tavoitteena tulee pitää laiteympäristön omaa hallittavuutta eikä tukitoimintojen ulkoistamista voida suositella pysyvänä ratkaisuna. Laadunvalvonta kuuluu tältä osin luontevana osana kuvadigitoijan toimenkuvaan ja mahdollisiin digitointiprosessissa tapahtuviin laatumuutoksiin tulee reagoida välittömästi eikä esimerkiksi ohjeellisen määräaikaistarkastusten yhteydessä. Mainittakoon, että monessa mielessä ongelmallinen ja työllistävä värinhallinta on kohdannut organisaatioissa myös vastustusta. Värinhallinnan specialisti Bruce Fraser on sanonut ymmärtävänsä kielteiset asenteet, sillä kuka haluaa työyhteisössä ottaa entisen reprokuvaajan työn ja vastuun. Fraser ennusti vuonna 2003, että värinhallinta saadaan kuluttajille sopivaan muotoon vasta 15 vuodessa.¹⁷

TIEDOSTOJEN NIMEÄMINEN

Virallisesti ohjeistetun nimeämiskäytännön tulee olla sellainen, että se mahdollistaa tiedoston ja sen originaalilähteen selkeän ja yksiselitteisen kytkemisen toisiinsa. Tiedostonimet tulisi johtaa arkistosignumeista tai muulla tavoin tiedostot selvästi identifioivista tunnuksista ja lyhenteistä. Esimerkiksi Lönnrotiana-kokoelmaa sidoksittain ja säilytysyksiköittäin digitoitaessa tiedostonimet muodostettiin arkistojen vakiintuneesta kokoelmalyhenteestä, sidoksen numerosta sekä tiedostojen juoksevasta teknisestä numeroinnista (LNA01_0001). Verkkoympäristössä tiedostonimiin voidaan lisätä organisaatiota koskevat ylätasen tiedot, jotka viittaavat säilyttävään arkistoon sekä toisaalta aineiston omistajuuteen ja/tai tekijänoikeuksiin (SKS_KRA). Lönnrotin kirjeenvaihdon osalta tiedostot nimettiin Kirjallisuusarkiston käytänteiden mukaisesti:

¹⁵ ICC-värinhallinta (<http://www.prepress-center.fi/icc/sivut/perusteeto1.html>) sekä Autiokari 1996–2005.

¹⁶ ICC-värinhallinta (<http://www.prepress-center.fi/icc/sivut/perusteeto6.html>)

¹⁷ Fraser, Murphy ja Bunting 2004.

- 1) Yksi kirje on saanut tunnisteeksi yhden numerosignummin eli kansioissa on juokseva numerointi kirjeittäin.
- 2) Kaikki kirjeen liuskat on numeroitu erikseen eli kansioissa on juokseva numerointi liuskoittain (myös tyhjät liuskat numeroitu).

Mikäli tiedostojen vaihdettavuus eri käyttöjärjestelmien välillä halutaan optimoida, tulisi tiedostonimen maksimipituus rajata 8 merkkiin. Tiedostonimen perusosaa seuraa pisteen jälkeen tiedostoformaatin ilmaiseva 3-merkinen laajennusosa (esim. TIF ja JPG). Ennen digitoinnin keskittämistä yhteen palvelupisteeseen tai keskukseseen eri osastojen tiedostojen nimeämisperusteet ja -käytänteet tulee selvästi ohjeistaa. Tämä ei kuitenkaan tarkoita aineistolähtöisten käytänteiden mekaanista yhdenmukaistamista.

Tiedostojen nimeämisen yleisiä periaatteita

Perussääntö: Tiedostonimen tulee yksilöidä tiedosto ja nimeämisperusteen tulee olla selkeän yksiselitteinen.

- Aiemmat nimeämiskäytänteet tulee dokumentoida, jotta myös vanhemmat sähköiset tallenteet ovat löydettävissä ja mahdollisilta epäselvyyksiltä ja ristiriitaisuuksilta välttyään.
- Käytössä olevista tunnisteista ja lyhenteistä tulisi olla erillinen ylläpitorekisteri (tiedostot, kansiot ja tietovälineet).
- Hierarkisia kansiorakenteita tulee välttää (tiedon vaikea hallittavuus) ja kansiot tulee nimetä sisältöä kuvaavasti (aineistokokonaisuus ja/tai sen osa). Tiedostojen nimeämisprosessia ei tule kytkeä kansiorakenteiden nimeämiseen. Eli siirrettäessä, kopioitaessa tai konvertoitaessa tiedostot uuteen ympäristöön informatiiviset kansiorakenteet eivät seuraa automaattisesti mukana.
- Tiedostonimien merkkimäärä tulee säilyttää kohtuullisissa rajoissa (luettavuus ja selailtavuus sekä tekniset rajoitukset). Periaatteessa tiedoston nimi voi sisältää enintään 255 merkkiä polkumääritykset mukaan lukien (Unix/Linux). Windows 2000 käyttöjärjestelmässä yläraja on 260 merkkiä.
- Käytännön esimerkki kirjapainoalan tiedostojen nimeämissäännöistä: Nimen pituus tunnisteineen korkeintaan 31 merkkiä. Rajoituksen perusteena on CD:lle ja DVD:lle tallennettavien tiedostojen nimien ohjelmistokohtainen enimmäispituus.
- Joissakin järjestelmissä/ohjelmissa ainoastaan tietty määrä merkkejä tiedoston nimen alusta on merkitseviä. Esimerkiksi "PaulaharjuS_kenttapiirroks_00001" ja "PaulaharjuS_071_00001" olisivat järjestelmän kannalta saman nimisiä tiedostoja, mikäli ainoastaan 11 ensimmäistä merkkiä olisivat merkitseviä. Erityinen huomio tulee kiinnittää helposti pitkäksi venyvien alkutunnisteiden muodostamiseen.
- Varsinaisen tiedostonimen (arkistolliset) tunnisteosat tulisi erottaa toisistaan alaviivalla (kirjain- ja numerotunnisteiden välillä alaviiva ei ole tarpeen).
- Tiedostot yksilöivän juoksevan numerotunnisteen perusteena voidaan käyttää eri aineistokokonaisuuksien sisältämien tiedostojen laskennallista enimmäismäärää. Numerosarjan pituus vaikuttaa tiedostojen oikeaan järjestymiseen mm. kansioissa ja erilaisissa listauksissa.
- Tiedostonimissä ei saa käyttää välilyöntejä, skandinaavisia fontteja eikä seuraavia merkkejä: * + = : ; " < > ? : / \ | -.
- Joukko lopputunnisteita on rekisteröity, ja niille on varattu tietty merkitys käyttöjärjestelmissä (esim. DOS, UNIX ja Windows). Näitä tunnisteita ovat muun muassa .BAT, .COM, .DLL, .EXE, .FIL, .HLP, .INI, .PFC, .REG, .WCM ja .WPT.

- Myös seuraavat laitenimet ovat kiellettyjä tiedostojen nimissä: CON, PRN, AUX, NUL, COM1, COM2, COM3, COM4, COM5, COM6, COM7, COM8, COM9, LPT1, LPT2, LPT3, LPT4, LPT5, LPT6, LPT7, LPT8 ja LPT9.
- Tiedostonimessä pisteen jälkeinen kolmikirjaiminen tunniste viittaa tiedoston tallennusmuotoon.
- Windows-ympäristössä ei ole merkitystä, onko tiedoston nimi kirjoitettu pienellä vai isolla. Sen sijaan Unix-maailmassa, johon esimerkiksi www-sivut yleensä siirretään, järjestelmä erottelee pienet ja isot kirjaimet. Epäselyyksien välttämiseksi on suositeltavaa käyttää ainoastaan pieniä kirjaimia.

METATIEDOT

Projektin tarkoituksena ei ollut luoda uusia rinnakkaisia standardeja ja ohjeistuksia koskien digitointimenetelmiä ja muistiorganisaatioiden metatietomäärittelyä¹⁸. Periaatteessa jokaisen organisaation tulee itse määrittellä omat tarpeensa ja edellytyksensä kansallisten ja kansainvälisten standardien noudattamiseksi sekä erityisesti niiden mielekkääksi soveltamiseksi. Seuraava metadatan eli metatiedon määrittely on vapaamuotoinen referaatti suomalaisesta Dublin Core -formaatin käyttöoppaasta¹⁹. Tässä yhteydessä ko. määrittelyn tarkoituksena on antaa yleiskäsitys siitä, mitä metadatalta tarkoitetaan ja mihin sitä käytetään.

Metadatalta tarkoitetaan tallenteen kuvausta eli tietoa tiedosta. Termin etuliite "meta" tulee kreikasta ja tarkoittaa muutosta. Metadata on näin tietoa dokumentin alkuperästä, muutoksista ja käytöstä. Metatietoa voidaan käyttää moniin tarkoituksiin: esim. tallenteen identifiointiin ja yhdistämiseen tiettyyn tiedontarpeeseen sekä dokumenttien muutosten tai käytön seuraamiseen. Eri tahoilla luodut metadata-standardit on suunniteltu ainakin osittain näitä tarkoituksia varten. Todettakoon, ettei Dublin Core-standardi sovellu erityisen hyvin esimerkiksi arkistolaitoksen määrittelemän rekisteriaineiston kuvailuun ja se on luotu alunperin verkkosivujen kuvailuun.

Metatiedot koostuvat joukosta elementtejä ja attribuutteja, jotka ovat tarpeen tallenteen kuvailussa. Metadata ja tallenne voidaan yhdistää kahdella eri tavalla:

- 1) Metadata-elementit sijaitsevat erillisessä tietueessa itsenäisenä kokonaisuutena (esim. arkistojen tietokannat ja tietojärjestelmät).
- 2) Metadata on osa tallennetta. Esimerkkinä upotetusta metadatasta voi käyttää CIP-luettelointia eli kirjaan on painettu sitä koskevat luettelointitiedot valmiiksi. Elektronisessa ympäristössä esimerkiksi sopii Text Encoding Initiative -kuvailu, joka on upotettu elektroniseen tekstiin. Useimmat metadata-standardit, mukaanlukien Dublin Core, eivät ota kantaa siihen, onko metadata tallennettu itsenäisenä tai erillisenä tietona vai osana tallennetta.

Metadata-standardien laajamittainen käyttöönotto tulee parantamaan tiedonhakuja erityisesti verkkoympäristössä ja se mahdollistaa metatietojen vaihdettavuuden ja siirrettävyyden eri tietojärjestelmien välillä. Jonkinasteiseksi ongelmalliseksi on muodostumassa eri organisaatioissa ja kansainvälisissä projekteissa luotujen standardien moninaisuus.

¹⁸ Arkistojen kuvailu- ja luettelointisäännöt 1997 sekä Sähköisten tietojärjestelmien ja aineistojen käsittely 2001.

¹⁹ Stenvall 2002.

Digitointiprosessin yhteydessä metatiedon voidaan käsittää olevan erilaisiin työvaiheisiin, tiedostoihin ja tietovälineisiin sekä laite- ja ohjelmistoympäristöön liittyviä teknisluontoisia tietoja. Digitaalisten aineistojen pitkäaikaissäilytyksessä metatiedoilla todennetaan eri tietovälineille tallennettujen tiedostojen alkuperä ja mahdollinen myöhempi elinkaari (kuka, miten ja milloin). Tätä kautta turvataan digitaalisten aineistojen käytettävyys ja hallittavuus eri yhteyksissä. Arkistoissa tekniset metatiedot tulisi voida syöttää tai siirtää suoraan arkistotietojärjestelmään tai sen yhteydessä olevaan ala- tai rinnakkaisjärjestelmään (moduulit). Näin originaaliaineistojen perinteisten kuvailutietojen ja digitaalisten tallenteiden välille voidaan luoda erilaisia loogisia kytkentöjä ja hakuominaisuuksia (mm. tietojen periytyminen).

Seuraavat metatietoelementit on lähtökohtaisesti muodostettu projektin ja digitointiprosessin käytännön tarpeisiin. Tässä vaiheessa kenttien nimet ja määrytykset ovat lähinnä viitteellisiä, sillä ne tullaan määrittelemään tarkemmin Kansanrunousarkiston ja Kirjallisuusarkiston tietokantauudistusten yhteydessä. Todettakoon, että arkistotietojärjestelmässä tulee jatkossa voida ilmaista millä tavoin syntyneestä tai tuotetusta digitaalisesta aineisto on kysymys. Perusjako tapahtuu tällöin organisaatiossa digitoidun tai ostopalveluna tuotetun aineiston sekä digitaalisena syntyneen aineiston välillä. Esimerkissä annetut kenttänimet ovat jälkikäteen helposti muutettavissa esim. XML-dokumenttien rakenteistamista ja standardointia varten. Edellä mainittuja tietojärjestelmiä kehitettäessä voidaan määrytykseen tarvittaessa lisätä uusia elementtejä, tämä koskee erityisesti aineistokokonaisuutta tai koko digitointiprosessia koskevia yleistietoja. Tekniset metatiedot tulee soveltuvin osin tallentaa sekä mastertiedostoista että niistä systemaattisesti valmistetuista käyttökopioista. Projektissa tuotetut luettelot käsittelevät erikseen SLR-nauhoille tallennettuja TIFF-tiedostoja ja CD:lle tallennettuja JPG-tiedostoja. Käytännössä digitaalisen aineiston luettelointi ja tallenteiden valmistaminen eri tietovälineille työllisti hankkeessa lähes kokopäiväisesti yhden työntekijän.

Toimivassa arkistotietojärjestelmässä digitaalisen aineiston teknistä osuutta edeltää ymmärrettävästi muistiorganisaation kuvailusääntöjen mukaiset sisältö-, identifiointi-, viite- ja käytettävyystiedot. Projektin kuluessa tutkija Pasi Klemettinen osallistui Kansanrunousarkiston arkistotietojärjestelmän kehittämiseen tähtävien työryhmien toimintaan. Pitkällä aikavälillä tarkasteltuna teknisillä metatiedoilla ja niiden aktiivisella hallinnoinnilla tulee olemaan tärkeä rooli tietojärjestelmissä. Digitaalisten tallenteiden tietosisällöllä ei ole suurtakaan merkitystä ellei tiedetä millaisia tallenteita ja tietovälineitä on olemassa, missä niitä säilytetään ja miten niitä voidaan käyttää.

Digitointiprosessissa syntyviä tietoja

- 1) *Arkisto / Projekti (vastuu- ja yhteyshenkilöt)*
- 2) *Originaaliaineiston tiedot*

Digitoitavasta aineistokokonaisuudesta voidaan antaa tunnistetiedot arkiston, kokoelman, sarjan, sidoksen ja/tai signumin tarkkuudella (esim. SKS KRA. Lönnrotiana 1:1-457). Muita mahdollisia relevantteja tietoja ovat mm. aineiston otsikko, aineistotyyppi (käsikirjoitukset), laajuus (sivumäärä), koko ja käyttöehdot. Tietojen tulisi periytyä automaattisesti arkistotietojärjestelmästä tms. viitetietokannasta.
- 3) *Toimenpiteet*

Yleiskuvaus toimenpiteistä ja menetelmistä, joilla digitointi suoritetaan (esim. linkki projektin raporttiin tms. dokumenttiin).

4) *Käsittelijä (digitoija)*

Henkilötietona annetaan skannauksen suorittajan ja/tai tiedostojen muokkaajan ja konvertoijan nimi. Tieto spesifoidaan tarvittaessa eri työvaiheiden kuvauksen yhteydessä.

5) *Käsittelyvaiheet*

Tietoja digitointiprosessissa olevien tallenteiden käsittelyvaiheista ja erityisesti jatkotoimenpiteitä vaativista aineistoista (esim. aineisto digitoitu ainoastaan osittain tai SLR-nauhavarmistus toistaiseksi tekemättä).

6) *Aikamääreet*

Aikamääreet on luonteva sijoittaa eri tietoelementtien ja käsittelyvaiheiden yhteyteen. Kenttätarkenteita: projektin kesto, digitointiajankohta, tiedostojen tallennus-, muokaus- ja konvertointipäivämäärät sekä eri tietovälineiden luontipäivämäärät.

7) *Laitte- ja ohjelmistoympäristö*

Keskeiset laitteet ja ohjelmistot, joita käytetään eri aineistotyyppien digitointiin (mallit, tyypit ja versiot).

8) *Tiedostojen tekniset tiedot*

Tiedostomuoto (formaatti ja versio)

Formaatti ilmaisee digitaalisen informaation tallennusmuodon. Digitaalisen tallenteen elinkaari alkaa master- eli päätiedoston luomisesta.

Datatyypit

Ilmaisee tiedoston tallennuksessa käytetyn värijärjestelmän ja värisyvyyden. Tieto annetaan tarvittaessa yksilöitynä eri formaateille.

Resoluutio

Tarkkuus, joka ilmaisee paljonko kuvassa on pikseleitä tai pisteitä tietyllä pituusmittayksiköllä (perusyksikkö tuuma).

Skaalaus

Kuvakoon prosentuaalinen muuttaminen ohjelmallisesti suhteessa alkuperäiseen tiedostoon. Toimintoa voidaan käyttää esimerkiksi näyttöpäätteelle parhaiten soveltuvien selailukuvien tuottamiseen (kuvien tiedostokoko pienenee ja samalla niiden selailtavuus nopeutuu).

Pakkaus (kompressointi)

Digitaalisen tallenteen pakkaamista käytetään muistitilan säästämiseen sekä tiedon nopeaan siirtämiseen ja hakemiseen verkkoympäristössä. Pakkausarvo ilmoitetaan ohjelmakohtaisesti (Photoshopissa arvoina 0-12).

Terävöitys

Terävöitystä käytetään kuvatiedostojen konvertoinnin yhteydessä tuotettaessa pienikokoisia mutta laadullisesti heikompia katselukuvia verkkokäyttöön. Terävöitystä ei suositella käytettäväksi skannausvaiheessa eli arkistoitavien mastertiedostojen tulisi olla käsittelemättömiä.

Väriprofiili (ICC-profiili)

Väriprofiili kuvaa työprosessissa käytettävien laitteiden kyvyn toistaa värejä.

Värihallinnan perusideana on korjata laitteiden aiheuttamia vääristymiä. ICC-profiili (ohjaustiedosto) voidaan liittää skannaamalla tuotettujen kuvatiedostojen yhteyteen.

9) *Tietoväline (formaatti ja tunnus)*

Fyysinen tietoväline, jolle tiedostot tallennetaan. Säilytyksen perusyksikkö, jolle annetaan digitaalisen aineiston käytön ja hallinnoinnin mahdollistava yksilöivä tunnus. Kenttään

voidaan lisätä maininta tietovälineen tallennuskapasiteetista. Tieto annetaan yksilöitynä eri formaateille (esim. käyttökopiot CD:llä).

10) *Tiedostolistaus*

Samassa digitoitierässä tuotettuja tiedostoja koskevat tiedot (esim. sidos tai kokoelma). Kenttä voi sisältää linkin tiedostolistaukseen, joka identifioi yksittäisten tiedostojen alkuperäisyyden (tiedostonimi, formaatti, tallennuspäivämäärä sekä koko kilotavuina ja/tai pikseleinä). Tällöin ko. tietoja ei tarvitse syöttää erikseen (manuaalisesti) ja tiedosto kerrallaan tietojärjestelmään.

11) *Säilytyspaikka*

Tallenteen fyysinen arkistointipaikka pitkäaikais-, tilapäis- ja/tai määräaikaissäilytyksessä. Samasta aineistosta on tavallisesti useita varmuus- ja käyttökopioita, joita säilytetään eri paikoissa. Kenttä on erityisen tärkeä aineiston hallinnoinnin ja käytettävyyden kannalta. Tallenteen tarkka sijainti voidaan antaa huoneen, salin, varaston, hyllyn tms. yksikön perusteella. Aineiston sijaintitieto tulisi antaa myös mahdollisen palvelinosoitteen perusteella. Kenttä voi sisältää tiedon vastuuhenkilöstä, joka hallinnoi digitaalista arkistoa tai sen alakokonaisuutta (esim. kuva-, käsikirjoitus-, äänite- ja videokokoelmat).

18) *Virkistäminen*

Seurantatietoja tiedostojen ja fyysisten tallennevälineiden määräajoin suoritettavasta virkistämisestä. Virkistäminen voi tapahtua tiedostoja uuteen formaattiin konvertoimalla sekä tallennevälineitä teknisesti käsittelemällä tai vaihtamalla.

Aineistoluettelot

Projektissa digitoiduista aineistoista tuotettiin seuraavat luettelot. Osa teksteistä luotiin tekstintunnistuksen ja tekstieditoinnin avulla (OCR). Kansanrunousarkiston osalta erilaiset luettelot, selvitykset yms. sähköiset dokumentit tultaneen soveltuvin osin liittämään KRA:n uudistettavaan arkistotietojärjestelmään.

Digitoitujen aineistojen perusluettelo (ks. [liite 1](#))

Teknisten metatietojen luettelot

- Tiedostoformaatit: TIFF ja JPEG

- Tietovälineet: SLR ja CD-R

Lönnrotiana 1-30: Tiedostokohtainen aineistoluettelo

- Tiedostot ja niiden sisältämät perinneyksiköt sidoksittain

Lönnrotiana 1-30: Originaaliaineiston kuntokartoitus (OCR)

- Konservoijan tekemä kuntokartoitus

Lönnrotiana 1-50: Perinnelajihakemisto (OCR)

- Kaisa-Marja Perttusen selvitys julkaisemattomasta SKVR-aineistosta²⁰

Tutkimuskäyttöön luovutettujen aineistojen luettelo

²⁰ Perttunen 1976–1977.

Digitoitujen aineistojen käytettävyydestä

Sähköisten palveluiden tuottamisen, tarjoamisen ja hyödyntämisen kannalta keskeisiä tietoja ovat arkistoaineistojen käyttöön liittyvät erilaiset säännöt, ehdot, sitoumukset, rajoitukset, tekijänoikeudet sekä julkaisuluvat. Arkiston yleiset käyttö säännöt koskevat koko arkistomateriaalia; tämän lisäksi yksittäisille aineistokokonaisuuksille voidaan määritellä erityisiä käyttöehtoja. Kansanrunousarkiston nykyisessä tietojärjestelmässä parametreinä toimivat:

1. ei käyttörajoituksia
2. vain tutkimuskäyttöön
3. luovuttajan tai oikeudenomistajan asettamat ehdot ja rajoitukset
4. julkaistavissa vain luovuttajan tai oikeudenomistajan luvalla sekä
5. suljettu kokoelma

Relaatiotietokantaan tallennettuja käytettävyystietoja voidaan muiden keskeisten tietojen tavoin tulevaisuudessa hyödyntää mahdollisessa digitaalisessa arkistossa (tietojen periytyminen). Periaatteena on, että arkiston asiakkaille aineistokopioita ja -jäljenteitä luovutetaan ainoastaan yhtä käyttötarkoitusta varten ja aineisto on palautettava sovittun määräajan jälkeen takaisin arkistoon (käyttö sidottu esim. julkaisu- tai tutkimushankkeen keston). Digitaalisten aineistojen palauttamisen ja tarpeettomien kopioiden hävittämisen hallinnointi tulee jatkossa muodostamaan oman haasteellisen osa-alueensa arkistotoimen kehittämisessä. Huomattakoon, että eri kokoelmia ja arkistosarjoja koskevat käytettävyystiedot tulevat niiden elinkaaren aikana muuttumaan ja esimerkiksi tekijänoikeudellisen suoja-ajan päätyminen mahdollistaa aineistojen avoimen ja vapaan verkkokäytön (70-vuotta tekijän kuolemasta). Tämä edellyttää luonnollisesti arkaluontoisten tietosuojasioiden muuten olevan hallinnassa.

Projektissa digitoitujen 1800-luvun arkistoaineistot ovat tekijänoikeusvapaita, eikä niiden verkkojulkaisemiselle ole olemassa erityisiä esteitä. Toisaalta arkistojen velvollisuutena on kontrolloida kaikkien hallinnoimiensa aineistojensa jatkokäyttöä. Tässä mielessä kuvatiestojen yms. tallenteiden kontrolloimaton kopiointi, levittäminen ja uudelleen julkaiseminen voi muodostua ongelmalliseksi. Verkkoympäristön osalta sähköisten aineistojen käyttöä ja erityisesti asiointia käyttöä voidaan pyrkiä valvomaan seuraavilla teknisillä menetelmillä:

1. Kuvatiedostoihin liitetään muistiorganisaation näkyvä vesileima tai tekijänoikeusleima.
2. Kuvatiedostoihin liitetään näkymätön digitaalinen vesileima, joka ilmaisee kuvan omistajuuden ja jonka avulla sen käyttöä voidaan jäljittää Internetissä.
3. Kuvatiedostot kryptataan (salakoodaus) ja ne voidaan avata verkon kautta ainoastaan rekisteröidyillä käyttäjätunnuksilla. Kyseinen menetelmä voi rajoittaa aineistojen kiinnostavuutta ja käyttöarvoa verkossa. Lisäksi menetelmä edellyttää aktiivista käyttäjärekisterin ylläpitoa ja valvontaa.
4. Kuvatiedostot julkaistaan verkossa matalaresoluutioisina katselukuvina (75 dpi), joiden käyttöarvo muihin tarkoituksiin on vähäinen (esim. tulostaminen).
5. Digitaaliset aineistot ovat ainoastaan tiettyjen yhteisöjen rekisteröityjen ja virallisten jäsenten käytettävissä (esim. tutkijayhteisöt ja muistiorganisaatiot).²¹

SKS:n sähköisiä palveluita tullaan lähivuosina kehittämään arkistoaineistojen esittelyiden ja verkkonäyttelyiden osalta (avoimet ja julkiset verkkoratkaisut). Toisena haasteellisena ja voimavaroja sitovana tehtävänä on kehittää räätälöityjä tietokantaratkaisuja, joissa

²¹ Good Practices Handbook 2004.

hyödynnetään mm. tekstikorpusten merkkimuotoisia hakuominaisuuksia kulttuurintutkimuksen erityistarpeisiin (esim. kalevalamittaisten runojen säetutkimus). Se kuinka laajalaiseen ja monimuotoiseen digitaalisten aineistojen hyödyntämiseen taloudelliset resurssit tulevat riittämään on avoin kysymys kaikissa muistiorganisaatioissa.

Digitointihankkeiden edistyessä haasteena tulee kaiken edellä mainitun ohella olemaan konekirjoitettujen viite- ja kopiokortistojen sisältämien eritasoisten tietojen siirtäminen digitaaliseen ympäristöön. Paradoksaalisesti vanhempien aineistojen tarkka haettavuus voi digitalisaation myötä jopa heikentyä, ellei aiemmin tehtyä arvokasta työtä voida täysimittaisesti hyödyntää. Mainittakoon, että yksinomaan Kansanrunousarkiston kortistot sisältävät noin 2 miljoonaa korttia. Arkistojen rajalliset resurssit huomioiden on selvää, ettei aiemmin manuaalisesti työstettyjä massiivisia aineistoja voida käsitellä alusta uudelleen nykypäivän tarpeiden ja kansainvälisten standardien mukaisesti. Arkistojen ja niiden käyttäjien on hyväksyttävä erilaiset tiedonhakumenetelmät, jotka ovat seurausta tietyistä historiallisista valinnoista ja kehitysprosesseista liittyen säilytettävien aineistojen erityispiirteisiin. Kansanrunousarkistossa hyvänä esimerkkinä spesifisestä tavasta jäsentää ja tuottaa tietoa toimii perinnelaji- ja tyyppiluokitus, joka kohdistuu yksittäisiin muistiinpanoihin ja perinnetuotteisiin. Edellä mainitut laajat kopiokortistot on muodostettu vuosikymmenten kuluessa alkuperäisestä tekstuaalisesta kontekstistaan erotetuista ja puhtaaksikirjoitetuista perinneyksiköistä, jotka on luokiteltu ja tilastoitu perinnelajien, kerääjien sekä alue- ja paikkakuntatietojen perusteella (ATK-listaukset). Rinnakkaisten ja parhaassa tapauksessa toisiaan täydentävien hakumenetelmien omaksumista ja hyödyntämistä voidaan luonnollisesti monin tavoin helpottaa ja edistää. Eräänä kustannusedullisena ja varsin nopeasti palveluita tuottavana vaihtoehtona voidaan pitää digitaalisten näköiskortistojen luomista kontrolloituun verkkoympäristöön, jolloin osittain jo huonokuntoisten manuaalikortistojen asiakaskäyttöä voidaan perustellusti rajoittaa. Aineistojen pitkäaikaissäilytyksen näkökulmasta manuaalikortistoista saadaan korkealaatuiset mastertiedostot, joita voidaan myöhemmin hyödyntää moniin eri käyttötarkoituksiin (mm. merkkimuotoiset tekstikorpukset ja tietokantaratkaisut).

TEKSTINTUNNISTAMINEN (OCR)

Lönnrotiana-kokoelman osalta merkkimuotoisina teksteinä oli mahdollista hyödyntää joitakin aineistoluetteloita, Lönnrotin kirjeenvaihtoa sekä julkaistuja ja julkaisemattomia SKVR-tekstejä (ks. Tietokantaratkaisut). Projektin alkuvaiheessa Pekka Allonen jatkoi Kansanrunousarkistossa aiemmin aloittamaansa työtä, jossa hän tunnisti ohjelmallisesti Elias Lönnrotin kuvamuotoon digitoituja runotoisintoja tekstitiedostoiksi (julkaisematon SKVR-kopiokortisto). Lönnrotin osuus kyseisestä digitaalisesta kopiokortistosta on 1865 kuvatiedostoa. Digitoitujen A5-kokoisten korttien tekstintunnistus tehtiin ABBYY Finereader -ohjelmalla. Tämän jälkeen raakatekstit oikoluettiin ja muokattiin yhdenmukaisiksi WP-teksteiksi, joista mm. poistettiin tekstien typografiassa ilmenevä korttimaisuus. Eli kortilta toiselle jatkuvat runotekstit yhdistettiin ja korttien ylälaidassa tarpeettomasti toistuvat arkistolliset metatiedot hävitettiin tarkastuksen jälkeen. Tämän ohella työprosessiin kuului WP-tekstien alustava rakenteistaminen XML-muotoon. Digitaalisen aineiston merkkimuotoisuus mahdollistaa tekstien systemaattisen rakenteistamisen tietokantaa varten XML-muodossa (eXtended Mark-up Language)²².

Suoritetuissa testeissä tekstintunnistusohjelman todettiin tekevän varsin paljon tunnistusvirheitä. Syynä tähän oli konekirjoitettujen merkkien huono näkyvyys ja lähekkäisyys

²² Extensible Markup Language (XML) 1996–2003 sekä Harold & Means 2001.

sekä erilaiset lyöntivirheet. Kuvamuodossa olevien merkkien optisen tunnistamisen kannalta on oletusarvon mukaista, että graafiselta ulkoasultaan toisiaan muistuttavat numerot ja isot/pienet kirjaimet sekoittuvat toisiinsa.

Oikolukuvaiheessa epäselvät ja epäilyttävät kohdat tarkistettiin soveltuvin osin alkuperäisistä käsikirjoituksista, sillä julkaisemattomia SKVR-tekstejä ei ole puhtaaksi kirjoittamisen jälkeen tarkastettu kuten julkaistuja tekstejä. Tekstien työstämisen yhteydessä havaittiin, että konekirjoitettujen tekstien ja alkuperäisten käsikirjoitusten välillä on paikoitellen suuria eroavuuk-sia. Kopiokorteissa havaitut virheet, sekaannukset ja puutteet pyrittiin korjaamaan suoraan digitoituihin teksteihin. Lisäselvitystä kaipaavat kohdat merkittiin kysymysmerkillä tai kommentteilla hakasulkeisiin. Tekstien rakenteistamisessa ja XML-koodauksessa noudatettiin soveltuvin osin samoja periaatteita kuin SKVR-aineiston kohdalla oli tehty Tartossa. Tekstien rakenteistamisen suunnittelusta on vastannut pääosin erikoistutkija Jukka Saarinen.

OCR-työskentelyyn liittyy vanhojen aineistojen osalta väistämättä eritasoisia teknisluonteisia ongelmia ja haasteita. Ennen varsinaista tekstintunnista on syytä tehdä valmistavia testejä skannauksen tarkkuuden ja merkkien tunnistettavuuden optimoimiseksi. Joissakin tapauksissa paperi voi olla esimerkiksi siinä määrin huokoista ja ohutta, että voimakas skannausvalo nostaa tarpeettomasti näkyviin sivun takana olevia merkkejä. Haittavaikutusta voidaan vähentää asettamalla systemaattisesti jokaisen skannattavan sivun taakse paperiarkki. Laajempien aineistokokonaisuuksien kohdalla tavoitteena tulee pitää mahdollisimman helppoa ja nopeaa oikolukuvaihetta, jolloin tarpeettomat virheet on karsittu pois jo ohjelmallisesti. Joissakin tapauksissa on suositeltavaa "kouluttaa" ohjelma tunnistamaan erilaiset merkit, fontit ja kirjasintyypit oikealla tavalla. Niin ikään tekstintunnistuksen tueksi voidaan tarvittaessa luoda uusi räätälöity taustakieli ja/tai sanakirja erikoismerkkeineen ja -sanastoineen.

Kokemusten perusteella samankin korpuksen sisällä testitulokset voivat vaihdella suuresti kerääjästä ja aineiston sisällöstä riippuen (esim. tarkemerkkien käyttö eri murteissa ja kielissä). Korostettakoon, että ongelmana on tavallisesti digitoitavan materiaalin heikko tai vaihteleva laatu eikä varsinaiset ohjelmalliset puutteet tai heikkoudet. Toisin erityisesti kielitieteellisten julkaisujen ja korpusten kohdalla myös perusohjelmien tekniset rajat tulevat nopeasti vastaan. Skannereiden mukana saatavia tekstintunnistusohjelmien ns. kevyt versioita ei voida suositella ammattimaiseen käyttöön.

Muistiorganisaatioissa tekstintunnistusta voidaan hyödyntää erityisesti konekirjoitettujen luetteloiden, hakemistojen ja kortistojen digitoinnissa. Tekstintunnistus mahdollistaa vuosikymmenten kuluessa eri tavoin tuotettujen dokumenttien hyödyntämisen mm. web-sovelluksissa ja tietokantaratkaisuissa, jolloin tietoa voidaan hakea vapaasanahaullla ja/tai rakenteistettujen tekstien tarkasti määritellyistä tietoelementeistä. Eri aineistotyypeillä testattujen ohjelmistojen ja laitteistojen osalta on tärkeää käytännön kokemusten ja tietotaitojen jakaminen organisaatioiden ja erilaisten sidosryhmien välillä (esim. SKS, Kotus ja Kalevala-instituutti). Sama tietojen jakamisen ja vaihdettavuuden vaatimus koskee luonnollisesti myös metatietoja ja digitointiprosessiin liittyviä käytänteitä.

Projektissa tuotettujen sähköisten aineistoluetteloiden pohjalta luotiin erilaisten XML-dokumenttien alustavia mallinnuksia (Klemettinen). Seuraava esimerkki käsittelee Lönnrotiana-kokoelman ensimmäisen sidoksen alkuosaa. Vastaava looginen tietorakenne on löydettävissä kaikista Kansanrunousarkiston vanhemmista sidoksista (konekirjoitetut etusivut). Tekstintunnistuksen ja XML-rakenteistamisen avulla voidaan kuvamuotoon digitoiduille aineistoille luoda suhteellisen helposti perusta tyydyttävälle hakujärjestelmälle (vrt. SKVR-korpus). Käsikirjoituksia koskevat perushaut voitaisiin näin kohdistaa kerääjä-, perinnelaji- ja

paikkakuntatietoihin. Ristiriitaiseksi tilanteen tekee se tosiseikka, että vastaavat perinnelaji-sidonnaiset tekstit on löydettävissä systemaattisesti luokiteltuina ja puhtaaksikirjoitettuina kopiokortistoista. Suoran linkityksen tekeminen kortistojen yksittäisten tekstien ja digitaalisten sidosten välille on käytännössä erittäin työlästä. Ihanneltilanteessa haku kohdistettaisiin mittavista kortistoista muodostettuihin tekstikorpuksiin (merkkijonohaut) ja hakutuloksien kautta olisi mahdollista tarkastella lähdekriittisesti varsinaisia originaalitekstejä ja niiden tekstuaalista kontekstia (värikuvatiedostot). Arkistoaineistojen kokonaishallinnoinnin kannalta digitoitavan tai jo digitoidun aineiston luettelointitiedot tulee jatkossa syöttää Kansanrunous-arkiston uudistettavaan arkistotietojärjestelmään, joka nykyisellään sisältää tietoja yksinomaan uudemmista keruuaineistoista.

Esimerkin loppuosassa kuvatiedostot ja sidoksissa juoksevasti numeroidut perinneyksiköt on linkitetty toisiinsa, tämä mahdollistaa jatkossa mm. erilaisten web-linkkisivustojen rakentamisen. Projektin kuluessa kaikki Kansanrunousarkiston Lönnrotiana-sidokset (1-30) luettelointiin kyseisellä menetelmällä. Arkiston kokoelmiin viitattaessa tulee mainita SKS ja Kansanrunousarkisto, kerääjä (luovuttaja), arkistosarja, sidoksen numero, muistiinpanon numero sekä vuosiluku. Arkistosarjaa S (Kerääjäsidokset) ei kuitenkaan merkitä erikseen. Lönnrotiana-kokoelman ensimmäiseen runotekstiin viittaava arkistosignum muodostetaan seuraavasti: SKS KRA. Lönnrotiana 1:1a (tiedosto: lna001_010.jpg).

```
<ITEM> (rakenteistettu yksikkö eli tässä tapauksessa sidoksen osa)

<SER>S</SER> (arkistosarja)
<COLLECTOR>Lönnrot, Elias</COLLECTOR> (kerääjä)
<COLLECTION> (kokoelma)
<NUM>Lönnrotiana 001</NUM> (sidos)
<PART>001a</PART> (sidoksen osa)
<UNITS>0001-0036</UNITS> (perinneyksiköt)
</COLLECTION>
<PERIOD>1828-1831</PERIOD> (aikarajaus)
<HEAD>Kantele taikka Suomen Kansan sekä Wanhoja että Nykysempiä Runoja ja
Lauluja. Wiides osa. (6) + 33 s.</HEAD> (otsikko)

<LOCATION> (paikkakuntatiedot)
<PARISH>Eno<UNITS>0008, 0012</UNITS></PARISH> (pitäjä)
<PARISH>Ilomantsi<UNITS>0030-0036</UNITS></PARISH>
<PARISH>Ilomantsi tai Tuupovaara<UNITS>0015?</UNITS></PARISH>
<PARISH>Juuka<UNITS>a: VI siv. 27-28</UNITS></PARISH>
<PARISH>Kajaanin mlk.<UNITS>0026</UNITS></PARISH>
<PARISH>Kerimäki<UNITS>0013, 0021?</UNITS></PARISH>
<PARISH>Kesälahti<UNITS>Ib, 0005, 0011, 0015, 0018</UNITS></PARISH>
</LOCATION>

<GENRE> (perinnelajit)
<a1> h: siv. 45 </a1> (kasku, vitsi, anekdootti, pilasatu, jokulaatti)
<a3> 25a </a3> (uskomustarina, memoraatti)
<b1> lb-3, 4a, 19-20, a: siv. 19-20, II (38) siv. 20-23, III (7) siv. 23-
25, IV siv. 25-27, V (44) siv. 27, VI siv. 27-28; 21, 22a-b, 23a-h, 24-
26, 31-37, 40-41, 43-66, 70-89, 90a-115, 117-121, 131, 133, 137-141, 149-
171, 173-174, 176-181, 183, 186, 188-206, 218, 223, 229-230, 234-237,
239a, 241, 251, 252, 254-258, 260-266, 269-270, 271-277, 278, 282-286,
291, 293, 293a, 295-300, 301-316, 318-320, 322-326, 329, 330a, 332-333,
335-336, 339, 351-352, 356, 358-359, 362-376, 380-404, 408, 1413-414,
415b-416, 420-425, 426-448, 450-452, 454-456 </b1> (kalevalamittainen
runo tai laulu)
<b2> 182, 184, 185, 187, 249, 259, 267, 268, 280, 290, 292, 317, 41 </b2>
(riimillinen kansanlaulu)
</GENRE>
```

```
<CONTENT> (tiedostonimet + arkistoyksiköt)
<LINK><FN>lna001_001.jpg</FN> <UNIT>Etukannen sisäpuoli</UNIT></LINK>
<LINK><FN>lna001_002.jpg</FN> <UNIT>Sidoksen etulehti</UNIT></LINK>
<LINK><FN>lna001_003.jpg</FN> <UNIT>Paikkakunta- ja perinnelajiluettelo
(1)</UNIT></LINK>
<LINK><FN>lna001_004.jpg</FN> <UNIT>Jatkoa (2)</UNIT></LINK>
<LINK><FN>lna001_005.jpg</FN> <UNIT>Jatkoa (3)</UNIT></LINK>
<LINK><FN>lna001_006.jpg</FN> <UNIT>Jatkoa (4)</UNIT></LINK>
<LINK><FN>lna001_007.jpg</FN> <UNIT>Sisällysluettelo</UNIT></LINK>
<LINK><FN>lna001_008.jpg</FN> <UNIT>a)-osan etulehti</UNIT></LINK>
<LINK><FN>lna001_009.jpg</FN> <UNIT>Nimiölehti</UNIT></LINK>
<LINK><FN>lna001_010.jpg</FN> <UNIT>Teksti 0001a; Esipuhe</UNIT></LINK>
<LINK><FN>lna001_011.jpg</FN> <UNIT>Esipuhe</UNIT></LINK>
<LINK><FN>lna001_012.jpg</FN> <UNIT>Sisällysluettelo; 0001b</UNIT></LINK>
<LINK><FN>lna001_013.jpg</FN> <UNIT>0001b-0004a</UNIT></LINK>
<LINK><FN>lna001_014.jpg</FN> <UNIT>0004a-0004c</UNIT></LINK>
<LINK><FN>lna001_015.jpg</FN> <UNIT>0004c-0006a</UNIT></LINK>
<LINK><FN>lna001_016.jpg</FN> <UNIT>0006b-0007</UNIT></LINK>
<LINK><FN>lna001_017.jpg</FN> <UNIT>0007-0010</UNIT></LINK> jne.
</CONTENT>

</ITEM>
```

AINEISTOJEN SUHTEELLISET DIGITOINTINOPEUDET

Kansanrunousarkistossa on digitoitu käsikirjoitusaineistoja ja niihin liittyviä kortistoja vuodesta 2000 lähtien. Digitoidut aineistot sisälsivät vuoden 2004 lopussa julkaisemattoman SKVR-kopiokortiston, osia tähtikokoelmasta (aineistot 1600-luvulta 1870-luvulle), Kansanrunousarkiston kirjeenvaihtoa sekä SKS:n julkaisuhankkeisiin liittyviä suppeahkoja aineistoja. Määrällisesti SKVR-kortiston osuus on noin 96 000 kuvatiedostoa ja muun digitoidun materiaalin 35 000 tiedostoa. Seuraavassa esitetään arvio eri aineistotyyppien suhteellisista digitointinopeuksista, laskennallisena perusyksikkönä toimii yhden kokopäiväisen digitoijan tasokannerilla tuottamien tiedostojen määrä työvuodessa. Työprosessi sisältää aineistomassan nimeämisen, järjestämisen ja tallentamisen sovittujen periaatteiden mukaisesti. Vuositasolla laskelmista on vähennetty kaksi kuukautta lomien sekä muiden työtehtävien ja tilaisuuksien perusteella (mm. kokoukset, neuvottelut ja koulutustilaisuudet). Aineistokohtaisia digitointimääriä voidaan radikaalisti nostaa erityisesti suuremmissa digitointikeskuksissa, joissa laiteympäristö- henkilöstö- ja työaikajärjestelyt voidaan optimoida. Esimerkiksi Ruotsin kansallisarkiston alaisuudessa toimivassa Media Konverterings Centrumissa työskentelee vuoden 2003 tiedon perusteella 70 henkilöä²³.

Julkaisemattoman SKVR-kortiston osalta arvio perustuu digitointihankkeessa vuosina 2000-2001 saatuihin käytännön kokemuksiin. Kyseisen työprosessin helpottamiseksi ja nopeuttamiseksi testattiin tasokannereihin (2) liitettäviä arkinsyöttölaitteita. Korttien paksuuden ja osittain heikon laadun seurauksena syöttövirheitä syntyi siinä määrin, ettei lisälaitteiden käyttöä katsottu enää mielekkääksi. Skannereiden tekninen lukunopeus ei ole kortistojen kohdalla avainkysymys vaan pitkäjänteiseen ja tavoitteelliseen digitointiin sitoutuvan työntekijän rekrytoiminen. Sama koskee pääsääntöisesti kaikkia muitakin "käsityönä" suoritettavia digitointeja.

²³ <http://www.mkc.ra.se/>

Digitointinopeuden arvioiminen tekstintunnistuksen yhteydessä on tapauskohtaista ja käytännössä erittäin vaikeaa. Projektissa tekstintunnistusta käytettiin ainoastaan joidenkin luetteloiden sekä testimielessä julkaisemattoman SKVR-aineiston digitoinnissa. Yhteistyöhankkeeseen liittyen Kotuksessa on tunnistettu merkkimuotoon Elias-digitoinnissa skannattuja kielitieteellisiä aineistoja ja julkaisuja. Käytännössä sujuva OCR-työskentely edellyttää konekirjoitetuilta tai tulostetuilta teksteiltä hyvää laatua (mm. merkkien tarkkuus, voimakkuus ja selvärajaisuus). Joidenkin materiaalien kohdalla tekstien uudelleen kirjoittaminen saattaa olla nopeampi vaihtoehto kuin niiden ohjelmallinen tunnistaminen ja oikolukeminen (kortistot, sisällysluettelot ja originaaliaineistot). Teknisesti tarkasteltuna OCR on suhteellisen nopea työvaihe ja tunnistusvirheiden määrä on oletusarvoisesti vähäinen. Tämä ei kuitenkaan poista sitä tosiseikkaa, että ko. digitointimuoto sitoo voimakkaasti työvoimaresursseja. Mainittakoon, ettei perinnearkistoissa ole juurikaan käyttöä automaattisen tekstintunnistuksen kautta varsin nopeasti tuotetuille ja virheitä sisältäville ns. raakatekstikorpuksille, joita voidaan käyttää hyväksi mm. laajojen sanomalehtiaineistojen sisältöön kohdistuvassa vapaasanahaussa²⁴. Pitkälle automatisoituun ja laajamittaiseen tekstintunnistukseen eri aikoina, eri menetelmin ja eri tarpeisiin tuotetut originaalikäsikirjoitukset ovat yksinkertaisesti liian haasteellisia.

