



Sukukielten digitointiprojekti Pilottihankkeen loppuraportti

**LAATINUT
JUSSI-PEKKA HAKKARAINEN**

SISÄLLYSLUETTELO

1. TIIVISTELMÄ	1
2. TAUSTAA	3
3. HANKKEEN KULKU	4
3.1. Projektiorganisaatio	4
3.1.1 Projektin työryhmä	4
3.1.2 Projektin ohjausryhmä	5
3.2 Tutkijayhteistyö	6
3.2.1 Aineistojen valinta osana yhteistyötä	6
3.2.2 Editorin kehittäminen tutkijoiden tarpeisiin	7
3.3. Yhteistyö venäläisten toimijoiden kanssa	8
3.3.1 Sopimukset	8
3.3.2 Tekijänoikeudet	8
3.3.3 Digitointi ja jälkikäsitely	10
4. PILOTTIVAIHEEN RISKIANALYYSI	11
4.1 Sopimukset	12
4.2 Venäläiset organisaatiot	13
4.3 Budjetti	14
4.4 Aikataulut	15
5. JULKAISUJÄRJESTELMÄ JA DIGITAALINEN FENNO-UGRICA –KOKOELMA	16
5.1 Fenno-Ugrica -kokoelma julkaisujärjestelmänä	17
5.2 Fenno-Ugrican metadata	18
6. OCR-EDITORI KIELENTUTKIMUKSEN TUEKSI	19
6.1 Yleiskatsaus tekniseen ratkaisuun	20
6.2 Aineiston vastaanotto	20
6.3 OCR-editori	20
6.4 OCR-editorilla työskentely	21
6.5 OCR-editorin hallinnointi	22
6.6 Aineistojen tuominen OCR-editorista	22
7. VIESTINTÄ	22
7.1 Verkkosivut	22
7.2 Kansalliskirjasto-lehti	23
7.3 Blogi	23
7.4 Esitelmät	24
7.5 Sosiaalinen media	24
7.6 Kick off -tilaisuus	24
8. BUDJETTI	25
9. TULOKSET JA YHTEENVETO	27
9.1 Asetettujen tavoitteiden saavuttaminen	27
9.2 Pilotoinnin aikana opittua	28
9.3 Tulokset suhteessa Koneen Säätiön Kieliohjelmaan	30
10. SUOSITUKSET JATKOHANKKEEN TOTEUTTAMISTA VARTEN	32
10.1 Toteutustapa	33
10.2 Jälkikäsitely	34
10.3 Tutkijayhteistyö ja kansalaistieteet	35
11. LIITTEET	37
12. YHTEYSTIEDOT	54

1. TIIVISTELMÄ

Kansalliskirjasto toteutti Koneen Säätiön maaliskuussa 2012 myöntämällä apurahalla Sukukielten digitointiprojektin pilottihankkeen ajanjaksolla heinäkuu 2012-lokakuu 2013. Koneen Säätiön Kieliohjelmassa tähän osahankkeeseen viitataan termillä *Pietarin aineistojen digitoinnin pilottihanke (2012–2013)*. Pilotoinnin tarkoituksena oli selvittää 1) edellytyksiä Venäjän Kansalliskirjaston [Российская национальная библиотека, Pietari] kokoelmissa olevien suomalais-ugrilaisia aineistojen digitaalisesta käyttöön saattamista, 2) aineistoihin kohdistuvien tekijänoikeuksia ja 3) aineistojen soveltuvuutta tutkimuskäytössä.

Pilottihankkeessa digitoitu aineisto sisälsi 17000 sivua marin-, mordvan-, inkeröisten- ja vepsänkielisiä julkaisuja, yhteensä 156 monografiaa, jotka ovat pääsääntöisesti neuvostoajan alun oppikirjoja ja sanakirjoja. Monografioiden lisäksi digitoitiin lähes 25 000 sivua marilaisia ja mordvalaisia sanomalehtiä lähinnä 1920- ja 30-luvuilta. Ensimmäiset pilottihankkeen digitoidut aineistot asetettiin avoimeen käyttöön kesäkuussa 2013 ja viimeiset materiaalit kuvailtiin Fennougrica -kokoelmaan [<http://fennougrica.kansalliskirjasto.fi/>] syyskuussa 2013.

Pilottiprojektissa on työskennelty niin kansainvälisen kirjastokentän kuin kotimaisen kielentutkijayhteisön kanssa. Kansainvälistä työnjakoa on ohjannut lähtökohtaisesti Suomen ja Venäjän Kansalliskirjaston välinen yhteistyösopimus, johon nojaten yhteistyö voitiin käynnistää. Hanketta varten neuvoteltiin syksyllä 2012 kirjastojen välinen sopimus, missä määriteltiin hankkeen puitteissa toteutettavat kummankin osapuolen vastuut ja velvollisuudet digitoitavien aineistojen suhteen.

Aineistojen tekijänoikeuksien selvittämisestä ja korvaamisesta sovittiin erikseen Kansalliskirjaston ja venäläisen kirjastoalan tekijänoikeuksia selvittävän toimiston National Library Resourcen (venäjäksi Национальный библиотечный ресурс, Moskova) välillä, joka suoritti digitoitaviin aineistoihin kohdistuvan tekijänoikeuksienselvityksen joulukuun 2012 ja maaliskuun 2013 välisenä aikana. Selvitykseen nojautuen, Kansalliskirjasto päätti julkaista digitoidun aineiston avoimella palvelimella, jotta projekti vastaisi Koneen Säätiössä Kieliohjelmassa esitettyjä vaatimuksia aineistojen saatavuudesta, saavutettavuudesta ja käytettävyydestä. Kansainvälisesti merkittävän hankkeesta on tehnyt se, että molemmat hankkeessa mukana olleet venäläiset partnerit olivat ensimmäistä kertaa toteuttamassa tämänsuuntaista digitointihanketta ulkomaalaisten yhteistyökumppanin kanssa. Digitointihanke selvitystöineen on kansainvälisesti ainutlaatuinen – koskaan aikaisemmin ei venäläisten kirjastojen kokoelmista digitoituja aineistoja ole julkaistu toisen maan alueella sijaitsevan kirjaston kokoelmissa tässä laajuudessa ja tekijänoikeudet huomioiden.

Kotimaiseen suomalais-ugrilaisia kieliä tutkivaan tiedeyhteisöön oltiin yhteydessä pilotoinnin aikana tiiviisti. Yhteistyö oli luonteeltaan keskustelevaa ja tutkijalähtöistä, ja hankkeessa pyrittiin ottamaan huomioon tutkijoiden mielipiteet ja toiveet mahdollisimman laajasti. Kieliohjelmassa esitettyihin vaatimuksiin pyrittiin vastaamaan myös luomalla tutkimusinfrastruktuuria, jonka avulla tutkijakunta pystyy työstämään digitoitua aineistoa omaa tutkimustaan varten. Yksi näistä tutkimusinfrastruktuurin työkaluista on pilottihankkeen aikana kehitelty OCR-editori, jonka avulla aineistojen sähköiseen muotoon tallennettua tekstiä voidaan muokata. Työkalun kehittämisessä Kansalliskirjasto ja tutkijakunta toimivat yhteistyössä. Digitoitavien aineistojen valinnassa ja niiden evaluoinnissa tutkijakunnalla oli keskeinen rooli.

Erityisesti suomalais-ugrilaisten kielten alkuperäispuhujille aineistot ovat tarjonneet monisyisen mahdollisuuden tarkastella menneisyyttään, kun taas muille loppukäyttäjille Fenno-Ugricassa julkaistut aineistot tarjoavat uuden mahdollisuuden tutustua suomalais-ugrilaisten kansojen kieliin ja kulttuureihin. Pilottihankkeen aikana oli tarkoitus saada avoimeen käyttöön sellaisia aineistoja, joita ei ole saatavilla toisaalla digitaalisessa muodossa ja etenkin tässä laajuudessa. Vaikka digitoitavan aineiston määrä tuntuu kovin pieneltä, niin kyseessä on tarkoin harkittuja kokonaisuuksia, jotka palvelevat laajempia tarkoituksia, ja kuten Koneen Säätiön Kieliohjelmassa sanotaan, ”tarkoituksena on pienten suomalais-ugrilaisten kielten, suomen sekä Suomen vähemmistökielten dokumentointi ja niiden aseman vahvistaminen.” Tämän edellytyksenä on ollut digitoivan aineiston mahdollisimman vapaa saavutettavuus ja tavoitteena olikin, että pilottihankkeen aineisto saatettaisiin avoimeen käyttöön myös tulevan Kansallisen digitaalisen kirjaston asiakasliittymän (Finna) ja vastaavien kansainvälisten palveluiden (mm. Europeana) avulla. Aineistot ovat löydettävissä myös Googlessa ja muilla yleisillä hakukoneilla. Lisäksi aineistoa voidaan etsiä Kansalliskirjaston ylläpitämästä Uralica-portaalista [uralica.kansalliskirjasto.fi], johon on linkitetty uralilaisilla kielillä digitoituja kieliaineistoja eri kirjastojen digitaalisista kokoelmista.

Sukukielten digitointiprojektin pilottihankkeessa on onnistuttu saavuttamaan kiitettävällä tavalla ne tavoitteet, joita pilottiprojektille oli asetettu. Aineistojen käyttöön saattamisen kannalta keskeiset tekijänoikeudelliset kysymykset ratkaistiin yhteistyössä venäläisten toimijoiden kanssa ja tämän saavuttaminen oli keskeinen edellytys hankkeen onnistumisen kannalta. Pilotoinnin aikana on todettu, että Venäjän Kansalliskirjaston kokoelmista voidaan digitoida ja asettaa avoimeen käyttöön sellaista aineistoa, joka on tutkimuksellisesti merkittävää. Tähän saavutukseen on päästy uupumatomalla yhteistyöllä venäläisten partnerien kanssa.

2. TAUSTAA

Koneen Säätiö käynnisti vuonna 2012 kieliohjelman, jonka yhtenä tavoitteena on dokumentoida pieniä suomalais-ugrilaisia kieliä, suomea ja Suomen vähemmistökieliä. Kieliohjelman osana Koneen Säätiö oli syksyllä 2011 käynnistämässä digitointihanketta, jossa tarkoituksena oli digitoida Venäjällä olevia suomalais-ugrilaisia julkaisuja. Tuolloin Koneen Säätiö tiedusteli Kansalliskirjaston ja Helsingin yliopiston kirjaston halukkuutta toimia digitointihankkeen suomalaisena tilaajana. Keskinäisten neuvotteluiden jälkeen Kansalliskirjasto päätti ensin selvittää, olisiko hanke ylipäänsä mahdollista toteuttaa ja millä ehdoilla.

Alkuvuodesta 2012 Jussi-Pekka Hakkarainen selvitti Kansalliskirjaston toimeksiannosta edellytyksiä toteuttaa nk. *Pietarin aineistojen digitoinnin pilottihanke (2012–2013)*, jonka tarkoituksena oli kehittää tuotantoprosessi ja luoda käytännöt eri osapuolten yhteistyönä toteutettavalle aineiston digitoinnille, käyttöön asettamiselle ja aineiston säilytykselle. Tavoitteena oli, että pilottihankkeen tulokset olisivat hyödynnettävissä suomalais-ugrilaisen kielikorpuksen kehittämisessä, siihen liittyvissä tutkimushankkeissa ja että aineistot ovat vapaasti saatavilla internetissä.

Selvityksestä syntyneessä projektisuunnitelmassa (ks. **Liite 1**, raportin kohdasta **11. Liitteet**.) hahmoteltiin se infrastruktuuri, missä edellä mainittu suomalais-ugrilaiden aineistojen pilottihanke voitaisiin toteuttaa. Selvitystyön aikana keskusteltiin niin Koneen Säätiön, tutkijayhteisön edustajien, Kansalliskirjaston, Helsingin yliopiston kirjaston kuin Venäjän Kansalliskirjaston (Pietari) edustajien kanssa hankkeen toteuttamisesta yksityiskohtaisesti.

Selvityksessä hahmoteltiin digitoinnin, digitoidun aineiston käyttöön asettamisen, julkaisujärjestelmän ylläpidon ja viestinnän prosessit. Selvityksessä kuvattiin pilottihankkeen tekninen toteuttamistapa ja määriteltiin digitoitua aineistoa koskevat vaatimukset, myös aineiston metadattaa koskien. Selvityksessä tarkasteltiin ja esiteltiin digitoitavaan aineistoon liittyviä tekijänoikeuksia ja niihin liittyviä kysymyksiä. Lisäksi selvityksessä otettiin kantaa yhteistyösapuolien keskinäiseen työnjakoon ja annettiin ehdotus hankkeen organisaatiosta (ohjausryhmä, taloushallinto ja viestintä) sekä sen järjestämisestä.

Selvityksen pohjalta muotoillulla projektisuunnitelmalla Kansalliskirjasto haki maaliskuussa 2012 Koneen Säätiöltä apurahan pilottihankkeen toteuttamiseksi vuosina 2012–2013. Apuraha myönnettiin Kansalliskirjastolle haetun, eli 356 040 euron suuruisena.

3. HANKKEEN KULKU

Sukukielten pilottiprojekti käynnistyi heinäkuussa 2012 ja loppui lokakuussa 2013. Hankkeen kulkua ja sen toteuttamista käsitellään tässä raportissa kolmen toimintaympäristön kautta. Ensimmäinen niistä liittyy projektin paikalliseen toteuttamiseen Kansalliskirjastossa. Toisen kehikon sisällä tarkastellaan venäläisten partnerien kanssa harjoitettua yhteistyötä, kun taas kolmas osa-alue tarkastelee projektin aikana harjoitettua tutkijayhteistyötä.

3.1 Projektiorganisaatio

Sukukielten digitointiprojektilla ei ole ollut varsinaista omaa, muusta Kansalliskirjaston toiminnasta erillistä työyksikköä ja hankkeen projektipäällikkö on työskennellyt projektille ilman Kansalliskirjaston yleiseen toimintaan tai muihin projekteihin liittyviä velvoitteita.

3.1.1 Projektin työryhmä

Hankkeelle koottiin elokuussa 2012 työryhmä, johon kuuluivat projektipäällikön lisäksi kirjastonhoitaja Marina Ivanova (Slaavilainen kirjasto), tietojärjestelmäasiantuntija Juho Vuori (Kirjastoverkkopalvelut) ja taloussuunnittelija Marja-Liisa Lonardi (Hallintopalvelut). Ivanovan ja Vuoren työpanos korvattiin heidän omalleen työyksikölle tuntikirjanpidon mukaisesti projektin varoista. Joulukuussa 2012 hankkeeseen palkattiin tietojärjestelmäasiantuntijaksi Wouter van Hemel, jonka palkkaukseen käytettiin sekä Sukukielten digitointiprojektin (6kk) että Kirjallisuuspankkihankkeen (9kk) varoja. Lisäksi hankkeeseen palkattiin kirjastosihteeriksi Viktoria Kurkina ajalle heinä-syyskuu 2013.

Lonardin työtehtävät hankkeessa keskittyivät projektipäällikön tukemiseen taloushallinnon, erityisesti budjetointiin liittyvissä tehtävissä. Vuori osallistui Helsingin yliopiston asettaman rekrytointikiellon ajan OCR-editorin suunnitteluun ja van Hemel jatkoi Vuoren työtä joulukuusta 2012 lähtien. Ivanovan ja Kurkinan työtehtävät keskittyvät digitoidun aineiston kuvailuun ja tarvittaessa käännöstehtäviin. Lisäksi Ivanova teki paljon asiantuntijatyötä liittyen kuvailuformaatteihin. Projektipäällikkönä toimineen Jussi-Pekka Hakkaraisen tehtäväkenttä oli laaja. Hänen työtehtäviinsä kuuluivat mm. yhteistyön koordinointi kotimaisten ja ulkomaisten toimijoiden kanssa, projektin viestintään, raportointiin, Fenno-Ugrica -kokoelmaan ja taloushallintoon liittyvät työtehtävät. Projektityöryhmä kokoontui useimmiten hankkeen alkuvaiheissa, ennen kaikkea syksyn 2012 ja alkutalven 2013 aikana. Kun työryhmän jäsenten työtehtävät eriytyivät projektin edetessä, yhteiset kokoontumiset lopetettiin ja projektipäällikkö tapasi työryhmän jäseniä lähes viikottain heidän työtehtäviensä tiimoilta.

Työryhmän jäsenten lisäksi projektissa ovat eri tavoin auttaneet muu Kansalliskirjaston henkilökunta. Kansalliskirjaston Digitointi- ja konservointikeskuksen (Mikkeli) henkilökunta on auttanut hanketta erityisesti OCR-editorin XML-formaatin hallintaan liittyvissä kysymyksissä. Kansalliskirjaston Kirjastoverkkopalvelut ovat tarjonneet asiantuntemustaan projektin käyttöön erilaisissa teknisissä ja metadatan hallintaan liittyvissä työvaiheissa. Kansalliskirjaston Tutkimuskirjastopalveluiden henkilökunta on auttanut projektipäällikköä liittyen yhteistyöhön sekä tutkijakunnan kuin ulkomaisten toimijoiden kanssa.

3.1.2 Projektin ohjausryhmä

Projektille kutsuttiin ohjausryhmä, jossa oli edustettuna niin Kansalliskirjasto, Helsingin yliopiston kirjaston, rahoittajan kuin tutkijoiden edustajia. Ohjausryhmää valittaessa ja sitä koolle kutsuttaessa Sukukielten digitointiprojekti ja Kansalliskirjasto halusivat, että mukana olisi mahdollisimman laaja-alainen osaajien joukko, jotka pystyisivät ohjaamaan hanketta useasta eri asiantuntijaperspektiivistä ja että ohjausryhmän tapaamisen toimisivat myös informaatiokanavana rahoittajan suuntaan. Ohjausryhmän jäsenet olivat:

- *Helsingin yliopisto*
Pauli Assinen
- *Koneen Säätiö*
Risto Alapuro
- *Tutkijakunnan edustajat*
Jack Rueter, Helsingin yliopisto
Jorma Luutonen, Turun yliopisto
- *Kansalliskirjasto*
Liisa Savolainen, johtaja, projektin omistaja, ohjausryhmän puheenjohtaja
Leena Saarinen, tuotantopäällikkö
Jussi-Pekka Hakkarainen, projektipäällikkö, esittelijä
Elli Salo, ohjausryhmän sihteeri

Ohjausryhmä kokoontui tässä kokoonpanossa yhteensä viisi kertaa syyskuun 2012 ja marras-kuun 2013 välisenä aikana. Ohjausryhmän vakinaisten jäsenten lisäksi kokouksiin kutsuttiin vierailevia asiantuntijoita tarpeen mukaan.

Pilottiprojektin etenemistä on seurattu ja ohjattu Kansalliskirjaston sisällä kahdella eri taholla: projektipäällikkö on esitellyt hanketta säännöllisesti Tutkimuskirjaston johtoryhmässä vähin-

tään kerran kuukaudessa. Tämän lisäksi projektipäällikkö on informoinut Kansalliskirjaston johtoryhmää kahteen otteeseen hankkeen etenemisestä.

3.2. Tutkijayhteistyö

Sukukielten digitointiprojektissa on pyritty rakentamaan tiivistä ja osallistuvaa yhteistyötä tutkijakunnan ja Kansalliskirjaston välille. Hanke on alusta lähtien ollut voimakkaasti tutkijalähtöinen ja se on tähdännyt tutkijakunnan kokonaisvaltaiseen palvelemiseen. Käytännön tasolla tämä on tarkoittanut sitä, että tutkijoita on sitoutettu mukaan hankkeen suunnitteluun, toteutukseen ja päätöksentekoon.

Hanketta suunnitellessa tutkijoiden aloitteet ja verkostot ovat olleet ratkaisevia. Kansalliskirjasto ei aikaisemmin ole toiminut näin voimakkaasti suomalais-ugrilaisen maailman ja kielentutkimuksen kentällä, joten hankkeen toteuttajaa voidaan pitää siinä mielessä uutena toimijana tällä kentällä. Tässä suhteessa tutkijakunnan kontaktit ja näkemykset ovat olleet hankkeen toteuttamisen kannalta oivallisia voimavaroja, joihin Sukukielten digitointiprojektissa on pyritty parhaan mukaan tutustumaan.

Tutkijoiden kutsuminen mukaan hankkeen ohjausryhmään ja lähemmäksi päätöksentekoa on ollut myös hankkeen toteuttamisen kannalta hedelmällistä ja tutkijakunnan edustajat ovat voineet täten vaikuttaa projektin etenemiseen heidän omista premisseistään lähtien. Tutkijoiden läsnäolo on lisännyt osapuolten keskinäistä ymmärrystä eri toimintakulttuureista. Seuraavassa esitellään kaksi toiminnan kenttää, missä tutkijayhteistyöllä on ollut ratkaiseva osuus.

3.2.1 Aineistojen valinta osana yhteistyötä

Aineistojen valinta ja valinnan kriteerien määrittely ei olisi ollut mahdollista ilman tutkijoiden osallistumista hankkeeseen. Tutkijat tuntevat parhaiten aineistoihin kohdistuvat tarpeet ja odotukset, joten tutkijoiden äänen kuuluminen oli tässä kohtaa ensisijaista.

Keskeisimpänä aineiston valintakriteerinä on ollut nykykirjakielten synty- ja vakiintumisajankohta. Komin, udmurtin, vuori- ja niittymarin, inkeroisen, ersän, mokšan ja vepsän osalta tämä prosessi tapahtui maailmansotien välisenä aikana, keskeisimmin vuosina 1930-luvun alkupuolella.

Sukukielten nykyisille vaalijoille kirjakielen syntykauden aineisto on tärkeää: 1920- ja 1930-luvun uudissanat ja niitä käyttävät tekstit toimivat meidän aikamme kielenkehittäjille yhtä lailla lähdemateriaalina kuin innovaatioiden ja innoituksen lähteenä. Digitoidut aineistot on valittu siten, että ne yhtäältä kuvastavat mahdollisimman hyvin 1920-luvun innovatiivista aikaa mutta toisaalta ilmentävät myös sitä kielipoliittista muutosta, joka tapahtui 1930-luvulla. Tällaisten aineistoihin kohdistuvien määrittelyjen tekeminen ilman tutkijoiden osallistumista ja heidän asiantuntemusta ei olisi ollut mahdollista.

Kieliaineistojen valintoja rajoittavia tekijöitä on ollut muutamia. Yksi keskeinen rajoitus on ollut aineiston saavutettavuus Suomesta. Mikäli teos löytyy joko Kansalliskirjaston tai suomalais-ten yliopistonkirjastojen kokoelmista, teosta ei ole otettu mukaan digitoitavaksi, sillä tutkijoiden on ajateltu pääsevän käsiksi aineistoon muutoinkin.

Toinen merkittävä rajaava tekijä on ollut aineistoihin kohdistuvat tekijänoikeudet. Ennen vuotta 1941 julkaistua aineistoa voidaan saattaa avoimeen tutkija- ja kansalaiskäyttöön verrattain helposti, mutta toisen maailmansodan aikainen ja sen jälkeinen aineisto edellyttää väistämättömästi sopimista tekijänoikeudenhaltijoiden kanssa. Tietoisuus tekijänoikeuksista ja niihin liittyvistä ongelmista on ohjannut aineistojen valintaa ja siten painostus on vahvasti 1920- ja 30-luvuilla julkaistuissa aineistoissa.