Elias-digitoinnissa saatiin myönteisiä kokemuksia myös erilaisten julkaisujen skannauksesta (SKS:n ja Kotuksen erillistyöt 2002-2004). Kotuksen osalta digitoinnit liittyivät merkkimuotoisten tutkimusaineistojen luomiseen sekä toisaalta kielitieteellisten teosten uudelleenjulkaisemiseen. Mikko Virtanen Kotuksesta digitoi 28.08.2003 Helsingin kaupungin rakennusviraston asiakirja-aineistoa A4-skannerilla (Epson 1200). Skannerissa oli lisälaitteena arkinsyöttölaite, joka mahdollisti noin 200-sivuisen tasalaatuisen erän häiriöttömän digitoimisen kahdessa tunnissa. Vastaavankaltaista edullista ja nopeaa peruslaitteistoa on mahdollista hyödyntää Kansanrunousarkiston uutta sitomatonta aineistoa digitoitaessa. Tällöin skannaus voidaan pääsääntöisesti suorittaa harmaasävyisenä, jolloin säästetään järjestelmien muistitilaa ja vältetään värienhallintaan liittyvältä problematiikalta. Digitoinnin tavoitteena on tällöin kuvamuodossa olevien sähköisten asiakirjojen luominen (mm. verkkokäyttö ja varmuustallennus). Samalla luodaan potentiaalinen mahdollisuus tekstiaineistojen tunnistamiseen myöhemmin merkkimuotoon (OCR). Matti Uusivirta digitoi Elias-digitoinnissa joulukuussa 2004 yhteensä noin 6000 sivua kielitieteellisiä julkaisuja, jotka tullaan jatkossa tunnistamaan merkkimuotoon ja uudelleenjulkaisemaan. Liitteenä on Matti Uusivirran kiinnostava raportti, jossa hän käsittelee seikkaperäisesti skannauksen ja tekstintunnistuksen projektisidonnaisia käytänteitä ja haasteita (ks. [liite 11](#)).

Kopio- ja viitekortistot

Originaalikoko: A5 (oletusarvo)

Skanneri: A4- tai A3-tasoskanneri

Resoluutio: 300 ppi

Formaatti: TIFF (harmaasävy)

Henkilöresurssit: 1

Tiedostoja päivässä: 500

Tiedostoja kuukaudessa: 10 000

Tiedostoja vuodessa: 100 000 (10 kk)

Käsikirjoitukset (vanhat käsikirjoitetut aineistot)

Originaalikoko: A4-A3 (oletusarvo)

Skanneri: A3-tasoskanneri

Resoluutio: 300 ppi

²⁴ <http://digi.lib.helsinki.fi/>

Formaatti: TIFF (väri)

Henkilöresurssit: 1

Tiedostoja päivässä: 250 - 300

Tiedostoja kuukaudessa: 5000 - 6000

Tiedostoja vuodessa: 50 000 - 60 000 (10 kk)

Huomautukset: Laitteen teknisellä skannausnopeudella on tässä yhteydessä merkitystä lähinnä A3-kokoisten ja sitä suurempien aineistojen kohdalla.

Konekirjoitetut tekstit ja tulosteet

Originaalikoko: A4 (oletusarvo)

Skanneri: A4- tai A3-tasoskanneri

Resoluutio: 300 ppi

Formaatti: TIFF (väri/harmaasävy)

Henkilöresurssit: 1

Tiedostoja päivässä: 400

Tiedostoja kuukaudessa: 8000

Tiedostoja vuodessa: 80 000 (10 kk)

Huomautukset: Kansanrunousarkistossa digitoitiin koemielessä 21.- 27.12.2000 Sanomalehti elämäkulussa -kilpakeruuaineisto (1449 sivua ja 30-sivuinen sisällysluettelo). Nykyisten tasoskannereiden tekninen lukunopeus ei vaikuta merkittävästi todellisen digitointinopeuden arviointiin vuositasolla.

Tekstiaineistojen tunnistaminen merkkimuotoon (OCR)

Originaalikoko: A4 (konekirjoitetut tekstit ja tulosteet)

Skanneri: A4- tai A3-tasoskanneri

Resoluutio: 300 ppi

Formaatti: TIFF (väri/harmaasävy)

Tekstintunnistusohejelma: Abbyy FineReader 4.0 - 7.0

Henkilöresurssit: 1

Tiedostoja ja sivuja päivässä: 50 - 100

Tiedostoja ja sivuja kuukaudessa: 1000 - 2000

Tiedostoja ja sivuja vuodessa: 10 000 - 20 000 (10 kk)

Edeltävien tietojen perusteella on suhteellisen helppo laskea yksittäisten kuvatiedostojen tuotantokustannukset suhteessa digitoijan vuosittaisiin palkkamenoihin. On arvioitu, että ainoastaan kolmannes digitoinnin kokonaiskustannuksista liittyy tiedostojen tuottamiseen skannaamalla tai kuvaamalla. Todellisia kustannuksia arvioitaessa on otettava huomioon mm. seuraavat keskeiset tekijät:

1. Digitointitilat ja niiden asianmukainen ylläpito
2. Laite- ja ohjelmistohankinnat
 - Laitteiden hankinta-arvo ja laskennallinen käyttöikä
 - Erilaiset ohjelmistot ja niiden päivitykset
3. Henkilöstökustannukset
 - Digitoijat, suunnittelijat, kouluttajat ja tekniset tukihenkilöt
4. Tietovälineiden säilyttäminen ja virkistäminen (tiedostojen arkistointi)
5. Digitaalisen arkiston kehittäminen ja ylläpitäminen (tietokantaratkaisut)
6. Koulutus- ja matkakustannukset

PROJEKTIN RINNAKKAISHANKKEITA

Projektin toiseen vaiheeseen liittyvien tietokantaratkaisujen kehittämisen ja suunnittelemisen keskeiseksi ongelmaksi muodostui SKS:n erilaisten kehityshankkeiden laajuus ja moninaisuus vuosina 2001–2004. Rajallisia voimavaroja ei ollut järkevää jakaa ja suunnata projektille räätälöitävän erillisjärjestelmän kehittämiseen ja ylläpitoon (digitaalinen arkisto). Projektin kuluessa vireillä olivat eri vaiheissa mm. SKS:n toimintastrategioiden laatiminen, kotisivujen uudistaminen, sähköisten asiakaspalvelujen kehittäminen, arkistonmuodostussuunnitelman laatiminen, Teamware-ohjelmistojen käyttöönotto, SKVR-tietokannan kehittäminen sekä KRA:n arkistotietojärjestelmän uudistaminen. Vastaavasti aineistoja digitoitiin ko. ajanjaksona eri laittein, menetelmin ja tavoittein Elias-digitoinnissa, KRA:n äänitarkistossa, Biografia-keskuksessa, Kustannusosastolla sekä Valistus-digitointiprojektissa (liite 5). Jatkossa erityistä huomiota tulee kiinnittää tietotekniikkaan kytkeytyvien ja eri osastoja yhdistävien hankkeiden koordinointiin sekä jo aiemmin käynnistettyjen kehitys- ja uudistushankkeiden loppuunsaattamiseen, jotta asiantuntijaresursseja voidaan vapauttaa ja osoittaa muihin suunniteltuihin tehtäviin. Tavoitteena tulisi olla yksittäisten hankkeiden realistinen ajoittaminen ja aikataulutaminen suhteessa organisaation muihin erityisosaamista vaativiin rinnakkaishankkeisiin.

Projektin toteuttamisen kannalta oli korvaamatonta ATK-päällikkö Matti Anttilan tietotekninen taustatuki, sillä suhteellisen lyhytkestoiseen ja pienimuotoiseen projektiin ei ollut mielekästä palkata erillistä IT-spesialistia. Tulevien laajempien hankkeiden osalta on suositeltavaa harkita erillisen IT-osaajan tms. konsultin palkkaamista ainakin projektin kriittisen suunnittelu- ja käynnistysvaiheen ajaksi. Ennen laajamittaisen digitointiprosessin käynnistämistä tulisi muistiorganisaatiossa laatia ohjeistus digitaalisten aineistojen tuottamiseksi ja säilyttämiseksi sekä toisaalta luoda tai kehittää arkistotietojärjestelmä, johon prosessin kuluessa syntyneet metatiedot voidaan suoraan syöttää. Kyseinen toimintajärjestys olisi ollut nyt käsiteltävän hankkeen kokonaisvaltaisen suunnittelun ja toteuttamisen kannalta ideaalinen, mutta sen toteuttamiseen ei ollut vielä vuosina 2001–2004 käytännön edellytyksiä. Mainittakoon, että KRA:n vuodesta 1992 tuotantokäytössä ollut arkistotietojärjestelmä ei nykyisellään sovellu digitaalisten tallenteiden hallintaan, tiedonhakuun ja reaaliaikaiseen käyttöön. On selvää, että yksityisarkistojen resurssit uusien web-palveluiden ja tietokantaratkaisujen suunnitteluun ja luomiseen ovat varsin rajalliset. Käsikirjoituskokoelmista puhuttaessa henkilöresurssit ja tietotaidot ovat pääsääntöisesti kohdentuneet perinteiseen ”analogiseen” arkistotoimintaan: kokoelmien kartuttamiseen ja hallintaan sekä tutkimustyöhön ja henkilökohtaiseen asiakaspalveluun.

Tällä hetkellä projektissa digitoitujen aineistojen käyttöön ovat käytettävissä SKS:n sisäverkossa sähköisinä sidoksina, jotka noudattavat käsikirjoitusten alkuperäistä järjestystä. Aineistojen asiakaskäyttö on mahdollista käyttösääntöjen puitteissa erilliseltä asiakaspalvelukoneelta, joka sijaitsee KRA:n pääarkistosalissa. Tutkimuskäyttöä varten aineistosta tallennettiin lisäksi CD-ROMit, jotka sisältävät sekä muokatut katselukuvat että korkeatasoiset kuvatiedostot. Digitaaliset käyttökopiot toimitettiin Kalevala-instituuttiin ja Joensuun perinnearkistoon, jossa aineistot ovat käytettävissä valvotuissa olosuhteissa. Tuleekin muistaa, ettei digitaalisten aineistojen etäkäyttö aina edellytä kalliiden verkkoratkaisujen rakentamista vaan palvelut voidaan toteuttaa erillisten palvelupisteiden ja tilauspalveluiden kautta (vrt. maakunta-arkistojen mikrofilmipalvelut). Varsinaisina ns. master-tallenteina toimivat TIFF-tiedostot, jotka tallennetaan SLR-magneettinauhalle. Kahtena kappaleena eri paikoissa säilytettävät SLR-nauhat tulee virkistää määräajoin (5 vuotta). Kyseiset luotettavat magneettinauhat korvataan jatkossa enemmän tallennuskapasiteettia omaavilla ja teknisesti edistyneemmällä tietovälineillä (esim. LTO). Mainittakoon,

että KRA:n vuoden 2005 toimintasuunnitelmassa on mukana www-palvelupaketin luominen digitoidusta Lönnrotiana-kokoelmasta. Käytännössä tämä merkitsee sidoskohtaisten linkkisivustojen luomista nyt jo olemassa olevien sähköisten aineistoluetteloiden pohjalta.

Lyhyenä yhteenvedona voidaan todeta, että XML-pohjaisia tietokantoja ja dokumentteja voidaan käyttää tiedonhakuun ja -hallintaan; digitaaliset käsikirjoituskokoelmat mahdollistavat puolestaan originaaliaineistojen suojelemisen ja aineistojen laadukkaan etäkäytön. Korostetakaan, että digitaalenteiden käyttövelvollisuus tulee ulottaa myös organisaation sisäiseen toimintaan, joka voi kuluttaa vanhoja hauraita aineistoja yllättävässä määrin. Digitaalisessa ympäristössä tiedonhaku tapahtuu parhaassa tapauksessa viitetietokannoista, niitä tukevista merkkimuotoisista tekstikorpuksista sekä digitoiduista sisällysluetteloista, hakemistoista ja kortistoista. Tämän jälkeen tiedonhakija voi halutessaan suoraan siirtyä katselemaan ja/tai kuuntelemaan digitaalisia tallenteita, jotka vastaavat olennaiselta tietosisällöltään originaaliaineistoja. Kehittyneiden järjestelmien luomisessa keskeisenä ongelmana tulee jatkossakin säilymään riittävän digitaalisen aineistomassan saavuttaminen, jolloin todella voidaan puhua sähköisistä arkistoista, jotka palvelevat tutkijoita ja laajemmin tietoyhteiskunnan kansalaisia nopeissa verkkoympäristöissä.

Itse asiassa on oletettavaa, että originaaliaineistojen käyttöpyynnöt lisääntyvät tietoverkoissa tapahtuvien aineistoesittelyiden myötä. Kokoelmista tuotettujen digitaalisten kuvatallenteiden laatuksiteriksi ei näin riitä niiden hyvä ja nopea selailtavuus näyttöpäätteeltä, vaan tutkijoiden tms. asiakkaiden käytettävissä tulisi tarvittaessa olla korkealaatuiset kuvatiedostot, jotka mahdollistavat kohteen yksityiskohtaisen luettavuuden ja katseltavuuden. Valokuvien osalta web-selailukuvia voidaan käyttää luontevasti laajoihin aineistomassoihin kohdistuvien hakujen ja valintojen lähtökohtana (esim. aiheet, henkilöt ja tapahtumat). Sen sijaan sähköisten käsikirjoituskokoelmien tulisi olla suoraan käytettävissä eli luettavissa jo ns. selailutilassa, jotta niiden esittely verkkoympäristössä olisi ylipäätään mielekästä. Korkearesoluutioisten kuvatiedostojen luvattoman kopioinnin ja levittämisen estämiseksi kyseiset erikoispalvelut voidaan tarjota esimerkiksi valvottujen palvelupisteiden kautta (sisäverkot ja/tai erilliset tietovälineet).

Digitointiprosessista saatuja kokemuksia voidaan jatkossa hyödyntää kehitettäessä Kansanrunousarkiston arkistotietojärjestelmää erityisesti digitaalisten aineistojen säilyttämisen, hallinnoinnin ja käytettävyyden kannalta. Käytännössä digitointiprosessin rinnalla tuotettiin jatkuvasti erilaisia metatieto- ja aineistoluetteloita, joista voidaan siirtää tai syöttää keskeisiä tietoelementtejä uudistettavaan arkistotietojärjestelmään. Niin ikään kyseisiä luetteloita tai listoja voidaan linkittää erillisinä sähköisinä dokumentteina tietojärjestelmiin (esim. tiedostolistaukset). Tarkennetakaan, ettei nyt käsillä olevaa projektia varten ollut tarkoituksenmukaista rakentaa kevyttä ja väliaikaista tietokantaratkaisua, jonka kautta kaikki digitoidut aineistot olisivat olleet heti haettavissa ja käytettävissä sisäverkon kautta. Projektissa tehty kehitystyö tuleekin nähdä osana SKS:n arkistotietojärjestelmien laajempaa kehittämistä ja arkistojen mahdollista integroimista ns. digitaalisen arkiston osalta (vrt. Kansallisarkisto).

I Elias-kirjetietokanta

Elias-kirjetietokannan kehittämiseen osallistuivat eri vaiheissa Hannu Hautala, Pasi Klemettinen, Raija Majamaa ja Pekka Tolonen. Oman kiinnostavan episodinsa merkki-muotoisen kirjekorpuksen luomisessa muodosti 1980-luvulla luotujen CPT-tekstinkäsittely-tiedostojen siirto ja konvertointi nykyjärjestelmien ymmärtämään muotoon. Korvaamattomina teknisinä tukihenkilöinä pelastusoperaatiossa toimivat Lasse Nirhamo ja Heikki Yliheikkilä Suomen Tietojenkäsittelymuseoyhdistyksestä²⁵. Raportin liitteenä on Nirhamon vapaamuotoinen selvitys haasteellisesta operaatiosta (liite 4).

Tietokannan nykyinen versio koostuu itse asiassa kahdesta tietokannasta. Ensimmäinen sisältää 1100 Elias Lönnrotin kirjetä tekstimuodossa. Mainittakoon, ettei tekstejä ole oikoluettu kriittisesti. Toinen tietokanta sisältää 2536 Elias Lönnrotin kirjeen metatiedot (vastaanottaja, päiväys jne.) sekä linkityksen käsikirjoituksista digitoituihin kuvatiedostoihin. Raija Majamaan liitteenä oleva raportti käsittelee tarkemmin Lönnrotiana-kokoelmaa ja erityisesti Lönnrotin kirjeenvaihtoa arkistollisesta näkökulmasta (liite 3). Tietokannat on toteutettu Turun yliopiston ja Åbo Akademin yhteislisenssillä ylläpitämän Trip-tietokantaohjelman avulla (TietoEnatorin omistama ohjelma). Samaa tietokantaohjelmaa käytetään yleisesti monilla yksityisen ja julkisen sektorin tahoilla (esim. Nokia, Saab, Suomen eduskunta, Muisti-projekti). Seuraava selvitys tietokannan luomisesta ja alustavista kehitysnäkymistä perustuu Pekka Tolosen projektille laatimiin raportteihin (2004).

Korpuksesta tietokannaksi

Elias-kirjetietokannan pilottiversio tehtiin Kalevala-instituutissa Turun yliopistossa (Pekka Tolonen). Käytännössä työ käynnistyi 2002 loppukesästä. Tätä ennen oli Kalevala-instituuttiin saatu SKS:sta käyttökopiot digitoidusta Lönnrotiana-aineistosta sekä Lönnrotin kirjeiden CPT-tekstinkäsittelylaitteistolla tehtyjen tiedostojen käännökset tekstitiedostoina.

Tietokannan toteuttaminen aloitettiin tutustumalla kyseisten CPT-tekstien sisältämään informaatioon. CPT-tekstit olivat 12 kansiossa, joissa jokainen tiedosto vastasi yhtä kirjetä. Joissakin tapauksissa kirje saattoi jatkua toisessa tiedostossa. Tiedostoja oli 1157 (?). Päällekkäisyyksien poistamisen ja osittaisten kirjeiden yhdistämisen jälkeen lopputuloksena oli 1100:n kirjeen tiedot.

Kirjeet oli kirjoitettu 1980-luvun lopulla puhtaaksi käsikirjoituksista (Raija Majamaa ja työryhmä). Kyseiset CPT-tekstit olivat työversioita, eikä niitä ole kriittisesti oikoluettu. Osa kirjeistä (508 kpl) on julkaistu tarkastetussa muodossa Elias Lönnrotin valituissa teoksissa²⁶.

CPT-tekstien rakenne:

- Ensimmäinen rivi on otsikkorivi
- Toiselta riviltä alkaen on tietoja kirjeestä eroteltuna pilkulla ja / -merkillä: vastaanottaja, lähettäjä, päiväys (kuusinumeroisena ppkkvv), kieli, signum, henkilöhakusanoja (lyhenteitä), paikkakuntahakusanoja (lyhenteitä), asiasanoja. (Eivät juuri edellä mainitussa järjestyksessä)
- Varsinainen kirjeen tekstiosa seuraa edeltäviä rivejä.

²⁵ <http://www.tietokonemuseo.saunalahti.fi>

²⁶ Majamaa 1990.

Tekniset työvaiheet:

1. Rakennettiin tekstitietokantaohjelman Trip avulla tietokanta ja siihen hakusivut www-ympäristöön (Triphighway). Tietokannan nimi on ELIAS ja hakusivut on suojattu salasanalla.
2. Tehtiin Microsoft Word-ohjelmalle makro, joka muokkasi CPT-tekstiedostot TFORM-muotoon.²⁷ Makrossa muodostettiin myös päiväyksestä erillinen vuosinumero erilliseen kenttään tietokannassa.
3. Muutamii tietueisiin liitettiin koeluontoisesti linkit käsikirjoituksista digitoituihin kuvatiedostoihin.
4. CPT-tekstit olivat erillisinä tietueina. Jokainen tiedosto muokattiin makron avulla erikseen, tallennettiin .tfo päätteisenä tekstitiedostona alkuperäisen tiedoston rinnalle. Lopuksi TFORM-muodossa olleet tekstit koottiin kahdeksi isoksi tiedostoksi. (002-580 ja 581-1160). Nämä kaksi tiedostoa ajettiin tietokantaan.
5. Tietokannasta tehtyjen listauksien avulla huomattiin, että mm. nimien kirjoitusmuodot vaihtelivat. Tietokannan tietojen harmonisoimiseen eli standardointiin (että esim. nimet ja hakusanat ovat samalla tavalla kirjoitettuja) ja oikolukemiseen kuluisi pitkä aika. Alkuperäisiä tekstejä ei oltu tarkoitettu tietokantaan soveltuvaksi korpuksiksi.

Tietokannan ensimmäinen versio valmistui syksyllä 2003. Seuraavaksi tavoitteeksi asetettiin käsikirjoituksista digitoitujen värikuvatiedostojen linkittäminen merkkimuotoisten kirjeiden tietoihin.

Lähtökohdaksi otettiin Hellevi Riittisen Lna 366 käsikirjoituksen pohjalta kirjoittama luettelo (1996) Lönnrotin lähettämistä kirjeistä. Pasi Klemettinen muutti kyseisen dokumentin tekstimuotoiseksi OCR-ohjelmiston avulla. Tätä tekstitiedostoa täydennettiin myöhemmin puuttuvilla noin 10:llä sivulla. Tekstitiedosto muunnettiin Microsoftin Wordin Etsi/Korvaa toimintojen avulla muodoltaan sellaiseksi tekstitiedostoksi, jotta se voitiin avata Microsoftin Excel-ohjelmalla.

Excel-ohjelmassa etsittiin ja korjattiin OCR-vaiheesta juontuvia kirjoitusvirheitä (esim. ohjelmisto saattaa tunnistaa kirjaimen l numeroksi 1 jne.).

Makrojen avulla lisättiin kirjeiden kirjoitusvuodet sekä päiväykset (muodosta 06.09.48 (=pp.kk.vv) muotoon 480906 (=vkkpp)). Tapauksissa, jossa tarkkaa päivämäärää ei tiedetty kirjattiin se muodossa 00. Esim. kirje, jonka (päätelty) päiväys oli helmikuu 1848, tuli muotoon 480200. Kirjeiden molemmat päiväystavat säilytettiin Excel-tiedostossa. Haut tapahtuvat jälkimmäisen päiväyksen merkintätavan perusteella. Ensimmäinen päiväystapa näkyy kuitenkin hakutuloksissa. Tapauksissa, joissa ei tiedetä vuotta (merkintä s.a., näitä tietokannassa on yhteensä 13 kpl) on tietokantaan numeromuotoisen päiväyksen kohdalle merkitty 0. Vuosikenttään ei ole merkitty mitään.

Tähän Lönnrotin lähettämien kirjeiden täydelliseen listaan tehtiin vertailu tekstimuotoisten (CPT) kirjeiden listan kanssa ja merkittiin omaan Excel-sarakkeeseen (sekä lihavoitiin) mitkä tekstit olivat myös tekstimuodossa.

Excel-tiedostoon lisättiin kuvatiedostojen nimet seuraavasti: ACDSee ohjelmalla tehtiin tekstimuotoinen tiedostolistaus kuvakansioista (lnak19, 20, 26 ja 32). Muita Lönnrotin lähettämien kirjeiden kansioita ei ole digitoitu 29.7.2002 päivätyn listan mukaan. Näistä tiedostolistauksista

²⁷ TFORM on Trip-ohjelman käyttämä tapa rakenteistaa tekstidokumentteja.

poimittiin kuvatiedostojen nimet, jotka on muodostettu suoraan kirjeiden arkistosignumeiden perusteella. Valitettavasti nimeämiskäytänteissä on häilyvyyttä eri kansioiden eli itse asiassa KIA:n alkuperäisten sidosten välillä. Esim. Lnak19 kansiossa yksi kirje on saanut yhden numeron ja kuvatiedostot on merkitty a,b,c jne kirjaimien avulla. Näin yksi kirje voi olla esim. lnak019_052a, lnak019_052b jne. Kansiossa Lnak26 jokainen sivu on numeroitu erikseen, jolloin yksi kirje voi olla esim. lnak026_751, lnak026_752 lnak, 026_750etulehti. Jälkimmäisessä tapauksessa signum Lna 366 käsikirjoituksen perusteella tehdyssä listassa on 26:751. Näiden aineiston erilaisten järjestämistapojen vuoksi kirjeet oli epävarmoissa tilanteissa tarkistettava manuaalisesti kuvatiedostoista.

Excel-tiedostoon liitettyistä kuvatiedostojen nimistä poistettiin tiedostopääte, siltä varalta, että joskus tiedostomuoto muutetaan eräajona joksikin muuksi kuin jpg. Tällöin riittää, kun muuttaa triphighway-tiedoston (post.thw) linkin muodostuksesta tiedostopääte. Tämä ratkaisu rajoittaa tietokantaa siten, että kuvatiedostojen tulee olla kaikki saman muotoisia.

Excel-tiedostoon lisättiin TFORM muotoilut ja tallennettiin tiedosto tekstimuodossa. Tämä tekstitiedosto ajettiin eräajona Trip-tietokantaohjelmaan Elias2 tietokantaan. Tietokanta sisältää tiedot 2536:sta kirjeestä.

Tietokantojen muodostamisen lopputuloksena on kaksi erillistä tietokantaa **Elias** ja **Elias2**. Elias sisältää CPT-tekstit. Elias2 pohjautuu Lna366 käsikirjoituksen sisältämään luetteloon Lönnrotin kirjoittamista kirjeistä.

Kuvaus Trip-tietokannoista Elias ja Elias2

Trip-tietokantaohjelman avulla tehdyt tietokannat koostuvat kolmesta osasta:

1. Itse tietokannasta, jota ylläpidetään Åbo Akademin koneella (unix).²⁸
2. Tietokannan tietoja www-hakusivuille kääntävästä Triphighway-järjestelmästä.
3. Turun yliopiston palvelimelle sijoitetuista kuvatiedostoista, joihin muodostetaan linkki www-hakujen avulla.

ÅA:n ja TY:n hakemistot vaativat tunnuksen ja salasanan. Tämä on toteutettu htaccess -menetelmällä. Kyseinen menetelmä ei ole mahdoton murtaa, mutta näin tehtäessä rikotaan yksiselitteisesti lakia.

Trip-tietokannassa olevien tietokantojen tietueiden kentät voivat olla eri muotoisia. Elias-kirjetietokannoissa on käytetty kolmea eri muotoa: Phrase, Text ja Number.

Phrase: Kentässä voi olla useita eri arvoja, joita voidaan hakea yhdessä ja erikseen. Näin esimerkiksi kirjeen kieli voidaan merkitä suomeksi, ruotsiksi tai kumpaakin kieltä sisältäväksi. Phrase kentistä on helppo muodostaa eri arvojen listauksia.

Text: Kenttä voi muodostua pitkäköstä tekstistä. Jokainen sana on indeksoitu erikseen ja sanojen osat ovat myös haettavissa. Tekstin kappaleet on merkitty erikseen.

Number: Numerokenttä mahdollistaa haut = ,> ja < määritteiden avulla. Numerot listautuvat matemaattisen arvonsa mukaan eivätkä aakkosjärjestyksessä (kuten phrase-kentät).

²⁸ trip.abo.fi / quattro.abo.fi

Elias-tietokanta muodostuu kentistä:

NUMBER	Päiväys
PHRASE	Vuosi
PHRASE	Lähettäjä
PHRASE	Vastaanottaja
PHRASE	Kieli
NUMBER	Kokoelma (Signum)
NUMBER	Yksikkö (Signum)
PHRASE	Paikkakunnat
PHRASE	Asiasanat
PHRASE	Julkaisutiedot
TEXT	Varsinainen tekstiosa
PHRASE	Henkilöt
PHRASE	Linkki kuvatiedostoihin
PHRASE	Pääkokoelma

Kenttien järjestyksellä ei ole hakujen tai tietokannan toiminnan kannalta merkitystä.

Huom. Kenttä nro 2 Vuosi (year) on laitettu tietokantaan Phrase-muodossa, joka mahdollistaa vuosilistauksien tekemisen.

Elias2-tietokanta muodostuu kentistä:

PHRASE	Lähettäjä
PHRASE	Vastaanottaja
PHRASE	pp.kk.vv ja tarkennuksia
NUMBER	Päiväys vvkpp
NUMBER	Vuosi
PHRASE	Kirjeen kieli
PHRASE	Kirjeen muoto (kirje / konsepti)
PHRASE	Kokoelma
PHRASE	Signum
PHRASE	Julkaisutiedot
PHRASE	Kuvatiedostojen nimet
PHRASE	Henkilöt
PHRASE	Paikkakunnat
PHRASE	Asiasanat
TEXT	Kirjeen teksti
PHRASE	Apukenttä ylläpitoon (esim. CPT)

Elias2-tietokanta sisältää enemmän kenttiä kuin Elias-tietokanta. Muutenkin tietokannan suunnittelu on kehittyneempi. Ajatuksena on ollut, että tietokantojen tiedot yhdistetään siten, että lopullinen tietokanta noudattaa rakenteeltaan Elias2-tietokantaa.

Elias2-tietokanta on suunniteltu siten, että siihen voidaan myöhemmin lisätä myös muiden lähettämiä kirjeitä. Tässä vaiheessa pitää vain hakusivuja muuttaa siten, että kenttään Lähettäjä voidaan kohdistaa hakuja.

Hakusivujen toiminta

Tietokantojen yhteinen aloitussivu sijaitsee osoitteessa <http://trip.abo.fi/tydb/elias>. Tältä sivulta valitaan kumpaa tietokantaa halutaan käyttää. Hakusivuilla on kahdenlaisia hakuja:

- 1) Linkiksi merkittyä sanaa seuraamalla saadaan listauksia kenttien arvoista. Nämä hakutulokset toimivat linkkeinä eteenpäin (mikäli listauksessa kyseistä arvoa esiintyy vain yksi tietokannassa päästään suoraan dokumenttinäkymään ja mikäli listauksessa on mainittu, että arvoa esiintyy useammassa dokumentissa päästään listaan näistä dokumenteista, joista päästään itse dokumenttinäkymään.)
- 2) Haettavat sanat tai arvot voidaan kirjoittaa niille varattuihin kohtiin ja lopuksi painaa HAE-painiketta. Tuloksena on joko suoraan yksi dokumentti tai usean hakutuloksen muodostama lista. Näiden listojen sisältämät tiedot voidaan muodostaa mistä tahansa tietueiden kentistä. Tällä hetkellä ne ovat: tuloksen järjestysnumero, vastaanottaja, signum, päiväys sekä tietueen ns. Record number²⁹.

Kumpaankin hakujärjestelmään on dokumenttinäkymässä merkitty päiväyksen kohdalle linkki toiseen tietokantaan. Tätä linkkiä seuraamalla käynnistyy haku toisesta tietokannasta kyseisen päiväyksen avulla. Hakutuloksena voi olla yksi dokumentti tai useiden dokumenttien lista, jos samaa päiväystä löytyy useista kirjeistä.

Elias2-tietokannan dokumenttinäkymästä on mahdollista tarkastella myös linkitettyjä kuvatiedostoja seuraamalla tiedostojen nimien muodostamia linkkejä Kuvatiedostot-kohdasta. Kuvat aukeavat erilliseen ikkunaan. Tässä yhteydessä (koska kuvat sijaitsevat TY:n palvelimella) kysytään tunnusta ja salasanaa uudestaan.

Elias2-tietokannan hakusivulle on liitetty tietoja kenttien sisältämistä arvoista ja hakujen tekemisestä pallossa olevien kysymysmerkkien taakse.

Huomioita, kokemuksia ja ajatuksia

Kirjallisuusarkiston signum-systeemien erilaisuus on tietokannan kehittämisen kannalta ongelmallista. Joissakin sidoksissa/kansioissa on sivulta sivulle juokseva numerointi, joissakin kirjekohtainen numerointi, joissakin ei ole numerointia ollenkaan (tai on merkitty tietokantaan vain kansion numerolla yksilöimättä tarkemmin).

Signumien merkintätapojen systematisointia tulisi miettiä. Tällä hetkellä tietokannassa on tätä varten kaksi kenttää: Kokoelma (col) ja Signum (sig). Mahdollisesti voisi käyttää myös erillistä kenttää, joka ilmaisisi missä on kopio originaalista. Esim. HYKissä säilytettävä kirje saisi varsinaisiin kokoelma ja signum kenttiin arvot HYKin arkistojärjestelmän mukaisesti: Kokoelma = HYK ja Signum = HYKin signum. Kopio-kenttään tulisi maininta, että kirje löytyy myös siitä ja siitä LNA-kansiosta, jos kirjeestä on olemassa kopio.

Yleisellä tasolla voidaan todeta, että tekstin tunnistuksella (OCR) ja valmiiden tekstitiedostojen avulla saadaan aineisto suhteellisen helposti normalisoitua ja rakenteistettua. Erityisesti tekstin siirto Excel-ohjelmaan (joka kylläkin on ensisijaisesti taulukkolaskentaohjelma) helpottaa manuaalista ja tietokoneavusteista tarkistustyötä. Se, mihin teksti siirretään Exceliltä, on tässä yhteydessä ollut Trip-tietokanta, mutta se voi olla periaatteessa mikä tahansa muukin ohjelma.

²⁹ Muuttumaton numero, jolla yksittäinen tietue voidaan identifioida tietokannassa.

Raija Majamaan suunnitelma Elias Lönnrotin koko tunnetun kirjeenvaihdon kriittisen edition toimittamisesta antaa haluttaessa mahdollisuuden myös tietokantaversioon jatkokehittelyyn ja loppuun saattamiseen.

II SKVR-tietokanta

Johdatus

SKVR-pilottitietokannan tekninen kehittäminen ja ylläpito ajoittui vuosille 2001–2004. SKS:n ja CSC:n välillä käytyihin neuvotteluihin yms. keskusteluihin osallistuivat pääsääntöisesti Jukka Saarinen ja Manne Miettinen sekä projektityöntekijöistä Pasi Klemettinen. Yhteistyöhankkeen tavoitteena oli SKS:n tuottaman digitaalisen runokorpuksen hyödyntäminen CSC:n ylläpitämässä tietokannassa³⁰. Neuvotteluissa keskusteltiin myös mahdollisuudesta hyödyntää CSC:n tietokantaa ja levytilaa Elias-digitoinnissa tuotettujen kuvatiedostojen osalta. Ajatuksena oli linkittää Lönnrotin alkuperäiskäsikirjoituksia (muistiinpanoja) vastaaviin julkaistuihin SKVR-teksteihin. Linkittäminen olisi mahdollistanut runojen tekstuaalisen kontekstin ja koostamistekniikan tarkastelun reaaliaikaisesti tutkijan omalta näyttöpäätteeltä (esim. säkeiden lisäykset ja poistot). Lisäyötä vaativien uusien ominaisuuksien lisääminen CSC:n tietokantaan ei kuitenkaan osoittautunut mahdolliseksi pilottihankkeen puitteissa.

SKVR-tietokanta annettiin tutkijoiden testikäyttöön kesällä 2002. Digitaalinen korpus tulee vuoden 2005 loppuun mennessä käsittämään Suomen Kansan Vanhat Runot -julkaisusarjan kokonaisuudessaan. Julkaisemattoman SKVR-aineiston osalta tietokantaratkaisut ovat toistaiseksi avoimia. Vuonna 2005 tietokanta siirretään SKS:n omalle palvelimelle jatkokehittäväksi ja sitä tultaneen ylläpitämään avoimena verkkopalveluna. Tietokannan avoimuus mahdollistaa oletusarvoisesti korpuksen helpomman ja laajemman hyödyntämisen mm. erilaisissa perinnettologioissa, sukututkimuksissa, pitäjähistorioissa ja taiteellisissa produktioissa. Pasi Klemettinen on esitellyt tutkimustyössään käyttämäänsä SKVR-tietokantaa Budapestissa, Riikassa, Joensuussa ja Turussa järjestetyissä tilaisuuksissa. Kotimaiset esittely- ja info-tilaisuudet järjestettiin perinteentutkimuksen jatko-opiskelijoille ja tutkijoille. Vastaavasti Jyrki Pöytä on hyödyntänyt ja esitellyt erikoistutkijan ominaisuudessa digitaalisia aineistoja Joensuun perinnettarkistossa. Tulevaisuudessa arkistojen e-aineistojen näkyvyyttä, markkinointia ja käyttäjäkoulutusta tulee ymmärrettävästi tehostaa, jotta niiden yleistä käyttöastetta saadaan nostetuksi. Erityisesti nuorempaa sukupolvea edustavat arkistoasiakkaat, joille tietoverkoissa toimiminen on osa arkipäivää, tulisi aktivoida e-palveluiden kriittisiksi käyttäjäksi ja tätä kautta myös kehittäjiksi (käyttäjäpalaute).

Suomen Kansan Vanhojen Runojen digitointi

Seuraavassa erikoistutkija Jukka Saarisen laatima erillisraportti monivuotisesta SKVR-hankkeesta.

Suomen Kansan Vanhat Runot (SKVR) -julkaisu on kalevalaisen kansanrunouden keskeinen lähde-teos. SKVR-teoksessa (34 sidosta) on pyritty julkaisemaan kaikki arkistoihin talletettu ja kirjallisista lähteistä löydetty "aito" kalevalamittainen runous. Suurin osa aineistosta on peräisin Kansanrunousarkiston kokoelmista. Tekstit on julkaistu perinnealueittain (14 aluetta, osat I–XIV) vuosina 1908–1948. Vuonna 1997 ilmestynyt täydennysosa XV sisältää aineistoja

³⁰ <http://www.csc.fi/suomi/tutkija/index.phtml.fi>

merkittävien varhaiskerääjien (mm. Kristfrid Gananderin ja Elias Lönnrotin) kokoelmista. Alueen sisällä runot on järjestetty pääryhmiin lajeittain ja alaryhmiin teeman ja tehtävän mukaan. Poikkeuksena on Inkeri (osat III–IV), joka on kokonaisuudessaan järjestetty kerääjittäin. Sivuja teoksessa on yli 27000, runotekstejä yli 100 000.

Suomalaisen Kirjallisuuden Seura aloitti vuonna 1998 mittavan hankkeen, jonka tarkoituksena on ollut saattaa koko teossarja digitaaliseen muotoon ja siten tutkijoiden paremmin hyödynnettäväksi. Hanke on toteutettu vaiheittain, ilman alusta loppuun kantavaa kokonaissuunnitelmaa. Ongelmana tässä järjestelyssä on ollut se, että eri vaiheiden vaatimia resursseja ei riittävän ajoissa ole aina osattu ennakoida, mikä on aiheuttanut hankkeen pitkittymistä. Toisaalta etuna on ollut se, ettei projektin varhaisessa vaiheessa, jolloin SKS:n piirissä ei vielä ollut kokemusta sähköisistä tekstikorpuksista, sitouduttu sellaisiin teknisiin ja periaatteellisiin ratkaisuihin, jotka olisivat olleet peruuttamattomia tai hankalia muutostilanteissa.

Projekti on edennyt seuraavissa vaiheissa:

1. Tekstin skannaus ja merkkimuotoon muuttaminen 1998–2000
2. Tekstin rakenteistaminen (xml) 2000–2005
3. Alustavan tietokannan luominen 2002–2004
4. Lopullisen hakemistoon yhdistetyn tietokannan suunnittelu ja luonti 2005–2006

1. Tekstin skannaus ja merkkimuotoon muuttaminen

Osahanke toteutettiin Eesti Kirjandusmuuseumiin tiloissa työskennelleen virolaisen Keele ja kirjanduse instituutin (myöh. Eesti Kirjandusmuuseumin) tutkimusryhmän toimesta. Vaiheen rahoitti kokonaan SKS, joka myös määritteli tavoitteet. Toteutuksesta huolehti ryhmä itsenäisesti. Ryhmän johtajana toimi akateemikko Arvo Krikmann ja ryhmään kuului pääosan ajasta 6 henkilöä.

Työn tavoitteena oli korpuksen perusdigitointi: teoksen sivujen skannaus, muunto merkkimuotoon (OCR), oikoluku ja erikoismerkkien koodaus. Tarkoituksena oli tuottaa mahdollisimman tarkkaan SKVR:n "näköinen" sähköinen teksti fontteineen, tekstityylein ja asettelui-
neen, jossa jokainen SKVR-teoksessa esiintyvistä merkistä (yli 400 erilaista) on sellaisenaan, jos kyseinen merkki on löytynyt tekstinkäsittelyohjelman merkistöstä, tai koodattuna omalla uniikilla tavallaan (esim. %a). Digitoinnin alkuvaiheessa keskusteltiin siitä, onko tarpeellista säilyttää kaikki diakriittiset merkit, vaikka ne useassa tapauksessa ovat vanhentuneita tai eivät sisällä erityistä lisäinformaatiota tekstien käyttäjille. Harkinnan jälkeen ja Kotimaisten kielten tutkimuskeskuksen asiantuntijoiden neuvon mukaan päätettiin ohjeistaa, että kaikki erikois-merkit on säilytettävä tavalla tai toisella lopputuloksessa. Ratkaisua voidaan perustella seuraavasti:

- "Yksinkertaistaminen" olisi palauttamaton muutos, merkkiä ei enää saataisi aikaan kuin uudella digitoinnilla.
- Lähdetekstiin puuttuminen olisi ollut uuden edition luomista, mikä olisi ollut helposti sekaannuksia aiheuttavaa, koska SKVR itsekkin on alkuperäislähteen editio.
- Erikoismerkin tulkinta on usein kontekstisidonnainen, henkilöstä ja ajasta riippuvainen.

Erikoismerkkien säilyttäminen aiheutti omat ongelmansa niin digitoinnissa (lisätyö oikoluvun yhteydessä) kuin myöhemmin tietokannan hakutoiminnossakin, jossa siihen kuitenkin kehitettiin varsin toimiva ratkaisu. Suurin erikoismerkeistä aiheutuva ongelma ei kuitenkaan ole liittynyt niinkään säilytysperiaatteeseen kuin tietokoneen käyttämien merkistöjen

puutteelliseen ymmärtämiseen. Tekstinkäsittelyohjelmalla (Word) luotujen tekstien siirto koneesta toiseen, osin myös ohjelmasta toiseen, on aiheuttanut runsaasti lisävaivaa, kun on pitänyt tarkastella miten eri merkit ovat siirroissa säilyneet. Kunnollinen suunnittelu ja Unicoden käyttöönotto heti projektin alussa olisivat olleet tarpeen. Unicode tuli kuvaan mukaan kuitenkin vasta vaiheessa II, kun tehtiin lopullisesti XML:n valinta rakenteistamisen työvälineeksi.

Koska perusdigitointi suoritettiin liki kokonaan ulkoistettuna, sen toteutusvaiheen ongelmat eivät aktuaalistuneet SKS:n piirissä. "Virallisena" lopputuloksena on SKVR sellaisenaan sekä merkkimuotoisena Word-tekstinä että mustavalkoisina TIFF-kuvina: tämä aineisto on CD:lle siirrettyä SKS:ssa. Vaihe saatiin valmiiksi vuoden 2000 lopussa.

2. Tekstin rakenteistaminen (xml)

Hankkeen alussa oli virolaisten kanssa keskusteltu myös tietokannan rakentamisesta, mutta hyvin pian ilmeni, että tietokantaa varten tarvittavien metatietojen eristäminen teksteistä ei olisi Virossa onnistunut ilman suomalaisten runotekstien asiantuntemusta. Katsottiin tärkeimmäksi, että teksti ensin rakenteistettaisiin siten, että eri tietoluokkiin kuuluvat tekstin osat erotettaisiin tietokoneiden ymmärtämällä tavalla. Merkintä päädyttiin tekemään mahdollisimman standardinomaisella tavalla, jotta oltaisiin riippumattomia ohjelmistoista ja järjestelmistä sekä varmistettaisiin aineiston konvertoitavuus tulevaisuudessa. Luonnollinen valinta oli XML. SKVR:n "natiivia" rakennetta noudattaen luotiin SKVR:n runoteksteille oma dokumenttityyppimääritelmä (DTD), koska valmiit olemassa olevat DTD-mallit eivät olisi ainakaan työstämisvaiheessa olleet tarkoituksenmukaisia. SKVR:n tekstien rakenne on periaatteessa varsin pysyvä läpi kaikkien osien: ne sisältävät teoksen toimittajien määrämuotoon laittamia taustatietoja (metatietoja) ja lähteestä (arkistomuistiinpanosta, julkaisusta) sellaisenaan kopioidusta tekstiosasta, jonka suurin eroavuus alkuperäisestä ovat lyhentämällä kirjoitettujen sanojen täydennykset: nämä on johdomukaisesti laitettu hakasulkeisiin. Huomattakoon, että rakenteistaminen kohdistettiin vain SKVR:n runoteksteihin, ei väliotsikkoihin, johdantoihin jne. Tähän johtaneista syistä tehdään selkoa jäljempänä vaiheen IV selostamisen yhteydessä. Koska yli 100 000 tekstin merkintä erikseen olisi ollut aivan liian työlästä, haettiin sellaista ratkaisua, jolla tekstien rakenteistaminen voitaisiin tehdä mahdollisimman automatisoidusti (massana). Lopulta tekstien koodaaminen tehtiin kahdessa vaiheessa:

Ensin akateemikko Krikmannille annettiin tehtäväksi luoda tiettyjen sovittujen periaatteiden mukaan alustava rakenteistus, "raaka-xml". Hän toteutti työn käyttäen hyväkseen Wordin makroja ja tekstiin skannauksen jälkeen jääneitä piirteitä: fonttien muutoksia, rivinvaihtoja, sisennyksiä jne., jotka tekstin lukijallekin kertovat tekstin erilaisista osista mutta jotka sähköisessä muodossa sellaisenaan eivät pysty välittämään tietoa oikealle paikalleen tekstikorpuksessa. "Raaka-xml:ään" kuului 13 eri elementtiä. Tekstiin sijoitettiin xml-tagit, mutta teksti tallennettiin edelleen Word-muodossa, jotta skannauksen jäljiltä olleet erikoismerkit säilyisivät. Lisäksi Krikmann sijoitti tekstiin korostuksella kysymyksiä ja huomioita ongelmakohdista. Raaka-xml kaikkiin osiin valmistui keväällä 2002.

Seuraavassa vaiheessa tiedostot siirrettiin Kansanrunousarkistoon, jossa tämän raportin laatijan tehtävänä on ollut muokata ne lopulliseen muotoonsa. Muokkaukseen kuuluvat seuraavat työvaiheet:

1. Merkistön tarkastus ja korjaus, erikoismerkkien koodaus Unicoden numeroentiteeteillä ja kursivointien muunto elementeiksi. Tarkoituksena on, että tietoa ei häviä, kun tiedostot

siirretään toiseen ohjelmaan jatkokäsittelyä varten (Word -> WP) ja lopulta tallennetaan xml:n käyttämään tekstimuotoon.

2. Krikmannin huomioiden läpikäynti, havaittujen säännöllisten virheiden korjaus, uuden osa- ja kommentti-elementtien lisääminen tekstiin sekä yksilöivän ID-numeron antaminen jokaiselle tekstille sekä tallennus tekstimuotoon.
3. Tekstimuotoon tallennetun tiedoston xml-koodauksen validisuuden tarkastus käyttäen XML Writer -ohjelmaa.

Vaiheen lopputuloksena on SKVR-DTD:n mukaan validia tekstiä, jossa erikoismerkit on koodattu joko Unicoden numeroentiteeteillä (jos merkki voidaan esittää myös näytöllä) tai omilla väliaikaisiksi tarkoitetuilla entiteeteillä (esim. &a8;), jos kyseistä merkkiä ei löydy Unicodesta tai sille ei löydy MS Arial Unicode -fonttia. Tekstiä voi lukea sellaisenaan esim. selaimella, tai sitä voidaan käyttää siirrettäessä tekstiä tietokantaan. Viimeistely xml on tällä hetkellä (maaliskuu 2005) tehty noin puoleen SKVR:n sidoksista.

Liitteessä 6 on kuvaus SKVR-tekstin rakenteesta ja korpuksen xml-elementeistä, liitteessä 7 on sen virallinen dokumenttityypin määritelmä eli DTD.

3. Alustava SKVR-tietokanta

Isojen, kokonaisen SKVR:n osan käsittävien, xml-tekstien käytettävyys on varsin rajallinen. Tarvittiin käyttöliittymä. Kotimaisten kielten tutkimuskeskuksen asiantuntijoiden suosituksesta syksyllä 2001 kysyttiin apua Tieteen tietotekniikan keskukselta (CSC). CSC:n kanssa sovittiin koeluontoisesta projektista: CSC kehittäisi korpukselle hakujärjestelmän ja www-pohjaisen käyttöliittymän. Samalla CSC saisi oikeuden sisällyttää korpuksen osaksi kielitieteen ja kieli-tekniikan palvelua, Kielipankkia. CSC:n kielitieteen asiantuntija Manne Miettinen rakensi MySQL-tietokannan, johon tehtiin käyttöliittymä hakua varten PHP:llä ja JavaScriptilla. XML-tiedostoja on käsitelty Perl-kielillä ja sen LibXML-modulilla.

Tietokannan demoversio valmistui huhtikuussa 2002 ja pilottiversio parinkymmenen käyttäjän testiä varten vuoden 2002 lopussa. Testauksesta saatujen kokemusten jälkeen julkistettavaa tietokantaa alettiin tehdä kesällä 2003 ja CSC:n Tutkijan käyttöliittymään sijoitettu tietokanta julkistettiin 10.12.2003.

Tietokannasta on mahdollista tehdä tekstihakuja sekä varsinaisesta runotekstistä että taustatiedoista (metatiedoista). Tekstiosuuden erikoismerkki-ongelma ratkaistiin merkkien normaalistuksella: jokaiselle suomen kielen 29 kirjainmerkistä poikkeavalle merkille annettiin "hakuarvo", ohje, minkälaisena sen haussa tulisi olla mukana. Käytännössä merkit riisuttiin kaikista tarkkeistaan (esim. š -> s) ja luotiin rinnakkaisteksti: tekstihaku kohdistetaan rinnakkaiseen pelkistettyyn tekstiin mutta tuloksena näytetään täydellinen teksti. Jonkin verran ongelmalliseksi muodostui tässä ratkaisussa se, että hakutuloksena on koko teksti, haun varsinaista osumaa ei voida näyttää, koska se ei itse asiassa ole näytetyssä tuloksessa mukana. Mainittakoon, ettei taustatietoja huomattu ottaa pelkistykseen mukaan.

Tekstihaun lisäksi metatiedoissa on tehty erikseen haku säännönmukaistettuihin kerääjä-, paikkakunta- ja vuosilukutietoihin. SKVR:ssä nämä tiedot, jotka sinänsä ovat säännöllisesti mukana kaikissa teksteissä, varioivat kirjoitusasultaan liikaa, jotta niitä voisi hyödyntää tekstihaussa. Jokainen näistä perustiedoista on muutettu numeeriseksi arvoksi, joka sijoitetaan XML-tiedostossa yksittäisen tekstin (ITEM-elementin) attribuutiksi:

- Kerääjä perustuu joko Kansanrunousarkiston kerääjäkortiston mukaiseen nimeen tai julkaistun tekstin tekijän nimeen. Nimistä ja niiden tunnusnumeroista pidetään erillistä taulukkoa, jota täydennetään työn kuluessa.
- Paikkakunta perustuu Suomen, Karjalan ja Inkerin alueen pitäjäjakoon, joka on sekä SKVR:n että Kansanrunousarkiston topografisen jaottelun lähtökohta. Pitäjien tunnusnumeroina hyödynnetään KRA:n arkistotietojärjestelmän kunta- ja aluenumeroita.
- Vuosiluku saadaan suoraan suoraan tekstiin liitetystä päiväyksestä. Vuosiluku ilmoitetaan nelinumeroisena ja se perustuu joko todelliseen tallennusvuoteen, arkistointivuoteen tai julkaistun tekstin ilmestymisvuoteen. Mikäli tekstissä annetaan vuosilukuväli, vuosiluvuksi on käytännön syistä valittu alkaen-luku (vaikka päättyen olisi haun kannalta loogisempi).

Säännönmukaistaminen toteutetaan tuottamalla vain tekstin ID-numeron ja ko. tiedot sisältäviä aakkostettuja taulukoita XML Writer -ohjelmassa käyttäen hyväksi tehtävää varten laadittuja XSL-tyylisivuja. Taulukot siirretään tekstinkäsittelyohjelmaan (WP), jossa makrojen avulla voidaan kerrallaan antaa tunnusnumerot suurelle joukolle yhtäläisiä tietoja. Lopuksi taulukko "siivotaan" kaikesta ylimääräisestä niin, että jäljelle jää vain tekstin ID ja siihen liittyvän tiedon tunnusnumero. Kustakin tiedosta (kerääjä, paikkakunta, vuosi) tulee omat tiedostot. Seuraavassa patkä SKVR I -osan kerääjätaulukkoa:

26	skvr1107380
26	skvr1107450
47	skvr1112220
47	skvr1110870
47	skvr1112520
64	skvr1108120
64	skvr1108180
64	skvr1108750
64	skvr1108960

Kerääjäluettelossa on vastaavasti tieto, että nro 26 on "Hannikainen, Lauri", 47 "Blomstedt-Ojoinen, Yrjö" ja 64 "Krohn, Kaarle".

Taulukkojen tiedot yhdistetään XML-tiedostoon aineistoa tietokantaan viettäessä. Käytännössä tämä tehdään suoraan SKS:sta CSC:n koneella olevan www-sivun kautta, johon ylläpitäjällä on oikeudet.

Tietokantaa varten on kehitetty systeemi, jossa säännönmukaistaminen tehdään varsinaisen perustekstin ulkopuolella eikä kosketa itse tekstin ilmiäsuun, tulee olemaan hyvin käyttökelpoinen SKVR-korpuksen jatkokehittämisessä ja myös muissa tekstikorpuksissa.

Tietokannan käyttö CSC:n tutkijan käyttöliittymän kautta on osoittautunut ongelmalliseksi. Käyttöä varten kaikkien käyttäjien on pitänyt hakea tunnus Tutkijan käyttöliittymää varten. CSC tarjoaa korkeakouluille ja tutkimuslaitoksille tietoteknistä tukea ja resursseja: muut potentiaaliset käyttäjät eivät varsinaisesti kuulu sen kohderyhmään. CSC:n kanssa sovittiinkin syksyllä 2004, että SKS vastaa jatkossa tietokannan ylläpidosta ja jatkokehittämisestä sekä avaa kannan vapaaseen käyttöön. Siirto toteutettiin maaliskuussa 2005.

4. SKVR-tietokanta – lopullinen versio

CSC:n kanssa yhteistyössä kehitelty SKVR-tietokanta on vielä väliaikainen: tekstihakuominaisuuksissa on runontutkijan kannalta joitain vakavia puutteita (haun kohdistaminen säkeisiin tai peräkkäisiin säkeisiin), hakutuloksessa ei ole osuman näyttöä, eikä SKVR:hin olennaisesti kuuluvaa sisällönmukaista järjestystä tai luokittelua ole otettu mukaan. Tähän on syynä se, että teoksen sisällysluettelot ovat vaikeasti käytettäviä, koska ne ovat hajallaan eri osissa ja laadittu eri aikoina erilaisin periaattein. Siten samat runot voivat esiintyä erilaisten otsikoiden alla. Ne eivät yksinkertaisesti sovellu yhtenäisen tietokannan haku-perusteiksi. Tilanteen korjaamiseksi Kansanrunousarkistossa on tekeillä yhtenäinen ja koko SKVR:n kattava runotyyppihakemisto. Se on järjestetty runolajeittain SKVR:n perusjaottelun mukaisesti (epiikka, lyriikka, loitsut, häärinot, lastenrunot jne.). Runolajin sisällä aineisto on jaoteltu runotyyppeihin. Kunkin runotyypin kohdalla luetellaan viitteet ko. runon kaikkiin SKVR:ssä julkaistuihin teksteihin. Usein liitteenä on kommentti, jossa luonnehditaan runon sisältöä ja kosketuksia muihin runoihin. Tämä hakemisto tullaan valmistuttuaan liittämään tietokantaan. Liittämässä voidaan käyttää hyväksi hakemistoprojektin tietojen avulla laadittavia taulukoita ja säännönmukaistetun tiedon tietokantaan yhdistämiseen luotua tekniikkaa.

Näiden uudistusten toteuttaminen vaatii todennäköisesti aivan uuden tietokantarakenteen. Vuoden 2005 aikana tullaankin tekemään uuden tietokannan määrittäminen, ja uudistus pyritään toteuttamaan vuoden 2006 aikana. Käytettävissä on aivan sama peruskorpus kuin nykyisessä tietokannassakin. XML-formaatti varmistaa aineiston siirrettävyyden uuteen käyttöympäristöön.

Digitaalisen tekstin (SKVR) ja arkistomuistiinpanojen suhde

SKVR sisältää alkuperäislähteestä kopioituja tekstejä. Valtaosa näistä alkuperäislähteistä kuuluu Kansanrunousarkiston kokoelmiin. Elias-projektissa digitoitu Lönnrotiana edustaa yhtä tärkeimmistä kalevalaisen runon lähdeaineistoista SKVR:ssä.

SKVR:ssä on pyritty toistamaan alkuperäislähteen teksti juuri sellaisena kuin se on, mutta sekä käsikirjoitetun tekstisivun muunto painetuksi tekstiksi että yksittäisten runotekstien poiminta käsikirjoituskontekstistaan hävittävät tietoa, jota teksteillä alkuperäisyhteydessään on. Tekstin asetelu, sijoittelu, kynänjälki jne. voivat kertoa paljonkin muistiinpanojen ajoituksesta ja korjailuista sekä muistiinpanotekniikasta ja sitä kautta perinteen tuottamisesta. Tekstien metatieto on usein eri paikassa kuin varsinainen muistiinpano: aiempien muistiinpanojen yhteydessä, erillisessä luettelossa tai siten, että tietojen yhteys juuri tiettyyn tekstiin on epävarma. Metatieto myös muokataan SKVR:n metatietoformaattiin. Itse Lönnrotianassa metatietoja on vähän, SKVR:hin päätyneistä suuri osa on päättelyitä. Koska SKVR ei erittele metatietojen lähteitä, alkuperäisen lähteen käyttö ja saatavuus on usein hyvin tärkeää.

Liitteessä 8 on selvitys mahdollisuuksista yhdistää SKVR:n tekstit alkuperäiskäsikirjoituksiin, sekä mahdollisuuksista käyttää SKVR:n digitoinnissa saatua merkkimuotoista metadataa digitoitujen käsikirjoitusten luettelointiin.

Mitä tietoa sitten on mahdollista saada käyttämällä rinnakkain SKVR-tietokantaa ja kuvamuotoon digitoituja Lönnrotianan alkuperäisaineistoja? Mitä "lähde" kertoo? Runojen analyysissa painetut tai digitoitut tekstit ovat käytettävyydeltään ylivertaiset. Jo Lönnrotin vaikea käsiala hankaloittaa tuntuvasti alkuperäisaineiston käyttöä. Lisäksi SKVR:ssä runot ovat sisällön eli runotyypin mukaisesti järjestettyinä: tätä mahdollisuutta ei pelkän käsikirjoituksen

käyttäjällä ole. Käsikirjoituksesta saadaan puolestaan esille yksittäisen tekstin tekstuaalinen yhteys: missä järjestyksessä runot yms. tekstit on esitetty (SKVR:ssä pitää vertailla yksikkö-numeroja, eikä SKVR sisällä kaikkea käsikirjoitusaineistoa). Käsiälästä ja tekstin sijoittelusta käsikirjoituksen sivuille voidaan tehdä päätelmiä muistiinpanotilanteesta. Olen vienälaisia runonlaulajia Arhippa ja Miihkali Perttusta tutkiessani käyttänyt Miihkali Perttuselta runsaasti keränneen A. A. Boreniuksen alkuperäisiä kenttämuistiinpanoja, joista on mahdollista rekonstruoida useita eri tekstin sanelu- ja laulukertoja, toisintoja³¹. SKVR, joka käyttää Boreniuksen puhtaaksikirjoitusta, hävittää näiden toisintojen eron. Lönnrotin runomuistiinpanot mm. Arhippa Perttuselta (Lönnrotiana 5:66–109. 1834) ovat kenttämuistiinpanoja. Ne sisältävät korjailuja ja lisäilyjä verrattuna Boreniuksen niin niukasti, että ne tuskin kertovat kahdesta eri esityskerrasta. Paremminkin muutokset on tehty sanelun kuluessa, joko kerääjän tai laulajan esittämästä selvennyksestä. Alkuperäismuistiinpanon kynänjälki viittaa tähän suuntaan.

SKVR:n tekstin kannalta ongelmallisinta on sen metadata: vaikka tiedot on huolellisesti harkittu ja sisältävät todennäköisesti vain vähän virheitä, niissä on tulkittavaa. Lönnrot ei merkinnyt laulajien nimiä johdonmukaisesti runomuistiinpanojen yhteyteen vaan antoi heistä tietoja muualla: kirjeissään, matkakertomuksissaan jne. Arhipan tapaamista Lönnrot selostaa julkaistun Kalevalan esipuheessa (1835), mutta Elias-projektin digitoimassa Kalevalan konseptissa (Lönnrotiana 35) sama kuvaus on painotuksiltaan hieman erilaisessa muodossa. Tapaamista Lönnrot selostaa myös Helsingfors Morgonbladetissa julkaistussa ruotsinkielisessä matkakertomuksessa. Konseptiteksti, jonka Lönnrot lähetti J. L. Runebergille julkaisua varten, on myös Lönnrotianassa, vaikka itse kirjettä ei enää olekaan tallella. Vertailemalla näitä eri versioita samasta tapahtumasta voidaan havaita, että Lönnrotin kuvaus muuttui koko ajan eikä näitä versioita voi pitää suorina etnografisena dokumentteina.³² Tällaisten vertailujen ja huomioiden tekeminen on digitoitujen käsikirjoitusten avulla huomattavasti helpompaa kuin käyttäen alkuperäisiä käsikirjoituksia: kuvaa on mahdollista suurentaa tekstin saamiseksi luettavammaksi, haurasta ja vaikeasti avautuvaa sidosta ei tarvitse selaila, työkopioiden valmistaminen on yksinkertaista.

Lönnrot ei yleensä merkinnyt runonlaulajien nimiä muistiinpanojen yhteyteen. Kuitenkin selailemalla ja lukemalla tarkasti Lönnrotin käsikirjoitusta keruumatkalta keväällä 1834 (Lönnrotiana 5) on mahdollista todeta, että hän jätti merkin Arhipasta dokumenttiin. Se ei ole ensimmäisen Arhipan esittämäksi otaksutun runon alussa (kuten metadatan yleensä mielletään olevan) vaan sen lopussa! Lemminkäisen virren (Lönnrotiana 5:66–67) jälkeen lukee selvästi Lönnrotin käsiälällä: "Af Arhippa i Latvajärvi by". Merkintä on todennäköisesti jälkikäteen laitettu kahta eri runoa erottavan tyhjän rivin ja siihen vedetyn poikkiviivan päälle (Elias-projektissa digitoitujen Lönnrotianan tiedosto lna005_045.jpg).

Lopuksi

Kalevalaisen runouden lähteiden ("Kalevalan lähteiden") julkaisu ja tutkimus ovat olleet SKS:n historiassa hyvin merkittävällä sijalla jo 1800-luvun lopulta saakka. Painettu SKVR 1900-luvun ensimmäisellä puoliskolla, julkaisemattoman runoaineiston kopiointi 1960–1980-luvuilla, runotyyppihakemisto 1980-luvulta asti, SKVR:n digitointi ja tietokanta sekä alkuperäislähteiden digitointi ovat osoituksia siitä valtavasta työmäärästä ja suuresta merkityksestä, joka asialle on SKS:n ja suomalaisen folkloristisen tutkimuksen piirissä annettu. Runotekstien ja lähde-

³¹ Saarinen, Jukka. *Variaatio Arhippa ja Miihkali Perttusen epiikassa*. Suomalaisen ja vertailevan kansanrunoudentutkimuksen pro gradu -työ. 1988. Julkaisematon käsikirjoitus.

³² Pasi Klemettinen on pitänyt aiheesta esitelmän Kuhmossa Jumin keossa 25.4.2004. Tekstiä ei ole julkaistu.

aineistojen (sekä keruuaineistojen että niiden taustamateriaalien) käytettävyyden ja hakumahdollisuuksien paraneminen ovat nykytutkimukselle välttämättömiä. Lisäksi ne avaavat nämä hauraat aineistot uudella tavalla niin eri tieteenaloja edustaville tutkijoille kuin muillekin käyttäjille.

III Valistus-digitointihanke

Amerin Kulttuurisäätiö ja SKS käynnistivät vuoden 2003 alussa hankkeen Amerin kulttuurisäätiön omistuksessa olevan lastenkirjataiteen digitoimiseksi (liite 10). Hankkeen koordinaattorina toimi Kirsi Keravuori SKS:n biografiakeskuksesta ja digitoinnin asiantuntijana Martti Kupiainen Kansallisarkistosta. Elias-digitoinnin projektityöntekijöistä Hannu Hautala ja Pasi Klemettinen osallistuivat eri vaiheissa vuosina 2002–2004 Valistus-digitointihankkeen suunnitteluryhmien toimintaan.

Amer-yhtymän toimitiloissa toteutetussa erillishankkeessa digitoitiin kustannusyhtiö Valistuksen kokoelman kuvitusoriginaaleja korkealaatuisiksi kuvatiedostoiksi ja kuvien luettelointitiedot tallennettiin kuvatietokannaksi (kuvapankki). Hankkeen tarkoituksena on parantaa arvokkaan kuvitusaineiston järjestystä, käytettävyyttä ja tunnettavuutta sekä saattaa kulttuurihistoriallisesti arvokas kokoelma paremmin sekä tutkijoiden että yleisön käyttöön verkkoympäristössä (web-palvelut). Hankkeen kiinnostavana osa-alueena mainittakoon kuvitusten keskeisten aihepiirien verkkohakua palvelevan asiansoituksen kehittäminen. Seuraavassa käsitellään tarkemmin Akatemian rahoittaman projektin kuluessa rinnakkain kehitettyä SKVR-tietokantaa ja Elias-kirjetietokantaa, joista molemmista on laadittu erillisraportit.

YHTEENVETOJA

Digitointilinjan käynnistämiseen tähtäävässä pilottihankkeessa saatiin arvokasta tietoa siitä, millaisia moninaisia ongelmia aineistojen systemaattiseen digitointiin, laiteympäristöön, tietovälineisiin, metatietojen käsittelyyn ja yleensä ajankäyttöön liittyy konkreettisella tasolla. On ilmeistä, ettei ilman käytännön kokemuksia ja tutkimustuloksia voida onnistuneesti laatia ohjeistuksia ja mallinnuksia, jotka tavalla tai toisella liittyvät tietotekniikkaan ja sen moninaiseen hallintaan. Yksityiskohtaiset kuvaukset työprosesseista ja -vaiheista ovat siten tarpeellisia, ja auttavat tulevaisuuden toimintojen suunnittelua. Toisaalta katse on suunnattu menneisyyteen, sillä kyseisen kaltaisissa hankekuvauksissa on kysymys myös sähköisten aineistojen, tietovälineiden ja laiteympäristöjen elinkaarten tärkeästä dokumentoinnista.

Arkistoaineistojen laajamittaisen digitoinnin näkökulmasta tavoitteena tulisi olla tuotantoyksikkö, jossa ammattitaitoa ja asiantuntijuutta olisi saatavilla kaikilta digitoinnin ja sähköisen arkistoinnin osa-alueilta. Käytännössä taloudellisia edellytyksiä suoranaisten tuotantoyksiköiden ylläpitämiseen on kuitenkin ainoastaan suuremmilla muistiorganisaatioilla. Kääntäen ilmaistuna pienten digitointiyksiköiden ongelmaksi voi muodostua henkilöstön vähäisyys, tietotaitojen kapea-alaisuus sekä teknisen infrastruktuurin yleinen heikkous. Riittämätön resursointi heijastuu mm. 1) digitointi- eli tuotantomääriin vuositasolla, 2) digitointiprosessin hallintaan ja sujuvuuteen sekä 3) laajemmin koko sähköisen arkistoinnin järjestelmällisyyteen. Kysymys ei ole yksiköiden tai organisaatioiden välisestä ”kuka ensin ja eniten” -kilpailuasetelmasta, vaan laajojen digitaalisten aineistomassojen teknisesti ja taloudellisesti mielekkästä tuottamisesta sekä pitkälle kehittyneestä hallinnoinnista (esim. pitkäaikaistallenteiden säilytys).

Erityisesti yksityisarkistojen realistisena vaihtoehtona on kaikkien tai joidenkin resursseja sitovien digitointitehtävien ulkoistaminen (esim. mikrofilmiskannaus). Mainittakoon, ettei organisaatioiden välisten yhteisten digitointikeskuksien perustaminen ole toistaiseksi saanut erityistä kannatusta. Tähän on epäilemättä syynä mm. alkuperäisaineistojen käsittelyyn ja siirtämiseen liittyvät kriittiset kysymykset sekä digitointiaikataulujen ylläpidon ongelmallisuus käytännön tasolla.

Kokopäiväisesti toimivat digitointilinjat edellyttävät aina mittavia alkuinvestointeja ja ylläpito-kustannuksia. Huomattakoon, ettei toiminnan laskennallinen tuottavuus ole sinänsä itse-tarkoitus ja se tulee aina suhteuttaa yksittäisen muistiorganisaation toimintastrategiaan ja digitointisuunnitelmaan. Jos organisaation tavoitteena on yksinomaan esitellä aineistojaan vaihtuvissa web-näyttelyissä, on mahdollista ja jopa suositeltavaa tukeutua kustannusedullisiin ja helposti hallittaviin tietoteknisiin ratkaisuihin. Toisaalta myös näennäisen pienin alku-kustannuksin käynnistetyt digitointiprosessit voivat laajetessaan ja sähköisten aineistomäärien kasvaessa muodostua organisaatiolle yllättävän kalliiksi, mikäli puutteellisesti ja/tai vähäisin resurssein suunnitellut, ohjeistetut ja toteutetut toiminnot sitovat jatkuvasti eri ammattilaisten työaika. Digitoinnin kokonaiskustannuksiin tulisikin laskennallisesti sisällyttää organisaatiossa toimivien IT-spesialistien, mikrotukihenkilöiden yms. tahojen neuvontaan, laitteiden ja ohjelmistojen asentamiseen sekä erilaisten teknisten ongelmien ratkaisemiseen käyttämä työaika. Periaatteessa kyseiset kustannukset on laskettavissa jälkikäteen, mikäli niistä pidetään kirjaa. Toimintojen suhteellisen tehokkuuden ja tehottomuuden selvittämiseksi voidaan tarvittaessa laskea tai arvioida myös yksittäisen kuvatiedoston tuotanto-, tallennus- ja hallinnointikustannukset. Kyseiset laskelmat on suoritettava aineistokohtaisesti, eivätkä eri laitoksien tai yksiköiden esittämät luvut ole aina vertailukelpoisia keskenään (laskentaperusteet vaihtelevat).

On selvää, ettei digitoinnin tuloksia voida mitata yksinomaan määrällisesti, sillä myös digitoinnin laadulliset kriteerit vaihtelevat suuresti. On täysin eri asia tuottaa korkealaatuisia ja julkaisukelpoisia värikuvia vanhoista kartoista kuin tuottaa harmaasävyisiä katselukuvia sanomalehdistä web-sovelluksiin. Laiterympäristösidonnaiset digitointimäärät voivat tällöin vaihdella muutamasta kymmenestä kuvatiedostosta tuhansiin päivässä. Muistiorganisaation on syytä arvioida ja suunnitella tarkkaan, minkä tyyppiseen ja tasoiseen digitointiin sillä on intressejä ja realistisia edellytyksiä.

Aineistojen systemaattisen ja samalla pitkäkestoisen digitoinnin käynnistäminen organisaatiossa edellyttää alla oleviin kriittisiin kysymyksiin vastaamista. Kysymykset eivät ole priorisointi-järjestyksessä ja ne nivoutuvat monelta osin toisiinsa. Esimerkiksi henkilö-, laite- ja tilaresurs-sien yksityiskohtainen suunnitteleminen edellyttää, että toiminnalle on olemassa taloudelliset edellytykset ja selkeät strategiset puitteet. Toisaalta tavoitteita ja puitteita on hallinnollisella ylätasolla vaikea asettaa, ellei digitoinnin perusteita ja käytännön realiteetteja ole etukäteen hahmotettavissa. Tässä mielessä digitointiin liittyvillä pilottihankkeilla on keskeinen rooli sekä kokemusten ja tietotaitojen kartuttamisessa että ennakoivassa suunnittelutyössä.

Miksi digitoidaan? (toimintastrategiat)

Kenelle digitoidaan? (kohderyhmät)

Mitä digitoidaan? (toimintasuunnitelmat)

Millä digitoidaan? (tietotekniset ratkaisut)

Miten digitoidaan? (käytänteet, ohjeistukset ja koulutustarpeet)

Ketkä digitoivat? (henkilöstöresurssit)

Missä digitoidaan? (tilasuunnitelmat)

Mistä taloudelliset voimavarat? (rahoitussuunnitelmat)

Kulttuurintutkimuksen kannalta muistiorganisaatioiden toteuttamalle digitoinnille voidaan asettaa seuraavia puitteita ja tavoitteita:

- 1) Keskeisenä tavoitteena tulisi olla analogisia ja digitaalisia tallenteita koskevien luettelointitietojen vieminen samaan viitetietokantaan, joka mahdollistaisi perustasolla kattavat haut koko arkistomateriaalista. Metatietojen keskitettyä sähköistä hallinnointia palvelevat erillishankkeet, joissa aiempia manuaalisia kortistoja, luetteloita ja sisällysluetteloita muutetaan tietojärjestelmien kautta helposti haettavaan ja käytettävään sähköiseen muotoon (verkkopalvelut).
- 2) Tutkimuskäytön kannalta on olennaista riittävän laajojen aineistomassojen systemaattinen digitointi sekä linkittäminen soveltuvien osin olemassa oleviin tietojärjestelmiin. Aineistojen massadigitoinnin tulee olla suunnitelmallista ja noudattaa muistiorganisaatiossa määriteltyä digitointisuunnitelmaa. Digitoitavat aineistot voidaan priorisoida niiden käyttökunnon, kysynnän sekä mahdollisten temaattisten tms. yhteyksien perusteella (arkistolliset rinnakkaisaineistot). Aineistojen valintaprosessiin voivat suoraan tai välillisesti vaikuttaa myös digitointiprojektien ulkopuolisten rahoittajien vaihtelevat intressit (digitoinnin painopisteet). Huomattakoon, ettei esimerkiksi viime vuosikymmeninä kerättyjen muistitietoaineistojen arvoa tai merkitystä voida määritellä samalla tavoin kuin digitoitavaksi priorisoitavien ainutlaatuisten historiallisten asiakirjojen ja museoesineiden.
- 3) Kolmantena painopistealueena mainittakoon erillisdigitointihankkeet, joissa voidaan toimia yhteistyössä arkistojen, yliopistojen yms. tutkimuslaitosten kanssa. Digitaalisten aineistojen tuottamisen ohella hankkeisiin liittyvät kiinteästi erilaiset räätälöidyt tietokantaratkaisut, jotka palvelevat suoraan käytännön tutkimustyötä (esim. tekstikorpuukset). Toisaalta on muistettava, että muistiorganisaatioiden tuottamien tietokantojen tulisi aina palvella myös laajempaa asiakaskuntaa. Tarvittaessa samaan tietokantaan voidaankin luoda eri käyttäjäryhmille esim. kaksitasoinen käyttäjäliittymä, joka mahdollistaa aineiston yleishaut sekä tarkemmat hakuvaihtoehdot ja niiden innovatiiviset kombinaatiot. Tutkimuksellisia digitointi- ja tietokantatarpeita voidaan arvioida ennalta yhteistyöryhmissä sekä laajemmin tutkijoiden yms. testiryhmien antaman palautteen ja arvioinnin perusteella (tietokantojen pilottiversiot). Sähköisten palveluiden kehittämiseen tähtävien käyttäjä-tutkimusten tekeminen tulee edellyttämään e-arkistoaineistojen nykyistä huomattavasti laajempaa ja systemaattisempaa (monitieteellistä) tutkimuskäyttöä, jolloin palveluiden toimivuutta ja käytettävyyttä voidaan todella kriittisesti arvioida.

Digitaalisesti syntyneiden ja/tai tuotettujen tallenteiden pitkäaikaissäilytys sukupolvelta toiselle tulee muodostamaan oman tietoteknisesti haastavan osa-alueensa arkistoinnissa. Aineistojen käytettävyys voidaan vaihtoehtoisesti turvata konvertoinneilla, eri datamuotojen käyttöön soveltuvien ohjelmistojen ja laitteistojen säilyttämisellä ja/tai vanhojen ja uusien laitteiden välisillä emuloinneilla. Kansainvälisiä suosituksia noudattavien tiedostoformaattien massakopiointia (siirto) uusille tietovälineille ja/tai massakonvertointia toisiin formaatteihin ei voida teknisessä mielessä pitää varsinaisena tulevaisuuden ongelmana. Pitkäaikaistallennukseen soveltuvien tietovälineiden tallennuskapasiteetti kasvaa jatkuvasti ja vastaavasti tietokoneiden suorituskyky paranee vuosittain, jonka seurauksena myös yksittäisen tiedoston käsittelystä ja säilyttämisestä aiheutuvat reaalikustannukset tulevat väistämättä laskemaan. Toisaalta on ilmeistä, että arkistoinnin kokonaiskustannukset nousevat mitä enemmän tallenteita on säilytettävänä ja hallinnoitavana. Aineistojen digitoinnista ja digitaalisesta arkistoinnista aiheutuvat kustannukset tuleekin suhteuttaa muistiorganisaation resursseihin ja muuhun toimintaan sekä eriasteisia toimenpiteitä vaativien aineistojen kokonaismäärään.

Digitaalisten aineistojen säilyttämisen rinnalla keskeisenä haasteena tulee olemaan ohjelmistosisonnoisten tietojärjestelmien toimivuuden takaaminen nopeasti muuttuvassa digitaalisessa maailmassa. Lähtökohtaisesti stabiileiksi suunniteltujen järjestelmien tulisi olla samanaikaisesti riittävän dynaamisia, jotta niissä voidaan hyödyntää uusia tietoteknisiä sovelluksia ja käyttäjäystävällisiä apuvälineitä. Arkistotietojärjestelmien kautta tulisi kyetä hallinnoimaan myös potentiaalisia uusia medioita ja aineistotyyppisiä, jotka eivät kuulu järjestelmän suunnitteluvaiheessa perinteisten arkistoitavien aineistojen ydinjoukkoon.

Arkistoaineiston elinkaaren näkökulmasta on tärkeää tietää missä formaatissa ja millä mediolla originaaliaineisto on toimitettu arkistoon, mitä teknisiä toimenpiteitä aineistolle on tehty ja missä eri muodoissa aineisto on saatavilla arkistossa (esim. sähköpostiviestit ja erilaiset liitetiedostot). Menneisyyden valikoivan haltuunottamisen ohella arkistojen tulisi strategisia suunnitelmia tehdessään kyetä ennakoimaan tulevia muutoksia ja niiden mukanaan tuomia haasteita. Tämä koskee osaltaan kulttuurin nykymuotojen aktiivista tallentamista; digitaalisessa verkkoympäristössä syntyvät, elävät ja nopeasti muotoaan muuttavat ilmiöt eivät ole useinkaan tallennettavissa takautuvasti esim. perinne-kyselyiden muodossa. Muistiorganisaatioiden eräänä strategisena tehtävänä onkin vastata kriittiseen kysymykseen: mitä tietoja haluamme ja kykenemme tallentamaan nykyisestä tietoyhteiskunnastamme ja sen tuottamasta mediavirrasta.

LÄHTEITÄ

Anttila, Aarne 1985. *Elias Lönnrot - elämä ja toiminta*. SKS, Helsinki.

Arkistojen kuvailu- ja luettelointisäännöt. [on-line] Arkistolaitos, Helsinki 1997.

Saatavilla:

<http://www.narc.fi/ohjeet.html>

Autiokari, Timo 1996–2005. *Accurate Image Manipulation for Desktop Publishing*. [on-line] Espoo.

[Viitattu: 9.2.2005.]

Saatavilla:

<http://www.aim-dtp.net/>

Blatner, David, Chavez, Conrad, Fleishman, Glenn & Roth, Steve 2004. *Real World Scanning & Halftones. Industrial Strength Production Techniques*. 3. ed. Peachpit Press, Berkeley, CA, USA.

Fraser, Bruce, Murphy, Chris ja Bunting, Fred 2004. *Värinhallinta*. IT-press, Helsinki 2004.

Extensible Markup Language (XML). [on-line] W3C 1996–2003. [Viitattu 9.2.2005.]

Saatavilla:

<http://www.w3.org/XML/>

Good Practices Handbook . Identification of good practices and competence centres. [on-line] Version 1.3.

Minerva Working Group 6 2004. [Viitattu 9.2.2005.]

Saatavilla:

http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf

<http://www.minervaeurope.org/listgoodpract.htm>

Harold, Elliotte Rusty & Means, Scott 2001. *XML in a Nutshell. A Desktop Quick Reference*. O'Reilly, Sebastopol, CA, USA.

ICC-värinhallinta. [on-line] PrePress Center Oy. [Viitattu 7.2.2005.]

Saatavilla:

<http://www.prepress-center.fi/icc/sivut/perusteet01.html>

<http://www.prepress-center.fi/icc/sivut/perusteet06.html>

Kulttuuriperintö tietoyhteiskunnassa. Strategiset tavoitteet ja toimenpide-ehdotukset. [on-line]

Opetusministeriön julkaisuja 2003:24. Opetusministeriö 2004. [Viitattu 9.2.2005.]

Saatavilla:

<http://www.minedu.fi/julkaisut/kulttuuri/index2003.html>

<http://www.minedu.fi/julkaisut/kulttuuri/2003/opm24/opm24.pdf>

Lehtinen, Outi ja Lounela, Mikko 2004: A model for composing and (re-)using text materials for linguistic research. Papers from the 30th Finnish Conference of Linguistics. Marja Nenonen (ed.). University of Joensuu 2004. Studies in Language, University of Joensuu Vol. 39.

Lounela, Mikko 2002: Aiming Towards Best Practices in XML Techniques for Text Corpora.

Annotation: City of Helsinki Public Works Department - A Case Study. Towards the Semantic Web and Web Services. Proceedings of the XML Finland 2002 Conference. Eero Hyvönen and Mika Klemettinen (eds.). Helsinki, Institute for Information Technology. HIIT Publications: 1458–9451; 2002–03, s. 123–135.

Majamaa, Raija (toim.) 1990. *Elias Lönnrot. Valitut teokset 1*. Kirjeet. SKS, Helsinki.

Perttunen, Kaisa-Marja 1976–1977. *Lönnrotiana 1–30. Selvitys Elias Lönnrotin käsikirjoituskokoelmasta.* Suomi 121, 1–4. SKS, Helsinki.

Sähköisten tietojärjestelmien ja aineistojen käsittely. Määräys ja ohje 126/40/01, 22.5.2001. [on-line] Arkistolaitos, Helsinki 2001.

Saatavilla:

<http://www.narc.fi/sahk>

Stenvall, Jani 2002. *Dublin Core -formaatin käyttöopas.* [on-line] Helsingin yliopiston kirjasto.

Saatavilla:

http://www.lib.helsinki.fi/dublin_core/dc-opas2002.html

Liite 1 Digitoitujen aineistojen luettelo

Lönnrotiana-kokoelma (SKS)

Lönnrotiana 1

Sidos sisältää 11 osaa, jotka on numeroitu juoksevasti 1 – 457.

Laajuus: 311 sivua

Kuvatiedostot: 183 (TIFF, JPG)

Tiedostonimet: Lna001_001 – 183

a) "Kantele taikka Suomen Kansan sekä Wanhoja että Nykysempiä Runoja ja Lauluja. Wiides osa."

Laajuus: (6) + 33 s.

Perinneyksiköt: 1-26

Keruu aika: 1828 – 1831

b) "Tässä seura Pistoksen sanat ja kateita vastan sanat Kokon pantu Michael Tuoman poialda 1846." (Luultavasti laulajan käsialaa)

Laajuus: 7 s.

Perinneyksiköt: 27 a-b

Keruu aika: 1846

c) Anders Vartiaisen runoja Kiuruvedeltä 30.8.1819. Anders Vartiaisen runoja ovat yhdeksän ensimmäistä, joista suurin osa (ellei kaikki) esiintyy myös Arwidssonin ja Crohnsin keräelmissä (28-29 h). Lönnrotin samalle arkille siististi kopioimat lienevät Ilomantsista J. Falckilta saatuja. Vartiaisen toisinnot ovat kaikki loitsuja, Falckilta saaduissa on lyyrisiäkin runoja.

Laajuus: 8 palstaa (luultavasti laulajan käsialaa) + 2 s.

(Lönnrotin käsialaa)

Perinneyksiköt: 28-39

Keruu aika: 30.8.1819

d) Runoja. "Begagnadt för Kalevala. A. J. Sjögrenin keräämiä" (ja hänen käsialaansa).

Laajuus: 16 s.

Perinneyksiköt: 40-50

Keruu aika: 1825

e) Vihko runoja. Lönnrotin kahdeksannen keruumatkan tuloksia (1838).

Laajuus: 20 s.

Perinneyksiköt: 51-141

Keruu aika: 1838

f) Vihko runoja. Lönnrotin kahdeksannen keruumatkan tuloksia (1838).

Laajuus: 24 s.

Perinneyksiköt: 142-206

Keruu aika: 1838

g) Runoja (Lönnrotin käsialan lisäksi ainakin kuutta tuntematonta). Lönnrotin käsialaa ovat numerot 229-242, jotka ovat hänen yhdeksännellä keruumatkallaan (1839)

Akonlahden seuduilla tekemiä muistiinpanoja. Mahdollisesti vielä numerot 243-247 ovat Lönnrotin omaa käsialaa. Muut (207-228) ovat eri henkilöiltä saatuja.

Laajuus: 58 s.

Perinneyksiköt: 207-247

Keruu aika: n. 1828-1839

h) Runoja Suomen Karjalasta, luultavasti Nurmeksesta. Lönnrotin ensimmäisellä keruumatkalla (1828)

puhtaaksikirjoitettuja tuloksia.

Laajuus: (2) + 45 s.

Perinneyksiköt: 248-329

Keruu aika: 1828

i) Runoja Ilomantsista, Taipaleelta, Enosta, Juvalta.

Ensimmäisen keruumatkan aineistoa. Numerot 330-340 on saatu Ilomantsista Falckilta. Suntiolta saatuja ovat ilmeisesti numerot 341-350 (joko kaikki tai osa). Numerot 351-371 on saatu Taipaleelta ja numerot 372-379 Pielisjärveltä.

Laajuus: (2) + 44 s.

Perinneyksiköt: 330-379

Keruu aika: 1828

j) Runoja Suomen Karjalasta. Ensimmäisellä keruumatkallaan 1828 Lönnrot alkoi saada runoja vasta Itä-Suomessa, Kerimäeltä lähtien. Käsialasta päätellen kaikki tekstit ovat puhtaaksikirjoitettuja.

Laajuus: (2) + 33 s.

Perinneyksiköt: 380-454

Keruu aika: n. 1828

k) Runoja. Mehiläisen 1838 tammikuun numeron käsikirjoitusta, jossa kaksi runoa "jälkimaineinen" ym.

Laajuus: 9 s.

Perinneyksiköt: 455-457

Keruu aika: 1838

Lönnrotiana 2

Runoja Suomen ja Venäjän Karjalasta. Kolmas keruumatka 1832. Runot ovat muokattuina ja puhtaaksikirjoitettuna sidoksessa 25.

Laajuus: 42 sivua

Perinneyksiköt: 1-31

Keruu aika: 1832

Kuvatiedostot: 28 (TIFF, JPG)

Tiedostonimet: Lna002_001 – 028

Lönnrotiana 3

Runoja Suomen ja Venäjän Karjalasta. Numerot 1-28 ovat kolmannelta keruumatkalta 1832 (ks. sidos 2). Numerot 30-81 ovat puolestaan neljänneltä matkalta 1833.

Laajuus: 62 s.

Perinneyksiköt: 1-85

Keruu aika: 1832-33

Kuvatiedostot: 41 (TIFF, JPG)

Tiedostonimet: Lna003_001 – 041

Lönnrotiana 4

Runoja Venäjän Karjalasta. Osa viidennen keruumatkan saaliista (ks. sidokset 5, 19 ja 20).

Laajuus: 98 s.

Perinneyksiköt: 1-91

Keruu aika: 1834

Kuvatiedostot: 54 (TIFF, JPG)

Tiedostonimet: Lna004_001 – 054

Lönnrotiana 5

Runoja Venäjän Karjalasta. Osa viidennen keruumatkan saaliista (ks. sidokset 4, 19 ja 20). Kaikki seitsemän rahvaanrunoa ovat muokattuina versioina sidoksessa 25.

Laajuus: 108 s.

Perinneyksiköt: 1-132

Keruu aika: 1834

Kuvatiedostot: 112 (TIFF, JPG)

Tiedostonimet: Lna005_001 – 112

Lönnrotiana 6

Runoja ja sananparsia. Sidokseen liitetyt paperit ovat hyvin eri ajoilta, vuosien 1839-73 väliltä. Alkumerot 1 a-u ja 2 a-n

(lyyriset runot) näyttävät olevan kopioita. Osalle niistä on löytynyt alkuperäinen käsikirjoitus (useimmiten sidoksessa 14).

Kopioiden käyttötarkoitus ei ole selvillä. Runot 3-8 ja 11-14 ovat yhdeksänneltä keruumatkalta 1839 Akonlahden seuduilta.

Laajuus: 12 s.

Perinneyksiköt: 1-14

Keruu aika: 1839-73

Kuvatiedostot: 16 (TIFF, JPG)

Tiedostonimet: Lna006_001 – 016

Lönnotiana 7

Vihko runoja Pohjois-Karjalasta. Pääasiassa riimillisiä kansanlauluja. "Tuupovaaran kansanlaulukokoelma". Kahdeksas keruumatka 1838 (ks. 1. sidoksen numerot 51-206).

Laajuus: 29 s.

Perinneyksiköt: 15-90

Keruu aika: 1838

Kuvatiedostot: 19 (TIFF, JPG)

Tiedostonimet: Lna007_001 – 019

Lönnotiana 8

Vihko runoja Ilomantsista. Kahdeksas keruumatka 1838. Muistiinpantu pääosin tai yksinomaan Mateli Kuivalattarelta.

Laajuus: 16 s.

Perinneyksiköt: 1-94

Keruu aika: 1838

Kuvatiedostot: 13 (TIFF, JPG)

Tiedostonimet: Lna008_001 – 013

Lönnotiana 9

Vihko Ritvalan helkavirsiä (käsiälä tuntematon) sekä Aug. Blåfieldin ja A. R. Niemen kirjeet. Kirjeet koskevat sidoksessa olevia Ritvalan helkavirsiä.

Laajuus: 4 s.

Perinneyksiköt: 1-3

Keruu aika: -1872

Kuvatiedostot: 13 (TIFF, JPG)

Tiedostonimet: Lna009_001 – 013

Lönnotiana 10

Runoja (käsiälä tuntematon).

Laajuus: 6 s. + 5 palstaa

Perinneyksiköt: 1-4

Keruu aika: Keruun on arveltu sijoittuvan vuosiin 1831-34

Kuvatiedostot: 11 (TIFF, JPG)

Tiedostonimet: Lna010_001 – 011

Lönnotiana 11

J. Juteinin julkaisemat sananlaskukokoelmat välilehditettyinä: "Uusia Sanan Laskuja, Toinen parannettu ja lisätty ulos-ando" (1817) ja "Walittuja Suomalaisten Sanan Laskuja" (1818). Sisältää runsaasti Lönnotin lisäyksiä vuosilta 1833-37.

Laajuus: 206 s.

Perinneyksiköt: 2673 kpl (c)

Aikarajaus: 1817, 1818, 1833-37

Kuvatiedostot: 107 (TIFF, JPG)

Tiedostonimet: Lna011_001 – 107

Lönnotiana 12

Arvoituskeräelmien puhtaaksikirjoitus v. 1836-37. Arvoitusten alku- ja hakusanaluettelaita. Katkelma Lönnotin "Suomen Kansan Arvoitusten" esipuheen luonnoksesta.

Laajuus: 166 s.

Perinneyksiköt: 1619 kpl (d)

Aikarajaus: 1836-37

Kuvatiedostot: 88 (TIFF, JPG)

Tiedostonimet: Lna012_001 – 088

Lönnotiana 13

Lönnotin julkaisun "Suomen Kansan Arvoituksia" käsikirjoituksen luonnos. 1840 (1. painos 1844, 2. painos 1851).

Laajuus: 190 s.

Perinneyksiköt: 1650 kpl (d)

Aikarajaus: 1840

Kuvatiedostot: 100 (TIFF, JPG)

Tiedostonimet: Lna013_001 – 100

Lönnotiana 14

Sidos runoja ja satuja sekä vuoden 1837 matkan kuvausten katkelmia. Sisältää runoja eri puolilta Karjalaa. Numerot 1-523 ovat suuren toisen vuoden 1837 tuloksia (Etelä-Karjalasta). Numerot 553-671a ovat mahdollisesti Vienasta seitsemänneltä keruumatkalta 1837. Sidos saattaa sisältää myös yhdeksän matkan 1839 saalista. Sidoksessa mukana 33 rahvaanrunoa, joista osa löytyy puhtaaksikirjoitetuina sidoksista 17 ja 25. Sidos on alkuperäinen ja osittain lyijykynällä kirjoitettu keräelmä. Siihen sisältyy monissa sidoksissa esiintyvien kopioiden alkuperäismuistiinpanoja.

Huomautukset: Aikojen kuluessa käsikirjoitus on kulunut ja nuhraantunut niin, että tulkinta saattaa usein tuottaa vaikeuksia.

Laajuus: 321 s.

Perinneyksiköt: 1-816

Aikarajaus: n. 1837-39

Kuvatiedostot: 166 (TIFF, JPG)

Tiedostonimet: Lna014_001 – 166

Lönnotiana 15

Lönnotin julkaisun "Suomen Kansan Sananlaskuja" (1842) käsikirjoitus (1840).