3.2.2 Editorin kehittäminen tutkijoiden tarpeisiin

Pilottiprojektissa toteutettiin web-käyttöliittymä digitoitujen kirjojen ja sanomalehtien kuvista tunnistetun tekstin korjaamiseen ja muuhun editointiin. Tekstintunnistus (OCR) tuottaa virheellisiä merkkijonoja erityisesti huonolaatuisista originaaleista, sekä kielistä, joiden korpuksia ei vielä ole saatavilla. Virheet on korjattava, että tunnistettua tekstiä voi jatkokäyttää.

Yhteistyö tutkijoiden kanssa voidaan sanoa olleen hankkeen onnistumisen kannalta keskeisellä sijalla ja tutkijat ovat olleet mukana edistämässä OCR-editorin suunnittelua eri kehitysvaiheissa. Ilman heidän kontribuutioitaan, OCR-editoriin liitetyt toiminnallisuudet eivät välttämättä vastaisi niitä tavoitteita, mitä sille asetettiin ja millaiseen työhön tutkijat editoria käyttäjät. Muutamia sellaisia toiminnallisuuksia, joihin aloite on tullut tutkijoilta, ovat mm.:

- virtuaalinäppäimistö, jossa näytetään teoksen kielen merkit
- ylläpidon ja peruskäyttäjän erottaminen
- automaattisten muutosten ilmoittaminen

- korpuksen näyttö
- linkit sivuihin, joilla editoitava sana tai sanamuoto on.
- sanojen järjestäminen aakkosjärjestykseen ja lukumäärän mukaan.

3.3 YHTEISTYÖ VENÄLÄISTEN TOIMIJOIDEN KANSSA

Sukukielten digitointiprojektin yksi isoimmista haasteista oli yhteistyö venäläisten toimijoiden kanssa. Pilottivaiheen aikana Kansalliskirjasto harjoitti tiivistä yhteistyötä ennen kaikkea moskovaalaisten tekijänoikeusjärjestö National Library Resourcen kanssa, joka selvitti aineistoihin koskevia tekijänoikeuksia ja Venäjän Kansalliskirjaston kanssa, joka digitoi ja jälkikäsittelee pilottihankkeessa tuotettavan aineiston. Molemmat partnerit olivat hankkeen aikana myös sopimusosapuolia. Edellä mainittujen toimijoiden lisäksi Kansalliskirjasto pyrki keskustelemaan Marin, Karjalan, Komin, Mordovian ja Udmurtian tasavaltojen kansalliskirjastojen kanssa yhteistyöstä mahdollisen jatkohankkeen aikana.

3.3.1 Sopimukset

Sukukielten digitointiprojektin pilottivaiheessa Kansalliskirjasto teki uusia sopimuksia venäläisten osapuolten kanssa. Sopimuksissa määriteltiin osapuolten vastuut ja velvollisuudet, ostopalveluiden hinnat ja aikataulut.

Syksyllä 2012 Kansalliskirjasto sopi aiesopimuksen National Library Resourcen kanssa mahdollisesti tekijänoikeuksien korvaamisesta ja tekijänoikeuksien selvitystyön kustannuksista. Kun National Library Resource sai maaliskuussa tekijänoikeuksien selvityksen valmiiksi, niin tuolloin ei ollut ilmennyt ainoatakaan tekijänoikeudenhaltijaa, jonka kanssa olisi voitu tekijänoikeuksien siirrosta ja niiden korvaamisesta sopia. Tekijänoikeuksista ei siis ole sovittu, vaan ainoastaan niiden selvittämisestä on maksettu korvaus National Library Resourcelle.

Venäjän Kansalliskirjaston kanssa sovittiin erikseen aineiston digitoinnista, sen jälkikäsitteystä ja näiden palveluiden hinnoista. Sopimus sisälsi aikataulun, osapuolten vastuut ja velvollisuudet sekä käytännöt mahdollisten kiistatilanteiden ratkaisemiseksi välimiesoikeudessa.

3.3.2 Tekijänoikeudet

Pilottiprojektin onnistumisen kannalta keskeisimmät kysymykset liittyivät venäläiseen tekijänoikeuslainsäädäntöön ja sen soveltamiseen Sukukielten digitointiprojektissa. Keskeinen periaate on se, ettei aineistoa voida digitoida ilman tekijänoikeudenhaltijoiden suostumusta. Sukukiel-

ten pilottiprojektissa tekijänoikeudenhaltijoiden suostumus aineistoa koskien haluttiin sopia mahdollisimman laajasti, jopa niin että aineistoa voitaisiin esittää rajoituksetta tietoverkoissa ja sitä voitaisiin jakaa edelleen kolmansille osapuolille.

Kansalliskirjastolta oltiin jo hankkeen suunnitteluvaiheessa yhteydessä National Library Resourceen, moskovalaiseen tekijänoikeusjärjestöön, jonka kanssa tuli keskustella ja mahdollisesti sopia tekijänoikeuksien selvittämisestä. Tämä järjestely johtui siitä, että Venäjän Kansalliskirjasto oli antanut mandaattinsa tekijänoikeuksien selvittämisestä tälle organisaatiolle.

National Library Resourceen oli oltu jo talven 2012 aikana selvitystyön yhteydessä, mutta tuolloin heihin ei oltu saatu aitoa keskustelukontaktia, vaikka Venäjän Kansalliskirjastolta oli heitä pyydetty vastaamaan tiedusteluihimme. Tekijänoikeuksien selvittäminen oli hankkeen toteuttamisen kannalta pullonkaula, sillä ilman niiden selvittämistä aineistoa ei voisi julkaista avoimella palvelimella ja tällöin Koneen Säätiön Kieliohjelman tavoitteet avoimuudesta eivät kokonaisuudessaan toteutuisi.

Keskusteluyhteys saatiin aikaan syyskuussa 2012, jolloin National Library Resource lopulta vastasi projektipäällikön tiedusteluihin. Lokakuussa 2012 projektipäällikkö vieraili järjestön luona Moskovassa, missä sovittiin digitoitavan aineiston tekijänoikeuksien selvittämisestä. National Library Resource ehdotti, että he suorittavat selvityksen jokaista digitoitavaa nimekettä koskien ja mikäli tekijänoikeuden haltijoita ilmenisi, tulisi heidän kanssaan sopia vasta selvityksen päätyttyä. National Library Resourceen puolelta selvitystä koordinoi Aleksandr Balatski ja selvityksen lainopillisena neuvonantajana toimi professori ja lakimies Vladimir Entin *Klishin & Partners* -nimisestä tekijänoikeuksiin erikoistuneesta lakiasiaintoimistosta.

Kansalliskirjasto hyväksyi National Library Resourceen ehdotuksen selvitystyön suorittamisesta ja osapuolet (Kansalliskirjasto ja National Library Resource) allekirjoittivat sopimuksen selvitystyöstä joulukuun 2012 alussa, jolloin selvitystyö myös aloitettiin. Selvitystyön kestoksi määriteltiin 90 arkipäivää. Selvitystyön aikana National Library Resource yritti tavoittaa tekijänoikeudenhaltijoita lehti-ilmoituksin, mm. *Literaturnaja gazetan* vuoden 2013 ensimmäisessä numerossa (ilm. 16.tammikuuta 2013) ja Joshkar-Olassa ilmestyvässä *Marii El* -lehdessä 12. tammikuuta 2013. Lisäksi National Library Resource selvitti tekijänoikeudenhaltijoiden mahdollisia elinvuosia kahdeksalta eri viranomaistaholta ja arkistoilta. Ks. **Liite 2** raportin kohdasta **11. Liitteet**.

Hankkeessa tehty selvitys oli poikkeuksellinen ja tällainen selvitys tekijänoikeuksienhaltijoista tässä laajuudessaan suoritettiin Venäjän nykyoloissa ensimmäistä kertaa. Koskaan aikaisemmin koti- tai ulkomainen kirjasto ei ollut selvittänyt aineistojen oikeuksia yhtä laajasti tavoitteenaan avoin aineiston käyttö.

3.3.3 Digitointi ja jälkikäsitteily

Tilattujen aineistojen digitointi ja jälkikäsitteily suoritettiin Venäjän Kansalliskirjastossa, jonka kokoelmiin teokset kuuluivat. Pilottihankkeen aikana digitoitujen aineiston teknisiä vaatimuksia määriteltiin yhdessä Venäjän Kansalliskirjaston kanssa jo selvitysvaiheen aikana talvella 2012 ja OCR-tekstin osalta uudestaan syksyllä 2012.

Koska aineistoja ei voida luovuttaa Suomeen digitoitavaksi, digitointi suoritetaan Venäjän Kansalliskirjastossa ja heidän digitoimansa aineistot toimitettiin edelleen Kansalliskirjastolle. Monografiat digitoitiin harmaasävykannattuna 300 dpi tarkkuuteen, kun taas sanomalehtien resoluutioksi määriteltiin 400 dpi. Primääriskannaus suoritettiin TIFF-kuvaformaattiin ja skannattujen kuvien yhteyteen lisättiin automaattisen tekstintunnistamiseen (OCR, optical character recognition) vaadittavat elementit. OCR-teksti liitettiin kuvatiedostoon ja nämä elementit tuotiin yhdeksi tiedostoksi, PDF-formaattiin, joka oli myös aineiston esitysformaatti julkaisujärjestelmässä. Mahdollista myöhempää käsittelyä varten myös TIFF-muotoiset kuvat luovutettiin Kansalliskirjastolle.

Varsinainen digitointityö onnistui Venäjän kansalliskirjastossa erinomaisesti ja kaikki hanketta varten digitoitua aineistot eräitä konservoitavia sanomalehtiä lukuun ottamatta olivat valmiina jo huhtikuun puolessa välissä. Aineistojen toimittaminen Helsinkiin kuitenkin viivästy huomattavasti, sillä OCR-tekstin lisääminen vei Venäjän kansalliskirjastolta enemmän aikaa kuin oli arvioitu. Tässä työvaiheessa ongelmia ilmeni sekä paikallisen työn koordinoinnissa että sen resursoinnissa. Venäjän Kansalliskirjasto oli järjestänyt työprosessien sisäisen koordinoinnin Ulkomaisten asioiden osaston kautta, joka pyrki koordinoimaan hankkeessa tehtävää työtä kirjaston eri osastoilla. Tämä ei kuitenkaan riittänyt työn laadukkaaseen tekemiseen kaikissa työvaiheissa osastoilla, vaan OCR-tekstin lisääminen viivästytti pahasti aineiston toimittamista. Touko-kesäkuussa tehtyjen tiedustelun perusteella ongelmaksi venäläisten puolelta ilmoitettiin ennen kaikkea työvaiheen aliresursointi. Viivästys koski ennen kaikkea sanomalehtiaineistoa, jonka viimeiset numerot saatiin kuvailuksi Fenno-Ugricaan vasta syyskuun 2013 lopussa, yli neljä kuukautta sopimuksessa olleen viimeisen toimituspäivän (9.toukokuuta 2013) jälkeen.

Yksi ongelma tämän työvaiheen laadukkaassa toteuttamisessa oli se, etteivät venäläiset kirjastot vielä ole tuottaneet paljoakaan OCR-tekstiä sisältäviä aineistoja. Tämä tuotannon tapa oli Venäjän Kansalliskirjastolle verrattain uusi ja kokemattomuus näkyi mm. siinä, että Sukukielten digitointiprojektin työntekijät kouluttivat paikallista henkilökuntaa käyttämään ohjelmistoa sopimukseen edellyttämällä tavalla. Toinen, sekä tähän työvaiheeseen että kirjastojen väliseen yhteistyöhön yleensäkin liittyvä aspekti on se, että suomalaisen ja venäläisen työkuulttuurin erilaisuus aiheutti esteitä keskinäisessä viestinnässä. Hankkeen aikana yleisestikin huomattiin, että venäläisen työkuulttuurin hierarkkisuus esti usein sikäläisten työntekijöiden omatoimisen ongelmanratkaisun.