Laajuus: 533 s.

Perinneyksiköt: 6751 kpl (c)

Aikarajaus: 1840

Kuvatiedostot: 273 (TIFF, JPG)

Tiedostonimet: Lna015_001 – 273

Lönnotiana 16

Chr. Gananderin "Suomalaiset Arwotuxet" (1783) välilehditettyinä, runsaasti Lönnotin lisäyksiä.

Laajuus: 73 s.

Perinneyksiköt: 250 kpl (d)

Aikarajaus: 1835

Kuvatiedostot: 41 (TIFF, JPG)

Tiedostonimet: Lna016_001 – 041

Lönnotiana 17

Sidos sisältää 9 osaa, jotka on numeroitu juoksevasti 1 - 262. Sisältää hyvin erilaista aineistoa: 1) alkuperäismuistiinpanoista keruumatkalla tehtyjä kopioita, joita voidaan alkuperäisten kadottua pitää niiden veroisena, 2) Lönnotin kopioita painetusta materiaalista, 3) saman sidoksen tai muiden sidosten runoista sekä 4) kopioita, joissa muokkaus- ja yhdistelytyö on jo alkanut. Sisältää 1. ja 2. keruumatkan runoja.

Laajuus: 114 s.

Perinneyksiköt: 1-262

Kuvatiedostot: 65 (TIFF, JPG)

Tiedostonimet: Lna017hi_001 – 065

a) Runoja ym. muistiinpanoja. (Lönnotin ja tuntemattomien käsiälää) Ainakin osaksi puhtaaksikirjoitettua. Perinneyksiköt: 1-39

b) Kuten edellä. Numerot 44-47 ovat 1. keruumatkalta 1828. Perinneyksiköt: 40-51 ja 143
Keruu aika: 1828

Laajuus (a-b): 56 s.

Kuvatiedostot: 57 (TIFF, JPG)

Tiedostonimet: Lna017ab_001 – 057

c) Kuten edellä. Suurin osa riimillisiä kansanlauluja tai rahvaanrunoja, joukossa muutama todennäköisesti Lönnotin sepittämä tilapäisruno.

Laajuus: 34 s.

Perinneyksiköt: 52-66

d) Vihko seksuaalirunoja puhtaaksikirjoitettuna sekä kansan- ja rahvaanrunoja. Alkuperäismuistiinpanot ovat 1. sidoksen loppupuolella sekä sidoksessa 17 (osat e ja h).
Laajuus: 20 s.

Perinneyksiköt: 67-84

e) Runoja. Lönnrotin tekemiä kopioita Schröterin teoksesta Finnische Runen sekä lehtiin Oskyldigt Ingenting ja Mnemosyne painetuista A. Poppiuksen ja A. I. Arwidssonin runoista ym. Välissä on 1. keruumatkalla Kiteeltä saadut runot (134-141). Gottlundin Pienistä runoista tuttuja ovat numerot 142 a-b, joten ne lienevät kirjasta opittuja tai kopioituja.

Laajuus: 59 s.

Perinneyksiköt: 85-142

f) Vihko runoja ja mytologista sanastoa. Runot ovat Lönnrotin tekemässä mytologisten nimien luettelossa "Tillökningar till Gananders Finska Mythologi", joka on 1. ja 2. keruumatkan väliseltä ajalta.

Laajuus: 33 s.

Perinneyksiköt: 144-151

g) Runo "Jälkipuheita Sihtierin Jakob Juteinin muutamista sopimattomista runoista ja kirjoituksista.

Laajuus: 13 s.

Perinneyksiköt: 152

Laajuus (c-g): 162 sivua

Kuvatiedostot: 95 (TIFF, JPG)

Tiedostonimet: Lna017cg_001 – 095

h) Runoja ensimmäiseltä keruumatkalta (mm. Juhana Kainulaisen runoja).

Laajuus: 104 s.

Perinneyksiköt: 153-256

Keruu aika: 1828

i) Paavali Remeksen "Walitusruno" vuodelta 1765 ym. kansanrunoja (257-262), Vihko lienee 2. keruumatkalta 1831.

Laajuus: 10 s.

Perinneyksiköt: 257-262

Aikarajaus: 1765, 1831 (?)

Lönnrotiana 18

Vihko runoja. Alkupuoli tuntematonta, loppupuoli Lönnrotin käsialaa. Sidos sisältää ainoastaan Lönnrotin saamia tai kopioimia runoja.

Laajuus: 18 s.

Perinneyksiköt: 3-22

Keruu aika: n. 1815-1828

Kuvatiedostot: 22 (TIFF, JPG)

Tiedostonimet: Lna018_001 – 022

Lönnrotiana 19

Vihko runoja sekä osa Alku-Kalevalan esipuheen konseptia. Runot kerätty 5. keruumatkalla 1834 Vuokkiniemen alueelta (ks. sidokset 4-5, 20).

Laajuus: 6 s.

Perinneyksiköt: 1-4

Keruu aika: 1834

Kuvatiedostot: 10 (TIFF, JPG)

Tiedostonimet: Lna019_001 – 010

Lönnrotiana 20

Vihko runoja Venäjän Karjalasta. Runot kerätty 5. keruumatkalla (huhtikuu 1834).

Laajuus: 11 s.

Perinneyksiköt: 1-20

Keruu aika: 1834

Kuvatiedostot: 15 (TIFF, JPG)

Tiedostonimet: Lna020_001 – 015

Lönnrotiana 21

Vihko runoja, suurin osa loitsuja (mahdollisesti kopioita ja koosteita). Vain viisi runoista on julkaistu SKVR:ssa.

Laajuus: 19 s.

Perinneyksiköt: 1-36

Keruu aika: n. 1834-36

Kuvatiedostot: 23 (TIFF, JPG)

Tiedostonimet: Lna021_001 – 023

Lönnrotiana 22

Runoja Venäjän Karjalasta. Alkuperäisiä muistiinpanoja 6. keruumatkalta.

Laajuus: 57 s.

Perinneyksiköt: 1-93

Keruu aika: 1835

Kuvatiedostot: 61 (TIFF, JPG)

Tiedostonimet: Lna022_001 – 061

Lönnrotiana 23

Vihko runoja Venäjän Karjalasta. 6. keruumatka. Vihkon loppuosassa mahdollisesti kopioita (SKVR:ssa julkaisemattomia).

Laajuus: 8 s.

Perinneyksiköt: 1-20

Keruu aika: 1835

Kuvatiedostot: 12 (TIFF, JPG)

Tiedostonimet: Lna023_001 – 012

Lönnrotiana 24

Vihko runoja, joista osa Pentti Hirvosen sepittämiä. Vihko sisältää ainoastaan rahvaanrunoja ja riimillisiä kansanlauluja.

Laajuus: 36 s.

Perinneyksiköt: 1a-28

Keruu aika: n. 1832-47

Kuvatiedostot: 39 (TIFF, JPG)

Tiedostonimet: Lna021_001 – 039

Lönnrotiana 25

Vihko useiden rahvaanmiesten, mm. Petter Kettusen sepittämiä runoja. Rahvaanrunot on muokattuja ja puhtaaksikirjoitettuja (34 kpl). Vihko on yhtenäiseksi suunniteltu kokonaisuus. Runojen pohjalla olevia muistiinpanoja on eri sidoksissa.

Laajuus: 60 s.

Perinneyksiköt: 1-38

Aikarajaus: n. 1804-30

Kuvatiedostot: 64 (TIFF, JPG)

Tiedostonimet: Lna025_001 – 064

Lönnrotiana 26

Koottuja loitsuja. Faksimilepainoksena Loihtoja julkaissut Aarne Anttila 1932. Sisältää aakkosellisen sisällysluettelon. Lönnrotiana 26 on Suomen kansan muinaisten loitsurunojen 1. versio.

Laajuus: (2) + 94 s.

Perinneyksiköt:

Aikarajaus: n. 1834-1880

Kuvatiedostot: 100 (TIFF, JPG)

Tiedostonimet: Lna026_001 – 100

Lönnrotiana 27

Vihko runoja ja sävelmiä ym. muistiinpanoja. Alkuosa mahdollisesti 7. keruumatkalta.

Laajuus: 18 s.

Perinneyksiköt: 1-88

Aikarajaus: n. 1836-37

Kuvatiedostot: 20 (TIFF, JPG)

Tiedostonimet: Lna027_001 – 020

Lönnrotiana 28

"Loihtoja." Suurimmaksi osaksi no:n 26 puhtaaksikirjoitusta.
Laajuus: (2) + 104 s.
Perinneyksiköt: 31 loitsua
Aikarajaus: n. 1834-45
Kuvatiedostot: 23 (TIFF, JPG)
Tiedostonimet: Lna028_001 – 110

Lönnrotiana 29

a) "Wanhoja lauluja. Begagnadt för Kalevala. D:o för Kanteletar och Loihtia." Runoja Kajaanin seudulta ja Venäjän Karjalasta. Kantelettaren 1. luonnon. Ei sisällä alkuperäisiä muistiinpanoja vaan koosteita.
Laajuus: (2) + 157 s.
Perinneyksiköt:
Keruuuika: n. 1828-45
Kuvatiedostot: 160 (TIFF, JPG)
Tiedostonimet: Lna029a_001 – 160

b) Satuja, suurin osa Venäjän-Karjalasta vuosien 1836-37 keruumatkalta. Puhtaaksikirjoitettujen satujen kokoelma.
Laajuus: 75 s.
Perinneyksiköt: 1-51
Keruuuika: n. 1836-37

Kuvatiedostot: 79 (TIFF, JPG)
Tiedostonimet: Lna029b_001 – 079

Lönnrotiana 30

Sidokseen on koottu sekalaisia muistiinpanoja, jotka on jaettu ryhmiin a-h.
Laajuus: 76 s.
Kuvatiedostot: 96 (TIFF, JPG)
Tiedostonimet: Lna030_001 – 096

a) "Runo Klaus Kurki (= Elinan surma) ja Murheruno Pielisen Kirkkoherra Jakob Steniuksen kuolemasta".
Laajuus: 8 s.
Perinneyksiköt: 1-3
Keruuuika: n. 1820-28

b) Hajamuistiinpanoja eri matkoilta.
Laajuus: 13 erikokoista lehteä
Perinneyksiköt: 1-18
Keruuuika: 1828-35

c) Kansanrunojen muistiinpanoja.
Laajuus: 11 + 3 s.
Perinneyksiköt: 1-58
Keruuuika: 1839?

d) Lemmen nosto. Kuvausta ja runoja.
Laajuus: 16 s.
Perinneyksiköt: 1-3
Keruuuika:

e) Runojen luetteloa ja kolme runoa. Runojen numerot viittaavat sidokseen 17, johon runot ja luettelo ovat varsinaisesti kuuluneet.
Laajuus: 4 s.
Perinneyksiköt: 151b, 262, 262a
Keruuuika:

f) Itkurunoja, sävelmä, sekalaisia muistiinpanoja.
Laajuus: 11 s.
Perinneyksiköt: 1-10
Keruuuika: 1835

g) Samojedilaisia sanoja venäläisin selityksin. Ilmeisesti vuoden 1842 matkalta.
Laajuus: 4 + 2 s.
Perinneyksiköt:
Keruuuika: 1842?

h) Sekalaisia muistiinpanoja.
Laajuus: 4 erikokoista paperia

Perinneyksiköt: 1-4
Keruuuika:

Lönnrotiana 31

Kantele taikka Suomen Kansan sekä Wanhoja että Nykysempiä Runoja ja Lauluja. I-IV. (Välilehditettyinä K. N. Keckmanin ruotsinnoksena, marginaaleissa Lönnrotin korjauksia ja selityksiä)

466 sivua
241 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna031_001 – 241

Lönnrotiana 32

Väitöskirjan De Väinämöine aineksia, 2 s. Kalevalan esitöitä: Lemminkäinen, Väinämöinen, Naimakansan virsiä, 88 s.

90 sivua
47 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna032_001 – 047

Lönnrotiana 33

Kalevalan esitöitä: Runokokous Väinämöisestä, 114 s. Sidos.

114 sivua
60 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna033_001 – 060

Lönnrotiana 34

Lisäyksiä Runokokoukseen Väinämöisestä, 7 s. (K. N. Keckmanin käsialaa) + 111 s. (Lönnrotin käsialaa). Sidos.

118 sivua
119 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna034_001 – 119

Lönnrotiana 35

Kalevala taikka Vanhoja Karjalan Runoja Suomen kansan entisistä ajoista. Edellisessä (alkuperäisessä päällysy-)lehdessä: N. L. Väinämöisestä. Väinölä. (1) + 351 s. Sidos.

352 sivua
358 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna035_001 – 358

Lönnrotiana 36

Kalevala taikka Wanhoja Karjalan Runoja Suomen kansan entisistä ajoista. (Edellisen puhtaaksikirjoitus), 313 s. Sidos.

313 sivua
317 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna036_001 – 317

Lönnrotiana 37

Toisinnoilla runsaasti lisäilty välilehtinen kappale Vanhaa Kalevalaa. (aluska kaksi A. A. Borenius-Lähteenkorvan 9.10.1873 ja 4.1.1895 päiväämää tiedonantoa). Sidos.

837 sivua
587 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna037_001 – 587

Lönnrotiana 38

Uuden Kalevalan käsikirjoitus (esipuhe ja vähäinen osa runoja Lönnrotin omaa käsialaa, muu kahden puhtaaksikirjoittajan), 60 + 501 + 300 + 132+47 s.Sidos.

1042 sivua
521 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna038_001 – 521

Lönnrotiana 39

Lönnrotin luennoillaan käyttämä ja huomautuksilla varustama Kalevalan toisen painoksen kappale. Sidos.

380 sivua
196 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna039_001 – 196

Lönnrotiana 40

Lisiä Uuden Kalevalan I ja II runoon, 17 s. Sidos.

17 sivua
10 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna040_001 – 040

Lönnrotiana 41

A. R. Niemen painoa varten toimittama kopio Lönnrotin lisistä Uuden Kalevalan I ja II runoon. 19s.

19 sivua
22 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna041_001 – 022

Lönnrotiana 42

Mytologinen nimiluettelo Kalevalaan ym. 59 s. Sidos.

59 sivua
61 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna042_001 – 061

Lönnrotiana 43

Kalevalan selityksiä. 8 s.

8 sivua
8 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna043_001 – 008

Lönnrotiana 44

Korjauksia Kalevalan toisen painoksen tekstiin, 1 s.

1 sivua
2 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna044_001 – 002

Lönnrotiana 45

Lyhennetyt, v. 1862 painetun Kalevalan käsikirjoitus. [Osa sanastoa vierasta käsialaa.] 593 s. Sidos.

593 sivua
288 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna045_001 – 288

Lönnrotiana 49

Kanteletar. I painoksen esitöitä. I (& II) kirja. 592 s. Sidos.

592 sivua
204 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna049_001 – 204

Lönnrotiana 50

Kuin ed. II (& I) kirja. 402 s. Sidos.

402 sivua
211 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna050_001 – 211

Lönnrotiana 51

Kuin ed. III kirja. 193 s. Sidos.

193 sivua
101 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna051_001 – 101

Lönnrotiana 77

Matkamuiistiinpanoja Ja -kertomuksia v:n 1835 ja 1837 keruumatkoilta. 53 s. Kirje SKS:lle 22.XI.1837. 3 s. Sidos.

56 sivua
55 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna077_001 – 055

Lönnrotiana 78-83

Lönnrotin kirjekonsepteja (ks. Lönnrotin kirjeenvaihto)

Lönnrotiana 95

Viikko kanteleella soitettavia laulunuohteja. [Nuotit on merkitty numeroin, L:n laatiman järjestelmän mukaisesti.] 79 s.

79 sivua
44 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna095_001 – 044

Lönnrotiana 99

a) Päiväkirjaa vuodelta 1836. [Lönnrotin ja Cajanin yhteiseltä keruumatkalta Uhtualle.] - Pro Memoria i Finska Syntaxen. - Aphorismer. [Miltei kokonaan Cajanin käsialaa.] 115 s.
b) Vienankarjalainen sana- ja arvoituskokoelma syksyltä 1836. [Cajanin käsialaa.] 174 s.

289 sivua
119 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna099_001 – 119

Lönnrotiana 121

Selityksiä Kalevalan runoihin 1, 2, 4, 5, 6, 7, 18-50. 154 s. Sidos.

154 sivua
80 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna121_001 – 080

Lönnrotiana 122

1) Muistiinpanoja E. Lönnrotin luennoista: Sana- ja asiaselityksiä Kalevalaan ja Kantelettareen. 21 s. [Vain 16 sivua] - Samoin Loitsurunoihin. 7 s.
- 2) Kalevalan yksityisten sanojen luetteloa. 44 s. [Vain 4 sivua]

27 sivua
29 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna122_001 – 029

Lönnrotiana 123

Kalevalan selityksiä. [L: n antamia vastauksia F. W. Rothstenin tiedusteluihin.] 81 s.

81 sivua
76 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna123_001 – 076

Lönnotiana 125

Alkajaisluento suomalaisten muinaisuskosta eli mytologiasta, päivätty 18.IX.1856. 13 s.

13 sivua
8 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna125_001 – 008

Lönnotiana 127

Muistiinpanoja Nurmijärveltä 30. X. 1831, Huitulasta 3. XI. 1831 ja Hämeenlinnasta 12. XI. 1831. 3 s.

3 sivua
4 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna127_001 – 004

Lönnotiana 128

Ödmjukt memorial [koleerataarkastuksesta Huitulassa marraskuulla v. 1831, konsepti]. 4 s.

4 sivua
4 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna128_001 – 004

Lönnotiana 129

Muistiinpanoja Aunuksen matkalta 12.III.-17.IV.1841. [Säämäjärvi - Petroskoi - Sortavala - Jaakkima.] 24 s.

24 sivua
14 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna129_001 – 014

Lönnotiana 130

Muistiinpanot Kyröskoskella käynnistä elok. 1841. 5 s.

5 sivua
6 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna130_001 – 006

Lönnotiana 131

Bref om en resa från Kola till Kem. 4. V. 1842. 40 s.

40 sivua
42 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna131_001 – 042

Lönnotiana 132

Päiväkirja-muistiinpanot Lapin matkalta v. 1841-42 [Haukipudas - Salla - Inari - Karasjoki - Inari - Kuola]. 20 s. - Tilasto-ym. tietoja Kuolan niemimaasta. 2 s. - Taulukko välimatkoista Kuolan niemimaan lukuisien paikkain välillä, 1 s. - Venäjänkielinen kyyditysmääräys. 1 s.

24 sivua
26 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna132_001 – 026

Lönnotiana 133

Päiväkirja-muistiinpanoja: Kemi, 4. toukok. 1842. 5 s. - Kirjeiden konseptteja, päivättyjä Arkangelissa kesä- ja heinäk. 1842, 8 s.

13 sivua
6 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna133_001 – 006

Lönnotiana 134

Kirje Rabbelle Arkangelista 24. ja 26. VI. 1842. 30 s.

30 sivua
31 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna134_001 – 031

Lönnotiana 135

Kirje Rabbelle Pertominskista 11. VII, Onegasta 15. VII, Kargopolista 23.VII.1842, 31 s.

31 sivua
33 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna135_001 – 033

Lönnotiana 136

Kirje Rabbelle Stantsien Polkovasta 3. VII, Vyitegrasta 4. VIII, Lotinapellostä 18. VIII, Koivikosta 8. X, 1842. 12 s.

12 sivua
12 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna136_001 – 012

Lönnotiana 139

Viron matkan aikaisia (1844) kirjeitä ja konseptteja sekä 2 venäjänkielistä passia. 34 s.
[6 kirjettä tai konseptia (26 sivua) siirretty kirjekokoelmaan.]

8 sivua
12 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna139_001 – 012

Lönnotiana 171

Suomen Kansan Muinaisien Loitsurunojen esitöitä. 274 s. Sidos.

274 sivua
138 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna171_001 – 138

Lönnotiana 172

Suomen Kansan Muinaisien Loitsurunojen esitöitä. 6 vihkoa. 724 s.

Lönnotiana 172a
54 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna172a_001 – 054

Lönnotiana 172b
53 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna172b_001 – 053

Lönnotiana 172c
76 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna172c_001 – 076

Lönnotiana 172d
82 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna172d_001 – 082

Lönnotiana 172e
82 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna172e_001 – 082

Lönnotiana 172f
43 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna172f_001 – 043

Lönnotiana 173

Loitsurunojen esitöitä. [Kanteleen ja Kantelettaren leikkeleit, joissa L:n tekemiä korjauksia.] 36 s.

36 sivua
20 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna173_001 – 020

Lönnotiana 174

Loitsurunojen esipuheen luonnosta ym. 46 s.

46 sivua
23 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna174_001 – 023

Lönnotiana 175

Loitsurunojen erisnimien ja loitsujen aakkosellinen luettelo. 52 s.

52 sivua
28 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna175_001 – 028

Lönnotiana 176

Loitsurunojen käsikirjoitus. 966 s. [176 ja 177 yhteensä]

534 sivua
536 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna176_001 – 536

Lönnotiana 177

Loitsurunojen käsikirjoitus. 966 s. [176 ja 177 yhteensä]

432 sivua
434 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna177_001 – 434

Lönnotin kirjeenvaihto***Lönnotin saamat kirjeet***

Suomalaisen Kirjallisuuden Seuran kirjekokoelmien ensimmäisiksi siteiksi on järjestetty Elias Lönnotin saamat kirjeet, joista vain suhteellisen harvat ovat aikoinaan kadonneet tai Lönnotin määräyksestä liian yksityisluontoisina hävitetty. Kirjeet on pantu aikajärjestykseen ja sidottu 18:ksi sidokseksi. Sidoksiin liittyy luettelo, jossa kirjeet ovat lähettäjiä ja päiväyksiä mukaisessa aakkos- ja aikajärjestyksessä (69 s). Kuvatiedostot on nimetty arkistosignumeiden mukaisesti (tyhjiä sivuja ja liuskoja ei ole digitoitu).

Lönnotiana kirjeet 1

1825-1833..... 653 s.

653 sivua
532 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak001_001 – 652

Lönnotiana kirjeet 2

1834-25.IX. 1835..... 561 s.

561 sivua
442 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak002_001 – 561

Lönnotiana kirjeet 3

27.IX.1835-15.VIII. 1837.. 602 s.

602 sivua
469 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak003_001 – 602

Lönnotiana 178

Loitsurunojen erisnimien luettelo. [Perusteksti vierasta käsialaa, korjaukset L:n.] 119 s.

119 sivua
62 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna178_001 – 062

Lönnotiana 179

Loitsurunojen Sisällyksen ja Esipuheen käsikirjoitus. 109 s.

109 sivua
59 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna179_001 – 059

Lönnotiana 180

Loitsurunojen "Liitoksia". II. Outoja sanoja. [Perusteksti puhtaaksikirjoittajan, korjaukset L:n käsialaa.] 147 palstaa.

147 sivua (palstaa)
74 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna180_001 – 074

Lönnotiana 368

Lönnot, E., Muistiinpanoja ja huomautuksia välilehditettyssä kappaleessa Gananderin Mythologia Fennicaa.

112 sivua
71 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna368_001 – 071

Lönnotiana 370

Lönnotin De Väinämöine välilehditettyinä kappaleena, joitakin tekijän huomautuksia.

16 sivua
17 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna370_001 – 017

Lönnotiana kirjeet 4

26.X. 1837-19. VII. 1839..... 568 s.

568 sivua
451 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak004_001 – 568

Lönnotiana kirjeet 5

21.VII.1839-9.IV.1843..... 609 s.

609 sivua
476 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak005_001 – 609

Lönnotiana kirjeet 6

19. IV. 1843-1845 669 s.

669 sivua
512 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak006_001 – 668

Lönnotiana kirjeet 7

1846-47..... 694 s.

694 sivua
544 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak007_001 – 694**Lönnotiana kirjeet 8**

1848-49..... 685 s.

685 sivua
537 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak008_001 – 685**Lönnotiana kirjeet 9**

1850-15.X. 1852 689 s.

689 sivua
504 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak009_001 – 689**Lönnotiana kirjeet 10**

16. X. 1852-30. IV. 1855 678 s.

678 sivua
494 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak010_001 – 678**Lönnotiana kirjeet 11**

7. V. 1855-16. IV. 1858..... 667 s.

667 sivua
478 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak011_001 – 666**Lönnotiana kirjeet 12**

17.IV. 1858-31. V. 1862..... 685 s.

685 sivua
465 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak012_001 – 685**Lönnotin lähettämät kirjeet**

Lönnotin lähettämät kirjeet eivät ole läheskään yhtä hyvin järjestettyinä ja lukuisina tallella, mutta aukkoja täydentävät Lönnotiana-kokoelman numeroidut 78-83 ja 321-322 konseptikirjeet.

Lönnotiana kirjeet 19

Lönnotin alkuperäisiä kirjeitä, 184 kpl. Vastaanottajina: F. Collan, J. F. Elfving, Cl. Collan, V. S. Schildt, C. P. Flander, Chr. E. Barck, F. F. Lönnot, K. Dahlberg, C. G. v. Essen, J. F. Cajan, A. E. Ahlqvist, C. G. Borg, J. Krohn, J. A. Cleve rouvineen, J. F. Bergh.

184 kappaletta
531 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak019_000asis – 246e**Lönnotiana kirjeet 20**

Lönnotin alkuperäisiä kirjeitä, 51 kpl [Anttilan teoksessa määräksi ilmoitettu 151 kpl]. Vastaanottajina: K. N. Keckman, F. J. Rabbe, C. D. v. Haartman, Rafn, A. F. Lönnot, C. H. Ståhlberg.

51 kappaletta
472 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak020_000sis_a – 166l**Lönnotiana kirjeet 13**15. VII 1862-23. V. 1866 713 s.
713 sivua
422 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak013_001 – 713**Lönnotiana kirjeet 14**

25. V. 1866-7. VIII. 1869..... 654 s.

654 sivua
405 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna014_001 – 654**Lönnotiana kirjeet 15**

24.VIII 1869-13. V.1 872.... 660s.

660 sivua
390 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak015_001 – 660**Lönnotiana kirjeet 16**

18. V.1872-6. XII 1875 652s.

652 sivua
375 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak016_001 – 652**Lönnotiana kirjeet 17**

I. III 1876-30.XII 1881 661s.

661 sivua
366 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak017_001 – 661**Lönnotiana kirjeet 18**

13. I 1882-14. III 1884..... 455 s.

456 sivua [Anttilan luettelossa 455 s.]
316 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak018_001 – 456**Lönnotiana kirjeet 26**

Lönnotin alkuperäisiä kirjeitä (osa kopioina), 24 kpl. Vastaanottajina: J. F. Granlund, J. F. Bergh, M. Akiander, K. Bergbom, K. G. Leidenius. [Kansiosta digitoitu etulehti ja sivut 653-758]

107 sivua
55 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak026_652 – 758**Lönnotiana kirjeet 32**

Lönnotin kirjeitä A. Vareliukselle. (Osa kopioina), 50 kpl. [Kansiosta digitoitu sivut 377-516]

140 sivua
115 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lnak032_377 – 516**Kirjekonseptit****Lönnotiana 78**

L:n kirjekonsepteja I. v. 1833 - 34. 264 s. Sidos (kansi digitoitu).

264 sivua
265 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna078_001 – 265 (tekninen juokseva numerointi)

Lönnotiana 79

L:n kirjekonsepteja II. v. 1835- 37. 268 s. Sidos (kansi digitoitu).
268 sivua
265 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna079_001 – 265 (tekninen juokseva numerointi)

Lönnotiana 80

L:n kirjekonsepteja III. v. 1837-39, 1842-46. 203 s. Sidos (kansi digitoitu).

203 sivua
170 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna080_001 – 170 (tekninen juokseva numerointi)

Lönnotiana 81

L:n kirjekonsepteja IV. v. 1847-49. 167 s. Sidos (kansi digitoitu).

167 sivua [Anttilan luettelossa 164 s.]
86 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna081_001 – 086 (tekninen juokseva numerointi)

Lönnotiana 82

L:n kirjekonsepteja V. v. 1867- 1883. 353 s. Sidos (kansi digitoitu).

353 sivua
183 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna082_001 – 183 (tekninen juokseva numerointi)

Lönnotiana 83

L:n kirjeitä ja konsepteja v. 1826-82, tilivihko v. 1830-31, ansioluettelo, yht. 29 kpl. Sidos (kansi digitoitu).

29 kappaletta
86 kuvatiedostoa (TIFF, JPG)
Tiedostonimet: Lna083_000kansia – 029b (numeroitu signumeiden mukaisesti)

Liite 2 Konferenssit ja työseminaarit

Konferenssi: The digitization and automated content exploration of Finno-Ugric cultural heritage

Aika: 14.09.2001

Paikka: National Czechényi Library, Budapest

Osallistuja: Pasi Klemettinen

Esitelmän aihe: The Ancient Poems of the Finnish People – Digitising Project

Projektin työseminaari: Digitoitu arkistoaines kulttuurintutkimuksessa

Aika: 07.11.2001

Paikka: Suomalaisen Kirjallisuuden Seuran juhlasali

Ohjelma: liite 2.1.

Konferenssi: Dainu Skapis - historical texts and cultural contexts

Aika: 30.10.2002

Paikka: Riika, Tiedeakatemia

Järjestäjät:

Archives of Latvian Folklore

Institute of Literature, Folklore and Art (University of Latvia)

Osallistuja: Pasi Klemettinen

Esitelmän aihe: Digitising the Finnish Folk Poetry

Projektin työseminaari: Digitoitu arkistoaineisto kulttuurintutkimuksessa

Aika: 28.3.2003

Paikka: Suomalaisen Kirjallisuuden Seuran juhlasali

Ohjelma: liite 2.2.

Työseminaari: Atk-yhteistyö Viron ja Latvian perinnearkistojen kanssa

Aika: 10.-12.6.2003

Paikka: Eesti Rahvaluule Arhiiv, Eesti Kirjandusmuuseum (Tartto)

Osallistajat: Pasi Klemettinen, Jukka Saarinen

Kansainvälinen seminaari: Folklore Archives and Cultural Dynamics

Problems of Digitizing and Management of Intangible Cultural Heritage

Aika: 19-20.11.2004

Paikka: SKS, Helsinki

Osallistuja: Pasi Klemettinen, Jukka Saarinen

Esitelmän aihe: Digitisation of Manuscripts and Archiving Processes

Muut tilaisuudet

SMA 32 -digitointilaitteiston esittelytilaisuus

Aika ja paikka: 18.5.2001, Riksarkivet (Tukholma)

Osallistajat: Hannu Hautala, Pasi Klemettinen

Järjestäjät: Myung-Soo Han (Mitema & HMS Mikrosystem AB)

Karl-Magnus Drake (IT-osaston johtaja, Riksarkivet, Tukholma)

SMA 21 -digitointilaitteiston testaaminen

Aika ja paikka: 24.01.2002, Elias-digitointi (SKS)

Osallistajat:

Karl-Magnus Drake (IT-osaston johtaja, Riksarkivet, Tukholma)

Peter Bergman (SVAR, Riksarkivet, Fränsta)

Bo Lindkvist (SVAR, Riksarkivet, Fränsta)

Myung-Soo Han (Mitema & HMS Mikrosystem AB)

Martti Kupiainen (digitointi, Kansallisarkisto)

Pirkko Pirilä (Kansallisarkiston teknisen yksikön johtaja)

SKVR-digitointihankkeen koordinointi ja suunnittelu

Aika: 14.-16.03.2002

Paikka: Tartto

Osallistajat: Jukka Saarinen ja Pasi Klemettinen
Digitointiryhmän edustajat: Arvo Krikmann ja Mart Orav

SMA 21 -laitteiston testaaminen

Aika ja paikka: 16.-18.03.2002, Hannover (CeBIT 2002)

Osallistuja: Matti Anttila (SKS:n ATK-vastaava)

Laitteiston edustajat:

Myung-Soo Han (Mitema & HMS Mikrosystem AB)

Josef Schaut (SMA-edocument)

Lasse Nirhamon vierailu

Aika: 15.04.2002

Paikka: Elias-digitointi (SKS)

Aihe: Elias Lönrotin kirjeenvaihdosta 1980-luvulla tehtyjen CPT-tekstinkäsittelytiedostojen konvertointi (Elias-kirjetietokanta).

Taustayhteisö: Suomen Tietojenkäsittelymuseoyhdistys (Jyväskylä)

Elias-kirjetietokannan suunnittelu

Aika: 11.9.2002

Paikka: Elias-digitointi (SKS)

Osallistajat: Pasi Klemettinen, Pekka Tolonen

Aika: 21.10.2002

Paikka: Kalevala-instituutti (Turku)

Osallistajat: Pasi Klemettinen, Pekka Tolonen

Liite 3 Lönnotiana ja Elias Lönnotin kirjeenvaihto tutkimuskohteeksi digitaaliarkistona

Raija Majamaa

Digitoitu arkistoaineisto kulttuurin tutkimuksessa -hankkeen taustaa

Suomalaisen Kirjallisuuden Seuran hallussa oleva Elias Lönnotin kirjallinen jäämistö on osa kansallisesti ja kansainvälisesti merkittävää kulttuuriperintöä. Suomen Unesco-toimikunta suositteli vuonna 2001 The Elias Lönnot Collection -kokoelmaa ehdolle Unescon Memory of the World-rekisteriin. Perusteluineen se on yhä nähtävillä Nominations 2001-listalla, vaikkei saavuttanutkaan rekisterinimitystä (<http://portal.unesco.org>). Sen ohittivat valinnassa mm. Gutenbergin Raamattu, Beethovenin 9. sinfonian (d-minor, op.125) käsikirjoitus, J. W. von Goethen kirjallinen perintö ja Henrik Ibsenin Nukkekodin käsikirjoitus.

Elias Lönnotin monitieteinen kirjallinen jäämistö sekä myöhäiskeskiaikainen runo ”Jos mun tuttuni tulisi” liittyvät läheisesti suomalaiseen kansalliseen herätykseen, jolla oli yleiseurooppalainen tausta. Napoleonin sotien jälkeisenä ajanjaksona eurooppalaiset oppineet kiinnostuivat kansan kielestä ja kulttuurista ja kävivät keskenään kirjeenvaihtoa seuraten Suomessakin tapahtuvaa valistuksen jälkeistä kulttuurista kehitystä. Kalevala (1835) herätti ilmestyttyään valtaisaan kansainvälistä kiinnostusta, joka yhä jatkuu epiikan tutkijoiden piirissä uusien näkökulmien noustessa esiin. Eepoksen vaikutus on nähtävissä itsenäisyyttään etsivien kansakuntien taistelussa kohti itseymmärrystä. Elias Lönnotin nimi tunnetaan maailmanlaajuisesti oraalista historiaa keräävissä ja arkistoisissa laitoksissa.

Suomalaisen Kirjallisuuden Seuran kirjallisuusarkiston kokoelmiin kuuluu ns. Zetterqvistin kokoelma, joka on sidos eri puolilta maailmaa kerättyjä käännöksiä ”Jos mun tuttuni tulisi”-runosta 1850-luvun puolivälistä. Tullimies Zetterqvist toteutti suuren keräilyideansa ns. ketjukirjeenä. Hän kääntyi eri puolilla maailmaa vaikuttavien kirjailijoiden ja tiedemiesten puoleen hyväksikäyttäen suomalaisia kuuluisuuksia ja pyysi heitä suorittamaan käännoksen omalla kielellään tai murteellaan. Vastajien joukossa on kansainvälisiä kuuluisuuksia, mm. J. W. von Goethe ja Pietarin tiedeakatemian jäsen, sheikki Tantawi. Käännöksiä tuli 467 kielelle ja murteelle, joukossa on myös jo kadonneita kieliä. Kokoelmaa ei ole aikomuksista huolimatta koskaan kyetty painamaan puuttuvien kirjasimien vuoksi. Käännökset ovat osa maailmanlaajuisesta kirjoituksen ja tapakulttuurin historiaa.

Suomalaisen Kirjallisuuden Seuran (1831) nuorten oppineiden tavoitteena oli kustantaa piirilääkäri Elias Lönnotin tutkimusmatkoja ja saattaa julkisuuteen kerätty kansanrunousaineisto. Hän oli jo omalla kustannuksellaan julkaissut ensimmäisten matkojensa keruusaaliita Kantele taikka Suomen Kansan sekä Wanhoja että Nykyempiä Runoja ja Lauluja I–IV –vihkosissaan 1829–1831. Seuran kauaskantoiseksi tavoitteeksi tuli kehittää suomen kielestä tieteen ja taiteen tarpeisiin soveltuva kieli ja suomalaistaa maan sivistyneistö. Suomenkielistä sanallista kulttuuriperintöä kartutettiin aihepiiristä riippumatta kokoelmiin, olivatpa tekstit syntyneet kirjoittajansa oman työn tuloksena tai saapuneet seuraan säilytettäväksi ns. löytöinä.

Elias Lönnot oli harvinainen tiedemieheksi, sillä hän luovutti auliisti julkaisematontakin aineistoa toisten tutkijoiden käyttöön ja hyödynnettäväksi. Hän määräsi testamentillaan käsikirjoituksensa Suomalaisen Kirjallisuuden Seuralle, jonka tehtäväksi tuli myös jakaa laaja kirjasto oman tarpeensa jälkeen yliopiston kirjastolle ja ympäri maata oleville oppilaitoksille. Lönnotiana –kokoelma muodostui sekä pertinenssi- että provenienssiperiaatteen pohjalta. Kokoelman alkunumeroihin 1–30 luettelointiin Lönnotin runokeräelmät, jotka muodostavat oman kokonaisuutensa. Kyseiset numerot sisältävät myös keruusiin liittyvää kontekstietoa. Tutkijoiden jäämistöistä karttui vähitellen alkuperäistä Lönnotin aineistoa, jotka luettelointiin saapumisjärjestyksessä Lönnotianaan. Kokoelmaa asiansanoitettiin, kortitettiin uudelleen vallitsevien käytäntöjen mukaan ja lopulta seuran satavuotisjuhlan aikoihin se hajautettiin. Kansanrunoutta koskevat alkunumerot sijoitettiin Kansanrunousarkistoon, muut käsikirjoitukset numerosta 31 lähtien sekä kirjeenvaihto seuran kirjaston kokoelmiin. Luettelokorttien käsialoista voidaan ajoitus tehdä.

Kirjallisuusarkisto eriytyi omaksi osastokseen 1960-luvulla ja itsenäiseksi, lakisääteistä valtionapua nauttivaksi arkistoksi 1974. Elias Lönnot oli tietokirjailija ja runoilija, jonka käsikirjoitukset ja tutkimusaineisto ovat kulttuurihistoriallisesti arvokkaita monitieteisyytensä vuoksi. 1800-luvun dokumentteja on myös yksityinen kirjeenvaihto, joka kertoo kulussientakaisista kulttuuriverkostoista, kehittämisajatuksista ja toteutuneista toimenpiteistä. Sensuurin kahlitsemana aikana ei painettuun aineistoon vetoaminen anna oikeata kuvaa aikakaudesta ja päättäjistä.

Alkuperäiset asiakirjat ovat haurautensa eli niiden huonon kunnon vuoksi käyttökiellossa. Tutkijat ovatkin voineet tutustua aineistoon mikroäljenteiden avulla, joiden laatu ei tyydytä nykypäivän tutkijaa. Käsialat ovat vaikeasti luettavia, mikä hidastaa tutkimusta. Kun Lönnrotiana ja kirjeenvaihto tulevat digitoituiksi, alkuperäisten dokumenttien käytettävyys epäilemättä lisääntyy. Korostettakoon, että mv-mikrofilmit eivät kestä kuvalaadullista vertailua tarkkojen digitaalisten kuvatiedostojen välillä.

Elias Lönnrotin vastaanottamat kirjeet luetteloiitiin ensimmäisiksi kirjesidoksiksi (numerot 1–18).

Arkisntonmuodostajalle kertyvät vastaanotetut kirjeet tarvitsevat parinsa, joten Lönnrotin kirjoittamia kirjeitä kerättiin jo 1850-luvulla, jolloin etenkin matkakirjeiden julkaisusuunnitelma syntyi. Hanke raukesi Paavo Tikkasen sairastuttua. Lönnrot oli itse lähettänyt matkakirjeitään ystävilleen ja arkkiatrinrouva Eva Agatha Törngrenille, joka seurasi kiinnostuneena suojattinsa keruumatkoja. J. L. Runeberg julkaisi niitä Lauantaiseuran lehdessä Helsingfors Morgonblad 1830-luvulla ja J. V. Snellman Saimassa ja Litteraturbladissa 1840-luvulla. Satavuotis syntymäpäivän lähestyessä herätettiin julkaisusuunnitelma, jota varten kirjeitä oli puhtaaksikirjoitettu jo Lönnrotin elinaikana. A. R. Niemen valikoima Elias Lönnrotin Matkat I-II ilmestyivät 1902 ja Jenny af Forsellesin Elias Lönnrots svenska skrifter I–II ilmestyivät 1908–1911. Elias Lönnrotin Valitut teokset 1–5 (toim. Raija Majamaa 1990–1993) sisälsivät vain vähän yli 500 Lönnrotin lähettämää kirjettä. Yksittäisiä julkaistuja kirjeitä sisältyy kokoomateoksiin. Kokonaiskirjeenvaihtoa ei ole koskaan julkaistu.

Digitoitu arkistoaineisto kulttuurin tutkimuksessa -projektin sisällöllinen kuvaus

Lönnrotiana –kokoelma edustaa klassista suurmiesaineistoa, jonka systemaattinen digitointi on ensisijaisen välttämätöntä. Projekti edistää digitaalisten tietovarantojen tuottamista, hallintaa ja käyttöä tutkimuksessa, joka on periaatteessa ajasta ja paikasta riippumatonta. Aineiston säilyvyyden varmistaminen ja palveluiden kehittäminen edellyttävät uudelleenluettelointia, jossa alkuperäiset luettelointimerkinnot säilyvät tunnisteineen. Tietoja Vakka-tietokantaan siirrettäessä alkuperäiset tunnisteet toimivat hakuelementteinä.

Digitoinnin koordinaattorina toimi FT Pasi Klemettinen (SKS) ja sisällöllisenä asiantuntijana erikoistutkija Raija Majamaa (SKS). Digitoinnin sisällöllinen suunnittelu ja valintakriteerit toteutettiin tutkijoiden toiveiden mukaisesti. Aluksi kartoitettiin Lönnrotiana-kokoelma ja valittiin digitointiin soveltuva aineisto. Projektin alkuvaiheisiin kuului vanhojen sidosten purkaminen ja uudelleennumerointi eri aikakausina tapahtuneiden luettelointitapojen yhdistämiseksi (Hellevi Riittinen, SKS). Kirjeenvaihto tuli digitointiohjelmaan myös professori Lauri Hongon erityisestä toivomuksesta.

Raija Majamaan valikointitehtävä oli poimia asiantuntijana tutkimuksen kannalta relevantti materiaali digitoitavaksi, sillä rahalliset resurssit eivät riittäneet koko aineiston digitointiin. Valikointi perustui kokoelmatuntemukseen, joka oli karttunut vuosina 1985–1993 Elias Lönnrotin Valittujen teosten 1-5 toimitustyössä sekä teossarjan ilmestyttyä jatkuneissa tutkimuksissa, Kirjallisuusarkiston tietopalvelussa ja kansainvälisen komiteatyöskentelyn pohjalta. Valikointi perustui suunnitelmallisuuteen ja kattavuuteen, eikä sillä pyritty ohjaamaan tulevaa tutkimusta.

Projektisuunnitelman budjetissa Raija Majamaalla ei ollut rahallista varausta budjettiin, ainoastaan tutkimuksellinen. Väliseminaareissa ja työkokouksissa hän esitteli digitaalisen korpuksen tuottamiseen liittyviä kysymyksiä arkistotutkijan näkökulmasta. Havaintojen pohjalta on syntynyt omaa julkaisutuotantoa ja kirjauksia hyvistä käytännöistä tietopalvelun kannalta.

Elias Lönnrotin arkisto ennen digitointia

1985–1993

Valittujen teosten projektin käynnistyessä kartoitettiin suomalaisten arkistojen, kirjastojen ja museoiden kokoelmat sekä Tukholman Kuninkaallisen kirjaston käsikirjoitusosaston kokoelmat. Elias Lönnrotin arkistoa täydennettiin valokopioilla, joita verrattiin varhempisiin, käsinkirjoitettuihin jäljenteisiin, joita professori Aarne Anttila oli kerännyt laatiessaan monografiaa Elias Lönnrot – elämä ja toiminta 1–2 (1931, 1935) Suomalaisen Kirjallisuuden Seuraa satavuotisjuhlaa varten. Vastaanotettuja kirjeitä oli huomattavasti enemmän kuin Lönnrotin kirjoittamia kirjeitä, sillä niitä oli kadonnut vastaanottajilta ja hän itse oli testamentillaan määrännyt hävitettäväksi sellaiset kirjeet, jotka hän oli saanut perheenisänä, ystävänsä tai lääkärinään ja jotka olivat sopimattomia säilyttää julkisessa arkistossa. Niinpä Lönnrotin kirjoittamia kirjeitä vaimolleen tai rakastetuilleen ei ole säilynyt, vaimolta on toistakymmentä kirjettä jäänyt hävittämättä virsikirjakomitean ajoilta 1860-luvulta, jolloin puoliset olivat toisistaan viikkoja erillään. Lönnrotin päiväkirjoja on säilynyt vain katkelmina, joko yksityisyyden suojan tai pitämättömyyden vuoksi. Lönnrotin tapana oli tehdä muistiinpanoja kirjoittamiensa kirjeiden aiheesta tai jäljentää kopio erityiseen sidokseen, jossa nämä dokumentit etenevät kronologisesti.

Teossarjan toimitustyötä varten kirjoitettiin osa Elias Lönnrotin lähettämiä kirjeitä merkkimuotoon CPT-tekstinkäsittelylaitteistolla, joka ei ole enää käytettävissä. Vuonna 1989 hankittu CPT-mikrotietokone 9386 räätälöitiin toimitustyötä varten mm. kokosivun näytöllä, joka oli siihen aikaan uutta. Laitteistossa oli 40 MB:n kiintolevy, uutuutena lasertulostin ja lisälevykeasema 8 ”:n levykkeitä varten. CPT-yritys lainasi vanhan työaseman lisäksi arkistokäyttöön ilmaiseksi neljä muuta työasemaa. Levykkeet (DD, HD ja 5 ”) oli alustettava ennen käyttöönottoa.

CPT-järjestelmän uutuuksia oli valintaohjelma, joka lajitteli kirjeitä (1131 kpl) toivotun kentän mukaisesti. Kirjeet asiansanoitettiin ja rakenteistettiin kenttien avulla. Valintarivillä oli 240 merkkiä, johon sisältyivät pilkkulla erotettuina valintakentät. Vaihtoehdot erotettiin tavuviivalla toisistaan.