Yleisesti Venäjän Kansalliskirjaston tuottamaan digitoitiin ja OCR-tekstin laatuun oltiin tyytyväisiä, mutta jatkohankkeessa OCR-tekstin tuottamista suositellaan poissirrettäväksi Pietarista, jotta työvaihetta voidaan kontrolloida laadukkaammin kuin mitä pilottihankkeessa on ollut mahdollista, kasvavien kustannuksien uhallakin. Mikäli työvaihe siirretään Suomeen, niin OCR-tekstin laatua voidaan parantaa prosessinomaisesti yhteistyössä tutkijoiden kanssa. Tämä tuotantotavan muutos tarkoittaisi sitä, että jatkossa vastaanotettaisiin digitoinnit ainoastaan TIFF-muotoisina kuvina ja jälkikäsitteilyt tehtäisiin Suomessa.

4. PILOTTIVAIHEEN RISKIANALYYSI

Ohjausryhmän kokouksessa 10. lokakuulta 2012 keskusteltiin pilottivaiheen projektisuunnitelmasta ja sen toteuttamisesta. Ohjausryhmän jäsen, Pauli Assinen (HYK), huomautti, ettei projektisuunnitelma varsinaisesti sisältänyt riskianalyysiä. Ohjausryhmän pyynnöstä hankkeen projektipäällikkö valmisti riskianalyysin, missä pyrittiin kartoittamaan ja paikantamaan mahdolliset projektin toteuttamiseen liittyvät riskit, arvioimaan niiden todennäköisyys ja esittämään miten näitä riskejä voidaan hallita tai poistaa kokonaan.

Riskien suuruuden arvioinnissa hyödynnettiin kuuden muuttujan kenttää. Riskitaulukossa seurausten vakavuudelle ja tapahtuman todennäköisyydelle on kolme eri tasoa. Tehtyjen selvitysten perusteella valitaan ensiksi seurausten vakavuus taulukon ylimmältä riviltä ja sen jälkeen tapahtuman todennäköisyys ensimmäisestä sarakkeesta. Riski on valittujen kohtien leikkauspisteessä olevan arvon suuruinen. Riskin suuruus saa pienimmillään arvon 1 (Merkityksetön riski) ja suurimmillaan arvon 5 (Sietämätön riski).

Tapahtuman todennäköisyys	Vähäiset seuraukset	Haitalliset seuraukset	Vakavat seuraukset
Epätodennäköinen	merkityksetön (1)	vähäinen (2)	kohtalainen(3)
Mahdollinen	vähäinen (2)	kohtalainen(3)	merkittävä (4)
Todennäköinen	kohtalainen(3)	merkittävä (4)	sietämätön (5)

Riskianalyysi esiteltiin ohjausryhmälle joulukuun kokouksessa ja ohjausryhmällä ei ollut siihen huomautettavaa – sietämättömiä riskejä ei ollut. Taulukon neljä vasemmanpuolimmaista saraketta ovat osa ohjausryhmälle esitettyä riskianalyysiä, kun taas viiden ja viimeinen sarake on loppuraporttiin kirjattu arvio kyseisen riskin hallinnasta ja vaikutuksesta projektille.

4.1 Sopimukset

Sopimukset venäläisten osapuolien kanssa solmittiin joulukuun ensimmäisellä viikolla, joten sopimusten aikaansaamiseen ei liittynyt enää suoranaisia riskitekijöitä, kuten syksyllä 2012 oli pelätty. Sen sijaan sopimuksen toteuttamiseen liittyi vielä tuolloin muutamia riskitekijöitä, joita tässä yhteydessä on tuotu esille.

National Library Resourcen kanssa sovittiin tekijänoikeuksien selvittämisestä. Sopimuksessa määriteltiin osapuolien vastuut ja velvollisuudet. Samoin selvitystyön korvauksista päästiin yhteisymmärrykseen ja ne määriteltiin joulukuun 2012 alussa allekirjoitetussa sopimuksessa. Silti tekijänoikeuksista sopimiseen liittyi riskejä, sillä tuolloin oletettiin, että mahdollisista tekijänoikeuksien korvaamisesta ja julkaisuluvan luovuttamisesta joudutaan sopimaan erikseen tapauskohtaisesti. Tekijänoikeuksiin liittyviä riskejä arvioitiin seuraavasti:

Riskin kuvaus	Projektipäällikön arvio	Perustelu	Mahdollinen toimenpide	Toteuma ja huomautukset
Tekijänoikeuskorvausten loppusumma	Vähäinen (2)	Digitoitava aineisto vanhaa ja julkaisemisesta pääosin yli 75 vuotta, tai se on julkaistu ennen 1939, joten on epätodennäköistä että tekijänoikeudenhaltijoiden määrä nousee suureksi.	Mikäli korvausten määrä nousee isoksi, voidaan a) budjettia allokoida uudelleen, b) mikäli summa nousee korkeaksi, voidaan anoa täydentävää apurahaa, c) tai uuden arvioinnin perusteella jättää aineisto hankkimatta.	Ei vaikutusta projektille. Tekijänoikeuksista ei ole maksettu korvauksia, mutta Kansalliskirjasto on varannut nimellisen summan, mikäli tekijänoikeuksienhaltijat esittävät vaateita.
Tekijänoikeuden haltijat eivät myönnä lupaa julkaista aineistoa	Vähäinen (2)	Vaikka digitoitavien aineistojen joukossa paljon rinnakkaisnimikkeitä ja tarvitaan sekä kääntäjän, että tekijän suostumus julkaisulupaan, niin tapausten määrä jää matalaksi.	Mikäli lupaa ei myönnetä, niin onko aineisto korvattavissa toisella materiaalilla? Digitoidaan, mutta aineisto jää Venäjän Kansalliskirjaston sisäiseen käyttöön	Ei vaikutusta projektille. Tekijänoikeudenhaltijoita ei tavoitettu National Library Resourcen selvitystyöstä huolimatta. Orpoteoksia koskeva lainsäädäntö mah-

				dollisti julkaisemisen.
--	--	--	--	-------------------------

4.2 Venäläiset organisaatiot

Venäläisten yhteistyökumppaneiden toimintaan liittyi joitain riskejä, jotka olivat olleet tiedossa jo projektin alusta lähtien, mutta myös sellaisia, joita ilmeni pilotoinnin aikana.

Riskin kuvaus	Projektipäällikön arvio	Perustelu	Mahdollinen toimenpide	Toteuma ja huomautukset
Digitointia ei saada valmiiksi aikataulusaan / lainkaan	kohtalainen (3)	Venäjän Kansalliskirjastossa on käynnissä parhaillaan organisaationmuutos – vähentämistarve on noin 350 henkilöä ja tämä vähennys saattaa vaikuttaa aikatauluihimme. Onko Pietari lupautunut toimittamaan aineiston aikaisemmin kuin pystyy toteuttamaan?	Vahvistettua tietoa vähennysten vaikutuksesta hankkeelle ei ole, mutta keskusteluista ja toimenpiteistä (mm. skannerin hankkiminen) päätellen he ovat sitoutuneita projektiin. Vaikutusta lähinnä sanomalehtien skannaamiseen? Yhteistyön koordinointi kärsii.	Riski toteutui osittain ja vaikutus projektille oli kohtalainen. Varsinaiset digitoinnit oli tehty ajoissa, mutta jälkikäsitteily Pietarissa viivästytti aineiston toimitamista yli sovitun aikarajan. Riskiä ei voitu ennakoida, sillä Venäjän kansalliskirjastosta vakuutettiin koko prosessin ajan, että kaikki aineisto tullaan toimittamaan ajoissa, mutta näin tapahtui ainoastaan monografioiden osalta. Kun palvelu julkistettiin 6. kesäkuuta, vain noin kymmenesosa sanomalehtiaineistosta oli Fenno-Ugricassa. Viimeiset sanomalehtiaineistot kuvailtiin tietokantaan syyskuussa, yli neljä kuukautta sovitun toimitusajankohdan jälkeen. Syynä tähän viivästykseen oli sen, että sanomalehtiaineiston OCR-työ kesti oletettua pidempään ja syynä tähän oli työvaiheen aliresursointi, osaamiskapeikko ja työn huono koordinointi Pietarissa.
Aineiston laatu digitointiin kelpaamatonta	vähäinen (2)	Kaikkea sanomalehtiaineistoa ei pystytä digitoimaan sen heikon kunnan vuoksi. Noin 2000 sivua jää toistaiseksi digitoimatta.	Aineiston entisöinti maksaisi 513 000 RUR + tuotantokulut, eli arviolta noin 943 000 RUR / 23500 EUR. Tämänhetkiseen budjettiin kuluera oli liian suuri, mutta voisiko tämän aineiston digitointiin hakea täydentävää rahoitusta? Olisi kartoitettava saadaanko materiaali muualta joko alkuperäisestä skannaamalla tai mikrofilmiltä digitoimalla? Vaatisi joka tapauksessa lisärahoitusta.	Vaikutus projektille vähäinen. Valtaosa aineistoista pystyttiin digitoimaan ja vain osa aineistosta vaati konservointia. Viimeiset konservointia vaatineet aineistot toimitetaan Kansalliskirjastolle huhtikuun 2014 loppuun mennessä. Osa digitoitavaksi esitetystä aineistoista tosin oli kärsinyt vuosia aikaisemmin vesivahingoista, eikä niitä sen vuoksi voitu digitoida. Tämä ei ollut tiedossa aineistoa valitessa.

National Library Resourcen asema vahvistuu talvella 2013 tekijänoikeusjärjestönä	vähäinen (2)	National Library Resource tulee jatkossa lisensoimaan / myöntämään tekijänoikeuden haltijoiden puolesta luvan nyk. Venäjän federaation alueella julkaistuihin aineistoon ja vie kansalliskirjastoilta mahdollisuuden sopia tekijänoikeudenhaltijoiden kanssa suoraan.	Keskusjohtoisuus lisääntyy, mutta meillä on nyt jo kokemusta NatLibin kanssa toimimisesta. Huonoa se, että joudumme käymään Moskovin kanssa keskusteluja, vaikka aineisto digitoitaisiin muualla. Pitäisi saada aikaan pieni projekti esim. Udmurtian kanssa ennen kuin laki astuu voimaan. Keskeisin vaikutus kuitenkin vuosille 2013–2016.	Vaikutus projektille vähäinen. National Library Resourcen asemaa venäläisen kirjastoalan tekijänoikeusjärjestönä asemaa ei oltu vahvistettu duumassa projektin loppuun mennessä. Venäläisten partnerien mukaan asia on edelleen käsitellyssä.
Tekijänoikeuksienhaltijoiden myöntämisen luvan jakaminen ei onnistu	vähäinen (2)	On allekirjoitettu aiesopimus, minkä perusteella Kansalliskirjasto saa luovuttaa edelleen kolmansille osapuolille oikeuden julkaista aineistoa, koskee ml. Venäjän Kansalliskirjastoa. Kyseessä on kuitenkin vain aiesopimus, mikä ei ole sitova millekään osapuolelle.	Mahdollinen konflikti Venäjän Kansalliskirjaston kanssa vältettävissä – olemme sitoutuneet avoimeen yhteistyöhön ja noudattamaan sekä sopimuksia että lainsäädäntöä. Pietarilla ei vastustusta, mutta pitää huomioida Pietarin heikko asema lisensoinnin suhteen. Vaikuttaa myös vuosina 2013–2016, mikäli digitoidaan Pietarista.	Vaikutus projektille vähäinen. Digitoidun aineiston jakaminen kolmansille osapuolille onnistuu, sillä aineisto on tekijänoikeuksien selvityksen myötä public domain Venäjän lainsäädännön mukaan. Digitointeja koskeva tekijänoikeuksien selvitys tehdään jatkossakin yhteistyössä NatLibin kanssa.

4.3 Budjetti

Budjettiin kohdistuvat riskitekijät kohdistuvat lähinnä kuluerien uudelleen kohdistamiseen ja mahdollisten tekijänoikeuskorvausten määrään.

Riskin kuvaus	Projektipäällikön arvio	Perustelu	Mahdollinen toimenpide	Toteuma ja huomautukset
Tekijänoikeuskorvausten suuruus ylittää budjetin	kohtalainen (3)	Emme tiedä mikä tulee olemaan tekijänoikeudenhaltijoiden korvausvaatimusten taso, mikäli sellaisia ylipäänsä ilmenee. Selvitystyön aikana ei saatu tietoja selvitystyön suuruudesta, eikä mahdollisista korvausmääristä, vaan tämä kuluera jouduttiin arvioimaan ilman parempaa tietoa. Kustannusten odottamaton	National Library Resource on ottanut vastuulleen korvata selvitystyön jälkeen esitettävät tekijänoikeuskorvausvaatimukset, jolloin Kansalliskirjasto ei joudu korvaamaan mitään jälkikäteen. Digitointi ja selvitystyö laskettu siten, että noin 5% kulueristä vielä käyttämättä. Budjetissa muuallakin liikkumavaraa (mm. koulutuksesta), jota voidaan allokoida toiseen toimintoon, mikäli kustannukset kasvavat. Mikäli tekijänoikeuksien korvaamisesta koituvat kustannukset kasvavat yli budjetin, neuvoteltava rahoittajan kanssa mahdollisesta lisärahoituksesta ja mikäli lisärahoitusta ei saada, tulee keskustella tutkijoiden kanssa siitä, mitä	Ei vaikutusta projektille. Tekijänoikeuksienhaltijoille ei ole maksettu korvauksia, sillä heitä ei ole selvityksen yhteydessä paikannettu. ks. myös luku Tekijänoikeudet ja niiden selvittäminen.

		kasvaminen euroalueen kriisin tai valuuttakurssien heittelyn seurauksena	aineistoja otetaan mukaan.	
Henkilöstö/budjetoitu työmäärä ei riitä hankkeen toteuttamiseen	vähäinen (2) / merkittävä (4)	Selvitysvaiheessa arvioitu työmäärä ylittyy ja projektille tehtävää työtä ei voida tehdä virkатыönä. Tietojärjestelmä-asiantuntijan palkkaaminen ja yhteistyö Kirjallisuuspankin kanssa parantaa nykytilannetta. Projektipäällikölle allokoitu työmäärä riittämättömän hankkeen täysimittaiselle toteuttamiselle.	Rekrytointikiellon vuoksi jouduttiin tekemään enemmän virkатыötä kuin oli ajateltu – riittävätkö korvaavalle työlle allokoitut varat pilottikauden loppuun? Riskinä resurssien ja osaamisen poistuminen hankkeesta. Jos resurssit lähtevät, mistä korvaavat resurssit tilalle? Riskinä hankkeen kaatuminen. Projektipäällikkö tehnyt suunniteltua enemmän töitä sopimusvaiheessa ja jäljellä olevilla resursseilla voidaan keskittyä lähinnä pilotin läpiviemiseen, mutta kehittämiseen ei suuriakaan mahdollisuuksia. Selvitystyön jälkeen (maaliskuuhuhtikuussa) uusi arvio työajan kohdentamisesta ja mahdollisista uusista allokoinneista.	Vaikutus projektille kohtalainen. Helsingin yliopiston rekrytointikielto hidasti projektiin rekrytoitujen tietojärjestelmäasiantuntijien palkkaamista yli kolmella kuukaudella. Rekrytointilupa saatiin marraskuussa 2012 ja haastattelut suoritettiin saman kuun lopussa. Tehtävään valittu henkilö aloitti työssään joulukuun puolessa välissä. Tämän viivästyksen vaikutukset ovat näkyneet ennen kaikkea OCR-editorin kehittämisessä. Vaikka Kansalliskirjaston kirjastoverkopalveluissa aloitettiin editorin kehittäminen virkатыönä, niin editorin täysipainoinen suunnittelu voitiin aloittaa vasta tietojärjestelmäasiantuntijan aloitettua työnsä. Projektipäällikön käyttämä työaika ylitti suunnitteluvaiheessa lasketun työaika-arvion lähes kaksinkertaisesti. Pilotti-hanke olisi voinut epäonnistua täydellisesti ellei Kansalliskirjasto olisi kokopäiväistänyt projektipäällikköä huhtikuun 2013 alusta alkaen.

4.4 Aikataulut

Hankkeen toteuttamisen kannalta tietyt tehtävät ja niihin liittyvät aikataulut ovat kriittisiä. Osa aikataulutukseen liittyvistä riskeistä syntyy venäläisten toimijoiden asettamista rajoista, kuten selvitystyön kestosta, kun taas Kansalliskirjaston sisällä tehtävä työ ja sen aikatauluttaminen voi muodostaa riskin tutkijayhteistyön täysimittaiselle toteuttamiselle.

Riskin kuvaus	Projektipäällikön arvio	Perustelu	Mahdollinen toimenpide	Toteuma ja huomautukset
OCR-editoria ei saada valmiiksi ajoissa	kohtalainen (3)	OCR-editoria ei saada käyttöön kun aineisto voidaan asettaa julkisesti saataville. Rekrytointikiellon vuoksi työ pääsee alkamaan yli	OCR-editoria on kehitetty virkатыönä ja lähtökohdat sen toteuttamiseen talven ja kevään aikana ovat olemassa. Se mitkä editorin toiminnot ovat tutkijoiden käytössä ja mistä alkaen, on vielä epäselvää. Tahtotila editorin kehittämiseen on olemassa. Jos OCR-editoria ei saada aikaiseksi, niin vaihtoehtona on mm. sen tuottamisen ulkoistaminen.	Vaikutus projektille on ollut ennakoitua suurempi. OCR-editoria ei ole vielä (lokakuu 2013) saatu tutkijoiden käyttöön kokonaisuudessaan. Yksi työtä viivästyttäneistä syistä on ollut Helsingin yliopiston rekrytointikielto, minkä vuoksi editorin kehittämisen aloitus viivästyti. Toinen keskeinen tekijä on ollut suunnittelu- ja koodaustyön aliresursointi - siitäkin huolimatta, että editorin kehit-

		kolme kuukautta myöhemmin kuin oli suunniteltu.		tämiseen on laitettu myös virkatyötä. Yhden tietojärjestelmäasiantuntijan sijaan olisi tarvittu enemmän tai jaettuja resursseja, jolloin työvaiheita olisi voitu jakaa useamman tekijän kesken.
Selvitystyön kesto viivästyttää hanketta	merkittävä(4)	Selvitystyö tulee saattaa päätökseen 90 työpäivän kuluessa, sopimuksen allekirjoittamisesta (4.12.2012) lähtien.	90 työpäivää on pitkä aika ja jos koko aika käytetään, mikä on erittäin todennäköistä, työ saadaan päätökseen vasta maalisi- ja huhtikuun vaihteessa. Tekijänoikeuksista vapaa materiaali voitaneen julkaista heti selvitystyön päätyttyä, mutta jos oikeudenhaltijoita ilmenee, pitää heidän kanssaan vielä sopia tekijänoikeuksien luovuttamisesta, korvauksista jne. Mitä työtä tutkijat voivat sillä aikaa tehdä?	Vaikutus projektille oli arvioitua vähäisempi. Tekijänoikeuksienselvitys valmistui sovitussa ajassa ja Kansalliskirjasto hyväksyi National Library Resourcen tekemän selvityksen huhtikuun alussa. Selvitystyö ja sen kesto eivät hidastaneet projektia ja aineistojen saattamista avoimeen ja tutkijoiden käyttöön.
Aineiston digitointi ja käyttöön saattaminen viivästyy	merkittävä (4)	Pietarissa olevat resurssit ovat vajaat ja he ovat suostuneet kovin tiukkaan aikatauluun. Pystyvätkö pitämään digitoinnin tason korkealla? Sanomalehtiaineiston käsittely ja konservointi vie aikaa. Ei riittävästi laitteistoa?	Pietari ei tule pysymään aikataulussa, mutta viivästysten vaikutuksen hankkeen toteuttamiselle todennäköisiä, mutta kuitenkin pieniä. Viivästykset tulevat kohdistumaan sanomalehtiaineistoon, jota ei pystytä digitoimaan yhtä paljon kuin Pietari on itse arvioinut, sillä työskentelevät kahdella sanomalehtien digitointiin soveltuvalla skannerilla, mutta toisen skannerin (A1-kokoinen) tuottama aineisto joudutaan yhdistämään toisiinsa, mikä vie aikaa ja mahdollisesti heikentää laatua. Testiaineisto tehty toisinaan viimeisen päälle, toisinaan huonosti. Mihin digitoinnin taso asetuu ja kuinka paljon voimme hyväksyä aineistoja?	Riski toteutui osittain ja vaikutus projektille oli kohtalainen. Varsinaiset digitoinnit oli tehty ajoissa, mutta jälkikäsitteily Pietarissa viivästytti aineiston toimittamista yli sovitun aikarajan. Riskiä ei voitu ennakoida, sillä Venäjän kansalliskirjastosta vakuutettiin koko prosessin ajan, että kaikki aineisto tullaan toimittamaan ajoissa, mutta näin tapahtui ainoastaan monografioiden osalta. Kun palvelu julkistettiin 6. kesäkuuta, vain noin kymmenesosa sanomalehtiaineistosta oli Fennougricassa. Viimeiset sanomalehtiaineistot kuvailtiin tietokantaan syyskuussa, yli neljä kuukautta sovitun toimitusajankohdan jälkeen. Syynä tähän viivästykseen oli sen, että sanomalehtiaineiston OCR-työ kesti oletettua pidempään ja syynä tähän oli työvaiheen aliresursointi ja huono työn koordinointi Pietarissa.

5. JULKAISUJÄRJESTELMÄ JA DIGITAALINEN FENNO-UGRICA -KOKOELMA

Fenno-Ugrica (fennougrica.kansalliskirjasto.fi) on toteutettu DSpace-ohjelmistolla. DSpace on avoimen lähdekoodin sovellu digitaalisten aineistojen hallinnointiin. Kansalliskirjasto käyttää sitä eräiden omien aineistojensa hallintaan sekä tarjoaa sen avulla tuotettuja maksullisia julkaisuarkistopalveluja.

Pilottihankkeessa räätälöidyn DSpace-instanssin avulla edistetään digitoidun aineiston saavutettavuutta ja käyttöä myös ulkomailla, niin kansainvälisen tiedeyhteisön kuin Venäjällä asuvien suomensukuisten kielten puhujien keskuudessa. Fenno-Ugrica toteutettiin myös venäjänkieli-

senä ja sen käyttöliittymä voidaan mahdollisesti julkaista myöhemmin myös muilla kielillä. Viestinnän ja tunnettuuden edistämiseksi pilottihankkeen DSpace-instanssille luotiin räätälöity ja tunnistettava ulkoasu.

Kuva. Fenno-Ugrican venäjänkielinen käyttöliittymä.



Фенно-Угрика

Фенно-Угрика – оцифрованная финно-угорская коллекция Национальной библиотеки Финляндии. Коллекция содержит публикации на ижорском, вепсском, марийских (горномарийский и луговомарийский) и мордовских (Эрзянский и Мокшанский) языках, а также газеты на марийских и мордовских языках, опубликованных в основном в 1920-ые и 1930-ые годы. В целом коллекция содержит более 120 монографий и почти 20 000 страниц текста газет.

Представленные здесь электронные ресурсы созданы в рамках [Проекта по оцифровке публикаций на финно-угорских языках](#) из фондов [Российской национальной библиотеки](#) в Санкт-Петербурге. Проект является частью [Языковой программы](#) Фонда Коне. Материалы оцифрованы в Российской национальной библиотеке и публикуются на основании исследования, выполненного [Национальным библиотечным ресурсом](#), по проверке наличия или отсутствия обладателей исключительных прав на включенные в Проект издания.