Valintaohjelma sisälsi seuraavat kentät:

- a-kenttä: lajittelua varten seuraavat kentät: vvkpp (kirjoitus- tai lähetysaika), kirjoittaja, saaja muodossa Sukunimi Etunimi, kirjoituspaikkakunta, mainitut henkilöt kolme ensimmäistä sukunimen kirjainta poimien, asiasana, tiedostonimi Lönxxx
- v-kenttä hakua varten: arkistosignum
- v-kenttä hakua varten: laaja asiasana
- v-kenttä koko kirjeen tekstin sisältäen.

Valintaohjelma oli hidas, mutta perushaut ajan ja saajan mukaisessa järjestyksessä autoivat tutkijoita alkuun ja valintarivin hakusanat olivat helposti silmin vertailtavissa. Teossarjan tekstit kyettiin omin avuin konvertoimaan wp-tiedostoiksi, mutta lajitteluohjelma toimi vain CTP-tekstinkäsittelymenetelmällä. Suomenkieliset tekstit lähetettiin Suomen kielen tutkimuskeskukseen liitettäväksi heidän kokoelmiinsa. Julkaisematta jääneet kirjeet jäivät oikolukua vaille, mistä on aiheutunut vaikeuksia kirjetietokantaa kehitettäessä. Laitteisto toimi moitteettomasti aina vuoteen 1999 saakka, jolloin vuonna Kalevalan ensipainoksen (1835) teksti konvertoitiin julkaistavaksi seuran Klassikko – sarjassa. Kiintolevy paljastui vaurioituneeksi vuonna 2001, kun digitointiprojekti alkoi. Kirjetekstit onnistuttiin pelastamaan Suomen tietojenkäsittelymuseoyhdistyksen avulla.

Lönnrotin kirjeenvaihto

Suomalaisen Kirjallisuuden Seuran kokoelmissa on Elias Lönnrotin lähettämiä kirjeitä 1124 kpl, lisäksi luonnoksia ja konsepteja on 1396 kpl. Tiedossa on 364 kirjettä, joista on säilynyt sekä lähetetty kirje että sen luonnos.

Alkuperäiseen verrattava lähdearvo on annettava konsepteille, jotka ovat Lönnrotiana–käsikirjoituskokoelmassa esim. päiväkirjasidoksen sivuina. Lönnrotilla oli tapana merkitä muistiin ainakin lähetetyn kirjeen saaja ja merkintä sisällöstä ellei lähes identtistä konseptia tarvittu vastaista kirjeenvaihtoa ajatellen. Lähetetyt kirjeet on digitoitu ja 1160 kpl konvertoitu, mutta alun perin CPT-tekstinkäsittelylaitteistolla tallennettuja, julkaisemattomia kirjeitä ei ole oikoluettu 1980-luvulla kirjoitetusta tekstimuodosta.

Elias Lönnrotin Valituissa teoksissa 1-5 (toim. Raija Majamaa 1990-1993, SKS, 3600 sivua kommentaareineen) julkaistiin vain 508 kirjettä ensimmäisessä volyymissä ja joitakin artikkeleiksi tai tausta-aineistoksi soveltuvia pitkiä kirjeitä muissa osissa. Yksittäisiä kirjeitä on julkaistu aiemmin tutkimuksissa ja mm. A. R. Niemen teoksessa Elias Lönnrotin Matkat I-II (1902).

Kirjeenvaihdossa on saapuneita kirjeitä 3500 kpl, jotka on digitoitu. Niitä ei ole kirjoitettu tekstiksi eikä julkaistu milloinkaan kokonaisuutena.

2000

Suomalaisen Kirjallisuuden Seuran IT-strategiassa ja Kansallisarkiston toteuttamassa kyselyssä kansallisen kulttuuriperinnön digitoinnin kohdentamisesta korostettiin Elias Lönnrotin arkiston merkitystä. Seuran järjestämä aineistoriikki 4.12.2000 ohjasi myös digitoinnin suunnittelua ja käynnistymistä.

Kansainvälinen toiminta:

- Raija Majamaa valittiin Kansainvälisen arkistoneuvoston (ICA) asiantuntijakomitean CLA (Kirjallisuusarkistojen komitea) jäseneksi kaudeksi 2000–2004. Komitea otti tehtäväkseen kirjeen aseman selvitystyön. (www.ica.org).

Julkaisutoiminta:

- Elias Lönnrot (1802–1884) creator of the Kalevala, professor of language, medical doctor. – Teoksessa 100 Faces from Finland. A Biographical Kaleidoscope. Ed. by Ulpu Marjomaa. Tr. by Roderick Fletcher. – Studia Graphica 2, s. 282–291. Finnish Literature Society: Helsinki 2000. Venäjänkielinen versio 2004.

Projektin kuvaus

2001

Suomen Akatemian rahoittama projekti käynnistyi. Lönnrotiana –kokoelman sisällöllinen arviointi ja digitoitavan aineiston valikointi toteutettiin. Projektin selvitystyön tuloksena päädyttiin taloudellisista ja ajallisista resursseista johtuen yksikkötason valintaan eli tutkimuksen tarpeita noudattaen vain osaan Lönnrotiana –kokoelmaa. Kirjeenvaihto kokonaisuudessaan kuului suunnitelmaan (Lauri Honko). Aineiston kartoitus, kunto ja sen vaatimat toimenpiteet suoritettiin.

Digitointi aloitettiin Raija Majamaan luettelon mukaisessa järjestyksessä 1.10.2001 (Pekka Allonen, Hannu Hautala). Vuoden loppuun mennessä oli digitoitu 1450 sivua (noin 2100 kuvatiedostoa) eli Lönnrotian numerot 1-4, 6-17hi, 31-33, 39-40, 43-44, 122, 171 sekä osittain kirjekokoelma 19. Aiheet liittyivät kansanperinteen tutkimukseen ja Lönnrotin lähettämiin kirjeisiin. Käyttöoikeus rajattiin toistaiseksi projektiin osallistuville tutkijoille. Suomen Kulttuurirahastolta anottu avustus Zetterqvistin kokoelman konservoimiseksi, digitoimiseksi ja saattamiseksi tietokannan välityksellä tutkijoiden ulottuville sai kielteisen päätöksen. Lönnrotiana –aineiston tutkimus- ja selvitystyön oheistuotteina valmisteltu Suomalaisen Kirjallisuuden Seuran kotisivulla 2002 julkaistavaa Lönnrot-pakettia (elämäkerta ja kuvitus).

Tutkimushankkeen seminaarissa 7.11.2001 Raija Majamaan esitys oli aiheesta ”Kirjallisuusarkiston Lönnrotiana – mitä se sisältää ja mitä siitä on digitoitu”. Tutkimustyön aiheesta: ”Elias Lönnrot botanistina”.

Digitaalinen kulttuuriperintö –aiheinen tutkimus ja koulutus:

- Osallistunut Kirjallisuusarkiston edustajana ”Tieteellisten äänitearkistojen digitointi ja tiedonhallinta”- yhteisprojektiin. Muuta osapuolet Tampereen yliopiston Kansanperinteen laitos, Kotimaisten kielten tutkimuskeskuksen Suomen kielen nauhoitearkisto ja Pop & Jazz Arkisto.
- Osallistunut aktiivisesti arkisto- ja kirjastoalan työryhmissä valmistelutyöhön, jonka liittyi EU:n Tietoyhteiskunnan tekijänoikeusdirektiivin 2001/29/EY implementointiin kansalliseen lainsäädäntöön. Valmistellut Arkistoyhdistys ry:n lausunnon opetusministeriölle.
- Osallistunut ”Tutkimus, tieto ja tekijänoikeudet” –seminaarin suunnitteluun ja toteuttamiseen 2.2.2001 Geologisessa tutkimuskeskuksessa.
- Osallistunut Arkistoyhdistyksen johtokunnan jäsenenä ”Uudet mediat – siunaus vai kirous?” –seminaarin suunnitteluun ja toteuttamiseen 2.11.2001 Tieteiden talossa.

Osallistunut seuraaviin koulutustilaisuuksiin:

- ”Long-term Preservations of Electronic Documents. Järj. Helsingin yliopiston kirjasto 23.4.2001 Viikin infokeskuksessa.
- ”Museokokoelmien digitoinnin priorisointiseminaari”. Järj. Suomen museoliitto 25.–26.4.2001 Valkoisessa talossa, Helsingissä.
- ”Elektronisten tiedonlähteiden käyttö” ja ”Tekijänoikeudet ja digitaalitekniikka opetuksessa”. Koulutusilta 25.4.2001 Helsingin yliopistossa. Järj. Suomen Tietokirjailijat.
- Sähköisen asioinnin teemapäivä 30.5.2001. Järj. Kansallisarkisto.
- Arkistotoimi ja sähköiset tietojärjestelmät 12.9.2001. Järj. Kansallisarkisto.

Esitelmää ja puheita:

- Juhlapuhe Elias Lönnrot ja huumori Elias Lönnrot –seuran Kohtuuden ystävien pidoissa 26.1.2001 Suomalaisella Klubilla Helsingissä.
- Elias Lönnrot – kuukauden kulttuurihenkilö. Vammalan kaupunginkirjaston tilaisuus 14.2.
- Suomen Tietokirjailijoiden liitto ry myönsi Elias Lönnrotin Valittujen teosten ja näyttelytoiminnan ansioista Warelius-palkinnon 2001. Vanhan kirjallisuuden päivien yhteydessä Vammalassa.
- Itkiskö vai nauraisko? – Arkistonhoitaja ja säädökset. – Arkistoaineiston elinkaari –seminaari. Kulttuurintutkimuksen juridisia ja eettisiä kysymyksiä. Turun yliopiston kulttuurintutkimuksen laitoksen seminaari 9.11.2001 Suomalaisen Kirjallisuuden Seurassa.

Kansainvälinen toiminta:

- The Elias Lönnrot Collection valittiin Unescon Memory of the World –rekisterin ehdokaslistalle.
- Litterära museer –pohjoismainen yhteistyöverkosto (ICLM:n verkosto) kokoontui Helsingissä 9.–11.3.2001 Hanasaassa. Järjestäjät: Svenska litteratursällskapet (Agneta Rahikainen), Suomalaisen Kirjallisuuden Seura (Raija Majamaa) ja Hanasaaren kulttuurikeskus. Pohjoismaisten kirjallisten museoiden ja arkistojen edustajat (48) pohtivat seminaarissaan mm. Z. Topeliuksen koottujen teosten hanketta, tekivät aloitteita henkilöarkistojen tieteellisen tutkimuksen käynnistämiseksi ja kansallisen kulttuuriperinnön säilyttämiseksi.

- Kansainvälisen museoliiton (ICOM) alaisen kirjallisuusmuseoiden ja säveltäjien museoiden komitea ICLM kokoontui 1.–8.7.2001 maailmankongressin yhteydessä Barcelonassa. Esitys Digital Archive Material in Cultural Studies -projektin toiminnasta 2.7. kytkettynä teemaan ”Managing Change: The Museum Facing Economic and Social Challenges”.
- ICA/CLA -komitea kokoontui Deutsches Literaturarchivin kutsusta 24.–28.10.2001 Marbachissa, Saksassa. Kokouksessa kartoitettiin alan kansainvälisten laitosten kirjetietokantoja ja niihin liittyviä ongelmia mm. tekijänoikeuslain ja yksityisyyden suojan kannalta. Henkilötietolainsäädännön ja tekijänoikeuslain eurooppalainen näkökulma muodostui, kun EU:n jäsenmaat vertailivat direktiivien kansallisen voimaansaattamisen eroja. Esitys Suomalaisen Kirjallisuuden Seuran ja yleisen suomalaisen arkistokäytännön näkökulmasta.

Julkaisutoiminta:

- Elias Lönnrot, Sammatti och jubileumsåret 2002. – Laboratorium för folk och kultur 4/2001, s. 8–11. Sama: Litterära museer i Norden. Seminarierapport, Hanaholmen, Esbo 9–11 mars 2001. Toim. Agneta Rahikainen 2002, s. 39–47.

2002

Vanhojen käsikirjoitus- ja kirjesidosten purkaminen digitointia varten saatettiin päätökseen (Hellevi Riittinen Kirjallisuusarkistosta). Lönnrotiana –aineistosta digitoitiin 18500 sivua eli 72 arkistoyksikköä: numerot 34–38, 41–42, 45, 49–51, 77–83, 95, 99, 121, 123, 125, 127–136, 139, 172a–172f, 173–180, 368, 370 sekä Elias Lönnrotin vastaanottamista kirjeistä: Suomalaisen Kirjallisuuden Seuran kirjekokoelmanumerot 1–20 sekä 26 ja 32 osittain (Pekka Allonen ja Hannu Hautala). Lönnrotiana –aineistoa valikoitiin digitoitavaksi.

Kirjetietokannan valmistelu- ja tutkimustyötä jatkettiin yhteistyössä Turun yliopiston Kalevalainstituutissa työskentelevän Pekka Tolosen toteuttaessa pilottitietokantaa (Raija Majamaa ja Pasi Klemettinen). Raija Majamaan toimenkuvaa muutettiin 1.12. 2002 lähtien osa-aikaeläkkeelle siirtymisen vuoksi eli toimenkuva kohdennettiin arkiston perustyöhön ja tutkijapalveluun. Opetusministeriö korvasi kahden kuukauden työpanoksen, joka kohdistui Elias Lönnrotin juhluvuoden toimintaan. Elias Lönnrot –seura Kajaanista myönsi arvonimen Vuoden Elias 2002.

Osallistuminen arkistoalan kehittämiseen koti- ja ulkomaassa, erityisesti tekijänoikeuslainsäädännön seuranta ja soveltaminen.

Esitelmöinyt Arkistolaitoksen, opetusministeriön ja Suomen Unesco-toimikunnan seminaareissa. ”Tekijänoikeudet 2000-luvulla: tekijät, välittäjät, käyttäjät” –seminaari 22.1.2002 Säätytalossa esitys arkistoalan puolesta.

Asiantuntijatehtävä opetusministeriön selvitysmiehen, asianajaja Pekka Tarkelan toimeksiannossa Tekijänoikeusdirektiivin kansallinen voimaansaattaminen /Aineiston käyttö arkistoissa, kirjastoissa ja museoissa. www.minedu.fi/tekijänoikeus. Esityksissä tähdennettiin kulttuuriperintöä säilyttävien arkistojen oikeutta valmistaa kokoelmistaan pysyvän säilytyksen tehtävänsä vuoksi jäljenteitä käytettävästä tekniikasta riippumatta. Digitaalinen tallennus- ja kopiointimenetelmä on välttämättä tuoreutettava säännöllisin väliajoin eheyden, alkuperäisyyden ja käytettävyyden säilymistä varmistamiseksi.

Lausunnot Tekijänoikeusdirektiivin kansallinen voimaansaattamisesta sekä HE 177/2002:stä annettu opetusministeriölle Arkistoyhdistyksen nimissä.

Tekijänoikeustoimikunnan mietinnössä ”Tekijänoikeudet tietoyhteiskunnassa” (Kom. 2002:5) ja laajassa muutospaketissa HE 177/2002 esitettiin, että tekijänoikeuslain 16 ja 16a §:ien nojalla asetuksella säädetään ne arkistot, jotka saivat valmistaa kappaleita omissa kokoelmissaan olevasta teoksesta (käsikirjoituksesta, kuvasta, äänitteestä) laitosten sisäisiä käyttötarpeita varten millä tahansa tekniikalla aineiston säilyttämiseksi, tekniseksi uudistamiseksi sekä aineiston turvaamiseksi. Kokoelmaan kuuluvan aineiston välittäminen laitoksen tiloissa mahdollistettiin. Suomalaisen Kirjallisuuden Seurat arkistot eivät kuuluisi asetuksen 1 §:ssä lueteltuihin laitoksiin, vaikka valtio ylläpitää niitä. Yksityisluontoisten arkistojen valtionavusta annetun lain (998/74) nojalla. Sivistyksellisesti merkittävän, kulttuuriperintöä säilyttävän tehtävänsä vuoksi kappaleen valmistamis- ja välittämisoikeus olisi laitosten valtakunnallisen toiminnan luonteen vuoksi perusteltua, esitettiin lausunnossa.

Puhutun suomen seminaarissa keskustelupuheenvuoro tekijänoikeudesta, tietosuojasta ja arkistoista. Arkistoalan seminaarien suunnittelu ja toteutus Arkistoyhdistyksen johtokunnassa sekä Suomen tieteellisen kirjastoseuran tekijänoikeusryhmässä.

Digitaalinen kulttuuriperintö-aiheinen koulutus:

- Suomen kuudennet arkistopäivät 16.5.–17.5.2002 Mikkelissä.
- Osallistunut seuraaviin seminaareihin ja kokouksiin:
- ”Tekijänoikeudet 2000-luvulla.” Järj. Opetusministeriö 22.1.2002 Säätytalossa.
- ”Käännöskirjallisuuden historia”. Järj. Suomalaisen Kirjallisuuden Seura 10.6.2002.
- ”Tekijänoikeusseminaari” Järj. Suomen tieteellinen kirjastoseura 30.8.2002.
- ”Kulttuurin Kaukametsä”. Järj. Kajaani 12.–13.9.2002.
- Osallistunut Arkistoyhdistyksen johtokunnassa ja Tieteellisten äänitearkistojen digitointi ja tiedonhallinta – projektin johtoryhmässä ”Auttaako digitointi kulttuuriperinnön säilyttämiseen?” –seminaarin 1.11.2002 suunnitteluun ja toteutukseen.
- Digimaan kartta –seminaari 20.12.2002 Säätytalossa. Järj. Opetusministeriön viestintäkulttuuriyksikkö. www.minedu.fi/julkaisut/pdf/Digimaan_kartta.pdf

Kansainvälinen toiminta:

- CLA-komitean kokouksessa Caenissa 23.-26.6., Ranskassa, esitys aiheesta The letter and copyright.

Julkaisut:

- Raija Majamaa: Elias Lönnrot, taitaja, tarkkailija, tiedemies. Yhdessä Väinö Kuukan ja Hannu Vepsän kanssa. SKS 2002. 150 s.

Artikkeleita ja kirjoituksia:

- Kaksisataa vuotta ja vielä voimissaan! - Pirta 1/2002, s. 6-9.
- Itsetuntoinen ja intohimoinen. (Elias Lönnrot 200 vuotta.) – Yliopisto 4/2002, s. 6-9.
- Botanisti Elias Lönnrot, tieteen moniottelija. – Elias Lönnrotin Flora Fennica I. Tammi/ Amanita: Karisto 2002, 32-68.
- Elias Lönnrot 200 vuotta. Helsingin yliopiston kirjaston vitriininäyttelyn 1.4.-31.8. suunnittelu ja tekstit yhteistyössä Inkeri Pitkärannan kanssa. 16 vitriiniä.
- Paikkarin torpan suurmies. Paikkarin torppa. Elias Lönnrot 1802-1884. – Paikkarin torppa. Opas. Toim. Elina Anttila. Museovirasto 2002, s. 5-33.
- Elias Lönnrot – siinä vasta mies! – Helsingin yliopiston kirjaston tiedotuslehti 2/2002, s. 20-21.
- Ei kieltä ilman Lönnrotia. – Bibliophilos 2/2002, s. 6-9.
- Elias Lönnrot. – Raunioyrtti ja muita kasveja Elias Lönnrotin Lammin talolla. Toim. Leena Hämet-Ahti. SKS 2002, s. 10-11.
- Elias Lönnrotin suhde luontoon. – Pakkasten kirjailijaperhe –seminaari 9.6.2002. Taipalsaaren kulttuuriyöryhmä. Toim. Jukka Luoto et alii. Etelä-Karjalan museo, Lappeenranta 2002. s. 36-42.
- Doctor's orders. Tr. David Hackston. – Books from Finland 2002, s.113-116.
- Elias Lönnrot, hengen rientomies 200 vuotta. – Tuusula-Seuran aikakirja XIV. Vuosijulkaisu 2002. Toim. Matias Korkiavuori et alii, s. 19-21.
- Hengen rientomiehet. – Lönnrotin hengessä. Toim. Pekka Laaksonen ja Ulla Piela. Kalevalaseuran vuosikirja. SKS 2002, s. 245-255.
- Kauneus Elias Lönnrotin elämänarvona. – Pirta 3/2002, s. 14-16.
- Elias Lönnrot – kulttuurivaikuttaja. – Suomen kulttuurihistoria 2. Tunne ja tieto. Toim. Rainer Knapas ja Nils Erik Forsgård. Tammi: Helsinki 2002, s. 441-444.
- Elias Lönnrotin juhluvuoden 2002 toiminnan raportointi sekä näyttelykierto. -
- Elias Lönnrotin juhluvuoden raportti 2002. Toim. Tellervo Krogerus. Helsinki:SKS. [www.finlit.fi/Elias Lönnrotin juhluvuoden raportti](http://www.finlit.fi/Elias_Lonnrotin_juhluvuoden_raportti).

Arkistoaineiston käyttöä koskevat julkaisut:

- Itkiskö vai nauraisko? Arkistonhoitaja ja säädökset. – Elore 2/2002. <http://cc.joensuu.fi/~loristi/>
- Juuret ja siivet – arkistonäyttely. – Arkisto 7. Toim. Markku Mäenpää et al. (Raija Majamaa). Arkistoyhdistys: Gummerus 2002, s. 61-72.

Näyttelyt:

- Elias Lönnrot 200 vuotta. Siinä vasta mies! Suomalaisen Kirjallisuuden Seuran kaksisataavuotisjuhlakiertonäyttely. Suunnittelu Raija Majamaa, visuaalinen toteutus Marja Vehkala. Oheisjulkaisuna kolme postikorttia.
- Elias Lönnrot 200 vuotta. Helsingin Yliopiston kirjaston vitriininäyttely 1.4.-31.8.2002. Suunnittelu Raija Majamaa ja Inkeri Pitkäranta.

2003

Hankkeeseen oli budjetoitu kolme kuukautta kokopäiväistä tutkimustyötä aiheesta digitaalisen korpuksen tuottaminen arkistotutkijan näkökulmasta. Elias-tietokannan lisenssi saatu 27.1.2003 kirjetietokannan tarkistustyötä varten. Työaika 60 %:ia, josta 10 %:ia (2 t/vk) varattu erikoistehtäviin, kuten Suomen Akatemian projektin tutkimustyöhön.

"Kirjeen aseman selvitystyöstä (CLA/Unesco) ja Elias Lönnrotin kokonaiskirjeenvaihdon toimitustyön haasteista" -esitys. Järj. Wanhan kirjasuomen seminaari 13.3.2003 /prof. Kaisa Häkkinen, Turun yliopisto.

"Elias-kirjetietokanta". Kalevalainstituutin kanssa yhteistyössä kehitetyn tietokannan esittely Pasi Klemettisen kanssa tutkimushankkeen työpajassa 28.3.2003 Suomalaisen Kirjallisuuden Seurassa.

Laati Arkistoyhdistyksen edustajana lausuntoja hallituksen esityksestä tekijänoikeuslain uudistajana (HE 177/2002) ja oli kuultavana 28.1.2003 eduskunnan Sivistysvaliokunnassa.

Osallistunut kansainvälisen Minerva-hankkeen pilottihankkeeseen yhdessä Kansanrunousarkiston Jukka Saarisen kanssa 14.1.2003.

Osallistunut keskustelutilaisuuteen 27.8.2003 "Tekijänoikeuslainsäädännön uudistaminen". Järj. Opetusministeriö.

Osallistunut koulutusseminaariin 10.10.2003 "Tutkimuksesta kirjaksi" Kansallisarkistossa. Järj. Tilaushistoriakeskus.

Osallistunut "Arkistot ja arvot" –seminaarin suunnitteluun ja toteuttamiseen 7.11.2003 Tieteiden talossa. Järj. Arkistoyhdistys.

Kansainvälinen toiminta:

- ICA/CLA:n kokous Madridissa 22.–25.5.2003. Esityksen aiheena suomalaista henkilötietosuojaa koskevan käytännön ja arkistojen käyttö sääntöjen esittely. Valmistellut luovutus- ja lainakaavakkeet Kirjallisuusarkiston aineistoa varten.

Esiintymiset:

- "Elias Lönnrot arkistojen valossa". Keskustelutilaisuus 30.1.2003 Topeliassa, Helsingissä. Järj. Arkistoyhdistys.
- "Elias Lönnrot ja nainen". Esitelmä 9.4.2003 Elias Lönnrot-Seuran vuosikokouksessa Kajaanissa.
- "Jaakko Juteini ja Elias Lönnrot tietokirjailijoina". Esitelmä Jaakko Juteini ja suomen kieli –seminaarissa 6.7.2003 Hattulan Ylikartanossa. Järj. Suomen Kulttuuriperinnön Säätiö.
- "Elias Lönnrotin persoona". Esitys 7.7.2003 kokouksessa Joutsenossa. Järj. Joutsenon Kalevalaiset Naiset.
- "Arkistoyhdistyksen kannanotto hallituksen esitykseen uudeksi tekijänoikeuslaiksi". Esitys Suomen tieteellisen kirjastoseuran tekijänoikeuspäivän ohjelmassa 5.9.2003 Tieteiden talossa.

Seminaarit ja kokoukset:

- Osallistunut "Käännöskirjallisuuden historia" –hankkeen seminaariin 2.6.2003 Suomalaisen Kirjallisuuden Seurassa (oma aihe Elias Lönnrotin kääntäjäprofiili).

Julkaisut:

- Elias Lönnrotin juhluvuoden raportointi ja näyttelykierto. – Elias Lönnrotin juhluvuoden raportti 2002. Toim. Tellervo Krogerus. Helsinki: SKS. sekä www.finlit.fi.

Lausunnot:

- Arkistoyhdistyksen edustajana antanut lausunnon eduskunnan Sivistysvaliokunnassa 28.1.2003 aiheesta "Tekijänoikeuskysymykset HE 177/2002".

2004

Projektiin oli ennakoitu 3 kk kokopäivätyötä arkiston perustyöstä irrallaan. Toimenkuvan perusteella käytössä oli erikoistehtäviin varattu 10 %:ia eli käytännössä 2 t viikossa (100 t/v). Hakuominaisuuksien testausta pilottikirjetietokannan avulla sekä sen kehittäminen jatkoprojektia suunnittelemana.

Tutkimuksellisesta näkökulmasta vertailtuna ovat digitaaliset kirjeet alkuperäisiä kirjeitä käyttökelpoisempia. Originaalitekstien alkuperäisyys ja eheys ovat dokumentoitavissa. Kirjoituslaturan repeämien aiheuttamat tekstikatkokset ovat todennettavissa. Käsialan muutokset,

yliviivaukset, värilliset korostukset ja muistiinpanot ovat suurennettavissa ja joustavammin saatettavissa merkkimuotoon. Kirjoituksen rytmin vaihtelut ja grafologinen tarkastelukulma antavat aiheita kulttuurihistoriallisiin tutkimusnäkökulmiin.

Projektin työkokouksia pidettiin 10.2. ja 22.3.2004. Vuoden 2004 työtehtäviin kuului myös osallistuminen Suomalaisen Kirjallisuuden Seuran digitointityöryhmän toimintaan: 8 kokousta. Puheenjohtajana toimi Lauri Harvilahti. Digitointityöryhmässä on tehtävänä Lönnrotiana –kokoelman jatkodigitoinnin sisällöllinen suunnittelu ja tekijänoikeuslain uudistamisen seuranta.

Asiantuntijalausunto (HE 28/2004) ja kuuleminen Arkistoyhdistyksen edustajana 7.5.2004 eduskunnan Sivistysvaliokunnassa. Asetusluonnoksessa mainitaan Suomalaisen Kirjallisuuden Seura oikeutetuksi Tekijänoikeuslain 16 §:n mukaiseen kopiointiin.

Raportti

Projektissa tehtävänä oli Elias Lönnrotin käsikirjoitusten (Lönnrotiana) valitseminen digitoitavaksi sisällöllisiä tutkimuksia varten. Elias Lönnrotin kokonaiskirjeenvaihdon tultua osaksi digitointisuunnitelmaa tehtävä laajeni käsittämään perinteisen ja digitoidun aineiston hakumenetelmien vertailun ja käytön kehittämisen. Jo kirjetietokannan pilottivaiheessa ovat haut osoittautuneet verrattomaksi apuneuvoksi tutkijalle, jolla on monitieteisiä intressejä. Kuvan ja tekstin linkitykset sekä listausten kehitystyö mahdollistavat toteutuessaan tutkimusaineiston yhteiskäytön ajasta ja paikasta riippumatta. Teoreettiselta kannalta on otettava huomioon Internet-ympäristössä tapahtuvat kehityslinjat tiedon esittämiseksi verkossa.

Selvitystyön tuloksena on kaksijakoinen raportti. Digitointihankkeen sisällöllinen toteutui tutkijoiden ensisijaisten tarpeiden mukaan. Suomen Akatemian käynnistämästä projektista kehittyi jatkohanke, joka on käynnistynyt Suomalaisen Kirjallisuuden Seuran johdon myötävaikutuksella.

Perusrahoituksen ulkopuolella toteutettava Elias Lönnrotin kirjeenvaihdon tekstikriittinen laitos hyväksyttiin syksyllä 2004 Suomalaisen Kirjallisuuden Seuran toimintasuunnitelmaan. Se käynnisti suomalaisen kirjallisuuden klassikkojen kriittisten editioiden suunnitelman. Tekstikriittisen laitoksen toteuttajaksi hyväksyttiin Raija Majamaa. Työ käynnistyi 1.2.2005 Alfred Kordelinin säätiön ja Jenny ja Antti Wihurin säätiön rahoituksella.

Kriittisen edition toimitustyön ensimmäisenä vaiheena on selvitystyö, jossa kartoitetaan Lönnrotiana –kokoelman sisällöllinen jatkodigitointisuunnitelma meneillään olevien muiden digitointihankkeiden priorisointiaikataulussa. Alkuperäisaineiston kunto tarkistetaan ja sidokset puretaan.

Toimitustyö käynnistyy vertailemalla Elias Lönnrotin kirjeenvaihdon alkuperäisiä ja digitoituja asiakirjoja. Digitoidun käsikirjoituksen ja kirjeen saattaminen tekstimuotoon edellyttää kuvan ja alkutekstin vertailua, tekstin säilyttämistä auktorisoidussa muodossa ja kieliasun tarkistusta. Toimitusperiaatteiden mukaisesti ei nykykielistäminen ole sallittua. Toimitustyöhön kuuluvat myös kommentaarien laatiminen tutkimuskirjallisuuden ja eri alojen asiantuntijoiden avustuksella.

Vuosina 1985-1993 toteutettu Elias Lönnrotin Valittujen teosten toimitustyö ja sen tausta-aineisto palvelevat yhä laajasti tiedeyhteisöä. Kun kriittinen editio julkaistaan, tiedeyhteisön olottuville saatetaan Elias Lönnrotin henkinen perintö, jonka merkitys on 1800-luvun kulttuuria ja historiaa valaisevana lähteenä erityisen suuri. Tietoverkon avulla edistetään keskeisten asiakirjojen käytettävyyttä, jotka tuovat tärkeän lisän tutkimukselle, opetukselle ja sisältötuotannolle. Elias Lönnrotin oma luova työ kuvastaa myös monialaista tutkimuskenttää kansanrunoudesta kasvitieteeseen, lääketieteestä kielitieteeseen, virsirunoudesta sanakirja-aineistoihin, sanoma- ja aikakauslehtien historiasta akateemisiin keskustelukysymyksiin. 1800-luvun kulttuurivaikuttajien verkosto tulee suurhankkeen avulla eläväksi kenen tahansa kulttuurihistoriasta kiinnostuneen olottuville.

Kansainvälinen toiminta:

- ICA/CLA:n kokous Wienissä kansainvälisen arkistokongressin yhteydessä 22.–28.8.2004 Wienissä. Kongressin ohjelmassa oli useita komitean organisoimia seminaareja ja pyöreän pöydän keskustelua, jossa Raija Majamaa piti esityksen aiheesta ”Correspondence series and problems of privacy and copyright”.

Julkaisut:

- J. L. Runeberg – Suomen runoilija. Yhdessä Marjut Paulaharjun kanssa. SKS, Helsinki 2004.
- Suomalaisuusmies Elias Lönnrot. Internetelämäkerta www.finnica.fi 2004.
- Elias Lönnrot. – Suuret suomalaiset. Toim. Jaana Iso-Markku. Otava 2004, s. 123–140.
-

- Elias Ljonnrot, sozdatel "Kalevalyi", professor finskogo jazyika, vratš. – Sto zametšatel'n'yih finnov. Kalejdoskop biografij. Pod. red. Timo Vihavainena. Studia biographica 5. SKS: Helsinki 2004, s. 333–345.
- Kirjeet ovat yksityisarkiston tärkein lähdeaineisto. – Faili 3/2004.

Esiintymiset:

- "Internet, tekijänoikeus ja arkistot". Seminaari Oulun maakuntakirjastossa 18.3.2004. Järj. OMA ja Suomen Kotiseutuliitto.
- "Terveisiä Euroopasta – kirjeen kulttuurihistoriaa". Suomen Tammi –koulutusseminaari 9.6.2004 Urajärvellä. Järj. Opetushallitus.
- "Elias Lönnrot ja naisten taudit". Taide ja taudit –luentosarja 8.10.2004. Järj. Tampereen yliopisto.
- "Runeberg – Suomen runoilija". Kirjallisuuskerho 22.9.2004 Helsingissä. Järj. Kalevalaiset Naiset.
- "Tekijänoikeuspuhvetti" 16.9.2004 Tieteiden talossa. Osallistuminen. Järj. Suomen tiedetoimittajain liitto.
- "Suurmiesten otteessa". Heikki Haapavaaran haastattelu Optio –lehdessä 16.9.2004.

LIITTEKSI

ELIAS LÖNNROTIN 200-VUOTISJUHLAVUOSI

Raija Majamaa 31.1.2003

Juhlavuosi käynnistyi Elias Lönnrot 200 vuotta –näyttelyllä, joka avattiin Suomalaisen Kirjallisuuden Seurassa 17.1. yhdessä kahden uutuusteoksen julkistamisen myötä. Sekä näyttely että teos Lönnrot, taitaja, tarkkailija, tiedemies (Väinö Kuukka & Raija Majamaa & Hannu Vepsä, SKS) avasivat uusia tarkastelukulmia kansallisen esikuvan monialaiseen toimintaan. Elias Lönnrot oli kansalliseepoksen luoja, suomen kirjakielen vankentaja, sanaston uudistaja, päämäärätietoinen tutkimusmatkailija, tehtävänsä sitoutunut tiedemies, itsetuntoinen kulttuurivaikuttaja, kansallistoimittaja ja tietokirjailija, taitava kirurgi ja luonnonlääkkeitten asiantuntija, valistava lääkäri, käytännöllinen lääkintöalojen uudistaja, avaramielinen virsirunoilija. Hän oli suomen- ja ruotsinkielisten henkinen eheyttäjä, samalla kertaa eurooppalainen kuuluisuus ja suomalainen kulttuuriorganisaattori.

Näyttelyn 10 muotokuvaa on otsikoitu teemoittain. Aloituspaneli "Siinä vasta mies!" viittaa ihailuun ja ystävyYTEEN, jota kansallisrunoilija J. L. Runeberg tunsu suomalaista, käsityöläissyntyperää olevaa tiedemiestä ja eeposlaulajaa kohtaan. Kehitysvuodet kertoo keskittymiskykyisen, opimhaluisen pojan urakehityksen kansallisen herätyksen toimijaksi ja nousevan suomenkielisen kulttuurin kehittäjäksi. Lääkäri tiesi tehtävänsä on Kajaanin piirilääkärinä 1833-1853 toimineen parantajan kyvyistä kansanlääkinnän tutkijana, epidemioiden taltuttajana, lääketieteellisen sanaston uudistajana ja terveydenhoito-oppaiden laatijana. Itse loime loihtijaksi, aikahtime laulajaksi on kansanrunouden kerääjän ja tutkijan omasta laulujuudestaan tietoinen taustasitaatti. Kulttuurimme keskeinen symbolilähde, Kalevala on osa kansainvälistä kulttuuriperintöä. The Elias Lönnrot Collection on maailmanlaajuisesti saavuttanut arvostuksen tullessaan esille Unescon portaalissa Memory of the World –ehdokkaana 2001 (<http://portal.unesco.org>). Ilo on käki metsässä, lapsi pieni lattialla –panelissa isyys, vastuu köyhien lasten kouluttamisesta ja kasvatukseen liittyvät kannanotot valaisevat miestä julkisuuden ja yksityisyyden rajamailla. Sujuvaa suomea esittelee "suomen kirjakielen toisen isän", jonka uudissanat kehittivät kaiken kansan ymmärrettävän, tieteen ja taiteen tarpeisiin soveltuvan kielen. Suhde luontoon tuo varhaisen ekologin ja luonnonsuojelijan esille. Kulttuurivaikuttaja –paneli korostaa yhteiskunnallista verkostoa, jonka avulla lehtimies loi suomalaista identiteettiä 1800-luvulla. Siinä vasta mies... jatkaa innovatiivisen, taiteellisesti lahjakkaan persoonallisuuden esittelyä. Hulluna Eliakseen palauttaa mieliin suurmieskultin ilmentymiä Lönnrotin elinajalta nykypäivään.

Näyttely kiersi 24 kotimaan kohteessa, joita olivat paikkakunnittain Helsinki, Hämeenlinna, Ilomantsi, Jyväskylä, Kajaani, Kokkola, Kouvola, Kuhmo, Kärsämäki, Lappeenranta, Lieto, Sammatti, Taipalsaari, Tampere, Turku, Vammala, Vuonislahti, Ylivieska, Äänekoski. Vakiintuneita esittelypaikkoja olivat yliopistokirjasto, yleinen tai maakuntakirjasto, maakunta-arkisto, kunnantalo, koulu tai juhlan tapahtumapaikka Juminkeko, Karjalatalo, Kaukametsä, Runonlaulajan pirtti. Sammatissa sijaitseva Lönnrotin vanhuudenkoti Lammin talo oli syntymä- ja asuinpaikkakunnan kirjaston lisäksi luonteva esittelypaikka. Kirjalliset tapahtumat, kuten Helsingin ja Turun kirjamesut, Vanhan kirjallisuuden päivät Vammalassa ja Vanhan kirjan talvi Jyväskylässä saavuttivat kymmentuhantisen yleisön.

Näyttelyn esittelyjen (3) lisäksi pidin esitelmiä (7) seminaareissa, kirjastoissa ja vanhusten palvelukeskuksessa. Esitelmät, luennot ja puheet olivat kohdennettuja kulloisellekin yleisölle ja joitakin henkilöitä kävi pääkaupunkiseudun tilaisuuksissa toistuvasti. Kalevalaisten Naisten Liiton seminaari Eduskunnassa puhemies Riitta Uosukaisen kutsumana oli mieleenpainuva elämys, samoin kirjallisuuspäivät Vihdin Päivölässä teemanaan

Ylivertainen Lönnrot. Leppävaaran kirjaston tapahtuma Elias elää – Lönnrot lives! oli vilkas, kaikenikäisen yleisön kohtaamispaikka.

Elias Lönnrot –aiheisia juhlapuheita (3) pidin Suur-Helsingin Kalevalaisten Naisten, Karjalan Liiton ja Uuden Kalevalan allekirjoittamisen 153-vuotisjuhlassa, jonka järjestivät Wesilahti-Seura, Vesilahden museoyhdistys ja Vesilahden kulttuuritoimi Narvan seurojentalolla Vesilahdella. Kirjalliset seurat kutsuivat puhujaksi (3), Kalevalanpäivänä pidin Kalevalaseuran puolesta puheen Elias Lönnrotin patsaalla järjestetyllä kunniakäynnillä Lönnrotinpuistikossa, Helsingissä. Puhe korosti Lönnrotin elämäntyön ensisijaista kohdetta suomen kieltä ja sen asemaa Euroopan yhteisössä. Tieteellisten seurain valtuuskunta pyysi oppaaksi kesäretkelleen, joka suuntautui Sammattiin. Sammatin kirjastossa korostin Lönnrotin lehtimiestoimintaa, Hämeenlinnan maakunta-arkistossa hänen apteekkioppilaskautensa vaikutusta vastaiseen elämänuraan. Oulun kaupunginkirjaston – maakuntakirjaston oman näyttelyn Konttiukko pohjoisessa avajaisjuhlassa toin esille yksityiskohtia syksyn 1832 kulkutautimuistiinpanoista ja ystävyys-suhteista kautta eliniän.

Yli tunnin kestäviä luentoja (6) pidin Ikääntyvien yliopistossa Kuopiossa, Helsingin seudun kesäyliopiston ja Suomen Tietokirjailijat ry:n yhteisessä luentosarjassa Tietokirjallisuus Suomessa, Ahjolan kansalaisopistossa Tampereella, Kansallismuseon luentosarjassa Elias Lönnrot, koko kansan suurmies, ja Opetushallituksen koulutustapahtumassa Suomen Tammen kesä 2002 – kulttuuriperinnön monimuotoisuus. Helsingin yliopiston kirjaston auditoriossa pidetty luento toteutettiin virtuaaliluentona.

Yleisradion Yle 1 –kanavalla esiinnyin kahdessa omassa ohjelmassa (Lönnrotin leikit – kulttuurihistoriaa sanoissa /Kielikorva ja Tasavalta tappelussa, lyöpi kuka kerkiää –ohjelmassa, joka uusittiin Suomen kielen muistomerkkejä –sarjassa. Ajankohtaishaastatteluja ja uutislähetystyksiä (8) tekivät Huomenta Suomi/MTV, Kulttuurituotiset, MTV, Radio Moreeni ja muut paikallisradiot, Radio Suomi, Ruotsin radio /Sisuradio sekä Sävelradio.

Pitkäkestoisia keskustelutilaisuuksia (2) olivat Radio Ilonan ohjelma Lönnrotin perheen naisista ja Luontoradion /Yle 1 ohjelma Elias Lönnrotista kasvitieteilijänä, johon osallistuivat toimittaja Risto Nurmen lisäksi myös teoksen loistopainoksen toimittaja Kai Linnilä ja professori Terttu Utriainen.

Koulutelevisiossa esitetyssä toimittaja Kirsi Mattilan ohjelmassa Mies nimeltä Lönnrot kerroin sammatilaisille koululaisille Lammin talossa Lönnrotin vanhuudenpäivistä ja elämänvaiheista.

Koululaisten käytettävissä olin suurissa tapahtumissa, kuten Helsingin ja Turun kirjamesseilla. Äidinkielenopettajat olivat antaneet tehtäviä, joihin vastaukset löytyivät näyttelystä. Pernon koulun 9. luokkalaiset haastattelivat Turussa Matkalla –aiheesta. Kirjallisissa tapahtumissa yleisöä oli kymmenissä tuhansissa, kuten Vanhan kirjallisuuden päivillä Vammalassa.

Kunnianosoituksiin (2) osallistuin Kalevalaseuran puolesta sepeleenlaskuun 28.2. Elias Lönnrotin patsaalla, Suomalaisen Kirjallisuuden Seuran edustajana haudalla 9.4. Sammattissa. Kirjan ja ruusun päivien avauspuheenvuoroni 23.4. Runebergin patsaan juurella sijaitsevalla lavalla käsitteli kansalliskirjailijan ja –toimittajan ystävyden merkitystä.

Esiintymispaikkakunnat jakautuivat seuraavasti: Espoo, Heinola, Helsinki (15), Hämeenlinna, Jyväskylä, Kuopio, Oulu, Sammatti (2), Taipalsaari, Tampere, Turku, Vammala, Vesilahti ja Vihti.

Esitelmien, puheiden ja luentojen aiheet käsittelivät kohdeyleisön mukaan suurmiespersoonaa, perhe-elämää, suhdetta kansaan ja ajan vaikuttajiin M. A. Castréniin, J. L. Runebergiin, Z. Topeliukseen, ystäväverkostoon, suosijoihin ja tukijoihin. Uusia, tutkimuksiin perustuvia näkökulmia avasin puhumalla Lönnrotin suhteesta kauneuteen elämänarvona, häneen tietokirjailijana, koulunuudistajana, sanaston luoja, lehtimiehenä sekä luonnon monimuotoisuutta korostavana ekologina, aikansa ”vihreänä”.

Julkaisuluettelossa on Elias Lönnrot, taitaja, tarkkailija, tiedemies –teoksen lisäksi artikkeleita (8 teoksessa, 5 aikakauslehdessä Bibliophilos, Books from Finland, Pirta, Yliopisto, Helsingin yliopiston tiedotuslehti).

Helsingin yliopiston kirjaston vitriininäyttely 1.4.-31.8. toteutui yhteistyössä Inkeri Pitkärannan kanssa (16 vitriiniä).

Tietopalvelu- ja asiantuntijatehtäviä

Kirjallisuusarkiston erikoistutkijana vastannut puhelimitse, sähköpostitse ja kirjeitse esitettyihin hakutehtäviin ja kysymyksiin, jotka liittyivät Lönnrotiana –kokoelman ja Elias Lönnrotin kirjeenvaihdon lisäksi 1800 –luvun aatemaailmaan. Tiedemiehet keskittivät tiedustelunsa asiakirjojen todistusvoimaisuuteen ja lähdearvoon, sukututkijat kirjeenvaihtoon, media ja yleisö suurmieskulttiin. Elias Lönnrotin asuinpaikkakuntien sekä

kulttuurimatkailuyrittäjien mielenkiinnon kohteina olivat ”suurmiehen jalanjäljissä” –tapaiset yksityiskohdat ja tarkistukset. Kajaanin kaupunginkirjaston – Kainuun maakuntakirjaston kutsusta tunnistin käsikirjoituksia ja kirjeitä Caiania-kotiseutukokoelmasta 1830-1850 -luvuilta.

Asiantuntijana ohjelmissa:

- Tid och rum: Elias Lönnrot. Toim. Wivan Fagerudd. 4.2. TV 2, FST/Kultur.
- Mies nimeltä Lönnrot. Koul-TV/TV 1. Runon taituri. 25.2. (uus. 27.2.) ja Lääkäri ja sanaseppo 8.4. (uus. 10.4.). Toim. Kirsi Mattila.

Henkilöhaastatteluja (lehdistö, isoja juttuja):

- Elias ja rouva Lönnrot. - Helena Hämäläinen/ Kodin kuvalehti 14.2., s.116-120.
- Elias Lönnrot – the father of Kalevala. – Pauliina Susi/ Universitas Helsingiensis 1/2002, s. 21-23.
- Lönnrotin juhlavuosi kuljettaa tutkijaa. - Terttu-Elina Kalaja/ Karjala 28.2., A3.
- Ruokaa ja rohtoja luonnosta. – Raija Kivinen/ Ruokalehti maaliskuu 2002, s.37.
- Lönnrot-tutkija Raija Majamaa tulkitsi piirilääkärien reseptejä. – Sinikka Viirret / Kainuun Sanomat 5.4.