В рамках проекта для лингвистов также разработан OCR-редактор, основанный на открытом исходном коде, который позволяет редактировать OCR-тексты на финно-угорских языках. Право редактировать ресурсы данной коллекции предоставляется, прежде всего, финноугроведам. Право пользования и редактирования можно получить у администрации Проекта по оцифровке финно-угорских языков. Дополнительная информация и контакты по электронному адресу: kk-fennougrica@helsinki.fi

Разделы и коллекции

- [Институт эстонского языка](#) [0]
- [Монографии](#) [156]
- [Газеты](#) [5101]

Весь архив

- [Заглавия](#)
- [Авторы](#)
- [Даты публикации](#)
- [Темы](#)
- [Свежие поступления](#)
- [Просмотр по языкам](#)
- [Карта сайта](#)

Мой профиль

- [Войти](#)
- [Зарегистрироваться](#)

KONEEN SÄÄTIÖ

1640
KANSALLIS
KIRJASTO

5.1. Fenno-Ugrica -kokoelma julkaisujärjestelmänä

Fenno-Ugrica on toteutettu virtuaalipalvelimeen, jota ajetaan Kansalliskirjaston Vmware-ympäristössä. Virtuaalipalvelimen levytilan koko on 337 Gt, ja siitä on tällä hetkellä käytössä noin 70 %. Fenno-Ugrican käyttöliittymä ja käyttöohjeet on toteutettu suomeksi, englanniksi ja venäjäksi. Sille on suunniteltu oma graafinen ulkoasu. Erityispiirteinä Fenno-Ugrican on latinalaisten ja kyrillisten julkaisujen rinnakkaisuus. Nimekkeiden ja tekijöiden selauksessa on käytössä rinnakkaiset listaukset molemmille järjestelmille. Lisäksi käyttöliittymään on toteutettu mahdollisuus selata aineistoja kielen mukaan.

Käyttöön asettamista varten aineistot on täytynyt kuvailla. Venäjän kansalliskirjastolla on saadusta aineistosta kuvailutietoja, mutta niiden automaattinen hyväksikäyttäminen osoittautui vaikeammaksi kuin uudelleen kuvailu. Lisäksi aineistojen syöttöprosessiin, jossa mm. lisätään

aineistoon metatietoja on tehty aineiston edellyttämiä räätälöintejä. Itse julkaisut asiakas saa käyttöön PDF-tiedostoina. Käyttöliittymään upotettua lukulaitetta ei ainakaan tässä vaiheessa ole katsottu tarpeelliseksi toteuttaa, koska se ei tuottaisi asiakkaalle merkittäviä uusia ominaisuuksia tai parempaa lukukokemusta. Kaikki saatu aineisto on siis tällä hetkellä asetettu yleisö-käyttöön siinä kunnossa kuin ne on saatu; kun tunnistetun tekstin korjaaminen on valmis, kukin dokumentti korvataan parannetulla versiolla.

Tällä hetkellä Fenno-Ugricasta ladataan tiedostoja keskimäärin 900 kertaa kuukaudessa. Kansalliskirjaston palvelut näkyvät erittäin hyvin hakukoneissa; aineiston luonteen huomioon ottamisen käyttö on vähintäänkin runsasta.

5.2. Fenno-Ugrican metadata

Pilottihankkeessa digitoitujen aineistojen kuvaillussa on käytetty DublinCore kuvailuformaattia. Venäjän Kansalliskirjasto käyttää aineiston kuvailuformaattina RusMarcia ja Suomessa painetun aineiston kuvaillussa on käytössä Marc21-formaatti. Ennen kuin projektissa digitoitu aineisto voitiin kuvailla Fenno-Ugricaan, tuli päättää, mitä DC-kenttiä tarvitaan tässä projektissa ja mitä tietoa mihinkään kenttään tallennetaan. DC-kentät määriteltiin yhdessä Marina Ivanovan, Jussi-Pekka Hakkaraisen ja Ulla Ikäheimon kesken.

Projektin alkuvaiheessa kirjastonhoitaja Marina Ivanova tutustui tarkemmin venäläiseen RusMarc-formaattiin ja hän osallistui DSpace-koulutukseen, joka järjestettiin 3.10.2012 Kansalliskirjaston Kirjastoverkkopalvelussa. Pystyäkseen työskentelemään DSpacessa käytettävällä metadataformaattilla, Ivanova perehtyi myös suomalaisiin ja ulkomaalaisiin DublinCore-formaatin käyttöoppaisiin ja käytäntöihin. DublinCore -formaatti on käytetty digitaalisina syntyneiden verkkodokumenttien kuvailuun jo 1990-luvusta alkaen, joskin digitoitujen aineistojen kuvaillussa se ei ole vielä vakiintunut ja DC-kenttiä käytetään vaihtelevasti. Esimerkkinä voi mainita julkaisija-kentän. DC-ohjeen mukaan julkaisija-kentän tarkoituksena on identifioida se taho joka on asettanut kyseisen tallentamiseen käyttöön. Monissa DSpace-järjestelmissä kuvailtaessa digitoituja dokumentteja tähän kenttään on kuitenkin tallennettu alkuperäisen (painetun) dokumentin julkaisutietoja.

Kuvailtu aineisto sisältää 156 digitoitua monografiaa inkeröisen, vepsän, ersän, mokšan, šokšan, niittymarin ja vuorimarin kielillä. Inkeröisen ja vepsän kirjaimisto on latinalainen, muiden kielten kirjaimisto on kyrillinen. Kyrillinen aineisto on kuvailtu käyttäen sekä alkuperäisiä merkkejä, että translitterointia. Translitterointistandardina on valittu ISO 9 -standardi (Transli-

teration of Cyrillic characters into Latin characters – Slavic and non-Slavic languages), koska se on käytetyin suomalaisissa kirjastoissa. Translitteroinnissa runsaasti käytetyt erikoismerkit eli diakriitit vaativat tarkkuutta ja kärsivällisyyttä, koska normaalista näppäimistöä niitä ei saa ja ne piti lisätä kopioimalla eri lähteistä.

Kirjastoalalla on paljon käytetty MARC-kielikoodoja, jotka perustuvat Library of Congressin ylläpitämään ISO-standardiin ISO 639-2. Tässä projektissa niitä ei voinut kuitenkaan hyödyntää, koska kaikille suomalais-ugrilaisille kielelle siinä ei ole omaa kielikoodia. Erilaisista kansainvälisistä kielikoodistandardista valitsimme ISO 639-3 standardin (<http://www-01.sil.org/iso639-3/default.asp>), koska se on täydellisempi ja sisältää kielikoodit kielille, joita tarvitsemme tässä projektissa.

Aineiston monikielisyys asetti haasteita kuvailuprosessille. Esimerkiksi tekijän nimimuodon valinnassa piti päättää, valitaanko alkuperäinen venäjänkielinen nimimuoto - Тетюрев vai käänöksissä esiintyvät muodot – ТЪЭТЬУРЭВ, ТЭТЬУРЭВ, Tetjurev ja Tetyrev? Kirjastotietokannoissa nämä ongelmat voidaan ratkaista luomalla auktorisoituja nimimuotoja, joiden avulla on mahdollista saada saman tekijän kaikki teokset yhdellä tekijähaulla. DSpace -ympäristö ja DublinCore-formaatti eivät anna tällaiseen auktorisoitujen nimimuotojen käyttöön mahdollisuutta. Tässä projektissa teoksien määrä on kuitenkin vähäinen ja haettu nimimuoto löytyy selaamalla tekijälistaa, joka ei ole pitkä. Projektissa digitoitu aineisto sisältää oppikirjoja ja sanakirjoja. Asiasanoituksen suhteen yleisiä ohjeita ja sovittuja käytänteitä oli helpompi seurata, sillä termeinä käytettiin vain ko. kieltä ja oppiainetta. Sisällönkuvailu tehtiin suomen- ja venäjänkielillä asiasanoilla.

6. OCR-EDITORI KIELENTUTKIMUKSEN TUEKSI

Sukukielten digitointiprojektilla on kiinnekohtia myös kieliteknologiseen tutkimukseen, sillä projektin yhdeksi päämääräksi voidaan laajasti ajatella digitaalisten kirjasto- ja arkistoaineistojen käyttötapojen ja käytettävyyden parantaminen. Projektissa on edistetty suomalais-ugrilaisen aineiston käyttöön saattamisen lisäksi menetelmiä, joilla digitoitua raakadataa voidaan jalostaa entistä käyttökelpoisemmiksi aineistoiksi ja joilla aineistoa voidaan hyödyntää.

Sukukielten digitointiprojektissa näillä menetelmillä tarkoitetaan digitoidun aineiston OCR-tunnistuksen lisäämistä, tunnistetun tekstimassan palstoittamista sekä ennen kaikkea kielenkor-

jaukseen tarkoitettua OCR-editorin kehittämistä, jonka avulla voidaan digitoinnin ja OCR-tunnistuksen yhteydessä jääneitä virheitä korjata tehokkaasti ja talkoistamalla. Kansalliskirjaston pilottivaiheen aikana kehittämä OCR-editori ainoa laatuaan maailmassa ja se tukee myös digitointialan kokonaisvaltaista kehittymistä.

6.1 Yleiskatsaus tekniseen ratkaisuun

Projektin teknisessä toteutuksessa on kolme pääkomponenttia:

- 1) aineiston vastaanotto sisällön toimittajalta
- 2) järjestelmä, jonka avulla aineisto esitetään yleisölle
- 3) järjestelmä, jonka avulla aineiston tekstitunnistusta korjailaan ja jossa siihen liittyviä prosesseja hallinnoidaan.

6.2 Aineiston vastaanotto

Projektin aineisto on digitoitu Pietarissa ja se siirrettiin Kansalliskirjastoon Pietarissa olevalta FTP-palvelimelta. Aineisto on tuotettu palvelimelle sovitunrakenteisina kokonaisuuksina, jotka ovat zip-pakattu. Yksi paketti sisältää yhden monografian tai yhden lehden numeron. Kukin paketti sisältää PDF-tiedoston, johon on upotettu automaattisen tekstintunnistuksen tulos, sekä kukin sivun erillisenä kuvatiedostona. Kansalliskirjasto on muodostanut PDF-tiedostoista ALTO XML -tiedoston. ALTO on vanhojen aineistojen digitoinnissa yleisesti käytetty tapa ilmaista sisältö rakenteisessa XML-muodossa. Tiedostosta käy ilmi dokumentin rakenne, mm. kaikkien sanojen tarkat sijaintikoodinaatit. Tässä vaiheessa myös ALTO XML-tiedosto siis sisältää tunnistetun tekstin tarkistamattomana. Paketissa olevat PDF-tiedostot esittävät julkaisun facsimilena, jossa on taustalla upotettuna automaattisen tekstintunnistuksen tulosteksti; nämä tiedostot on kopioimaan Fenno-Ugrica -julkaisuarkistoon julkista esittämistä varten. Aineisto varmistetaan nauhoille. Tähän mennessä aineiston koko on kaikkiaan noin 1,8 teratavua. Jäljellä olevien latausten toteuttua tilan tarve on noin 4 teratavua.

6.3 OCR-editori

OCR-editori koostuu kahdesta pääosasta:

- 1) editorin käyttöliittymästä, jota tekstin korjaajat käyttävät sekä
- 2) sen taustalla olevasta järjestelmästä, jossa hallinnoidaan tietokantoja, aineistoja ja niiden versiointia, käyttäjiä, selausnäkyymiä ja muita editoinnin vaatimia toimintoja.

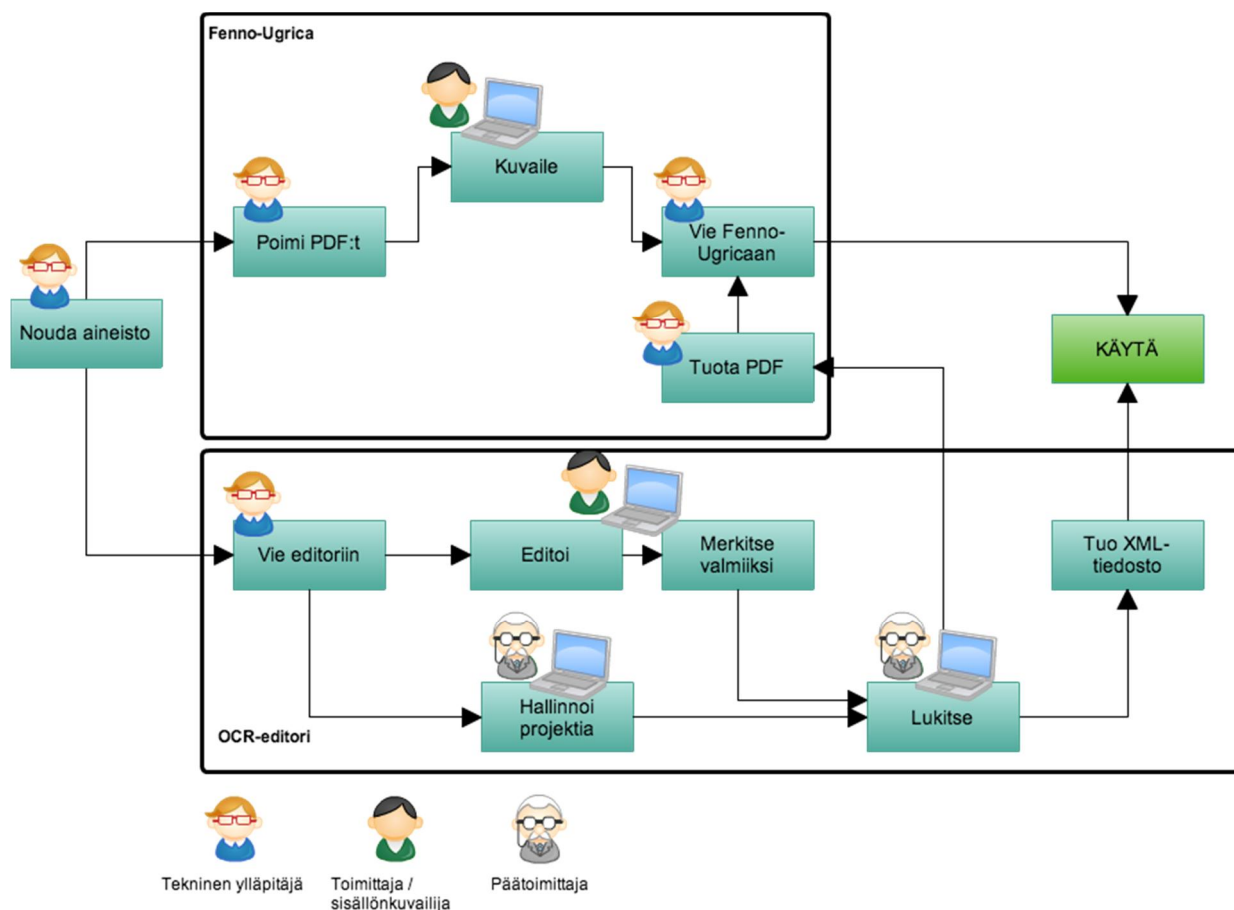
Käyttöliittymä on toteutettu JavaScriptillä ja taustajärjestelmä Pythonilla. Tiedonsiirto tapahtuu JSON API:lla toteutetun REST-rajapinnan avulla. Kansalliskirjasto tulee piakkoin julkaisemaan OCR-editorin avoimena lähdekoodina. OCR-editoriin ladataan saaduista paketeista ALTO XML -

tiedosto ja kuvatiedostot sekä paketeista saatava metadata; nimeke, tekijä ja pääasiallinen kieli. Näistä muodostetaan aineistoluettelo, josta aineisto voidaan valita avattavaksi editointikäyttöliittymässä.

6.4 OCR-editorilla työskentely:

- Toimittaja valitsee haluamansa dokumentin aineistoluettelosta.
- OCR-editorissa toimittaja näkee näytöllä rinnakkain dokumentin yhden sivun facsimilena ja sen konetunnistetun tekstin.
- Toimittaja voi pitää kuvan ja tekstin rinnakkain tai päällekkäin
- Kuvaa voi liikutella hiirellä, pienentää ja suurentaa
- Klikkaamalla sanaa joko kuvassa tai tekstissä, sana korostuu molemmissa versioissa
- Tekstiä voi korjailta tavanomaiseen tapaan. Korjatut sanat saa näkymään korostettuina.
- Erikoisten merkistötärpeiden takia toimittaja saa käyttöönsä myös merkivalikon, josta valitsemalla tekstiin voi lisätä merkkejä, jotka ovat hankalia käytössä olevalla näppäimistöllä.
- Toimittaja voi valita yksittäisen sanan tai tekstinosan ja merkitä sen kielen.
- Toimittaja voi tallentaa tekemänsä muutokset
- Toimittaja voi palata aiemmin tallennettuun versioon (työn alla per 17.10.2013).
- Toimittaja voi merkitä dokumentin mielestään valmiiksi korjatuksi (työn alla per 17.10.2013).

Kuva. Työvuorot Fenno-Ugricassa ja OCR-editorissa



6.5 OCR-editorin hallinnointi

Hallinnointitehtävillä tarkoitetaan editointiprojektin päätoimittajan tehtäviä. Niitä ovat mm.

- Aineiston kokoelmarakenteen luominen
- Uusien käyttäjien luominen
- Käyttöoikeuksien antaminen käyttäjille
- Aineistojen lukitseminen enemmiltä muutoksilta.

Taustajärjestelmä on suunniteltu niin, että se mahdollistaa toisistaan täysin irralliset projektit, so. aineistot ja käyttäjät eivät näy yli projektin rajojen. Käyttöoikeuksien hallinnointi perustuu kokoelmarakenteeseen. Toimittaja voi editoida aineistoja ainoastaan niissä kokoelmissa, joihin hänelle annetaan riittävät oikeudet. Päätoimittaja voi myös delegoida hallinnointioikeuksia. Hallintokäyttöliittymän toteutus on vielä työn alla (per 17.10.2013).

6.6 Aineistojen tuominen OCR-editorista

Aineistoja voi tuoda OCR-editorista sivu tai dokumentti kerrallaan. Käytännössä jokaista sivua vastaa yksi ALTO XML -tiedosto; tarvittaessa ne voidaan esittää peräkkäin. Editorin suunnittelun yhteydessä on keskusteltu siitä, missä muodossa aineiston pitäisi olla saatavissa. On ilmeistä, että XML-tiedosto on jatkokäsittelyn kannalta kiitollisin. Sen käyttö edellyttää hieman taitoa mutta on kaiken kaikkiaan vain tekstinkäsittelyä. XML-tiedostosta voi varsin helposti tuottaa ihmisluettavan version, siis pelkän tekstin; tai sanaluettelon; tai muutakin.

7. VIESTINTÄ

Pilottiprojektin viestintä on ollut projektipäällikön vastuualueella ja hän on vastannut viestinnällisistä sisällöistä ja tietojen paikkaansa pitävyydestä. Hänen käytössään ovat olleet Kansalliskirjaston ja Helsingin yliopiston yleiset viestintäkanavat. Viestintää on harjoitettu pääsääntöisesti suomeksi, mutta tarvittaessa sisältöjä on myös käännetty englanniksi ja venäjäksi.

7.1 Verkkosivut

Kansalliskirjaston verkkosivujen nk. siniselle kaistalle, jonne on kerätty tietoja Kansalliskirjaston kokoelmista ja palveluista, luotiin Sukukielten digitoointiprojektille oma osio Digitoitujen kokoelmien -välilehden alle. Verkkosivut löytyvät osoitteesta: [\[http://www.kansalliskirjasto.fi/kokoelmatjapalvelut/digitaalisetkokoelmat/finnougric.html\]](http://www.kansalliskirjasto.fi/kokoelmatjapalvelut/digitaalisetkokoelmat/finnougric.html)

Verkkosivuilla tarjottiin perustietoja hankkeesta ja sieltä pystyy lukemaan mm. pilottivaiheen projektisuunnitelman sekä digitoitavien aineistojen listat. Sisällöt toteutettiin myös englannin- ja venäjänkielisinä.

7.2 Kansalliskirjasto-lehti

Sukukielten digitointiprojektin pilottihanketta esiteltiin projektipäällikön kirjoittamien juttujen avulla myös Kansalliskirjasto-lehden numeroissa 4/2012 ja 4/2013 (tulossa). http://www.kansalliskirjasto.fi/attachments/5xaqbsRw7/6D1jBI3bh/Files/CurrentFile/KK_lehti_4_2012.pdf

7.3 Blogi

Sukukielten digitointiprojektin blogi on ollut tärkeä viestinnän väline, sillä se on tarjonnut Kansalliskirjaston verkkosivuja joustavamman foorumin keskustella ja tiedottaa hankkeen etenemisestä. Blogitekstejä on kirjoittanut pääsääntöisesti projektipäällikkö, mutta myös tutkijat ja venäläiset yhteistyökumppanit ovat voineet jakaa näkemyksiään blogissa. Blogi on englanninkielinen ja se sijaitsee osoitteessa <https://blogs.helsinki.fi/fennougrica/>

Kuva. Sukukielten digitointiprojektin blogi

Fenno-Ugrica
The Blog of the Digitization Project of Finno-Ugric Languages – National Library of Finland

1640
THE NATIONAL LIBRARY OF FINLAND

Home Contact Details Introduction to Fenno-Ugrica

New Interface for Uralica
Posted on 31.10.2013

Yes, finally, we have launched the new and fancy interface for the Uralica portal. As you may notice, its attractive design has something in relation with her siblings - the Finna information search service, which brings together the collections of Finnish archives,

RECENT POSTS

- New Interface for Uralica
- Udmurt material available in Uralica
- Uralica Beta Launched
- Planning for the future

7.4 Esitelmät

Projektipäälliköllä on ollut mahdollisuus esitellä hanketta erilaisissa kirjastoalan, tieteenhistorian ja digitaalisen kulttuurin tapahtumissa ja seminaareissa. Esitelmiä on pidetty mm. seuraavissa tilaisuuksissa:

- 8.8.2012, IFLA, Helsinki
- 22.9.2012, Venäjän kansalliskirjasto, Pietari
- 13.3.2013, Kirjaston tori, Helsinki
- 10.4.2013, Nekrasov-kirjasto, Moskova
- 12.4.2013, Udmurtian tasavallan kansalliskirjasto, Iževsk
- 16.5.2013, Karjalan tasavallan kansalliskirjasto, Petroskoi
- 5.6.2013, Variantti-kollokvio 2013: KORPUS-TIETOKANTA–EDITIO, Helsinki
- 6.6.2013, Fenno-Ugrica –kokoelman julkaiseminen, Helsinki
- 26.7.2013, International Congress of History of Science, Technology and Medicine, Manchester
- 23.10.2013, Viron kielen instituutti, Tallinna
- 14.11.2013, Historiallisen yhdistyksen seminaari, Helsinki.

7.5 Sosiaalinen media

Kansalliskirjaston Facebook-sivu toimi hyvänä viestintäkanavana Sukukielten digitointiprojektille. Pienemmälle ja rajatummalta yleisölle hankkeen etenemisestä tiedotettiin myös projektipäällikön oman Twitter-tilin kautta, jossa hankeviestintä tavoitti ainakin jonkin verran asiantuntijayleisöä nk. digital humanities -sektorilta. Hankkeen etenemisestä ilmoitettiin niin Facebookissa kuin Twitterissä uutisluontoisesti, eli ilmoitettiin uusien aineistojen saatavuudesta ja linkitettiin hankkeen blogiin. Uutisia hankkeen edistymisestä jaettiin sidosryhmien (mm. Uralistica, M.A. Castrénin seura, Sukukansojen ystävät jne.) Facebook-sivuilla.