LIITTEET

Esitelmät, puheet, alustukset ja luennot

Raija Majamaa:

- Elias Lönnrot 200 vuotta –näyttelyn julkistaminen. Näyttelyn esittely. SKS 17.1.
- Elias Lönnrot 200 vuotta –näyttelyn avajaisesitys 8.2. Jyväskylän yliopiston kirjasto.
- ”Siinä vasta mies!” Esitelmä Vanhan kirjan talvi-tapahtumassa 9.2., Jyväskylän kaupunginkirjasto.
- Ikääntyvien yliopisto/ Kuopion yliopisto, 14.2. (Elias Lönnrotin ystäväverkosto)
- Minna Canthin Seuran jäsenkokous 21.2., Helsinki (Lönnrotin perhe)
- Kalevalaisten Naisten Liiton Elias Lönnrot -seminaari 23.2. Eduskunnassa (Kauneus Elias Lönnrotin elämänarvona)
- Suur-Helsingin Kalevalaiset Naiset ry:n Kalevalajuhla 24.2., Helsinki (Elias Lönnrot – aikojen sankari)
- Kalevalaseuran seppeleenlasku Elias Lönnrotin patsaalla 28.2., Helsinki (Suomalaisen kulttuurin suurmies)
- Karjalan Liiton Kalevalapäiväjuhla 28.2., Helsinki (Elias Lönnrotin suhde kansaan)
- Elias Lönnrot, koko kansan suurmies /Suomen kansallismuseon tiistaisarja 5.3. (Elias Lönnrot kulttuurivaikuttajana)
- Topelius-seuran tilaisuus 7.3., Helsinki (Topelius ja Lönnrot – kaksi kansallistoimittajaa)
- Lönnrot –virtuaaliluento, Helsingin yliopiston kirjaston auditorio 3.4. (Elias Lönnrot – siinä vasta mies!). <http://video.helsinki.fi/HYK.html>
- M. A. Castrénin Seuran vuosikokous 11.4. , SKS. (Kaverusten keikka - M.A. Castrén ja Elias Lönnrot matkantekijöinä).
- Uuden Kalevalan allekirjoittamisen 153-vuotisjuhla. Wesilahti-Seura, Vesilahden museoyhdistys, Vesilahden kulttuuritoimi. Juhlapuhe. Narvan seurojentalo 17.4. (Elias Lönnrot Laukossa).
- Kirjan ja ruusun päivä 23.4. Esplanadilla, (Elias Lönnrot, siinä vasta mies! – J.L. Runeberg).
- Tieteellisten seurain valtuuskunnan kesäretki 28.5., Sammatti (Elias Lönnrotin elämästä Sammatissa).
- Taipalsaaren kulttuuriyöryhmän kirjallisuusseminaari 9.6., Taipalsaaren kunnantalo (Elias Lönnrotin suhde luontoon).
- Vanhan kirjallisuuden päivät ry 29.6., Vammala. Bibliofiilien Seuran Kirjankerääjän ABC-kurssi. (Lönnrotin ja Castrénin Lapin retkeltä)
- Suomen Tammen kesä 2002 - Kulttuuriperinnön monimuotoisuus. Opetushallitus. 7.8. Heinola. (Lönnrot, monipuolinen kulttuurivaikuttaja).
- Tietokirjallisuus Suomessa-luentosarja 22.8. Helsingin seudun kesäyliopisto / Suomen Tietokirjailijat ry. (Elias Lönnrot – tieteen renki vai toimija?).
- Opetushallituksen vierailu 27.8. SKS (Lönnrot ja koulu).
- Sammatista Päivölään; Ylivertainen Lönnrot (1802-1884). Kirjallisuuspäivät 31.8.-1.9. Vihdin Päivölässä (Elias Lönnrot sanaston luojana – kompromissien mies; Lönnrot luonnon tutkijana ja suojelijana).
- Helsingin Rotaryklubin kokous 3.9. (”Hengen rientomiehet”. Elias Lönnrot ja M. A. Castrén Lapissa).
- Konttiukko pohjoisessa –näyttelyn avajaiset 3.9. Oulun kaupunginkirjasto. (Elias Lönnrot – kulttuurivaikuttaja).
- Elias elää – Lönnrot lives! Leppävaaran kirjaston tapahtuma 5.9. (Elias Lönnrotin suhteesta luontoon).
- Kampin palvelukeskus 18.9. (Elias Lönnrot ihmisenä).
- Ahjolan kansalaisopisto 9.10., Tampere. (Lehtimies Lönnrot).
- Turun kirjamesut 3.-5.10. Koululaisten haastattelu aiheesta Matkalla. Pernon koulu 9 lk.
- Sammatin kirjasto 29.10. (Elias Lönnrot lehtimiehenä).

- Hämeenlinnan maakunta-arkiston näyttelyn avajaiset 11.11. (Elias Lönnrot 200 vuotta – Hämeenlinnassa).

Radio- ja televisioesitykset:

- Elias Lönnrot. Haastattelu 17.1. MTV:n uutisissa ja Kulttuuriuutisissa.
- Ida Lönnrot ja matkustavat naiset. Haastattelu, toim. Maritta Alander-Valtonen. Kulttuuriuutiset /Yle 27.2.
- Elias Lönnrot Kalevalanpäivänä 28.2. Haastattelu, toim. Liisa Syrjälä. Radio Suomi.
- Lönnrotin perheen naiset 2.3. (uus. 3.3.) Haastattelu, toim. Maritta Alander-Valtonen. Radio Ilona / Yle.
- Elias Lönnrot 200 vuotta. Haastattelu, toim. Inkeri Lamér. Ruotsin radio/Sisuradio 7.4.
- Monitietäjä Elias Lönnrot. Huomenta Suomi –lähetyksen vieraana 8.4. , MTV 3.
- Mies nimeltä Lönnrot. Lääkäri ja sanaseppo. Toim. Kirsi Mattila. 8.4. Koulu-TV /TV 1.
- Lönnrotin leikit – kulttuurihistoriaa sanoissa 9.4. Kielikorva /Yle 1.
- Elias Lönnrot Kirjan ja ruusun päivänä 23.4. Haastattelu, Sävelradio.
- ”Ihminen voi olla hyvä myös kukittuaan”. Flora Fennican pohjalta keskustelevat Elias Lönnrotin kasvitieteellisestä perinnöstä Kai Linnila, Raija Majamaa ja Terttu Utriainen 23.10. (uus. 11.12.) Toim. Risto Nurmi. Luontoradio /Yle 1.
- Tasavalta tappelussa, lyöpi kuka kerkiää. – Suomen kielen muistomerkkejä-sarja 9.12./ Yle 1.
- Elias Lönnrot 10.12. Haastattelu, Radio Moreeni/Tampere.

Julkaisut

- Elias Lönnrot, taitaja, tarkkailija, tiedemies. Yhdessä Väinö Kuukan ja Hannu Vepsän kanssa. Helsinki: SKS. 150 s.
- Kaksisataa vuotta ja vielä voimissaan! – Pirta 1/2002, s. 6-9.
- Itsetuntoinen ja intohimoinen. (Elias Lönnrot 200 vuotta). – Yliopisto 2/2002, s. 6-9.
- Botanisti Elias Lönnrot, tieteen moniottelija. – Elias Lönnrotin Flora Fennica I. Tammi/Amanita: Karisto, s. 32-68.
- Elias Lönnrot 200 vuotta. Helsingin yliopiston kirjaston vitriininäyttelyn 1.4.-31.8. suunnittelu ja tekstit yhteistyössä Inkeri Pitkärannan kanssa. 16 vitriiniä.
- Paikkarin torpan suurmies. Paikkarin torppa. Elias Lönnrot 1802-1884. – Paikkarin torppa. Opas. Toim. Elina Anttila. Museovirasto, s. 5-33.
- Elias Lönnrot – siinä vasta mies! – Helsingin yliopiston kirjaston tiedotuslehti 2/2002, s. 20-21.
- Ei kieltä ilman Lönnrotia. – Bibliophilos 2/2002, s. 6-9.
- Elias Lönnrot. – Raunioyrtti ja muita kasveja Elias Lönnrotin Lammin talolla. Toim. Leena Hämet-Ahti. SKS, s. 10-11.
- Elias Lönnrotin suhde luontoon. – Pakkasten kirjailijaperhe –seminaari 9.6.2002. Taipalsaaren kulttuuriryöryhmä. Toim. Jukka Luoto et alii. Etelä-Karjalan museo, Lappeenranta 2002. s. 36-42.
- Doctor's orders. Tr. David Hackston. – Books from Finland 2002, s.113-116.
- Elias Lönnrot, hengen rientomies 200 vuotta. – Tuusula-Seuran aikakirja XIV. Vuosijulkaisu 2002. Toim. Matias Korkiavuori et alii, s. 19-21.
- Hengen rientomiehet. – Lönnrotin hengessä. Toim. Pekka Laaksonen ja Ulla Piela. Kalevalaseuran vuosikirja. SKS, s. 245-255.
- Kauneus Elias Lönnrotin elämänarvona. – Pirta 3/2002, s. 14-16.
- Elias Lönnrot – kulttuurivaikuttaja. – Suomen kulttuurihistoria 2. Tunne ja tieto. Toim. Rainer Knapas ja Nils Erik Forsgård. Tammi: Helsinki, s. 441-444.
- Elias Lönnrot, Sammatti och jubileumsåret 2002. – Litterära museer i Norden. Möte IV. Toim. Agneta Rahikainen. ISBN 952-91-4013-4, s. 39-47.
- Kirjallisuusarkiston 3 vuosikorttia (aiheina Elias Lönnrot, visuaalinen toteutus Marja Vehkala).

Juhlavuoden kiertonäyttely

Elias Lönnrot 200 vuotta (suunnittelu Raija Majamaa & visuaalinen toteutus Marja Vehkala). Kiertonäyttely, 10 panelimuotokuvaa Elias Lönnrotin monipuolisesta toiminnasta.

Kiertoaikataulu (24 kohdetta, 2 identtistä kappaletta kierrossa):

2002

- | | |
|---------------|--|
| 17.1. | Avajaiset Suomalaisen Kirjallisuuden Seura |
| 1.2. - 15.2. | Jyväskylän yliopiston kirjasto, Jyväskylä |
| 18.2. - 28.2. | Karjalatalo, Helsinki |
| 4.3. - 27.3. | Juminkeko, Kuhmo |
| 2.4. - 30.4. | Helsingin yliopiston kirjasto, Helsinki |

6.5. - 20.5.	Kärsämäen lukio, Kärsämäki
23.5. - 10.6.	Taipalsaaren kunnantalo, Taipalsaari
15.6. - 29.6.	Vammalan kaupunginkirjasto, Vammala
Vanhan kirjallisuuden päivät 28.-29.6.	Sylvään koulu, Vammala
30.6. - 28.7.	Lammin talo, Sammatti
30.7. - 12.8.	Lappeenrannan maakuntakirjasto, Lappeenranta
13.8. - 1.9.	Kouvolatalo, Kouvola
3.9. - 27.9.	Kaukametsä, Kajaani / Oulun yliopisto/ Lönnrot-instituutti
1.10. - 30.10.	Sammatin kunnankirjasto, Sammatti
11.11. - 24.11.	Hämeenlinnan maakunta-arkisto, Hämeenlinna
2.12. - 8.12.	Kiviniityn lukio, Kokkola
12.12.2002 –18.1.2003	Tampereen kaupunginkirjasto – Pirkanmaan maakuntakirjasto, Tampere

Kierto jatkuu 2003

Toinen näyttelykappale

2002

5.5. - 15.9.	Parpeinvaara, Ilomantsi
20.9. - 22.9.	Vuonilahti, Suomen Kotiseutuliitto
4.10. - 6.10.	Turun kirjamesut, Turku
7.10. – 20.10.	Liedon kirjasto, Lieto
24.10. - 27.10.	Helsingin kirjamesut, Helsinki
4.11. - 27.11.	Ylivieskan kaupunginkirjasto, Ylivieska
2.12. - 18.12.	Äänekosken kaupungintalo, Äänekoski

Kierto jatkuu 2003, päättyy 1.4.2003 Suomen Kulttuuriperinnön Säätiön omistukseen.

Liite 4 CPT-tiedostojen pelastusoperaatio

Lasse Nirhamo
Suomen Tietojenkäsittelymuseoyhdistys ry
ATK-museo, Jyväskylä
<http://www.tietokonemuseo.saunalahti.fi>

CPT-tiedostojen pelastusoperaatio

Suomalaisen Kirjallisuuden Seurassa tallennettiin 1980-luvulla 1160 Elias Lönnrotin kirjettä CPT 9000 tekstinkäsittelylaitteistolla 8 tuuman levykkeille. CPT poistui markkinoilta ja SKS:ssa oli käytössä 386-pc-laitteisto, johon oli asennettu 8 tuuman levykeasema ja konvertointiohjelma. Kun laitteistoa yritettiin vuonna 2001 käyttää, osoittautui se vialliseksi eikä levyjen luku ja siirto onnistunut. Kiintolevy oli viallinen eikä tiedostojen pelastaminen ollut mahdollista.

SKS (Elias-digitointi) otti yhteyttä Suomen tietojenkäsittelymuseoyhdistykseen levykkeiden lukemiseksi ja tiedostojen konvertoimiseksi. Museon kokoelmista löytyi tarkoitukseen sopiva laitteisto, InterMedia konverteri lisälaitteineen. Ne ovat olleet käytössä Otava Oy:n kirjapainossa Keuruulla. Aikaisemmin oli tyypillistä, että kirjapainoon tulevat painotyöt olivat lähes kaikilla mahdollisilla tallennusformaateilla. Konverterilla onkin luettu yli 400 erilaista diskettiformaattia ja kirjoittaakin voitiin yli 250 formaatille.

Laitteistossa on Zenith Z118 keskusyksikkö 8088 prosessorilla ja Dos 2.01 käyttöjärjestelmä. Ohjelmana on englantilainen InterMedia yleiskommunikointi- ja konvertointijärjestelmä versio 3.03 vuodelta 1987. Levyasemia on neljä, 8 tuuman, kaksi 5,25 tuuman ja 3,5 tuuman asemat.

Laitteet ovat viimeksi olleet varsinaisessa käytössä yli 10-vuotta sitten eikä niiden käynnistäminen onnistunut ilman ongelmia. Käyttöohje oli ajan hengen mukaisesti ylimalkainen eikä se auttanut yhtään. Laitetta oli tuonut maahan Grafilaser Oy, joka sitten muuttui SysPar Oy:ksi. Yhteystiedot löytyivät, yhteyksiä ei. InterMedia toimii edelleenkin Englannissa ja tarjoaa konvertointipalvelujaan, budjetissamme ei kuitenkaan ollut varoja heidän palveluidensa käyttämiseen. CPT Datakin toimi vielä 1992 Espoossa mutta sen jälkeen ei siitä saatu mitään tietoja. Tyypillinen tilanne atk-alan muutoksien kentässä. Kaikki asiat ovat helppoja ja yksinkertaisia jälkikäteen, kun tietää miten kaikki piti ja pitää tehdä. Jokainen joka on ollut tekemisissä eri sukupolvea olevien laitteiden ja ohjelmistojen kanssa tietää, että ongelmat lisääntyvät vaihtoehtojen lukumäärän mukaan n:n:teen (nn).

Ryhdyimme siis tutkimaan laitteiston sielunelämää, toimintaa ja käyttöä. Levyasemien formaatit olivat pc-standardista poikkeavia. Levykkeelle tallennettuja tietoja ei pystynyt lukemaan pc:llä eikä laite lukenut pc:n levykkeitä. Levykeasemien välinen siirto toimii vain InterMedia-ohjelman kautta, ei siis suoraan siirtona tai kopiointina. Tulipa sitten tutustuttua perusteellisemmin levykkeiden eri parametreihin. Suurin ongelma oli 8 tuuman aseman epävarma toiminta. Eri kerroilla tuli erilaisia tuloksia, usein kuitenkin "seek error", "data error". Varmaa ei ollut johtuivatko virheet ohjelmien erilaisuuksista vaiko laitteiden epävarmasta toiminnasta. Vanhaan ja ainoaan laitteeseen piti kuitenkin suhtautua kunnioituksella. Liian innokkaat toimenpiteet saattaisivat lopettaa toiminnan lopullisesti. Onneksi museoyhdistyksessä on jäsenenä Vaalassa asuva Heikki Yliheikkilä, jolla on kokemusta monien laitteiden huollosta vuosikymmenten ajalta. Kun saimme hänet paikalle, hän onnistuikin huoltamaan ja säätämään levyaseman toimintakuntoon ja osoitti edelleenkin työaikojen lempinimen "kultasormi-Heikki", olevan edelleenkin voimassa.

Kun laitteet saatiin toimimaan, päästiin itse asiaan, tiedostojen siirtoon. InterMedia-ohjelmassa on konvertointi CPT8000 mutta ei CPT9000 varten. Kun ensimmäinen siirto saatiin onnistumaan oli lopputulos lähinnä siansaksaa. Tekstissä olivat kaikki ääkköset ja muotoilumerkit tulostuneet aivan käsittämättömänä tekstinä. Koodauksessa käytettäviä ASCII-merkkejä on 256 ja niistä muodostettuja jonokodeja CPT:ssä useita. Valmiita muunnostaulukoita on koneessa myös useita, mutta ei tietoa niiden sisällöstä eikä toisaalta ollut tietoa myöskään CPT:n koodauksesta. PC-Toolsin avulla pystyttiin selvittämään oikeat heksadesimaalimerkit ja tarvittavat muunnokset. Ajettiin läpi suora siirto ilman mitään muunnoksia ja siitä sitten selvitetiin tarvittavat muunnokset yksi kerrallaan ja niin saatiin lopulta aikaan siirto, jonka tulos avautuu oikein myös Windowsissa.

Tämän jälkeen työ oli vain "rahtiajao". Rahtasin ensiksi laitteet vaimon kauhistikseksi kotiin koska 1160 tiedoston siirto on sen verran aikaa vievä toimenpide, etten viitsinyt koko pääsiäistä viettää kylmässä varastossa. Tiedoston siirsin ensiksi InterMedia laitteella 8 tuuman lerpulta 720 kt:n korpulle. Toisessa pc-koneessa muunnettiin siirretyt

Liite 5 Digitoinnin teknisiä lisätietoja

- Teknisiä metatietoja digitointiprosessin eri työvaiheista ovat selvitystä varten toimittaneet Pekka Allonen ja Hannu Hautala.
- Elias Lönnrotin alkuperäiskäsikirjoitusten skannaaminen kuvatiedostoiksi aloitettiin lokakuussa 2001. Skannauksessa käytettiin korvaavana laitteistona Epson GT10000 -skanneria (Epson Twain pro 2.1 – ajuriohjelma).
- Käsikirjoituksia skannattaessa jokaisen käsikirjoitussidoksen aukeamasta tehtiin pääsääntöisesti yksi TIFF-kuvatiedosto. Joidenkin aineistojen osalta digitointi suoritettiin käytännön syistä sivu tai arkki kerrallaan omaksi tiedostokseen (esim. sitomattomat aineistot ja kirjeet). Työaseman kovalevyllä kuvatiedostot järjestettiin kansioiksi käsikirjoitussidosten mukaisesti. Skannauksen jälkeen tiedostojen laatu ja oikea järjestys tarkastettiin kuvanselailuohjelmalla.
- Keskusteluja kuvien laadun osalta käytiin runsaasti ja työn kuluessa testattiin erilaisia laatukriteerejä. Kaikki pakkaamattomat TIFF-tiedostot tuotettiin pääsääntöisesti samoilla teknisillä asetuksilla ja arvoilla. Kansainvälisten standardien mukaisia tallennusformaatteja käyttämällä voidaan turvata tiedon säilyvyys, jaettavuus ja konvertoitavuus tulevaisuudessa. On selvää, että TIFF-formaattia tullaan jatkossa tukemaan ja nyt luotujen tiedostojen siirtäminen täysin uuteen kuvamaailmaan tullaan varmasti tavalla tai toisella ratkaisemaan (esim. konversiot ja emuloinnit). Kaikissa arkistotalenteita koskevissa muutosvaiheissa on ratkaisevaa digitaalisen tiedon häviöttömyys, joka mahdollistetaan esimerkiksi tietovälinettä vaihdettaessa samoja pakkaamattomia formaatteja käyttämällä.
- Mastertiedostoista luotujen käyttökopioiden suhteen voidaan olla huomattavasti joustavampia ja huomioida tapauskohtaiset erityistarpeet (esim. JPEG, JPEG 2000, GIF ja PNG).

Esimerkkejä TIFF-kuvien keskimääräisestä tiedostokoosta sidoksittain (osa digitoitu aukeamittain ja osa sivuttain/liuskoittain):

- Lönnrotiana 1 = 38,5 Mt (183 tiedostoa)
- Lönnrotiana 30 = 13,5 Mt (96 tiedostoa)
- Lönnrotiana kirjekokoelma 1 = 15,5 Mt (532 tiedostoa)
- Lönnrotiana kirjekokoelma 32 = 13 Mt (115 tiedostoa)

Tekniset tiedot:

- Tiedostoformaatti: TIFF 6.0 (mastertiedostot)
- Kuvatiedostojen skannausresoluutio: 300 dpi
- Kuvatiedostojen datatyypit: RGB -True Color 24-bit

- Työkansioiden nimen eteen sijoitettiin Lönnrotiana-kokoelmasta johdettu lyhenne Lna ja sen jatkeeksi sidoksen numero (esim. Lna007). Yksittäisten tiedostojen kohdalla etuliitteen jälkeen tulee alaviivan jälkeen massakonvertoinnissa automaattisesti muodostuva kolminumeroinen järjestysnumero (esim. Lna007_005). Tiedostot eivät voi näin sekaantua toisiinsa eri kansioissa vaan jokaisella tiedostolla on yksilöivä tunniste. Katsottiin myös, että jatkokäsittelyn kannalta on mielekästä säilyttää kansiorakenteena Lönnrotianan alkuperäisistä sidoksista juontuva rakenne. Sidokset ovat vakiintuneita kokonaisuuksia, joten niiden perusteella on helppo etsiä aineistoja samannimisistä työkansioista.
- Mainittakoon, että selvityksessä annetusta suosituksesta poiketen tiedostonimissä (Lna) käytettiin rinnakkain sekä pieniä että suuria kirjaimia. Nimeämiskäytännö ei ole tässä tapauksessa ongelmallinen, sillä kaikki Lönnrotiana-aineistot on nimetty samalla periaatteella eikä kirjainkoko ole tarkoitettu tiedostoja erottavaksi tekijäksi.
- Käsikirjoitusten skannausta on vaikeuttanut käsikirjoitusten kirjava laatu, koko ja huono kunto. Käsikirjoitusten sidonta on myös tehty tavalla, joka vaikeuttaa niiden skannausta (ns. tiukat sidokset). Liukuhihnamainen skannaus ei ole ko. käsikirjoitusten osalta täysin mahdollista.
- Tuotettujen tiedostojen varmuuskopiointi tulee tehdä päivittäin palvelimelle tai toiselle kovalevylle.
- Varmistuksen merkitys kävi konkreettisesti ilmi, kun SLR-ajoa valmistellessa ja tiedostomassoja lähdekansioon siirrettäessä toinen kovalevyistä tallennettiin liian täyteen ja tiedostojärjestelmän rakenteellisen virheen seurauksena levyn tiedostokirjanpito katosi. Näin menetettiin arviolta kahden viikon työt, ja kadotetut tiedostot oli korvattava skannaamalla sidokset toistamiseen (Lna 13-17h-i sekä Lönnrotin teoksia).
- Tilaongelma helpotti hieman, kun työasema 1:een asennettiin 60 Gt-kokoinen kovalevy täyttyneen ja rikkoutuneen 40 Gt -kokoisen tilalle. Vieläkään SLR-nauhoja ei kuitenkaan voitu purkaa tallennuksen varmistamiseksi kovalevylle, sillä kovalevyltä ei uskallettu hävittää "alkuperäisiä" TIFF-tiedostoja ennen kuin SLR-nauhat oli tarkastettu. Periaatteessa tarkastaminen tulisi tehdä siten, että puretaan kaikki nauhoille

tallennetut tiedostot takaisin kovalevylle ja testataan tiedostojen yhä olevan käytettävissä ja samanlaisia kuin nauhoille aikanaan tallennettaessa.

- Projektin kuluessa nauhojen sisältämän datan tarkistamisesta eli kaikkien tiedostojen purkamisesta takaisin kovalevylle päätettiin työajan ja muistitilan säästämiseksi luopua. Tämän jälkeen listattiin ainoastaan nauhojen sisältämät tiedostot ja luotettiin teknisessä mielessä siihen, että kaikki TIFF-tiedostot saadaan aikanaan nauhoilta asianmukaisesti käyttöön. Listauksen ohella SLR-nauhoilta voidaan purkaa myös satunnaisotoksia, esimerkiksi joka kymmenes tiedosto.

Tiedostojen siirto ja konvertointi

- Käytännössä yksittäisten TIFF-tiedostojen etsiminen ja palauttaminen pitkäaikaistallennukseen käytettäviltä SLR-nauhoilta on varsin työlästä. Päivittäiseen tutkija- yms. asiakaspalveluun tulee luoda mastertiedostoista erilliset käyttökopiot. SKS:ssa käyttökopiot yms. toistaiseksi säilytettävät tiedostot tallennetaan yhteiskäytössä olevalle palvelinkoneella, jossa on automaattinen nauhavarmistus.
- Tehokkaiden verkkoyhteyksien ohella eri työpisteiden ja yhteistyötahojen välillä tapahtuvaan turvalliseen tiedonsiirtoon soveltuu hyvin siirrettävät ulkoiset kovalevyt. Elias-digitoinnissa tiedonsiirtoon ja -varmistukseen käytettävässä ulkoisessa kovalevyssä on muistitilaa 250 Gt.
- Tietovälineistä CD-R- ja DVD-levyjä ei voida suositella arkistoissa tapahtuvaan varmuus- ja pitkäaikaistallennukseen.
- Alkuvaiheessa TIFF-tiedostoja ei työstetty lainkaan ensimmäisessä työasemassa. Skannauksen ja tiedostojen nimeämisen jälkeen ja ennen nauhavarmistusta tiedostot kopioitiin verkkoyhteyden ja jaettujen kovalevyjen kautta työasemaan 2. Näin meneteltiin, jotta työaika jakautuisi tasaisemmin molemmille työasemille ja TIFF-tiedostoista olisi olemassa kopiot kahdella eri kovalevyillä ennen nauhavarmistuksen tekoa.
- Tiedostojen siirron jälkeen työasemalla 2 olevista TIFF-tiedostoista konvertoitiin sekä ”muokkaamattomat” (pakkaussuhde 100%) että käyttökopioiksi soveltuvat JPEG-kuvat.
- Massakonvertointi: 1) JPEG-käyttökopiot pakattiin ja pakkaussuhteeksi valittiin testien jälkeen 80 prosenttia. 2) Kuvatiedostoille tehtiin 20 prosentin terävöitys (ohjelmistokohtainen). 3) Kuvatiedostoja skaalattiin siten että originaalitekstin luettavuus olisi 19 tuuman näytöltä mahdollisimman hyvä. Tämä tarkoittaa yhden luettavan käsikirjoitusvun sovittamista näytölle leveyssuunnassa (skannatun aukeaman enimmäiskoko marginaaleineen noin 2500 x 2000 pikseliä).
- Käytännön kokemuksen perusteella tiedostojen siirtäminen verkossa vei liiaksi koneaikaa ja marraskuussa päädyttiin tekemään JPEG-konvertointi työasemalla 1 ja tallentamaan näin syntyneet tiedostot työasemalle 2. Niin ikään muokatut JPEG-tiedostot päätettiin konvertoida vasta JPEG-formaatista eikä enää suoraan TIFF-kuvista. Mainittakoon, että eri tavalla tuotetuissa käyttökopioissa ei voida havaita suuria laadullisia eroavuuksia. Uudella työkäytännöllä säästettiin työaika, mutta menetettiin samalla TIFF-tiedostojen kaksoisvarmistus eri työasemille.
- Projektissa käytettiin ja testattiin kuvatiedostojen käsittelyssä muutamia varsin toimivia ilmaisohjelmistoja, joita käytettiin eräskannaukseen ja tiedostojen massakonvertointiin. Suositeltavaa on kuitenkin turvautua kuvankäsittelyn ammattilaisten suosimiin laadukkaisiin ohjelmistoihin, jolloin mahdollistetaan yhteneväisten standardien, käytänteiden ja laatuksien käyttö. Näin säästetään myös työaika eri ohjelmistojen vertailulta sekä vältytään mahdollisilta ikäviltä yllätyksiltä, mikäli ilmaisohjelmistot ja niiden eri versiot eivät toimikaan tietystä käyttöympäristössä toivotulla tavalla. Seuraavassa Minerva-projektin työryhmän suosituksia kuvankäsittelyohjelmistojen toiminnallisiksi minimivaatimuksiksi (tekstiä on editoitu) :

- Suurten tiedostojen avaaminen.
- Resoluution ja värisyvyyden määrittäminen.
- Useiden erilaisten versioiden tallentaminen (eri kokoiset tiedostot).
- Kuvan rajaaminen, kopioiminen ja tallentaminen toiseksi tiedostoksi.
- Erilaisten formaattien tuottaminen (esim. terävöitetty ja pakatut verkkokuvat).

CD-R -käyttökopiot

- Muokkaamattomat ja käyttökopioiksi muokatut JPEGit päätettiin tallentaa 2. työaseman kovalevylle sekä CD-levyille (levyn kapasiteetti 650 Mb).
- Levyjen tallennusnopeutena käytettiin 16 x 2400 Kb/sec. Mikäli digitaalinen aineisto olisi tallennettu ainoastaan ko. tietovälineille olisi tallennusnopeutta ollut syytä laskea mahdollisten kirjoitusvirheiden minimoimiseksi (suositus 4 x nopeus). Käytännön ohjenuorana oli, ettei levyjä tallennettu aivan täyteen, sillä suurimmat tallennustekniset ongelmat esiintyvät tutkimusten mukaan levyn ulkolaidalla, johon tietoa poltetaan eli tallennetaan aina viimeiseksi.

- Pakkaamisen merkitystä muistikapasiteetin käytön kannalta kuvastaa se, että digitoidusta Lönnrotiana-kokoelmasta tehtiin muokkaamattomista JPEG-tiedostoista yhteensä 184 kpl CD-levyä ja muokatuista JPEG-tiedostoista ainoastaan 11 kpl.
- Kaikista muokatuista JPEG-kuvista annettiin CD-käyttökopiot Kalevala-instituutille.

SLR-nauhavarmistus

- Lönnrotiana-kokoelman käsikirjoituksista skannatut TIFF-tiedostot tallennettiin SLR-100 magneettinauhoille (muistikapasiteetti 50 Gt). Nauha-asema asennettiin samaan tietokoneeseen, johon myös skanneri oli kytketty. Työprosessien organisoinnin ja sujuvuuden kannalta suositeltavana voidaan pitää erillisen palvelinkoneen hankkimista varmuustallenteiden tekemiseen.
- Koko digitoitu aineisto (n. 25 000 sivua/liuskaa) mahtui 9 nauhalle, joista tehtiin samalla myös varmuustasnauhat (yht. 18 nauhaa).
- SLR-tallennukseen päädyttiin SKS:n kansanrunousarkiston äänitearkistossa sekä Kotuksessa saatujen myönteisten kokemusten perusteella.
- SLR-asemaa käytetään Linux-käyttöjärjestelmän kautta Unix-komennoilla ja TAR-ohjelmalla. Samaa työasemaa käytetään kuitenkin skannausvaiheessa Windows-ympäristössä, joten PC:ssä on käytössä ns. kaksoisbootaus.
- Varmuustallennukset pyritään tekemään yö- ja iltaiikaan jättämällä laitteet työskentelemään yksin. Tämä oli yksi painava seikka Linux-järjestelmän valitsemiseksi nauhavarmistuksen käyttöympäristöksi, sillä se on Windows-käyttöjärjestelmää vakaampi ja vähemmän altis toimintavirheille.
- Käytännön ongelmana on ollut rajallinen koneaika: TIFF-tiedostot tallennetaan vakiintuneen käytännön mukaisesti kahdelle identtiselle SLR-nauhalle. Tämä aiheutti käytännössä pulaa koneajasta, sillä nauhojen ja tiedostojen valmistelu tallentamista varten sekä toisaalta valmiiden nauhojen tarkistaminen vie aikaa. Niinpä työpäivät venyivät usein pitkiksi, kun varsinainen digitointi ja skannaaminen lopetettiin virka-ajan päättyessä ja tämän jälkeen opeteltiin ja hiottiin nauha-aseman käyttörotiinia mahdollisimman varmaksi ja tehokkaaksi.
- Koska nauhojen tallennuskapasiteetti on peruskiintolevyihin verrattuna varsin suuri, on tallentaminen tehtävä useammassa kuin yhdessä erässä, mikä jonkin verran heikentää työskentelyn tehoa ja lisää virhemahdollisuuksia. Työasema 1:n kovalevytila oli digitoinnin alkaessa 2 x 40 Gt, joten kumpikaan levy ei yksinään voinut toimia lähdekansiona kokonaisen SLR-nauhan täyttämiseksi kerta-ajolla. Kovalevytilan puutteen vuoksi TIFF-tiedostoja jouduttiin säilyttämään joskus useita päiviä vain yhtenä kopiona yhdellä kovalevyllä. Korostettakoon, että digitoinnissa tarvittavan muisti- ja tiedonsiirtokapasiteetin osalta lyhytkestoisenkin projektin tarpeet tulee suunnitteluvaiheessa aina mieluummin ylimitoittaa kuin alimitoittaa. Projektin aikana tehtävät laite- ja ohjelmistoasennukset voivat hidastaa digitointia sekä aiheuttaa uusia yllättäviä ongelmatilanteita (esim. yhteensopivuusongelmat).
- SLR-nauhoilla oleva data tullaan jatkossa siirtämään uusille ja enemmän muistikapasiteettia omaaville tietovälineille (esim. LTO). Magneettinauhojen virkistäminen suositellaan tehtäväksi viiden vuoden välein, virkistysvaiheessa voidaan harkita myös datan siirtoa suoraan toiselle medialle.
- Erillisten levypalvelimien merkitys varmuus- ja pitkäaikaistallennuksessa tulee varmasti jatkossa yhä korostumaan. Periaatteessa siirrettävistä tietovälineistä voidaan jopa osittain luopua, mikäli tehokkaiden palvelimien muistikapasiteettia voidaan jatkuvasti laajentaa digitoinnin edistyessä. Palvelimien kovalevyille arkistoidun tietomassan tulee luonnollisesti olla kopioitavissa tai siirrettävissä seuraavan sukupolven massatallennusvälineille.

Liite 6 SKVR-korpuksen rakenne

Jukka Saarinen

SKVR-korpuksen perusyksikön muodostaa yksittäinen runomuistiinpano, teksti. Jokainen numeroitu teksti muodostaa yhden yksikön (ITEM), jolla on aina sama toistuva rakenne. ITEM-elementissä on kolme elementtiä:

- metadata (META)
- tekstiosa (TXT)
- viitteet (REFS)

Metadata (META) on peräisin SKVR:sta virheineen ja kirjoitustapojen epäsäännöllisyyksineen. Sen elementit ovat:

- osa (OSA), kaikkien osien tekstejä yhdessä käsiteltäessä välttämätön lisäys
- tekstinumero (ID)
- topografinen tieto (LOC)
- kerääjän nimi (COL)
- tekstin arkistotunnus (SGN)
- ajanmääre (TMP)
- muuta taustatietoa (INF), joukossa yleisesti mm. esittäjän, informantin nimi.

Painetussa tekstissä henkilönimien, paikkakuntien ja ajanmääreiden kirjoitusasut vaihtelevat, minkä vuoksi automaattista tiedonhakua varten metadataa säännöllistetään käyttämällä numeerisia attribuutteja ilmaisemaan eksakteja arvoja :

- tekstinumerolle (nro)
- kerääjälle (k)
- paikkakunnalle (p)
- ajanmääreelle (y)

Rakenteellisesti hankala on informantin nimi, jota ei voida tarpeeksi virheettömästi automaattisesti erottaa INF-elementistä.

Tekstin (TXT) elementit ovat:

- runosäkeet (V)
- suorasanaiset jaksot (L)
- otsikot (CPT)
- toimittajien yksittäiset kommentit (K)

Viiteosa (REFS) sisältää alaviitteet, jotka ovat sekä toimittajien kommentteja ja selvennyksiä että tietoja käsikirjoituksissa olevista korjauksista, lisäyksistä ja poistoista.

Seuraavassa kuva tekstistä SKVR VII2 2168 ja sen XML-koodaus:

```
<ITEM nro="7221680" p="4360" k="35" y="1902">
  <META>
    <OSA>VII2</OSA>
    <ID>2168.</ID>
    <LOC>Rääkkylä.</LOC>
    <COL>Hyvärinen,</COL>
    <SGN>A. n. 323.</SGN>
    <TMP>-02.</TMP>
    <INF>
      Rasivaara. Loviisa Asikainen, 76 v. Kuullut lapsuudessaan Kiteen Potoskavaarassa.
    </INF>
  </META>
  <TEXT>
    <L>Päivännäkemättömällä kun paineltiin, sanottiin:</L>
    <V>Puuhun#1 muhkat, muahan mahkat,</V>
    <V>kantoin#2 ves'näräpät;</V>
    <V>elä immeiseen tähän</V>
    <V>ennee nosta muhkii!</V>
  </TEXT>
  <REFS>
    #1 kk:ssa pieni alkukirjain.
    #2 r. (kantoihin).
  </REFS>
</ITEM>
```

Liite 7 SKVR-korpuksen datatyyppin määritelmä (DTD)

```
<!-- DTD SKVR-korpusta varten -->

<!ELEMENT KOKONAISUUS (ITEM+)>
<!ELEMENT ITEM (META, TEXT, REFS?)>
<!ELEMENT META (OSA, ID, LOC, COL, SGN, TMP, INF?)>
<!ELEMENT OSA (#PCDATA)>
<!ELEMENT ID (#PCDATA)>
<!ELEMENT LOC (#PCDATA)>
<!ELEMENT COL (#PCDATA)>
<!ELEMENT SGN (#PCDATA)>
<!ELEMENT TMP (#PCDATA)>
<!ELEMENT INF (#PCDATA|I|KA)*>
<!ELEMENT TEXT (CPT|V|L|K)*>
<!ELEMENT CPT (#PCDATA|I|KA)*>
<!ELEMENT V (#PCDATA|I|KA)*>
<!ELEMENT L (#PCDATA|I|KA)*>
<!ELEMENT K (#PCDATA|I|KA)*>
<!ELEMENT REFS (#PCDATA|I|KA)*>
<!ELEMENT I (#PCDATA|KA)*>
<!ELEMENT KA (#PCDATA)>
<!ATTLIST ITEM nro ID #REQUIRED>
<!ATTLIST ITEM p CDATA #IMPLIED>
<!ATTLIST ITEM y CDATA #IMPLIED>
<!ATTLIST ITEM k CDATA #IMPLIED>

<!ENTITY a3 "!a">
<!ENTITY a4 "!a">
<!ENTITY a5 "!a">
<!ENTITY a7 "!"&#227;">
<!ENTITY a9 "!a">
<!ENTITY a10 "!"&#229;">
<!ENTITY a11 "!a">
<!ENTITY a12 "!"&#224;">
<!ENTITY a13 "!"&#228;">
<!ENTITY a15 "!"&#228;">
<!ENTITY a16 "!"&#228;">
<!ENTITY a17 "!"&#228;">
<!ENTITY a18 "!"&#228;">
<!ENTITY a19 "!"&#228;">
<!ENTITY a21 "!"&#228;">
<!ENTITY a22 "!"&#228;">
<!ENTITY a23 "!"&#228;">
<!ENTITY b1 "!b">
<!ENTITY d1 "!d">
<!ENTITY d2 "!d">
<!ENTITY d3 "!d">
<!ENTITY d4 "!d">
<!ENTITY e1 "!e">
<!ENTITY e2 "!e">
<!ENTITY e3 "!e">
<!ENTITY e4 "!e">
<!ENTITY e7 "!"&#275;">
<!ENTITY e8 "!"&#283;">
<!ENTITY e10 "!"&#235;">
<!ENTITY e11 "!"&#279;">
<!ENTITY e12 "!"&#279;">
<!ENTITY e13 "!"&#279;">
<!ENTITY e14 "!"&#279;">
<!ENTITY e15 "!"&#279;">
<!ENTITY e17 "!"&#279;">
<!ENTITY e18 "!"&#279;">
<!ENTITY e20 "!E">
<!ENTITY e21 "!"&#275;">
<!ENTITY e22 "!e">
<!ENTITY e23 "!"&#279;">
<!ENTITY h2 "!h">
<!ENTITY h3 "!h">
<!ENTITY h4 "!h">
<!ENTITY h5 "!h">
<!ENTITY h6 "!h">
<!ENTITY h7 "!h">
<!ENTITY i1 "!i">
<!ENTITY i2 "!i">
<!ENTITY i3 "!i">
<!ENTITY i4 "!"&#299;">
```


<!ENTITY i5 "!ī">
<!ENTITY i6 "!i">
<!ENTITY i7 "!i">
<!ENTITY i8 "!i">
<!ENTITY i9 "!î">
<!ENTITY i10 "!ĭ">
<!ENTITY i11 "!ĭ">
<!ENTITY i13 "!i">
<!ENTITY i14 "!i">
<!ENTITY i15 "!i">
<!ENTITY i16 "!i">
<!ENTITY i17 "!I">
<!ENTITY i18 "!I">
<!ENTITY j1 "!j">
<!ENTITY j2 "!j">
<!ENTITY j3 "!j">
<!ENTITY j4 "!j">
<!ENTITY j6 "!j">
<!ENTITY j8 "!j">
<!ENTITY j10 "!J">
<!ENTITY k2 "!k">
<!ENTITY k3 "!k">
<!ENTITY k4 "!k">
<!ENTITY k6 "!k">
<!ENTITY l1 "!l">
<!ENTITY l2 "!l">
<!ENTITY l3 "!l">
<!ENTITY l4 "!l">
<!ENTITY l5 "!l">
<!ENTITY l6 "!l">
<!ENTITY l7 "!l">
<!ENTITY m1 "!m">
<!ENTITY m3 "!m">
<!ENTITY m4 "!m">
<!ENTITY m5 "!m">
<!ENTITY n1 "!nn">
<!ENTITY n2 "!ň">
<!ENTITY n3 "!n">
<!ENTITY n4 "!n">
<!ENTITY n5 "!n">
<!ENTITY n6 "!n">
<!ENTITY n7 "!n">
<!ENTITY n8 "!n">
<!ENTITY n9 "!n">
<!ENTITY n10 "!n">
<!ENTITY n11 "!n">
<!ENTITY n12 "!ŋŋ">
<!ENTITY n13 "!ŋ">
<!ENTITY o1 "!o">
<!ENTITY o2 "!O">
<!ENTITY o3 "!o">
<!ENTITY o4 "!o">
<!ENTITY o5 "!O">
<!ENTITY o6 "!o">
<!ENTITY o7 "!o">
<!ENTITY o8 "!o">
<!ENTITY o9 "!o">
<!ENTITY o10 "!ō">
<!ENTITY o11 "!o">
<!ENTITY o12 "!o">
<!ENTITY o13 "!o">
<!ENTITY o14 "!o">
<!ENTITY o15 "!o">
<!ENTITY o16 "!o">
<!ENTITY o17 "!ö">
<!ENTITY o18 "!ö">
<!ENTITY o19 "!ö">
<!ENTITY o20 "!ö">
<!ENTITY o21 "!ö">
<!ENTITY o23 "!ö">
<!ENTITY p2 "!p">
<!ENTITY p4 "!p">
<!ENTITY p5 "!p">
<!ENTITY q1 "!Q">
<!ENTITY r1 "!r">
<!ENTITY r2 "!r">
<!ENTITY r3 "!r">
<!ENTITY r4 "!r">
<!ENTITY r5 "!r">
<!ENTITY r6 "!r">
<!ENTITY s1 "!s">

```
<!ENTITY s3 "!s">
<!ENTITY s4 "!s">
<!ENTITY s6 "!s">
<!ENTITY s8 "!&#353;">
<!ENTITY s9 "!&#353;">
<!ENTITY s10 "!&#353;">
<!ENTITY s11 "!&#352;">
<!ENTITY s12 "!&#352;">
<!ENTITY z1 "!z">
<!ENTITY z2 "!z">
<!ENTITY z4 "!&#382;">
<!ENTITY z5 "!&#381;">
<!ENTITY t1 "!t">
<!ENTITY t2 "!t">
<!ENTITY t3 "!t">
<!ENTITY t5 "!t">
<!ENTITY t6 "!t">
<!ENTITY t7 "!t">
<!ENTITY t8 "!t">
<!ENTITY t9 "!t">
<!ENTITY t10 "!t">
<!ENTITY t11 "!T">
<!ENTITY t12 "!T">
<!ENTITY u1 "!u">
<!ENTITY u2 "!u">
<!ENTITY u4 "!&#365;">
<!ENTITY u5 "!u">
<!ENTITY u6 "!u">
<!ENTITY u7 "!u">
<!ENTITY u10 "!&#252;">
<!ENTITY u11 "!&#252;">
<!ENTITY u14 "!&#252;">
<!ENTITY u15 "!&#252;">
<!ENTITY v1 "!v">
<!ENTITY v2 "!v">
<!ENTITY v3 "!v">
<!ENTITY v5 "!v">
<!ENTITY v6 "!v">
<!ENTITY v7 "!v">
<!ENTITY v8 "!v">
<!ENTITY y1 "!y">
<!ENTITY y2 "!y">
<!ENTITY y4 "!y">
<!ENTITY ayla "!a">
<!ENTITY byla "!b">
<!ENTITY eyla "!e">
<!ENTITY iyla "!i">
<!ENTITY kyla "!k">
<!ENTITY myla "!m">
<!ENTITY Nyla "!N">
<!ENTITY oyla "!o">
<!ENTITY tyla "!t">
<!ENTITY uyla "!u">
<!ENTITY vyla "!v">
<!ENTITY aumlyla "!&#228;">
<!ENTITY oumlyla "!&#246;">
<!ENTITY uumlyla "!&#252;">
<!ENTITY oala "!o">
```

Liite 8 Kuvamuotoisena ja merkkipohjaisesti digitoitujen aineistojen yhteiskäyttö – selvitys mahdollisuuksista yhdistää alkuperäisaineistot kopioihinsa

Jukka Saarinen

1. Johdanto

Tässä raportissa selvittelen niitä tapoja ja mahdollisuuksia, joilla eri tavoin digitoituidut aineistot saadaan toimimaan yhdessä. Varsinaisesti en puutu tekniikkaan - tietokantaratkaisuihin tms. - vaan pyrin tarkastelemaan tällaisen käytön kannalta kriittisiä tietoelementtejä ja niiden säännönmukaistamista konelukaiseen muotoon. Kohteena ovat merkkimuotoon digitoituidut ja xml-muotoon rakenteistetut Suomen Kansan Vanhat Runot (ja niihin läheisesti liittyvä julkaisemattomien kalevalamittaisten runojen kortisto) ja kuvamuotoon digitoituidut alkuperäiskäsikirjat, erityisesti Elias-projektissa digitoitu Lönnrotiana.

Käsittelen aluksi signumin käsitettä Kansanrunousarkistossa ja sitä, miten sitä julkaistussa SKVR-teoksessa käytetään. Sen jälkeen esittelen yhden mahdollisuuden signumin säännönmukaistamiseen konelukaiseen numeeriseen muotoon. Seuraavaksi tarkastelen numeeristetun signumin liittämistä digitointeihin (sivukohtainen luettelointi) ja mahdollisuuksia tällaisen signumin hyödyntämiseen alkuperäisiä ja kopioituja aineistoja yhteen liittäessä. Lopuksi tarkastelen mahdollisuuksia tuoda metadataa merkkimuotoisista digitoitujen alkuperäisaineistojen luettelointiin.

2. Signum ja SKVR-korpus

Linkki alkuperäiskäsikirjoituksen ja kopioiden välillä on Kansanrunousarkiston arkistosignum, joka on mukana molemmissa, sekä SKVR:ssä että julkaisemattomien kortistossa. SKVR:n signum muodostuu kerääjän nimestä, mahdollisesta kokoelmatunnuksesta ja yksikkönumerosta. SKVR-korpuksen XML-teksteissä signum sijaitsee elementeissä COL ja SGN. Pari esimerkkiä valaissee asiaa (SGN-elementin osuus kursiivilla)

Rytkönen A, n. 1037. -> johtaa A. Rytkösen kokoelman nroon 1037
Lönnrot S, n. 195 -> johtaa Elias Lönnrotin S-kokoelman nroon 195
Eur. G, n. 479. -> johtaa D.E.D. Europaeuksen G-kokoelman nroon 479
Mikkonen Ida, n. 246. -> johtaa Ida Mikkosen kokoelman nroon 246

Lönnrotin kohdalla "S" viittaa S-kokoelmaan, Europaeuksen "G" hänen G-kokoelmaansa, jotka fyysisesti ovat erillisiä sidoksia. Lönnrotin kokoelmissa on nykyään käytössä ns. Lönnrotiana-numerot, jotka voidaan suoraan johtaa vanhoista tunnuksista. Kuten kursivoinnista huomaa, koneellisesti ei ole onnistuttu erottelemaan kerääjän nimeen kuuluvia osia kokoelma- ja numerotunnuksista (kaikilla ei ole useampia kokoelmia). "Rytkönen A" tarkoittaa Antti Rytköstä, mutta "Lönnrot S" ei tarkoita Seppo Lönnrotia! Automaattinen rakenteistaminen ei voi tätä eroa tehdä. COL-elementti määriteltiin käsittämään pelkästään pitäjänimeä seuraava sana, ja mahdollinen etunimi tai sen alkukirjain on tullut osaksi SGN-elementtiä. Näihinkin on jouduttu tekemään korjauksia, koska satunnaisesti SKVR:ssä kerääjän nimi on alkanut etunimellä, jolloin normaalistamisen kannalta olennainen sukunimi on hukkunut SGN-elementtiin.

3. Signumin säännönmukaistaminen numeeriseksi

Signumin muuntaminen atk:lla tulkittavaksi edellyttää sen säännönmukaistamista. Puran signumin osiinsa.

1. *Sarjatunnus*. Sidossarjalla tarkoitetaan aineistokokonaisuutta, joka arkistoidaan erikseen. Sarjassa on tavallisesti useamman kerääjän aineistoa. SKVR-aineistossa sarjatunnus on hyvin harvinainen, koska systeemi ei ollut vielä siihen aikaan vielä nykyiselleen muotoutunut. Oletusarvoisesti liki kaikki aineistot ennen vuotta 1935 kuuluvat ns. kerääjäsidoksiin, jonka tunnus on nykyään S. Todennäköisesti jotkut käsikirjoiukset on sijoitettu ns. VK-sarjaan, mutta arkistotunnukseen SKVR:ssä tätä tietoa ei ole viety, koska numerointi on joka tapauksessa kerääjänmukainen, ja VK ilmaisee vain säilytyspaikan. Saman kerääjän kerääjäsidosten ja VK:n numerot eivät ole päällekkäiset. Periaate säilyi luultavasti vielä kaikissa muissakin SKVR:n julkaisuajan sarjoissa, mutta tätä periaatetta ei noudateta enää, joten sarjatunnus on pakollinen osa signumia. 1930-luvulla ja sen jälkeen julkaistuissa SKVR:n osissa on sarjoja jo enemmän.

Sarjatunnus on kirjainyhdistelmä tai sana (S, KRK, VK, KT, Alko, Rakentajat, Posti ...) - jokaiselle sarjalle voidaan antaa tunnusnumero.

2. *Kerääjä eli luovuttaja*. Henkilö, jonka nimiin aineisto on tallennettu ja joka siten määrää myös aineiston sijoituspaikan. Voi olla todellinen kerääjä, tai henkilö tai yhteisö, jonka nimissä keruu on tapahtunut ja joka aineiston on luovuttanut. SKVR-ajan luovuttajayhteisöistä merkittävien ryhmä ovat opistot ja koulut ("Hämeenlinnan alakansakouluseminaari"), joiden kerääjinä toimivat opiskelijat ja koululaiset.

Kerääjä on nimi. SKVR-korpuksessa nimien kirjoitusasut vaihtelevat suuresti ja jokaiselle nimelle on annettu numeerinen tunnus, joka on sijoitettu ITEM-elementin attribuutiksi "k".

3. *Kokoelma*. Tarkoittaa, että kerääjällä on useampia päällekkäisiä numerosarjoja, jotka eroavat kokoelmatunnuksen tms. avulla. Esim. Lönnrot S, Borenius A II., vanhojen aineistojen a), b), c) ja d) - kokoelmat. Yhdellä kerääjällä kokoelmia on hyvin rajallinen määrä.
4. *Yksikkönumero*. Juokseva numero, jolla voi olla alakohtia (4 a, 4 b ...). SKVR:ssä yksikkönumeroa edeltää usein lyhenne, n. tms.

Yksi tapa muodostaa määrämuotoinen signum olisi neljän numeron sarja: ensimmäinen numero ilmaisee sidossarjan, toinen kerääjän, kolmas kokoelman, neljäs yksikkönumeron. Esim.

Eur. G, n. 479. ==> 1-17-7-4790

Sama voitaisiin ilmaista myös yhtenä numerona, mutta silloin jokaiselle osalle olisi varattava riittävä määrä numeroita: sarja - n3, kerääjä - n5, kokoelma - n3, yksikkönumero - n6:

Eur. G, n. 479. ==> 00100017007004790

Sarjoille ja kerääjille tarvitaan omat taulukot, jossa on vastaavuudet numeroiden ja sarjan tai kerääjänimen mukaan:

Sarjat	Kerääjät
1 = S	14 = Collan, Anni
2 = VK	15 = Engelberg, R. S.
3 = E	16 = Erkko, J. H.
4 = KRK	17 = Europaeus, D. E. D.
...	...

Tarvitaan myös taulukko, jossa on kokoelmat kerääjittäin:

Kerääjän tunnus	Kokoelman tunnus	Kokoelman nimi
7	5	E
7	6	F
7	7	G

On myös mahdollista, että kokoelmat noudattavat pääpiirteissään yleisiä nimeämisperusteita, jolloin ne voitaisiin nimetä yhdessä taulukossa:

a)	1
b)	2
...	
F	16
G	17

4. Signumin liittäminen alkuperäiskäsikirjoituksen digitalisointeihin

Käsikirjoituksessa signumit ovat numerojärjestyksessä. Siksi määrämuotoinen (numeerinen tai vastaava) signum voidaan liittää esim. alkuperäiskäsikirjoituksen digitoituun kuvaan suhteellisen yksinkertaisesti seuraavasti¹:

Esim. Lönnrotiana 5

¹ Toinen vaihtoehto: jokaiseen signumiin liitetään kuvat, jotka vastaavat sitä: 1-77-5-20 <=> lna005_005.jpg, lna005_006.jpg.

<i>Kuvatiedosto (= sidoksen aukeama)</i>	<i>Aukeamalla olevat yksiköt</i>
Ina005_005.jpg	1-77-50-10 1-77-5-20
Ina005_006.jpg	1-77-50-20 1-77-5-30
Ina005_007.jpg	1-77-50-40 1-77-5-50
Ina005_008.jpg	1-77-50-50
Ina005_009.jpg	1-77-50-60 1-77-5-70
Ina005_010.jpg	1-77-50-70
Ina005_011.jpg	1-77-50-80

Elias-projektissa Lönnotianan digitoinnin yhteydessä on tehty jokaiseen digitoituun sidokseen xml-sisällysluettelo, joka ilmaisee, mitkä yksiköt sijaitsevat digitoituilla sivuilla:

```
<R><FN>lna001_015.jpg</FN> <SGN>Tekstit 4c-6a</SGN></R>
<R><FN>lna001_016.jpg</FN> <SGN>Tekstit 6b-7</SGN></R>
<R><FN>lna001_017.jpg</FN> <SGN>Tekstit 7-10</SGN></R>
<R><FN>lna001_018.jpg</FN> <SGN>Tekstit 10-13</SGN></R>
```

Numeerinen signum voitaisiin haluttaessa lisätä jokaisen rivin/kuvan/sivun yhteyteen. Sarja, kerääjä ja kokoelma ovat vakiot (1-77-1-), yksikkönumero muuttuu:

```
<R><FN>lna001_015.jpg</FN> <SGN>Tekstit 4c-6a</SGN><ID>1-77-10-43 1-77-10-50
1-77-10-61</ID></R>
<R><FN>lna001_016.jpg</FN> <SGN>Tekstit 6b-7</SGN><ID>1-77-10-62
1-77-10-70</ID></R>
<R><FN>lna001_017.jpg</FN> <SGN>Tekstit 7-10</SGN><ID>1-77-10-70 1-77-10-80
10-77-1-901-77-10-100</ID></R>
<R><FN>lna001_018.jpg</FN> <SGN>Tekstit 10-13</SGN><ID>1-77-10-100 1-77-10-110
1-77-10-120 1-77-10-130</ID></R>
```

Yksikkönumeroiden ilmaiseminen numerovälinä (esi. "4c-6a") aiheuttaa tarpeen tarkistaa tieto alkuperäisestä käsikirjoituksesta, joten tietojen kirjoitustapoihin on vastaisuudessa alun alkaen syytä kiinnittää enemmän huomiota, koska työ on yksinkertaisinta tehdä alkuperäisessä tallennusvaiheessa.

5. Signumin liittäminen SKVR- ja julkaisemattomiin teksteihin

SKVR:n tyyppiset kopioaineistot on järjestetty muilla kuij signumin mukaisilla perusteilla, joten signumin lisääminen suoraan edellyttäisi jokaisen kirjoittamista erikseen. Signumiin kuuluvat elementit (COL ja SGN) + SKVR-numero voidaan kuitenkin ottaa erikseen ja järjestää kerääjän nimen (tai attribuuttina olevan numeron) mukaan oikeaan järjestykseen ja sen jälkeen ainakin suurimmaksi osin automatisoidusti antaa signum:

Esim. (oikea) A. E. Ahlqvistin runoukeräykset (sarja= S=1, kerääjä=Ahlqvist=1, kokoelma=B=2)

SKVR:stä suoraan:	edellisistä osin automatisoidusti		
Ahlqvist	B, n. 10.	7126340	1-1-2-100
Ahlqvisi	B, n. 19.	7115280	1-1-2-190
Ahlqvisi	B, n. 21.	7119400	1-1-2-210
Ahlqv.	B, n. 64.	7125330	1-1-2-640
Ahlqv.	B, n. 67.	7124970	1-1-2-670
Ahlqv.	B, n. 76.	7128220	1-1-2-760
Ahlqvist	B, n. 104.	7118910	1-1-2-1040

Ahlqv.

B. n. 206.

7109300

1-1-2-2060

Numerisoidun signumin avulla voidaan yhdistää SKVR:n teksti (tai mikä tahansa alkuperäislähteestä oleva teksti, jossa on signum mukana) digitoituun kuvatiedostoon:

SKVR I 55 => skvr1100550 = Lönnrot All6 (Lönnrotiana 5) 68 . 1834 => 1-77-50-680

yhdistetään kuvatiedostoon haulla, jolloin saadaan tulokseksi, että ko. signum esiintyy kuvatiedostoissa

lna005_045.jpg, lna005_046.jpg, lna005_047.jpg, lna005_048.jpg, lna005_049.jpg, lna005_050.jpg, lna005_052.jpg, lna005_052.jpg.

Ongelmia signumin numerisoinnissa:

1. SKVR-teksteissä käytetyt arkistoviitteet ovat useassa tapauksessa vanhentuneita tai eivät sisällä kaikkea informaatiota, joka olisi tarpeen tietyn tekstin yksiselitteiseksi tunnistamiseksi. Lönnrotin viitteinä käytetään vanhoja kokoelmanumeroita, jotka vastaavat nykyisin käytettyjä Lönnrotiana-numeroita. Joissain tapauksissa (esim. Borenius) yksikkönumerointi on uusittu kokonaan. Näissä tarvitaan konversiotaulukoita, jotka ilmaisevat, mitkä kokoelmanimet, mitkä yksikkönumerot viittaavat samaan.
2. Nykyisen järjestelmän kannalta olennainen sarjatunnus ei vielä samalla tavalla ole olemassa SKVR:n tekoaikana.
3. Kokoelma-tieto on kerääjäkohtainen ja pitää jokaisen osalta määrittää ensin erikseen.
4. SKVR-tekstiin signum on kirjoitettu epäsäännöllisesti. Kerääjä, kokoelma ja yksikkönumero voivat esiintyä monissa muodoissa.

Numerisointi merkitsisi joka tapauksessa jonkin asteista käsityötä, automatisoidusti sen toteuttaminen ei suoraan onnistuisi. Sen sijaan niissä tapauksissa, joissa aineisto on suoraan signumin mukaisessa järjestyksessä, teksienn signumien numerisointi olisi varsin helposti toteutettavissa. Signumin mukaisesti järjestettyjä aineistoja ovat SKVR:n osat III ja IV ja koko julkaisematon digitoitu aineisto.

Numerisointi ei selailukäytössä kuitenkaan ole aivan välttämätön toimenpide, mikäli jokaisesta digitoidusta sidoksesta on hyvät yksikkönumeron käsittävät sisällysluettelot. Sisällysluettelon etuna on se, että se näyttää myös numeroinnin ulkopuolelle tippuvan aineiston ja antaa helpomman pääsyn ko. yksikön sivuja ympäröiviin sivuihin.

6. Digitoinnit ja metadata

Numerisoinnin suuri etu olisi siinä, että sen avulla voitaisiin valmiista tekstikorpuksista (kuten SKVR) siirtää yksikkökohtaista metadataa muihin järjestelmiin, esim. digitoituihin sisällysluetteloihin tai luettelointitietoja sisältäviin tietokantoihin. Näin voitaisiin hyödyntää erikoisluetteloihin tallennettua valmista ja usein aika perusteellista tietoa arkiston yleistietokannoissa. Mahdollisuus olisi liikkua myös toisin: arkiston yleistietokantojen tietoja voidaan hyödyntää erikoisluetteloissa ja digitoinneissa, mikäli tiedot kirjataan sovittujen käytäntäiden mukaan. Seuraavassa keskityn kuitenkin ensimmäiseen tapaukseen, erikoisluettelojen tiedon hyväksikäyttöön yleistietokannoissa.

On otettava huomioon seuraavat seikat:

1. SKVR:n tapauksessa (ja todennäköisesti muidenkin erikoisluettelojen suhteen) on muistettava, etteivät merkkimuotoiset digitoinnit kata kaikkea käsikirjoituksessa olevaa, jolloin osa käsikirjoituksesta ei saa metadataa SKVR:n kautta. Myös Lönnrotianassa on muuta kuin kalevalamittaista materiaalia, joka ei ole siten tullut merkkimuotoisen digitoinnin piiriin SKVR:n tai julkaisemattoman runouden kortiston kautta.
2. Yksikköinä digitoitujen SKVR- ja julkaisemattomien aineistojen metadata on sellaisessa muodossa, että jokainen yksikkö on periaatteessa ymmärrettävissä sellaisenaan, irrotettuna käsikirjoituskokonaisuuskontekstistaan. Tämä vuoksi toisaalta toistetaan tietoja, jotka ovat samat koko käsikirjoitukselle (kerääjän nimi) tai useammalle peräkkäiselle yksikölle (esim. paikkakunta, informantti, keruuajankohta.), toisaalta jätetään pois tietoa, jota ei tarvita kyseistä aineistoa omana kokonaisuutenaan

käsiteltäessä (esim. tieto perinnelajista, joka on oletusarvoisesti joku kalevalaisen runouden alalajeista b1, b3, b4, b5,7). Kokonaista käsikirjoitusta koskevat vakiot voidaan poistaa ja ilmaista hierarkisesti ylemmällä tasolla (kerääjä), mutta muiden tietojen toistaminen jokaisen yksikön yhteydessä lienee aivan tarkoituksenmukaista.

3. SKVR:ssä (ja osin julkaisemattomassa) on metatietoina käytetty tietoja, jotka

- kiistatta kuuluvat jonkun yksikön yhteyteen
- esiintyvät samassa käsikirjoituksessa mutta joiden yhteys ko. yksikköön on oletettu
- metatiedot ovat ainakin osin peräisin toisesta lähteestä, yhteys perustuu tutkimukseen.

Ongelmana SKVR:ssä on, että tekstien metatiedoissa eroa eri vaihtoehtojen välillä ei ilmoiteta. Jos luettelointitiedoista on helppo pääsy digitoituun alkuperäisaineistoon, käyttäjän on kuitenkin mahdollista itse päätellä, mihin tiedot perustuvat ja päättää, minkä arvon tiedoille antaa.

Otan esimerkiksi yhden xml-tekstin:

```
<ITEM nro="skvr1110520" p="5996" k="77" y="1835" s="1-77-220-420" type="b1-12345">
  <META>
    <OSA>I 2</OSA>
    <ID>1052.</ID>
    <LOC>Uhut.</LOC>
    <COL>Lönnot</COL>
    <SGN>A II 9, n. 42.</SGN>
    <TMP>27/4 35.</TMP>
    <INF>
      Varahvontta Sirkei&#324;i l. Jama&#324;i.
    </INF>
  </META>
  <TEXT>
    <V>Iivana iso isäntä,</V>
  ...
```

Esimerkissä on jo korpuksessa olevien normaalistettuja tietoja ilmaisevien attribuuttien (nro=SKVR-id, p=paikkakunta, k=kerääjä, y=vuosi) lisäksi kaksi mahdollista attribuuttia (s=signum, type=perinnelajikoodin sisältävä runotyyppi). Attribuutin "s" avulla kyseinen teksti voidaan sijoittaa aineiston luettelointitietoihin oikealle paikalleen. Muista attribuuteista saadaan säännönmukaistetut tiedot luettelointia varten.

Tietojen tarkennukset, esim. kylään ja keruuajankohtaan liittyvät, ovat ongelmallisia. Esim. laulajan nimeä (Varahvontta Sirkeinen eli Jamani) ei ole säännönmukaistettu, koska elementin INF sisältö ei ole säännönmukainen: laulajan nimi ei läheskään aina ole mukana eikä se aina ole ihan samassa paikassa. Myös nimen kirjoitusasu vaihtelevat. Yksi mahdollisuus olisi sijoittaa SKVR:n tiedot säännönmukaistettujen "kommenteiksi". "Kommentti"-mallissa elementin sisältö voisi olla myös sellaisenaan mukana. Jos elementin teksti on pitkä, otetaan mukaan esim. 80 ensimmäistä merkkiä, jotka kaiken todennäköisyyden mukaan sisältävät olennaisimmat kontekstitiedot. Näin myös INF-elementti tärkeine tietoineen pääsisi mukaan.

Jokaisessa yksikössä voisi SKVR:n kautta siis olla seuraavat tiedot siirrettäväksi SKVR:stä toiseen yleisluontoisempaan tietokantaan (Lönnot A II 9 = Lönnotiana 22):

Otsikko	attr	säännönm.	ed. sanallinen arvo	SKVR:n teksti
Signum	s	1-77-220-420	S Lönnotiana 22:42	Lönnot A II 9, n. 42.
SKVR-numero	nro	skvr1110520	SKVR I2 1052	
Vuosi	y	1835	1835	27/4 35.
Paikkatieto	p	5996	Uhtua	Uhut
Muut tiedot				Varahvontta Sirkeini I. Jamani
Sisältö	type	b1-12345	b1 kalevalamittainen runo: (epiikka) livanan virsi	livana iso isäntä [1. säe]

7. Päätelmät

Merkkimuotoisena olevan tekstin ja sen kuvamuotoisen alkuperäiskäsikirjoituksen välille voidaan perustaa yhteys. Yhteyden luominen voidaan perustella kahdesta eri näkökulmasta:

1) Merkkimuotoiset tekstiyksiköt voidaan palauttaa kontekstiinsa (lähdekriittinen lähestymistapa),

2) Merkkimuotoisista tekstiyksiköistä voidaan saada metadataa alkuperäiselle aineistolle (metadatakeskeinen lähestymistapa)

Käytännössä merkkimuotoisten ja kuvamuotoisten digitoitien suhteessa Kansanrunousarkiston aineistoissa ensimmäinen näkökulma on tutkimuksellisesti tärkeämpi. Yksikkökohtaisten vastaavuuksien automatisointi (linkkaus suoraan tekstistä vastaavaan kuvamuotoiseen sivuun) ei ole välttämätöntä, koska käyttäjä voi itse varsin vaivattomasti tämän yhteyden luoda, mikäli hänellä on pääsy käsikirjoituksen tasolle saakka. Etuna on lisäksi se, että kokonaisuus (mitä muuta käsikirjoituksessa on) tulee automaattisesti esille.

Toisessa näkökulmassa on kaksi suurta ongelmaa: arkistosignum numerisoinnin monimutkaisuus ja metadatan vaihtelevuus (erityisesti laulajan/informantin nimi), mikä aiheuttaa ongelmia, jos eri merkkimuotoisista korpuksista peräisin olevia tietoja yritetään yhdistää. Mutta jos dataa on paljon, se on muodoltaan suhteellisen homogeenista ja signumin numeeristaminen onnistuu automatisoiden, näkökulma on käyttökelpoinen. Täydelliseen tulokseen sillä ei päästä, mutta osittaiseen luettelointiin se voi olla hyvä apuneuvo.

Liite 9 Joitakin huomioita SKVR-tietokannasta käyttäjän näkökulmasta

Niina Hämäläinen

SKVR-tietokanta saatiin tutkijoiden koekäyttöön vuodenvaihteessa. Vuoden alusta ulkomailla, Luxemburgissa pääasiassa asuvana tutkijana tietokannasta on ollut minulle suunnatonta apua. Olen saanut runoaineiston suoraan työpyödylleni. Tämä ei olisi ollut mahdollista aikaisemmin, sillä SKVR:n saatavuus on ollut melko heikkoa. Teokset löytyvät kirjastoista, joitakin osia on voinut ostaakin, mutta yliopiston ulkopuolella en olisi koskaan saanut yhtä aikaa koko SKVR-antologiaa käyttöni. Vaikka tietokanta on vielä aineistoltaan vaillinainen, eikä sitä sen vuoksi voi täysin hyödyntää, ei tämä seikka poista tietokannan ansiota tutkijan apuvälineenä.

Ensimmäinen, ilmeinen etu tietokannasta on jo mainitsemani saatavuus. Tähän liittyy olennaisesti myös tasapuolisuus tutkijoiden kesken. Materiaali on saatavilla yhtä helposti Helsingissä, Oulussa tai Luxemburgissa. Tämä tulee ajankohtaisesti erityisesti sitten, kun tietokanta tulee täydelliseksi ja sinne koodataan myös julkaisematonta arkistomateriaalia. Tutkijan ei tarvitse välttämättä mennä SKS:ään tutustumaan materiaaliin, joskaan tietokanta ei voi täysin korvata alkuperäistä arkistomateriaalia.

Kun valtava runoaineisto on muutettu sähköiseen muotoon, on selvää, että sen käytettävyys helpottuu ja nopeutuu. Kullervo-runojen tutkijana saan yhdellä hakuehdolla näkyviin kaikki Kullervo-nimellä SKVR:n toimituksessa jaotellut runot. Tieto löytyy nopeasti. Haun suhteen merkittävää on se, että SKVR:n kieltä on normalistettu. Esimerkiksi saan hakusanoilla "kalev* poi* merkkijonot calevan poica, kalewan poika, kalevan poika ilman että sama haku olisi uusittava kirjaimia vaihtamalla (kuten joutuu tekemään Kalevala-instituutin Kalevala-tietokannassa).

Aineiston selaaminen on kätevää ilman että täytyy kahlata tuhansia sivuja läpi. Tietokannan yksi eduista liittyy aineiston järjestämiseen. Olen pystynyt suoraan tekstejä ja tietoja kopioimalla luomaan itselleni jonkinlaisen Kullervo-korpuksen. Myös yleiskuvan saaminen valtavasta aineistosta on nopeutunut. Tietokanta on laajentanut ja samalla muuttanut Kullervo-tietouttani. Esimerkiksi olen löytänyt Kullervon/ Kalevan pojan hahmon loitsuteksteistä, kontekstista, jossa en tiennyt sennimisen hahmon esiintyvän.

Käytettävyyteen kuuluu lisäksi tietokannan keveys. Pieni mutta merkittävä seikka. Nyt painavia SKVR-opuksia ei tarvitse nostella ja siirtää kirjastonhyllyltä työpyödylle.

Hakuja tehdessäni hankaluuksia aiheutti tiettyjen hakutoimintojen puuttuminen. Jos valitsin alueelliset kriteerit, mutta haun jälkeen halusin vaihtaa hakuehtoja, jäivät alueelliset ehdot voimaan. Kaipaisin hakuehtoihin sellaisia toimintoja kuin "tyhjennä hakukenttä/hakutulos", "säilytä hakukenttä/hakutulos", "aloita uusi haku", jotka antaisivat käyttäjälle mahdollisuuden valita, haluaako hän tallentaa, säilyttää hakuehtoja ja/tai tuloksia vai ei. Nyt toiminto "aloita alusta" tyhjentää automaattisesti niin edelliset hakuehdot kuin hakutuloksetkin. Lisäksi hakukenttää voisi monipuolistaa niin, että hakuja voisi tehdä halutessa vain yhden hakuehdon varassa, jolloin muut hakuehdot olisivat "jäissä".

Hakujen aikana on tullut esiin myös joitakin puutteita, jotka varmasti korjaantuvat, kun tietokanta tulee täydelliseksi. Esimerkiksi tietokanta ei suostu antamaan yhtään tulosta hakuehdolla "kaikki" perinnealueet.

Toinen seikka, josta Jukka Saarinen jo mainitsi tietokannan esittelykirjeessä, liittyy sanahakuun. Tiettyä sanaa ei ole korostettu tekstissä, vaan se täytyy hakea sieltä itse. Kokeillessani hakuja viime kesänä runoteksteissä oli erillisellä värillä alleviivattu haettu sana tai merkkijono. Varsinaisesta koeversiosta korostukset olivat poissa. Aluksi haku tuntui hankalalta ilman tähdennettävää tekstinkohtaa, mutta toisaalta silmä on jo tottunut löytämään melko vaivattomasti haettavan sanan. Yksittäisen sanan etsiminen runokorpuksesta saattaa olla hankalaa silloin, kun runoteksti on ennestään tuntematon.

Kalevala-instituutin ns. Kalevala-tietokantaan on luotu säehaun lisäksi erillinen sanahaku. Yksittäistä sanaa voidaan hakea korpuksesta melko tyhjentävästi. Hakuehto "kalev*" antaa listan kaikista sanoista, jotka alkavat kale-tavuisina. Sanojen kohdalla on ilmoitettu suluissa, kuinka monta kertaa sana esiintyy haetussa Kalevalan versiossa. Jokaisen sanan kohdalta voi myös hiirtä näpäyttämällä saada esiin runokontekstin. Tämäntapaista hakumuotoa toivoisin myös SKVR-tietokantaan.

Tietokantaan liittyvä yleisempi puute on eräänlainen kontekstittomuus. SKVR on digitoitu noudattaen tarkasti SKVR-julkaisussa käytettyjä tapoja, esim. erikoismerkkejä. Vaikka runojen yhteydestä löytyvät kontekstiedot keruusta sekä kerääjän mahdolliset lisähuomautukset, on aineistoa kirjaversioita vaikeampi hahmottaa pelkän luettelon avulla. Tietokannassa pääsee luettelosta yksittäiseen runotekstiin saakka, jolloin sivuruutu näyttää kokonaisuudessaan

runon sekä siihen liittyvät lisätiedot. Hakuja kokeillessani kaipasin laajempaa kontekstia, sitä mihin yksittäinen runoteksti SKVR:n kulloinkin kyseessä olevassa osassa sijoittuu. Samoin SKVR-antologian toimitukselliset erikoispiirteet ovat kadonneet. Esimerkiksi kunkin osan sisällysluettelo ja toimittajan alkusanat täytyy nyt etsiä kirjaversioista. Samoin jos ennen julkaisematonta arkistomateriaalia syötetään tietokantaan, olisi toivottavaa, että arkistointiin liittyvät piirteet säilyisivät myös sähköisessä muodossa, mikäli se on mahdollista.

Lopuksi, digitoitu SKVR toimii erinomaisesti kirjaversioiden rinnalla täydentämässä ja helpottamassa aineiston saatavuutta ja käsittelyä.

Liite 10 Valistus-digitointihanke

Kirsi Keravuori

Amerin Kulttuurisäätiön arkiston aarteet digitaaliseen muotoon

Amerin Kulttuurisäätiö ja Suomalaisen Kirjallisuuden Seura ovat vuoden 2003 alussa käynnistäneet hankkeen Amerin kulttuurisäätiön omistuksessa olevan lastenkirjataiteen digitoimiseksi. Hankkeen aikana kustannusyhtiö Valistuksen kokoelman kuvitusoriginaaleja digitoidaan korkealaatuiseksi kuvatiedostoiksi, digitoitujen kuvien luettelointitiedot tallennetaan kuvatietokannaksi ja kuvista muodostetaan toimiva kuvapankki. Tarkoitus on parantaa arvokkaan kuvitusaineiston järjestystä, käytettävyyttä ja tunnettavuutta sekä saattaa kulttuurihistoriallisesti arvokas kokoelma paremmin sekä tutkijoiden että yleisön käyttöön. Kulttuurisäätiö on esitellyt kuva-arkiston aarteita sekä järjestämällä itse näyttelyitä että lainaamalla töitä muiden näyttelyihin.

Lastenkirjataidetta Valistuksen kokoelmasta

Lasten- ja nuortenkirjojen, oppikirjojen ja lastenlehtien julkaisemiseen keskittyneen Osakeyhtiö Valistuksen kuva-arkisto sisältää suomalaisen lastenkirjataiteen parhaimmistoa. Kokoelman helmiin kuuluu noin tuhat Rudolf Koivun kuvitustyötä, joukossa Koivun kuvitukset Tuhannen ja yhden yön satuihin, Grimmin kauneimpiin satuihin sekä värikkäät kuvitukset Valistuksen joululehdissä ilmestyneisiin kertomuksiin ja runoihin. Koivun lisäksi Valistuksen kuva-arkistossa on yhteensä tuhansia kappaleita muiden merkittävien kuvitustaiteilijoiden töitä. Taiteilijoista tunnetuimpia lienevät mm. Venny Soldan-Brofeldt, Martta Wendelin, Risto Mäkinen, Helga Sjöstedt ja Maija Karma.

Valistuksen kokoelman kuvat ovat kirjojen ja lehtien sivuilta tulleet tutuiksi usealle sukupolvelle lukevia suomalaislapsia. Koululaisten käsissä kuluivat yli puoli vuosisataa opetuksen apuna olleet Valistuksen Lukukirja ja Isänmaallinen Lukukirja. Kustantamon bestsellereihin kuuluivat myös Lukemisto Suomen lapsille sekä yli 60 vuotta yhtäjaksoisesti ilmestyneet joululehdet Joulupukki ja Pieni Joulupukki. Joululehdet sisälsivät korkeatasoista kirjallista materiaalia. Lauri Pohjanpään joulurunot kuuluivat useita vuosia Joulupukin sisältöön, ja satuja lehteen kirjoittivat mm. Raul Roine, Anni Swan, Jorma Mäenpää ja Marjatta Kurenniemi. Myös joululehtien kuvitus oli erityisen tasokas, ja lehdet tarjoavatkin ainutlaatuisen läpileikkauksen suomalaisen joulun kuvamaailmaan.

Digitointihankkeeseen on kokoelmasta valittu mukaan koko säilynyt Joulupukki ja Pieni Joulupukki –lehtien kuvitusaineisto sekä kirjankuvitusaineiston parhaimmistoa. Kaikkiaan on tarkoitus digitoida noin 7000 originaalikuviusta korkealaatuiseksi, painokelpoiksi kuvatiedostoiksi. Lisäksi osana hanketta digitoidaan Joulupukin ja Pienen Joulupukin painetut numerot katselulaatuiseksi kuvatiedostoiksi. Täten osalle kuvitusaineistosta saadaan myös sen todellinen julkaisu-ympäristö digitaalisesti käytettäväksi. Hankkeen on määrä valmistua vuoden 2004 kuluessa.

Kokoelman digitointi

Digitaalisina kuvat ovat helpommin hallittavissa ja nopeammin toimitettavissa niitä haluaville. Digitointi edistää myös erittäin arvokkaan kokoelman säilymistä. Kuvien käyttö sähköisessä muodossa vähentää alkuperäisen aineiston käsittelyä ja kulumista. Aineistoa voidaan selata ja kuvia hakea ja saada käyttöön kokonaan koskematta alkuperäisiin kuviin.

Kuvituskuvat digitoidaan käyttämällä korkealaatuista A3-kokoista tasoskanneria. Digitointi tehdään suoraan alkuperäisestä kuvasta kalibroidussa ja profiloitussa työympäristössä. Kalibroinnissa ja profiloinnissa tietokoneen monitori ja skanneri säädetään mittalaitteiden ja värikarttojen avulla tuottamaan värit juuri oikean sävyisinä. Tuloksena saadaan erittäin korkealaatuisia kuvatiedostoja, joiden koko on keskimäärin 50 Mb. Kuvatiedostot tallennetaan siirrettäville kovalevyille ja varmistusnauhoille pakkaamattomina TIFF-muodossa. Suurista paljon tallennustilaa vievistä TIFF- kuvista tehdään pienemmät JPEG-katselukuvat, joita käytetään selailtaessa kuvia tietokannassa.

Kokoelmaan kuuluvia Rudolf Koivun kuvia on saatavissa osoitteesta
<http://www.finlit.fi/lehdisto/>

Liite 11 Sanakirjojen skannaus

Kotimaisten kielten tutkimuskeskus
Matti Uusivirta

SANAKIRJOJEN SKANNAUS

Raportti työstäni Kotimaisten kielten tutkimuskeskuksessa
12.11.- 31.12.2004

- I Johdanto
- II Huomioita ajasta
- III Testivaihe
- IV Työvaiheet skannatessa – eräs esimerkki ja kommentteja
 - 1. - 6. käynnistysvaiheet
 - 7. - 20. skannausvaiheet
 - 21. - 28. tallennus- ja tarkistusvaiheet
 - 28. - 30. varmuuskopiointi
- V Tallennusmuoto
- VI Asetukset

I Johdanto

Kun aloitin työni Kotimaisten kielten tutkimuskeskuksessa 12.11.2004, oli pääasiallinen tehtäväni työsopimuksen mukaan ”Karjalan kielen sanakirjan kolmen ensimmäisen osan muuntaminen sähköiseen muotoon”. Asiaa oli suunniteltu pitkään, mutta kellään Kotuksessa ei ollut kokemusta vastaavasta projektista, ei selvää käsitystä siitä, paljonko aikaa työ vaatisi, eikä liioin tarkkaa käsitystä työn yksityiskohdista. Niinpä sain varsin vapaasti harkita, opetella, kokeilla ja kehittää tehtävään sopivia työtapoja, Kotuksen vakinaisen henkilökunnan avulla ja yhteistyössä heidän kanssaan. Tietohuolto-osaston johtaja Elisa Stenvall antoi minulle työn, tukensa ja huomattavan vapaat kädet, ATK-osaston kaikki asiantuntijat auttoivat niin käytännön ongelmissa kuin tiedoillaankin aina kun apua pyysin (ja pyysinhan minä), Karjalan kielen sanakirjan tekijät ja lukuisat muut henkilöt uhrasivat kallista työaikaansa edistääkseen projektia niin paljon kuin mahdollista.

Varsinainen skannaus tapahtui SKS:n tiloissa Mariankadulla, missä Pasi Klemettinen ystävällisesti opasti minua laitteiden käytössä ja auttoi tarvittaessa. Testivaiheen jälkeen työ sujui varsin nopeasti ja ongelmitta.

Koska Karjalan kielen sanakirjan skannaus kuvatiedostoiksi vei aikaa vähemmän kuin oli odotettu, sain skannattavakseni myös Ordbok över svenska folk mål -sanakirjan vihkot 1-6. Lisäksi kokeilin skannattujen sivujen käsittelyä tekstinlukuohjelmalla ja yritin ratkoa sanakirjoissa käytettyjen erikoismerkkien aiheuttamia ongelmia.

II Huomioita ajasta

Päivässä sanakirjoja pystyi skannaamaan 300-360 sivua, ruotsin murteiden sanakirjaa parhaimmillaan yli 400 sivua.

Karjalan kielen sanakirjan skannaaminen oli hitaampaa, koska sivujen läpikuultavuuden vuoksi asetin skannattavien sivujen taakse mustat kartongit. Ruotsin murteiden sanakirjassa näitä ei tarvittu, koska paperi oli paksumpaa. Lisäksi vihkot taipuivat paremmin, asetelu oli helpompaa eikä rajauksia tarvinnut siirtää kuin harvoin. Myös hiukan pienempi rajaus nopeutti skannausta.

Huonoiten taipuva Karjalan kielen sanakirjan 3. osa oli selvästi muita hitaampi skannattava. Sitä joutui väänteleämään usein, sivut pyrkivät kääntymään vinoon, rajauksia oli siirrettävä, kirjan asentoa korjailtava ja kirjan päälle oli aseteltava ylimääräisiä painoja.

Pelkkä ruotsin murteiden sanakirjan yhden aukeaman skannaus vei nopeimmillaan aikaa kaksi minuuttia tai pari sekuntia alle. Karjalan kielen sanakirjassa vastaava aika oli noin kaksi ja puoli minuuttia, toisinaan selvästi enemmän. Lisäksi aikaa kuluu nimeämisiin, tallentamisiin, tarkistuksiin, mahdollisiin korjauksiin ja moniin muihin välttämättömiin työvaiheisiin, joten keskimääräinen tahti on selvästi hitaampi.

Koko prosessia ajatellen aikaa tarvitaan myös suunnitteluun, kokeiluun, neuvotteluihin ja paljon muuhun, joten nämä ajat ovat vain esimerkkejä. Sellaisenaan niitä ei voi soveltaa jonkin uuden projektin suunnitteluun. Jo kirjan koko, painatus, erilainen resoluutio tai mikä tahansa pieneltä tuntuva yksityiskohta voi vaikuttaa yllättävän paljon.

III Testivaihe

Koska sen enempää itselläni kuin kellään muullakaan Kotuksen työntekijällä ei ollut aiempaa kokemusta kokonaisten kirjojen digitoinnista ja jälkikäsitteystä tekstinlukuohjelmia käyttäen, oli ennen varsinaiseen urakkaan ryhtymistä kokeillen pyrittävä varmistamaan työn onnistuminen parhaalla mahdollisella tavalla.

Ennen SKS:n tiloihin siirtymistä tein pienimuotoisia kokeita Abby Fine Reader – tekstinlukuohjelmalla käyttäen materiaalina muutamia Kotuksen pienellä skannerilla digitoituja Karjalan kielen sanakirjan sivuja. Näin sain jonkinlaisen käsityksen kyseisen OCR-ohjelman toiminnasta ja tietoa siitä, mitkä asiat voisivat olla ongelmallisia ja mihin seikkoihin testeissä pitäisi kiinnittää erityistä huomiota.

Eräs erityisesti sanakirjojen digitointiin liittyvä ongelma nousi esiin jo miltei ensi vilkaisulla: erikoismerkit. Karjalan kielen sanakirjan leipätekstissä on 57 erilaista erikoismerkkiä, ja johdanto-osasta niitä löytyy vielä runsaasti lisää. Oli selvää, että sanakirjojen siirtäminen normaalien tekstinkäsittelyohjelmien ymmärtämään muotoon vaatisi paljon aikaa, ajattelua ja kokeilua. Digitointia kuvatiedostoiksi näiden ongelmien ratkaisemattomuus ei toki estäisi tai edes hidastaisi, mutta vaikutti lähtökohtiin: vaikka ohjelma ei (toistaiseksi) osannut lukea merkkejä eikä kyennyt korvaamaan niitä hyväksyttävällä tavalla, sen täytyi ehdottomasti kyetä erottamaan toisistaan toistensa kaltaiset merkit. Hattupää-äs piti tunnistaa eri merkiksi kuin äs, jonka ”hattu” onkin vain hieman erilainen kaari, eivätkä nämä saaneet sekoittua ässään, jonka päällä on sekä ”hattu” että kaari, tai ässään, jonka päällä on tästä vaivoin erottuva koukero – ja mieluiten tunnistuksen pitäisi onnistua silloinkin, kun kirjan painojälki on heikoimmillaan, paperissa ehkä jokin tummentuma, tai kirjain joltain osaltaan kulunut tai muuten vahingoittunut.

Jotta tarkka merkkien erottuminen edes periaatteessa olisi mahdollista, täytyi skannatun kuvan resoluution olla riittävän hyvä, ja mieluummin vielä hieman parempi. Mutta mikä resoluutio olisi ”riittävän hyvä”? Miten resoluution suurentaminen vaikuttaisi muihin asioihin? Mitkä olisivat mahdolliset raja-arvot, ja resoluutio lopulta juuri tämän kirjan kohdalla olisi ihanteellinen, kun otettaisiin huomioon kaikki mahdolliset tekijät niin pitkälle kuin mahdollista? Asia piti ehdottomasti testata, ettei myöhemmin 1864 sivun skannaamisen jälkeen tarvitsisi jonkun todeta, että väärin tehty, hukkaan meni sekin homma.

Tavoitteista ja ongelmista keskusteltiin sekä sanakirjan tekijöiden että ATK-asiantuntijoiden kanssa. Kun ensimmäistä kertaa tutustuin SKS:n laitteisiin, läsnä olivat Pasi Klemettisen lisäksi johtaja Stenvall ja ATK-osaston edustajat. Pienten alkuongelmien jälkeen laitteet saatiin viritettyä toimintakuntoon ja skannasimme muutamia näytesivuja erilaisilla resoluutioilla. Samalla todettiin myös, että resoluution huomattava suurentaminen selvästi hidasti skannausta. Näytteet kopioitiin CD-levylle ja niitä testattiin Kotuksessa tekstinlukuohjelmalla. Kokeiltavia asioita ja arvoja ehdottivat myös Pasi Klemettinen ja Kotuksen ATK-asiantuntijat. Ongelmia saattaisivat aiheuttaa esimerkiksi sellaiset tekijät kuin kirjan taitekohdan kaarevuus ja varjo, tekstin pieni vinous ja sivun toiselle puolelle painettujen tummien kohtien kuultaminen paperin läpi. Periaatteessa kaikki kuitenkin toimi. Keskustelut sanakirjan tekijöiden, ATK-ihmisten ja eräiden muidenkin kanssa johtivat siihen, että laadin tarkemman testisuunnitelman, jossa pyrittiin huomioimaan kaikki eri tahojen ja henkilöiden esittämät tarpeet, tavoitteet, näkökohdat, ehdotukset ja ongelmat.

Aloitin testaamisen selaamalla läpi kaikkia sanakirjan osia, pyrkimyksenä löytää tekstinlukuohjelman kannalta kaikkein vaikeimmat sivut ja yksityiskohdat. Yritin löytää sivuja, jotka olisivat vaikeita eri tavoin ja eri syistä, haaleita sivuja, kuvasivuja, pienimpiä mahdollisia merkkejä sisältäviä sivuja, niin monipuolisesti kuin suinkin. Sen jälkeen skannailin näitä sivuja erilaisilla resoluutioilla mutta samoilla SKS:n laitteilla ja ohjelmilla, joita lopullisessa työssäkin tulisin käyttämään. Samalla mittailin skannaamiseen kuluvia aikoja erilaisilla resoluutioilla yms. Koeotosta ei kannattanut paisuttaa järjettömän suureksi, mutta kaikki testaamisen arvoiset asiat yritin saada tavalla tai toisella mukaan. Ensin tarkastelin skannattuja sivuja yksin ja osin Pasi Klemettisen kanssa SKS:n laitteilla, minkä jälkeen kiinnostavista asioista otettiin lisänäytteitä, tarpeetonta aineistoa karsittiin ja merkittävämät sivut kopioitiin CD-levyille. Valitun testimateriaalin siirsin Kotukseen, missä lopullinen vertailu ja tarkastelu tehtiin Fine Reader – ohjelmalla.

Kokeilin ensin itse kaikkia testinäytteitä ja totesin mm. sen, että suurimmilla resoluutioilla otetut näytteet eivät enää toimineet OCR-ohjelmassa, koska ne veivät liikaa muistitilaa, ja yleensäkin resoluution suurentaminen tietyn rajan yli alkoi hidastaa toimintoja selvästi. Kovin pienillä resoluutioilla taas kuvan laatu heikkeni selvästi. Joissakin tapauksissa sivujen läpikuultavuus aiheutti häiritseviä tummentumia, mutta yllättäen keskittaitteen varsin suurikaan

taipuma ei näyttänyt häiritsevän ohjelmaa juuri lainkaan. Käytyäni eri asiat mahdollisimman huolellisesti läpi keskustelin tuloksista ATK-ihmisten ja sanakirjan laatijoiden kanssa, katsoimme yhdessä joitain näytteitä, ja kun uusia näkökohtia ei enää tullut esiin, oli johtopäätösten aika.

Resoluutioksi valittiin 500, koska sitä suuremman resoluution käyttö aiheutti jo muistiongelmia, OCR-ohjelma alkoi kangerrella, ja itse skannausvaiheessa toiminnat hidastuivat liikaa. Resoluutio 300 oli laadultaan selvästi huonompi; jotkut pienimmät merkkien osat alkoivat vääristyä muodoltaan niin paljon, että se ilmeisesti lisäisi ohjelman lukuvirheiden määrää. Resoluutio 400 olisi ehkä sinänsä riittänyt, mutta kun 500 oli parempi eikä vielä aiheuttanut liiallista hidastumista eikä muita ongelmia, se oli niin itseni kuin muidenkin mielestä varmin valinta.

Testien perusteella ratkaistiin myös muita asioita ja päädyttiin tiettyihin toimintamenetelmiin, joita en käsittele tässä sen enempää. Joitain olennaisia ja pienempiäkin käytännön seikkoja kommentoin seuraavassa skannauksen työvaiheita esittelevässä luvussa.

IV Työvaiheet skannatessa – eräs käytännön esimerkki ja kommentteja

1. Käynnistetään *skanneri*.
 - Käynnistettävä ehdottomasti ennen tietokonetta, muuten ei toimi!
2. Käynnistetään *tietokone*.
3. Valitaan *käyttöjärjestelmä*.
 - Käyttämäni SKS:n tietokone oli ohjelmoitu siirtymään automaattisesti Linuxiin, ellei jotain muuta käyttöjärjestelmää (Windows 1.) valittu hyvin nopeasti, joten tässä oli oltava tarkkana. Usein tämä vaihe puuttuu tai on erilainen.
4. Valitaan *kirjoittautumisalue*.
 - SKS:n tähän tietokoneeseen järjestelmänvalvoja oli määritellyt ulkopuolisille käyttäjille oman erikseen nimetyn kirjoittautumisalueen, joka salli vain tiettyjen ohjelmien käytön. Muualla voidaan toimia toisin.
5. Kirjoitetaan *käyttäjätunnus* ja *salasana*.
6. Käynnistetään ohjelma *XN View*.
7. Valitaan Tiedosto –valikosta vaihtoehto *Erä skannaus* (näin se oli kirjoitettu tässä kohden!).
 - XN View –ohjelmaan sisältyvä Eräskannaus –ominaisuus helpottaa ja nopeuttaa suurehkojen sivumäärien skannausta siten, että se ennakkoehtojen määrittämisen jälkeen (ks. seuraavat vaiheet) tallentaa, nimeää ja numeroi sivut automaattisesti oikeassa järjestyksessä. Kun sopiva erä materiaalia (esim. 50 sivua) on skannattu ja työvaihe keskeytetään, sivut näkyvät kuvaruudulla omina kuvakkeinaan. Hiiren avulla kuvakkeita valitsemalla tiedostoja (eli skannattuja sivuja) voi sitten käsitellä eri tavoin, katsella yksittäin tai sopivaan tahtiin etenevänä kuvasarjana tarkistuksia varten, nimetä uudestaan yms.
8. Määritetään Eräskannaus –ikkunasta *Kohdehakemisto*, esim. F:\Kuvasiirto.
 - Määrää paikan, mihin skannatut kuvat tallentuvat.
 - Oletushakemistoksi ohjelma määrittää automaattisesti sen, mitä viimeksi on käytetty, joten saman työn jatkuessa vaikka viikosta toiseen ei tähän asetukseen tarvitse puuttua muuta kuin silloin, jos asetusta väillä on muutettu.
9. Määritetään Eräskannaus –ikkunasta *Tiedoston nimi*, esim. KKS302.
 - Ohjelma antaa oletusnimeksi aina viimeksi käytetyn nimen, joten uuteen skannaukseen siirryttäessä nimeä kannattaa muuttaa, esimerkiksi nimen viimeinen numero. Jos nimeä ja aloitusindeksiä (vaihe 10.) ei muuta, uudet skannattavat sivut (tiedostot) tallentuvat edellisten, samannimisten päälle, ja edelliset häviävät (ellei niitä ole erikseen tallennettu kansioon)!
 - XN View –ohjelma ei huomauta eikä varoita päälle tallentamisen vaarasta! Myöhemmässä vaiheessa, kansioihin tallennettaessa ja niitä käsiteltäessä Windows-käyttöjärjestelmä kyllä varoittaa päälle tallentamisesta, pyrkii estämään sen ja kysyy ”korvataanko?” tms., mutta XN View –ohjelman sisällä toimittaessa nämä varoitukset ja estot eivät jostain syystä toimi.

10. Määritetään Eräskannaus –ikkunasta *Aloitusedeksi*, esim. 401.
 - Aloitedeksi määrittää ensimmäisenä skannattavan sivun tiedostonimen loppuosan, jonka ohjelma automaattisesti liittää edellisessä kohdassa määritetyn Tiedoston nimen perään. Jos Tiedoston nimi on KKS302 ja Aloitedeksi 401, tulee ensimmäisen skannattavan sivun nimeksi KKS302401, seuraavan sivun nimeksi KKS302402 jne., edellyttäen että ikkunan seuraavassa kohdassa valitaan vaihtoehto Kaikki sivut.
 - Aloitedeksin luvussa ei voi olla nolliä edessä. Esim. Aloitedeksi 001 ei kelpaa, vaan ohjelma korvaa sen luvulla 1. Jos halutaan käyttää järjestelmää, jossa tiedostonimet ovat jatkuva sarja esim. muotoa KKS101001 – KKS113598 (eikä KKS1011 – KKS113598), on sarjan alkuvaiheessa joko lisättävä nollat etukäteen edellisessä kohdassa Tiedoston nimi (ensin KKS10100, sitten KKS1010, sitten KKS101), tai ”väävät” nimet on muutettava jälkikäteen. Itse käytin molempia tapoja.
 - Aloitedeksi kannattaa määrittää niin, että lukema (tiedostonimen loppuosa) vastaa teoksen sivunumeroa. Alussa nimiä joutuu muuttamaan puuttuvien nollien vuoksi ja toisaalta siksi, että painetun kirjan alussa on usein tyhjiä sivuja, joita ei skannata, mutta myöhemmissä vaiheissa työ on huomattavasti helpompaa, jos tiedostonimet ja sivunumerot vastaavat toisiaan.
 - Aloitedeksin oletusarvoksi ohjelma antaa aina edellisellä kerralla tässä ikkunassa määritetyn aloitedeksin. *Tämä aloitedeksi on ehdottomasti aina tähän ikkunaan palattaessa muutettava!*
 - Jos Aloitedeksiä ei muuta, ohjelma skannausta jatkettaessa koneen tyyneydellä tallentaa uudet sivut edellisten päälle, täysin rauhallisesti, mitään varoittamatta, mistään huomauttamatta, ja hävittää alta kaikki ne samalla tiedostonimen loppuosalla varustetut sivut, joita ei *erikseen* tätä ennen ole tallennettu erilliseen kansioon.
 - Itse opin ymmärtämään XN View –ohjelman tämän ominaisuuden ”kantapään kautta” onneksi jo työni alkuvaiheessa, skannatessani ohjelman avulla testisivuja. Katsottuani välillä jo tehtyjä sivuja aloin skannata lisää, ja tarkistaessani lopputulosta totesin kaikkien aiempien sivujen hävinneen lopullisesti. Onneksi niitä ei ollut paljoa, eikä tappio ajallisesti laskettuna ollut suuri. Jos olisin törmännyt asiaan myöhemmässä vaiheessa, olisin voinut joutua tekemään isommankin työn kahteen kertaan.
11. Valitaan Eräskannaus –ikkunasta vaihtoehto *Kaikki sivut*.
 - Tämä on oletusarvo, jota ei tarvitse muuttaa. En tutkinut muita vaihtoehtoja.
12. Valitaan Eräskannaus –ikkunasta *Formaatti* jokin tallennusmuoto.
 - Oma valintani oli: TIF – TIFF Revision 6. Vaihtoehtoja on useita. Perustelut valinnalle ovat edellä luvussa Tallennusmuoto.
 - Kerran valittu tallennusmuoto säilyy muistissa oletusvaihtoehtona koneen sammuttamisen jälkeenkin, joten myöhemmin tämän kohdan voi ohittaa, ellei asetusta ole välillä muutettu.
13. Asetetaan *kirja skanneriin* (ellei tätä ole tehty sopivassa välissä aiemmin) varmana siitä, että kyseessä on oikea aukeama ja oikea sivu, ja että sivunumero täsmää sekä aiemmin tallennetun aineiston että Tiedostonimien ja Alaindeksien kanssa.
 - Skannerin lasin puhtaus on syytä tarkistaa aina työtä aloitettaessa ja aika ajoittain työn kestäessä. Tarvittaessa puhdistetaan varovasti puhaltamalla, käyttäen pehmeää liinaa tai muuta sopivaa välinettä niin, että mahdolliset roskat tai pölyhiukkaset eivät hankaa tai naarmuta pintaa.
 - Skannattavan aukeaman sivujen reunat on asennettava tarkasti skannerin lasin reunojen mukaisesti, esim. lasin vasemman yläkulmaan. Yli kahden millimetrin heitto alkaa jo olla liikaa.
 - Toisinaan kirjan sidos pyrkii vääntämään sivuja vinoon niin, että vierekkäiset sivut (ja tekstirivit) ovat keskenään erisuuntaisia. Itse jouduin toisinaan käyttämään melko rajuakin väkivaltaa saadakseni sivut samansuuntaisiksi. Jos skannattavana on harvinainen, hyvin vanha tai muuten arvokas alkuperäiskappale, on menettelytavoista (ja yleensäkin teoksen käsittelystä) etukäteen sovittava teoksen säilytyksestä vastaavien henkilöiden kanssa. Vinottain taipuvat sivut on mahdollista skannata yksi kerrallaan niin, että käännetään vuoroin toinen ja vuoroin toinen sivu suoraan. Tämä hidastaa työtä, joten vähemmän arvokkaita teoksia kannattaa tarpeen vaatiessa käsitellä kovakouraisemmin. Teoksen arvo ja työhön käytetyn ajan arvo on pyrittävä suhteuttamaan toisiinsa järkevällä tavalla.
 - Jos skannattava kirjajaksilö voidaan valitaan useiden eri tavoin sidottujen tai eri kuntoisten joukosta, kannattaa valita halvin sellainen kappale, jossa tallennettava teksti- ja kuvainformaatio on vahingoittumaton (ei kulumia, repeämiä, ylimääräisiä merkintöjä, läikkiä tms.), täydellinen (mitään ei puutu) ja mahdollisimman korkealaatuinen. Pehmeäkantainen laitos voi olla parempi kuin kovakantainen, ja ohuiksi vihkoiksi sidottu laitos parempi kuin paksu kirja. Hyvä taittumisen on tärkeää. On hyvä, jos teos kestää ehjänä jokaisen aukeaman tiukan painamisen skannerin pintaa vasten, joten esimerkiksi heikkolaatuinen liimasidonta voi tuottaa ongelmia. Sivujen irtoaminen hidastaa työtä, vaatii ylimääräistä asettelua ja lisää virheiden todennäköisyyttä (jokin sivu puuttuu, jokin tulee kahteen kertaan, järjestys muuttuu).
 - Huonosti taittuva kirja tai kirjan kohta vaatii lujaa puristusta skannerin pintaa vasten, hyvin taittuvan voi usein asettaa pinnalle sellaisenaan, ilman mitään ylimääräistä painoa. Kokeilemalla selviää, tarvitaanko puristusta vai

ei. Prosessin alussa painolaadultaan heikoimmista ja mahdollisimman ”vaikeista” sivuista on syytä tehdä koeskannauksia vaihtoehtoisilla menetelmillä, asetuksilla ja työtavoilla, ja tarkistaa testisivujen toimivuus myöhemmissä vaiheissa käytettävien (tekstinluku-, kuvankäsittely- ym.) ohjelmien kanssa. Kun tiedetään, missä huonoimman sallitun jäljen rajat menevät ja säilytetään sopiva turvamarginaali, riittää työn aikana yleensä pelkkä silmämääräinen skannausjäljen tarkistus. Mahdolliset ”riskisivut” oppii nopeasti näkemään jo esikatselu- (Prewiev-) ikkunan kuvasta, mutta aika ajoon kannattaa tarkastella myös valmiiksi skannattuja sivuja suuremmissa koossa.

- Kirjan puristaminen yleensä siirtää hivenen ainakin aukeaman toista sivua. Tämä vaatii ylimääräistä asetelua tai skannausalan muuntelua, mikä hidastaa työtä. Karjalan kielen sanakirjan kolmas osa oli muita jäykemmin sidottu ja vaati puristusta lähes joka aukeamalla, paremmin taittuvat ensimmäinen ja toinen osa vain aika ajoon. Yleensä käytin painona teoksen muita osia; vain pari jäykintä kohtaa vaati suurempaa puristusta. Ohuemmiksi ”vihkoiksi” sidottu ruotsin murteiden sanakirja taittui niin hyvin, että lisäpainoa tarvittiin lähinnä vain ensimmäisten ja viimeisten sivujen kohdalla. Tämä nopeutti työtä selvästi.
 - Skannerin kantta en kääntänyt alas skannattavan kirjan päälle missään vaiheessa. Tässä tapauksessa se olisi ollut tarpeetonta, hidastanut työtä. Kannen kääntäminen alas ja taas ylös lisää prosessiin kaksi ylimääräistä työvaihetta, minkä lisäksi kannen puristus usein siirtää kohdistusta kannen alla, piilossa, mikä vaatii taas ylimääräistä skannausalueen siirtelyä ja muuntamista.
 - Jos kirjan paperi on ohutta, lehden toisen puolen painatus voi kuultaa läpi niin, että se näkyy skannatulla sivulla tummina läikkinä. Nämä läikät saattavat häiritä tekstinlukuohjelman toimintaa ja kuvankäsittelyohjelmia käytettäessä ne voivat heikentää lopputulosta hyvinkin pahasti. Karjalan kielen sanakirjan joissain testiotoksissa läpikuultavuus tuntui niin häiritsevältä, että päätin (Pasi Klemettisen neuvosta) käyttää skannattavan sivun takana aina mustaa kartonkia. Se keino auttoi. Työtä kartonkien asetteleminen hidasti jonkin verran, mutta toisaalta ei tarvinnut pelätä, että osa työstä olisi tehtävä ehkä myöhemmin uudestaan tästä syystä. Jokaisen sivun kohdalla kartonkeja ei välttämättä olisi tarvittu, mutta sama käytäntö läpi koko tekstin piti laadun tasaisena. Erillinen harkinta kunkin sivun kohdalla olisi vienyt enemmän aikaa ja nopeaan rutiiniin pyrittäessä oikeaan ratkaisuun päätyminen olisi toisinaan voinut olla varsin sattumanvaraista. Ruotsin murteiden sanakirjan paperi oli paksumpaa eikä kartonkeja tarvittu.
14. Klikataan hiirellä: *Scan*. Skannausikkuna avautuu.
- Näkyviin tulevat myös skannausasetukset, joihin ei tarvitse puuttua sen jälkeen, kun ne on testivaiheessa säädetty (ellei niitä ole välillä muutettu). Asetuksista tärkeimmät (useimmin muutettavat) ovat kuvatyyppi (Image Type: 8-bit Grayscale), resoluutio (500 dpi) ja kuvakoko (Target Size: Original). Asetuksista lisää edellisissä luvuissa Testivaihe ja Asetukset.
15. Klikataan hiirellä: *Preview*. Skannattava kirjan aukeama ilmestyy esikatseluikkunaan.
- Esikatseluikkunassa rajataan kirjan aukeamasta ne osat, jotka halutaan skannata. Jos rajausta ei tehdä, kuvatiedostoon tulee koko skannauspinta, josta kirja peittää vain pienen osan.
 - Rajaus piirretään hiirellä siten, että ensin klikataan haluttu vasemman yläkulman piste ja sitten ”venytetään” vastakkaisen alakulman pistettä niin, että muodostuva suorakulmio kehystää skannattavan alueen. Näin piirrettyä skannausaluetta voi siirtää yhdenmuotoisena paikasta toiseen tai sen muotoa ja kokoa voi muuttaa siirtämällä haluttuja reunaviivoja erikseen.
 - Skannausalueita voi olla useita. Kirjan aukeaman voi skannata yhtenä kokonaisuutena tai sivu kerrallaan piirtämällä kummallekin sivulle oman rajauksensa. Sivun osia (kuvia, palstoja tai haluttuja yksityiskohtia) voi myös skannata erikseen.
 - Itse rajasin aukeaman molemmat sivut erikseen. Tästä on seuraavia etuja:
 - Kunkin sivun voi nimetä omana tiedostonaan, jonka numerointi (tiedostonimen loppuosa) vastaa sivun numeroa. Näin oikea sivu löytyy kansioista nopeasti, mikä helpottaa sekä tarkistustyötä että myöhempää jälkikäsittelyä muilla ohjelmilla.
 - Kirjoja käsiteltäessä yksi sivu on luonteva ja perinteinen yksikkö; kirjoissa ei numeroida aukeamia, vaan sivut.
 - Samalla resoluutiolla yhden sivun tiedosto on kooltaan alle puolet koko aukeaman tiedostosta. Kaikessa myöhemmässä tiedostojen käsittelyssä tällä on suuri merkitys. Testatessani koko aukeaman tiedostoja tekstinlukuohjelmalla sain koneen pari kertaa tiltaamaan tai kieltäytymään avaamasta tiedostoa. Ilmeinen syy oli se, että yhtä aikaa avattujen tai käsiteltävien tiedostojen koko ylitti koneen tai ohjelmien käsittelykyvyn. Resoluution nostaminen yhden sivun kokoisilla tiedostoilla taas johti toimintojen voimakkaaseen hidastumiseen. Jos tai kun käytetään nyt käytössä olevia laitteita ja ohjelmia, tuntuu järkevältä pysytellä niiden hallitseman kapasiteetin rajoissa, huolimatta siitä, että sekä tietokoneiden että ohjelmien kapasiteetti edellisinä vuosina on kasvanut hyvin nopeasti. Vaikka itsellä olisikin tarkoitus ostaa uusimmat laitteet ja ohjelmat sitä mukaa kuin niitä kehitetään, ei järkevasti (ja aina epävarmoilla budjeteilla) toimiva kulttuurilaitos voi lähteä siitä, mikä on paras mahdollinen ennustettava tulevaisuus, vaan siitä, mikä nykyhetkellä käytettävillä resursseilla nyt tehtynä antaa

parhaan mahdollisen tuloksen, riippumatta siitä, ovatko resurssit jatkossa huomattavasti paremmat, samat, vai huonommat kuin nyt (tämä on oma mielipiteeni, ei totuus, mutta näkökulmana ehkä huomion arvoinen).

- Koko aukeama kerrallaan voi olla järkevää skannata sellaiset teokset, joissa on sidontalinjan ylittäviä kahdelle sivulle ulottuvia kuvia, joissa aukeaman taitto muodostaa yhtenä kokonaisuutena käsiteltävän taiteellisen kokonaisuuden tai joissa on sellaisia kuviin kiinteästi liittyviä merkkejä, tekstejä tai muita selittäviä elementtejä, joiden käsittely yhtenä kokonaisuutena on ymmärtämisen tai jatkotyöskentelyn kannalta tarpeen. Jos kyse on vain taitekohdan kahden puolen ulottuvista kuvista, saattaa tällöinkin olla järkevää skannata eri sivujen kuvaosat erikseen sen vuoksi, että painotuotteissa katkoksen eri puolten kohdistuksessa usein on epäjatkuvuutta ja muita virheitä. Erikseen skannattuja kuvanosia yhdistäen hyvällä kuvankäsittelyohjelmalla on joskus mahdollista saavuttaa alkuperäistä painotuotetta parempi kohdistus. Tällaiset tapaukset on syytä harkita ja kokeilla huolellisesti ennen kuin skannataan hyvin laajoja ongelmallisia opuksia. Yksittäisten aukeamien tapauksissa voidaan niistä skannata useita versioita eri tavoin ja muuta työtä hidastamatta siirtää niihin liittyvien ongelmien ratkaisu myöhempiin aikoihin.
 - Rajaukset säilyvät ohjelman muistissa (myös koneen sammuttamisen jälkeen), joten skannausta jatkettaessa niitä ei periaatteessa tarvitse tehdä uudestaan. Jos teksti- tai kuva-ala eri aukeamilla on vakio ja kirjan asettelu tehdään riittävän tarkasti aina samoin, ei korjauksia tarvita. Kirjan taipuminen kuitenkin usein hieman vaihtelee ja painoala suhteessa sivun leikkausreunaan voi siirtyä, joten rajausta on jatkuvasti seurattava ja silloin tällöin muutettava.
 - Rajaus on tehtävä niin, että kaikki teksti, sivunumerot ja mahdolliset muut leipätekstialueen ulkopuolelle sijoittuvat merkit ja kuvanosat kokonaisuudessaan tulevat mukaan ja ympärille jää kaikkiin suuntiin riittävä, ainakin muutaman millimetrin suuruinen marginaali.
 - Jos marginaali on kovin pieni, joudutaan rajausta siirtämään usein, ja työ hidastuu sen vuoksi. Jos marginaali on turhan suuri, tiedostokoko kasvaa tarpeettomasti ja työ hidastuu siksi, että kone käyttää aikaa kunkin sivun tyhjien alueiden skannaamiseen.
 - Tapauksesta riippuen noin puolen sentin perusmarginaali riittää hyvin, se sallii parin kolmen millimetrin siirtymät eri sivujen kohdalla ilman, että rajausta tarvitsee siirtää. Siirto on tarpeen, kun marginaali jossain kohden jää pienemmäksi kuin noin kaksi millimetriä. Eri kirjojen kohdalla ihanteellinen marginaali voi vaihdella, mutta työtä tekemällä sen löytää. Alussa voi ottaa hieman varman päälle ja sitten kokemuksen karttuessa marginaaleja voi kaventaa, mikäli vaihtelu eri aukeamien välillä on vähäistä.
16. Valitaan aukeaman se *sivu* (tai muu rajattu alue), joka skannataan ensin, klikkaamalla hiirellä esikatseluikkunasta sivun kuvaa rajatun alueen sisäpuolella. Valitun alueen rajaviiva muuttuu katkoviivaksi, joka prosessin eri vaiheissa voi myös välkkyä tai ”kiertyä”.
- Ohjelman valitsema skannausalue on aina se, jota viimeksi on käsitelty (esimerkiksi rajausta siirtämällä) tai käytetty. Skannauksen jälkeen ohjelma ei itse siirry toiselle alueelle (seuraavalle sivulle), joten *yleensä skannattava sivu on joka kerta valittava erikseen*. Muuten voi käydä niin, että aukeamalta skannataan aina ensin väärä sivu (jolloin tiedostot tulevat väärään järjestykseen) ja ellei uutta valintaa tehdä tämänkään jälkeen, skannataan jokaiselta aukeamalta joka toinen sivu kahteen kertaan joka toisen jäädessä kokonaan väliin.
17. Klikataan hiirellä: *Scan*. Valittu sivu (alue) skannataan.
- Skannauksen aikana esikatseluikkuna säilyy näkyvissä, ja skannattavan alueen katkoviivaraja tekee kiertoliikettä.
 - Skannauksen aikana näkyviin tulee uusi ikkuna, josta klikkaamalla skannauksen voi keskeyttää. Tämä mahdollistaa sen, että skannauksen voi aloittaa heti skannausalueen valitsemisen (kohta 16.) jälkeen tuhlaamatta aikaa sekunnin osia enempiä esikatselukuvan tarkasteluun. Tarkemman katselun voi tehdä skannauksen aikana, ja mikäli esikatseluvassa tällöin on havaittavissa jotain vikaa, voi klikata hiirellä keskeytyspainiketta, korjata virheen ja aloittaa skannauksen uudestaan. Kun työssä on saavuttanut sellaisen rutiinin, että virheitä (rajaus, sivun vinous, liian laajalle ulottuva sidoslinjan varjo, heikko erottumien tms.) on vain harvoin, tämä toimintatapa nopeuttaa prosessia huomattavasti, verrattuna siihen, että esikatseluun joka aukeaman kohdalla käytettäisiin aikaa sekuntikaupalla.
18. Skannauksen valmistuttua valitaan koko ajan auki olevasta esikatseluikkunasta hiirellä *toinen sivu* (alue).
19. Klikataan hiirellä: *Scan*. Aukeaman toinen sivu (alue) skannataan.
20. Skannauksen valmistuttua käännetään kirjasta *seuraava aukeama*, asetetaan kirja skanneriin ja toistetaan vaiheita 13.-19. tarpeellisin osin kunnes sopiva erä sivuja on skannattu.
- Sopiva erä käyttämälläni laitteilla, ohjelmilla ja asetuksilla (resoluutio, tiedostokoko) sanakirjojen skannauksessa näytti olevan noin 50 sivua.

- Jos erä on liian suuri, eivät tiedostojen (sivujen) kuvakkeet mahdu kerrallaan kuvaruudulle ja niiden käsittely (tarkistus, tallentaminen, nimien muuttaminen jne.) vaikeutuu.
 - Jokainen skannauserä on erikseen tallennettava (ks. vaihe 22.) omaan erikseen määriteltävään paikkaansa (esim. Windows-kansioon) ennen kuin skannausta eräskannausohjelman keskeytyksen jälkeen voi jatkaa (paluu vaiheeseen 7.), tai ainakin aloitusindeksi (vaihe 10.) on muutettava. *Jos tallennus tai aloitusindeksin muuttaminen laiminlyödään, tallentuvat uudet skannattavat sivut edellisten päälle ja aiemmin skannatut (erikseen tallentamattomat) sivut menetetään.*
 - Koska jokainen skannauserä on käytännön vuoksi järkevintä tallentaa omana kokonaisuutenaan omaan kansioonsa (muuten sivuja on siirrettävä ja nimettävä hankalasti yksi kerrallaan), on otettava huomioon myös tiedostojen tuleva käyttö (ohjelmat, tallennusformaatit, tiedonsiirto) niin pitkälle kuin se on mahdollista. Jos esimerkiksi on ilmeistä tai todennäköistä, että jossain vaiheessa tiedostojen siirtoon tai käsittelyyn käytetään CD-levyjä, on yhden skannauserän (tiedostokansion) hyvä mahtua yhdelle CD-levylle. Jos CD-levyn kapasiteetti on 700 MB, jokin tietty sivu vie 13,54 MB (eräs todellinen arvo käyttämälläni asetuksilla, lukema näkyy esikatseluikkunassa) ja pienistä rajojen siirtelyistä huolimatta sivut keskimäärin vaativat tilaa varmasti alle 14 MB/kpl, voidaan helposti päässä laskea, että yhdelle levylle mahtuu noin 50 sivua. Jos sama sivun kokoinen tiedosto leveämpiä marginaaleja käyttäen vaatii tilaa 15, 64 MB ja keskiarvo on samaa luokkaa, ei yhteen skannauserään kannata mahdollistaa kuin nelisen kymmentä sivua, ei kovin paljoa yli (jos halutaan käyttää tallennuskapasiteetti viimeistä rajaa myöten, on syytä laskea ja tarkistaa täsmälliset lukemat).
 - Liian suuria skannauseriä kannattaa välttää senkin vuoksi, että yksi erikseen tallentamaton skannauserä on käytännössä suurin yksikkö informaatiota, mikä pienen laitehäiriön, sähkökatkoksen tai oman virheen vuoksi voidaan hetkessä menettää.
 - Jos skannauserää kasvatetaan hyvin suureksi, rajoittaviksi tekijöiksi tulevat jossain vaiheessa myös ohjelmien ja laitteiden kapasiteetti, esimerkiksi käyttömuistin koko. Pahimmassa tapauksessa kone voi juuttua (tiltata) ja sitä uudelleen käynnistettäessä saatetaan menettää tallentamaton skannauserä kokonaisuudessaan.
 - Jos skannauserät tallennetaan omiin kansioihinsa, mikä yleensä lienee käytännöllisin ja järkevin ratkaisu, kannattaa ottaa huomioon myös se, millaisia laitteita ja ohjelmia myöhempään tekstin lukuun ja jälkikäsitteilyyn huonoimmassa mahdollisessa tilanteessa saatetaan käyttää. Mikäli on mahdollista, että materiaalia tullaan käsittelemään myös vanhoilla tai muuten vähäkapasiteetisilla ohjelmilla tai laitteilla, on järkevää ottaa huomioon näiden asettamat rajoitukset. Vaikka ylimääräiset työvaiheet (kansioiden pilkkominen, tiedostojen uudelleentallennus toisessa formaatissa tai pienemmällä resoluutiolla) usein auttavat ylittämään tällaiset ongelmat, ne samalla vievät työaikaa ja lisäävät kustannuksia.
 - Kun teoksen tietyt osat (johdanto, erilliset luvut, kuvallitteet tms.) halutaan tallentaa omiin kansioihinsa, kannattaa sopivaan jakoon pyrittäessä ehkä käyttää mieluummin keskimääräistä pienempiä kuin ylisuuria skannauseriä.
 - Jos työ keskeytyy mistä tahansa syystä niin, että skannausohjelma suljetaan ja palataan johonkin aiempaan vaiheeseen, on usein järkevää samalla tallentaa skannattu erä omaan kansioonsa (vaihe 22), vaikka se olisi pienikin. Näin varmistetaan jo tehdyn työn säilyminen. Ylimääräistä aikaa ei kulu, koska edelliset vaiheet kohdasta 7. alkaen (tarpeellisin osin) on joka tapauksessa käytävä läpi ja myös tallennus on jossain vaiheessa tehtävä.
 - Sivun 99 on järkevä paikka lopettaa skannauserä, koska seuraavan skannauerän aloitusindeksiksi tulee kolminumeroinen luku ja tarve muuttaa tiedostonimiä tältä osin vähenee (ks. vaihe 10).
 - Turhan pienien skannauserien käytöstä ei kovalevylle tallennettaessa koidu muuta haittaa kuin vaiheiden 7.-14. tarpeettoman usein tehdystä osittaisesta toistamisesta johtuva ajanhukka. Mikäli tiedostot jossain vaiheessa siirretään esimerkiksi CD-levylle, voi pienien (tai muuten ”väärän” kokoisten) kansioiden käyttö aiheuttaa tallennuskapasiteetin vajaakäyttöä (levyjä tarvitaan enemmän ja kustannukset nousevat) ja/tai ylimääräistä työtä sekä ajanhukkaa tiedostojen siirtovaiheessa (mikä myös lisää kustannuksia).
21. Skannauserän tultua täyteen *skannausohjelma* suljetaan hiirellä klikkaamalla. Skannaus- ja esikatseluikkunat häviävät tiedostopuun päältä ja näkyviin tulevat skannattujen sivujen (tiedostojen) kuvakkeet.
- Mikäli oikeat kuvakkeet eivät näy, ollaan väärässä kohden tiedostopuuta (esim. jonkin aiemmin tallennetun kansion sisällä), ja kuvakkeet saadaan näkyviin klikkaamalla tiedostopuussa oikeaa kohtaa (jonka nimi voi olla esimerkiksi Eräskannaus.)
22. Tallennetaan skannattu erä omaan *kansioonsa* (samalla tavoin kuin missä tahansa Windows-ohjelmassa; näiden tapojen tai muissa käyttöjärjestelmissä noudatettavien periaatteiden opettaminen ei mahdu tähän raporttiin).
- Jos kansiota ei vielä ole, luodaan ja nimetään tässä vaiheessa uusi kansio noudattaen normaalia Windows-käytäntöä (muissa käyttöjärjestelmissä niiden käytäntöjä).
 - Seuraava tarkistusvaihe (23.) on mahdollista tehdä jo ennen kansioon tallennusta, samoin tiedostojen uudelleen nimeämiset, mutta mitään hyötyä toisessa järjestyksessä toimimisesta en näe. Jos tarkistuksia ja mahdollisia korjauksia tehdään ennen kansioon tallentamista, voi pienikin virhe korjaustyössä (esimerkiksi

joidenkin tiedostojen uudelleen nimeämisen unohtaminen) aiheuttaa uusia virheitä ja uusia ongelmia, joiden selvittämiseen kuluu aikaa, vaivaa ja rahaa.

- Itse nimesin kansiot (kokeilujen jälkeen) niin, että alun kirjainosa (KKS) on teoksen nimen lyhenne, seuraava numero (2) kertoo teoksen osan ja sitä seuraavat numerot nimeävät kansion (11). Tässä systeemissä kansioden sisällä olevien tiedostojen nimiin tulee vielä kolminumeroinen loppuosa, joka on sama kuin vastaava kirjan sivunumero (534). Tässä kansion nimi olisi siis KKS211 ja koko tiedostonimi KKS211534. Sanakirjoissa johdannot oli sivunumeroitu erikseen roomalaisin numeroin; näissä käytin johdannon tunnuksena kirjainta J, muutin sivunumerot arabialaisiksi, ja kun näitä kansioita tuli alle kymmenen, saattoi kansion nimenä olla esimerkiksi KKSJ2 ja tiedostonimenä KKSJ2064.
 - Koska ohjelma ei hyväksy aloitusindeksin alkuun nollia (ks. vaihe 10) ja koska kirjan alussa nimilehden jälkeen yleensä on ainakin yksi tyhjä sivu, jota ei skannata, on sivuvastaavuuden ja logiikan säilyttämiseksi osa tiedostonimistä muutettava (sivut 1-9 ja 10-99 sekä puuttuvista sivuista aiheutuvat korjaukset).
23. *Tarkistusta* varten tallentamisen jälkeen kansio avataan. Tiedostojen kuvakkeet tulevat näkyviin.
24. Valitaan hiirellä kuvakkeita klikkaamalla ne sivut (tiedostot), joita halutaan tarkastella lähemmin, suuressa koossa, niin että nähdään *lopullinen tallenne*.
25. Kun sivut on valittu, klikataan hiirellä *Kuvaesitys*-kuvaketta. Avautuu ikkuna, jonka vaihtoehtoista valitaan sopivat. Sivuja (kuvatiedostoja) voi katsella sarjana niin, että kuvat vaihtuvat automaattisesti esimerkiksi kolmen sekunnin välein (ajan voi vapaasti itse asettaa) tai manuaalisesti hiirellä klikkaamalla. Erilaisia valintamahdollisuuksia on paljon, mutta kaikissa on oletusarvot eikä itselle yhäntekeviin asioihin tarvitse puuttua. Tehdyt valinnat säilyvät muistissa oletusarvoina seuraavaa katselukertaa varten.
- Itse tein alkuvaiheessa tarkistuksia enemmän; myöhemmin valitsin ilmeisen ongelmatomasta skannuserästä lähempää tarkastelua varten yleensä neljä tai viisi sivua satunnaisotannalla alusta, keskivaiheilta ja aina ehdottomasti lopusta. Tärkeintä on varmistaa, että sivunumerot ja tiedostonimien loppuosat täsmäävät.
 - Jos erän lopussa sivunumero ja tiedostonimen loppuosa eivät täsmää, on virheen (esimerkiksi puuttuvan sivun) kohta etsittävä haarukoimalla viimeisen "oikean" ja ensimmäisen "väärän" sivun väliä. Jos jokin virhe on tapahtunut, löydetty ja korjattu, on korjauksen jälkeen syytä vielä tavallista huolellisemmin uudestaan tarkistaa pitempi yhtenäinen sivujatkumo sen varmistamiseksi, että kaikki on oikein.
 - Satunnaisotanta riittää varmistamaan sen, ettei sivuja lukumääräisesti puutu (ei ole harpattu jonkun sivun tai aukeaman yli korvaamalla sitä toisella). Satunnaisotanta ei riitä paljastamaan sellaista virhettä, että jokin sivu on skannattu kahteen kertaan ja esimerkiksi saman aukeaman toinen sivu on jäänyt väliin.
 - Sivunumeroiden ja tiedostonimien lisäksi on tarkistettava kuvan laatu (tarkkuus, heikoimpien/vaikeimpien merkkien erottuvuus, riittävä kontrasti, tekstin suoruus, ongelmat kirjan selän kohdalla, rajaus, ei roskien, lian tai sivun läpikuultavuuden aiheuttamia häiriöitä).
 - Mikäli skannaus on välillä keskeytynyt mistä tahansa syystä (juttelu toisen henkilön kanssa, kahvinkeitto, vessassa käynti, puhelinsoitto), kannattaa keskeytyshetken molemmin puolin skannatut sivut tarkastaa tavallista huolellisemmin. Etenkin ulkopuolisen häiriön yhteydessä virheiden todennäköisyys kasvaa selvästi. Itse tekemistäni virheistä, joita omissa tarkistuksissani löytyi vain muutama, lähes jokainen oli tullut työn keskeytyessä yllättävästi ja huomion kiinnittyessä johonkin muuhun.
 - Myös väsymys lisää virheriskiä.
 - Tavallista hankalimmat sivut (kuvia, karttoja, tavallista pienempi kirjasinkoko, mitä tahansa poikkeavaa) kannattaa tarkastaa huolellisesti erikseen.
 - Pitkän prosessin aikana kannattaa aina välillä tarkistaa, että rutiinin lisääntyessä kehittyvät aikaa ja vaivaa säästävät työtavat eivät huomaamatta heikennä työn laatua.
26. Korjataan mahdollisesti havaitut *virheet*.
- Jos sivu on visuaalisesti virheellinen, se poistetaan, skannataan uudestaan (käydään tarpeellisin osin läpi vaiheet 7.-21., muistaen asettaa oikea aloitusindeksi vaiheessa 10.), tallennetaan oikealla nimellä oikeaan kohtaan oikeaa kansiota ja tarkistetaan, että jatkuvuus toimii oikein.
 - Jos sivu puuttuu, muutetaan ensin virhekohtaa seuraavien sivujen nimet oikeiksi (niin että tiedostonimi vastaa sivunumeroa), ja vasta sen jälkeen skannataan puuttuva sivu toimien kuten edellä.
 - Jos jokin sivu on skannattu kahdesti, poistetaan ylimääräinen versio ja muutetaan muiden sivujen tiedostonimet oikeiksi.
 - Aina korjausten jälkeen on syytä tarkistaa, että kaikki on kansiossa oikein. Huolimattomuus tiedostonimien käsittelyssä johtaa helposti uusiin virheisiin kansion sisällössä.
27. Suljetaan kansio (palataan käytettyyn kohdehakemistoon, esim. F:\Kuvasiirto), luodaan seuraavaa erää varten *uusi kansio* ja nimetään se oikein.

- Uuden kansion luominen on mahdollista monessa muussakin vaiheessa, mutta sekä työn sujuvuuden että virhemahdollisuuksien minimoinnin kannalta se mielestäni kannattaa ehdottomasti tehdä tässä kohden.
 - Useampia tyhjiä kansioita ei kannata tehdä valmiiksi, koska silloin seuraavan erän voi helposti tallentaa väärään kansioon.
 - Uuden kansion nimeäminen tässä vaiheessa auttaa muistamaan oikean tiedostonimen muuttamista vaiheessa 9.
 - Ennen skannauksen jatkamista kannattaa varmistaa, ettei nyt auki olevassa Eräskannaus-hakemistossa ole muuta kuin kansioita. Jos hakemistossa on kansioihin siirtämättömiä kuvatiedostoja, pitäisi niiden kuulua samaan erään seuraavana skannattavien sivujen kanssa, ja tällöin on vaiheessa 10. aloitusindeksiksi muutettava järjestyksessä seuraava sivunumero (tiedostonimen loppuosa). Mikäli muutosta ei tehdä, ohjelma antaa uusille tiedostoille edellisen erän tiedostojen nimet, ja uusia sivuja skannattaessa vanhat samannimiset kansioon tallentamattomat tiedostot katoavat.
28. Jatketaan *skannausta* toistaen vaiheita 7.-27. tarpeellisin osin.
- Vaiheisiin 9. (Tiedostonimi) ja 10. (Aloituseindeksi) on muutettava oikeat lukemat. Muissa kohdin oletusarvot ovat oikeat (ellei niitä ole välillä muutettu) ja tällaiset vaiheet voidaan ohittaa hyväksymällä ne.
 - Scan- ja Prewiev-ikkunat eivät peitä näytön koko alaa. Skannauksen aikana ikkunoiden alla kannattaa pitää auki käytettyä kohdehakemistoa (esim. F:\Kuvasiirto), johon uudet tiedostot tallentuvat. Tällöin uusien sivujen kuvakkeet tiedostonimien aika ajoin näkyvät ikkunoiden peittämän alan ulkopuolella, ja silloin on mahdollista tarkistaa, että viimeisen sivun numero ja vastaavan kuvatiedoston nimen loppuosa täsmäävät. Jos ne eivät täsmää, on tapahtunut virhe – jokin sivu tai aukeama puuttuu, tai jotain on skannattu kahdesti.
 - Virhettä ei tarvitse korjata heti – korjaus vaatii paluuta vaiheeseen 21., vaiheiden 7.-26. osittaista toistoa ja tiedostonimien muuttamista, mikä vie aikaa. Mielestäni yleensä korjaukset kannattaa tehdä vasta sitten, kun koko skannuserä tallennetaan omaan kansioonsa, tehdään tarkistukset ja tarpeelliset korjaukset. Silti mahdollisia virheitä on syytä tarkkailla jatkuvasti skannauksen aikana – näin saadaan tietoa siitä, missä kohden virhe ehkä on, ja mihin tarkistusvaiheessa kannattaa kiinnittää erityistä huomiota.
29. Sopivin välein tallennettu ja tarkistettu aineisto *varmuuskopioidaan*.
- Itse käytin ensi vaiheen varmuuskopiointiin siirrettävää Matrix-kovalevyasemaa, jonka avulla materiaali myös kuljetettiin Kotuksen tiloihin kopioitavaksi toisiin tietokoneisiin. Siirron jälkeen materiaalia säilytettiin varmuuden vuoksi jonkin aikaa myös skannaukseen käytetyn tietokoneen muistissa.
 - Suurten tiedostomäärien varmuuskopiointi voi kestää pitkähkön tovin. Työaikaa säästyy, jos kone jätetään kopioimaan ruoka- tai kahvitaun ajaksi.
 - Varmuuskopioimaton osa aineistoa voidaan menettää esimerkiksi laitevirian vuoksi. Jos menetetään muutaman tunnin työ, se ei ole kovin vaarallista, mutta riskiä päivien tai viikkojen työn valumisesta hukkaan ei kannata ottaa.
30. Työn valmistuttua on materiaalista käyttökopioiden lisäksi tallennettava sopivassa muodossa ainakin yksi *arkistokopio*, jota tulipalon, tulvan tai muun suuronnettomuuden varalta säilytetään eri paikassa kuin muita kopioita ja alkuperäisteoksia. Hyvin arvokkaasta ja vaikeasti korvattavasta materiaalista eri paikoissa säilytettäviä kopioita on hyvä olla useita.
- Arkistokopiot on syytä tallentaa parhaassa mahdollisessa muodossa kuvatiedostoina (TIFF). Käyttökopioita voidaan tallentaa ja siirtää myös muistitilaa säästävässä formaateissa (kuten JPEG) ja mahdollisesti huomommalla resoluutiolla.

V Tallennusmuoto

Tallennusmuodosta neuvottelin itseäni asiantuntevampien Kotuksen ATK-ihmisten ja SKS:n Pasi Klemettisen kanssa, ja muodoksi valittiin TIFF. Käytännössä tietokoneelle muoto määritetään työvaiheita käsittelevän luvun kohdassa 12, ja asetus oli: Formaatti: TIF – TIFF Revision 6.

TIFF (Tagged Image File Format) on kuvantallennusmuoto, joka säilyttää alkuperäisdokumentin (tässä kirjan sivun) kuvan piste pisteeltä niin tarkasti kuin laitteisto, ohjelmat ja niiden asetukset (esim. 8-bittinen harmaasävykuva, resoluutio 500 dpi, jne.) sallivat. Formaatin tärkein etu on se, että tallennuksen laatu on niin hyvä kuin se käytetyillä laitteilla, asetuksilla jne. voi olla. Formaatin huono puoli on se, että muistitilaa kuluu erittäin paljon (tietokoneelta vaaditaan suurta kapasiteettia, tiedon siirto ja käsittely on hidasta, Internetissä suurten TIFF-tiedostojen siirto helposti vie aikaa tuhottomasti).

TIFF on se tallennusmuoto, johon (nykytekniikalla) on viisainta tallentaa ensivaiheessa kaikki arvokas, pitkään säilytettävä, tuhoutuva, suuritoinen tai muuten vaikeasti uudelleen tallennettava aineisto useistakin syistä:

- TIFF tallennusmuotona, periaatteena ja muiden tilaa säästävempien tallennusmuotojen lähtökohtana tulee epäilemättä säilyttämään arvonsa ja käyttökelpoisuutensa hyvin pitkään, ainakin kymmeniä vuosia, samalla kun TIFF-tiedostoja hyödyntävät ja muuntelevat ohjelmat ja pakkausformaatit kehittyvät ja vanhenevat hyvin nopeasti.
- Muun muassa EU-maiden kulttuuriperinnön digitaalisen tallentamisen yhteisiä periaatteita ja hyviä käytäntöjä etsivän Italian kulttuuriministeriön käynnistämän Minerva-projektin raportti (good practices handbook, version 1.3, 3 March 2004, http://www.minervaeuropa.org/publications/goodpractices1_3.htm) useissa kohdin (ks s. 25 ja 38) esittää, että ennen muuta käsittelyä materiaalin tallennusmuotona tulisi olla TIFF.
- Kun materiaali alun perin on tallennettu mahdollisimman tarkassa muodossa (TIFF), sitä voi aina jälkikäteen muuntaa, tiivistää (tiedonsiirtomielessä) ja heikontaa (tai ohentaa, tiedontallennusmielessä) mihin tahansa muuhun muotoon. Päinvastaiseen suuntaan siirtäminen ei ilman informaation häviämistä ole mahdollista, ei aina myöskään tiivistetystä (pakatusta, kompressoitua, käsitelystä) tallennusmuodosta toiseen siirtäminen niin, etteikö jotain (joissain tapauksissa ehkä ainutlaatuisen tärkeää) muuttuisi.
- TIFF on arvokkaan materiaalin tallennusmuotona hyvä ("alkuperäiskappale" ja tapauksesta riippuva määrä varmuuskopioita), mutta myöhemmässä käytössä kannattanee käyttää hyvinkin pitkälle tiivistettyjä (pakattuja, kompressoituja, resoluutiota heikentäviä) tallennusmuotoja siten, että tiivistetyn kopion informaatiomäärä (muistin kulutus) on mahdollisimman vähäinen ja silti riittävä siihen tarkoitukseen, mihin sitä sillä hetkellä käytetään. Tämä nopeuttaa työskentelyä ja on kaikin tavoin järkevää, jopa isossa mitassa koko Maapallon energiataloutta ajatellen. Olennaista on se, että tarvittaessa tarkempaa informaatiota, sitäkin on olemassa ja saatavissa, jos ei muuten, niin TIFF-tiedostoina.

VI Asetukset

Käyttämäni skanneri oli Epson GT-10000. Skannerin asetukset (työvaiheita käsittelevän luvun kohta 14.) olivat seuraavat:

Mode: Professional Mode
Settings
Name: Current Setting
Original
Document Type: Reflective
Document Source: Document Table
Auto Exposure Type: Document
Destination
Image Type: 8-bit Grayscale
Resolution: 500 dpi
Target Size: Original
Adjustments
Unsharp Mask Filter (valittu)
Descreening Filter (ei valittu)

Lisäksi ikkunassa on väri- ja valkotasapainon säätöön yms. liittyviä kuvakkeita. Näihin laitteesta vastaava Pasi Klemettinen kehotti olemaan koskematta, enkä koskenut. Ilmeisesti säädöt olivat kohdallaan.

Ainoat asetukset, joissa mietittiin eri vaihtoehtoja, olivat kuvatyyppi (Image Type) ja resoluutio. Kuvatyypiksi valittiin 8-bittinen harmaasävykuva, koska asiantuntijat sitä yksimielisesti ehdottivat. Resoluutioon 500 dpi päädyttiin kokeilemalla useita eri vaihtoehtoja, kuten olen kertonut edellä luvussa Testivaihe.