7.6 Kick off -tilaisuus

Fenno-Ugrica -kokoelma julkaistiin 6.6.2013 Kansalliskirjaston Auditoriossa. Tilaisuudessa pidettiin kolme esittelyä: projektipäällikkö Jussi-Pekka Hakkarainen kertoi yhteistyöstä yleistietoja hankkeesta ja esitteli kokoelmaa, tutkija Jack Rueter puhui aineistoon kohdistuvasta tutkimuksesta ja tietojärjestelmäasiantuntija Juho Vuori esitteli OCR-editoria. Tilaisuuteen osallistui paikanpäällä yhteensä 67 henkilöä ja etäyhteyden välityksellä tilaisuutta seurasi yhteensä 43 henkilöä, niin Suomesta kuin Venäjältäkin.

Fenno-Ugrica -kokoelman julkaisemisesta lähetettiin STT:n kautta tiedotusvälineille suomen-, englannin- ja venäjänkielinen lehdistötiedote, mutta yksikään suomalainen media ei huomionnut

kokoelman julkaisua uutisissaan. Yhtenä syynä medianäkymättömyydelle oli, että tuolla samaisella viikolla ja samana päivänä julkaistiin useita avoimen datan palveluita ja julkisuus täyttyi pikemminkin Ilmatieteenlaitoksen ja Yleisradion hankkeista kuin Fenno-Ugrica -kokoelmasta: <http://essetter.blogspot.fi/2013/06/det-regnar-data.html#.UndrP1N9H-5>

8. BUDJETTI

Pilottihankkeen kustannukset vuosille 2012–2013 muodostivat Koneen Säätiöltä haettavan apurahan suuruuden. Koneen Säätiön myöntämän apurahan suuruus oli 356 040 euroa. Pilottihankkeen kustannukset esiteltiin maaliskuussa 2012 Koneen Säätiölle jätetyssä apurahahakemuksessa laskemina, jotka perustuivat selvitystyön aikana tehtyihin arvioihin, Ks. **Liite 1** raportin kohdasta *11. Liitteet*.

Vaikka arviot lähtökohtaisesti pitivät paikkansa hankkeen aikana, niin kaikkia kustannuseriä, kuten tekijänoikeuksista koituvia kustannuksia, ei tuolloin selvitystyön yhteydessä voitu esittää yksityiskohtaisesti, sillä selvitysvaiheessa ei ollut tietoa mahdollisista tekijänoikeuskorvauksien määrästä ja selvitystyön kustannuksista. Todellisista kustannuksista saatiin parempi ja yksityiskohtaisempi käsitys pilottihankkeen aikana ja budjetointiin liittyvät muutokset tuotiin aina ohjausryhmän keskusteltavaksi.

Projektilla ei ollut Kansalliskirjaston omarahoitusosuutta, sillä hankkeessa ei tuotettu aineistoja kirjaston omista kokoelmista. Omarahoitusosuudeksi voidaan katsoa Kansalliskirjaston osallistuminen projektin mm. tarjoamalla palvelintila Fenno-Ugrica -kokoelman tiedostoille ja ottamalla vastuun niiden pitkäaikaissäilytyksestä.

Keskeisimmät muutokset budjetoinnissa koskivat tekijänoikeuksienkorvauksia ja koulutuksia ja ostopalveluita. Korvauksia tekijänoikeuksista ei pilottiprojektin puitteissa maksettu, eikä suunniteltua koulutusta venäläisille kollegoille koskaan järjestetty. Ostopalveluja kertyi budjetoitua enemmän, mutta aineistoja digitoitiin enemmän kuin alun perin oli tarkoituksena, sillä marinkielisten rinnakkaisnimekkeiden määrä nosti hieman kustannuksia tekijänoikeuksien selvityksen ja digitoinnin osalta. Lisäksi projektin varoista maksettiin harvinaisen mordvalaisen sanomalehtiaineiston konservoinnista ja digitoinnista – tätä kuluerää ei oltu allokoitu alkuperäiseen budjettiin, mutta koska budjetissa oli joustonvaraa, niin konservoitavat sanomalehdet päätettiin projektin loppuvaiheessa tilata Venäjän kansalliskirjastolta.

Kuva. Budjettisuunnitelma toiminnoittain, kesä 2012.

	Vuosi 1.	Vuosi 2.	Yhteensä	KOKO PROJEKTI
	Täydentävä rahoitus (anotaan ulkopuoliselta taholta)	Täydentävä rahoitus (anotaan ulkopuoliselta taholta)	Täydentävä rahoitus (anotaan ulkopuoliselta taholta)	Perusmääräraha
PROJEKTIN KULUT TOIMINNOITAIN				
LASKUTUS VENÄJÄLTÄ	39150	114600	153750	153750
Digitointi	23000	100000	123000	123000
Tekijänoikeuskorvaukset ja -selvitystyö	10000	14600	24600	24600
Sopimukset ja lisenssit	6150		6150	6150
LAITTEISTO JA JÄRJESTELMÄ	0	20000	20000	20000
Julkaisujärjestelmä - ylläpito	0	13000	13000	13000
Julkaisujärjestelmä - kehittäminen	0	7000	7000	7000
KANSALLISKIRJASTON PALKKAKUSTANNUKSE	30706,25	62093,75	92800	92800
Tietojärjestelmäasiantuntija 10/2012-03/2013	15800	15800	31600	31600
Kuvailu- ja formaattiasiantuntija	0	21450	21450	21450
Projektipäällikkö 07/2012-10/2013	14906,25	24843,75	39750	39750
MUUT KULUT JA PALKKIOT	6000	37050	43050	43050
Verkkosivut	0	2460	2460	2460
Käännöspalkkiot	1000	3920	4920	4920
Seminaarit	0	2460	2460	2460
Matkat ja päivärahat	5000	15910	20910	20910
Koulutukset	0	12300	12300	12300
YHTEENSÄ	75856,25	233743,75	309600	309600
Tämä summa muodostuu projektien kuluista ilman yleiskustannuksia, 15%.				
306900 + 46440 = 356 040 euroa				

Budjetin vuositoteuma ei vastannut täysin kesällä 2012 tehtyä budjettisuunnitelmaa, ks. kuva yllä. Keskeisimmät syyt tähän olivat sekä rekrytoinnin että ostopalvelujen viivästyminen. Helsingin yliopiston asettaman rekrytointikiellon vuoksi hankkeeseen palkattu tietojärjestelmäasiantuntija pystyi aloittamaan työnsä vasta joulukuussa 2013, jolloin palkkakustannuksia kertyi vuoden 2012 osalta suunniteltua vähemmän. Lisäksi Venäjältä hankittujen palveluiden (digitointi ja tekijänoikeudet) maksut olivat vuoden 2012 osalta alhaisemmat kuin projektia aloittaessa oli allokoitu.

Kuva. Projektin toteuma, SAP-raportti vuoden 2013 osalta (per 31.1.2014)

WBS		Kustannuslaji	EUR
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Tuotot	-356 040,00
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Kulut	356 141,61
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Palkat yhteensä	97 442,46
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Henkilösivukulut	21 990,13
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Tilakustannukset	
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Aineet ja tarvikkeet	1 854,59
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Koneet ja laitteet	
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Ostetut palvelut	173 292,14
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Matkat	9 529,81
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Muut kulut	
4702964	Koneen Säätö/Suom-Ugrilaisen aineistoje	Yleiskustannukset	52 032,48
Kokonaistulos			101,61

9. TULOKSET JA YHTEENVETO

Sukukielten digitointiprojektin pilottihankkeen lähtökohdat olivat yhtä aikaa sekä kiinnostavat että ongelmalliset. Erityisen kiinnostavaksi aloitteen teki se, ettei koskaan aikaisemmin oltu pyritty saattamaan entisen Neuvostoliiton alueella painettuja aineistoja yhtä laajasti avoimeen kansalaiskäyttöön. Ongelmalliseksi aloitteen teki se, ettei sen paremmin Kansalliskirjastolla kuin kenelläkään muullakaan toimijalla ollut aikaisempaa kokemusta tämänluontoisen hankkeen toteuttamisesta ja sen reunaehdoista. Keskeiseksi haasteeksi koettiin ennen kaikkea aineistoihin koskevat tekijänoikeudet ja niiden selvittäminen, sillä aikaisempaa kokemusta Venäjän ja Suomen lainsäädännön soveltamisesta tällaisessa tapauksessa ei ollut. (ks. myös luku 4. Pilottivaiheen riskianalyysi). Haasteista huolimatta, hanke onnistui vähintäänkin kiitettävästi, ennen kaikkea hyvän valmistelevan suunnittelun ja osapuolten välisen yhteistyön ansiosta.

9.1 Asetettujen tavoitteiden saavuttaminen

Mikäli hankkeen onnistumista ja sen tuloksia tarkastellaan sille asetettujen tavoitteiden kautta, niin hanke saavutti sille asetetut odotukset lähes kokonaisuudessaan. Pilotoinnin tarkoituksena oli selvittää

- edellytyksiä Venäjän Kansalliskirjaston kokoelmissa olevien suomalais-ugrilaisia aineistojen digitaaliseen käyttöön saattamista,
- niihin kohdistuvien tekijänoikeuksien selvittämistä, ja
- aineistojen soveltuvuutta tutkimuskäytössä.

Näistä ylätasoon tavoitteista keskeisimmät, eli aineistojen käyttöön saattaminen ja tekijänoikeuksien selvitys onnistuivat erinomaisesti. Ainoastaan viimeisen tavoitteen osalta voidaan keskustella kriittisesti, missä suhteessa tavoite on täyttynyt, sillä projektin päättyessä tutkijat eivät ole voineet hyödyntää OCR-editoria omassa tutkimuksessaan. OCR-editorin hidastunutta suunnittelua ja sen valmistumista tuskin olisi voitu näissä olosuhteissa vauhdittaa, sillä pilottihankkeen käytössä olleet resurssit eivät mahdollistaneet kehitystyön jakamista useamman työntekijän kesken. Jatko-hankkeen osalta tällaisen työnjakoon on kuitenkin varauduttu. Tämä esiintuotu puute ei kuitenkaan kiistä sitä seikkaa, että pilotointi oli kokonaisuudessaan erityisen onnistunut.

9.2 Pilotoinnin aikana opittua

Mikäli hankkeen onnistumista tarkastellaan myös muutoin kuin yllämainittujen tavoitteiden valossa, niin pilottihankkeen voidaan sanoa ylittäneen sille asetetut tavoitteet. Tätä argumenttia tukevat ainakin seuraavat huomiot:

- tutkijayhteistyö kielentutkimuksen yhteisön kanssa koti- ja ulkomailla kasvoi hankkeen edetessä
- yhteistyö venäläisten kirjastojen kanssa muuttaa tekemisen standardeja pidemmällä aikavälillä
- kansainvälinen tekijänoikeusyhteistyö sai uuden ulottuvuuden
- Kansalliskirjaston asema luotettavana toimijana vahvistui venäläisten partnerien keskuudessa
- oman henkilökunnan ammatillista osaamista voitiin hyödyntää ja lisätä pilotoinnin aikana
- budjettikuri toteutui IT-projekteille harvinaisen hyvin – sitä ei ylitetty.

Hankkeen aikana on harjoitettu tiivistä tutkijayhteistyötä Kieliohjelman turvin työskentelevien kielentutkijoiden kanssa. Tämä yhteistyö on ollut hedelmällistä myös Kansalliskirjastoa ajatellen, joka ei ole aikaisemmin toiminut yhtä tiiviissä yhteistyössä kielitieteen edustajien kanssa. Ottaen huomioon, että Kansalliskirjasto on tällä saralla uudehko toimija, niin tiivis yhteistyö tutkijoiden kanssa on luonut vankkoja edellytyksiä yhteistyön jatkamiselle. Myös sellaiset kielentutkimuksen tieteelliset verkostot, jotka ovat olleet aikaisemmin Kansalliskirjastolle vieraita, ovat tulleet hankkeen aikana tutuimmiksi. Sukukielten digitointiprojektin pilottihankkeen kautta kotimainen tutkijakenttä on alkanut mieltää Kansalliskirjaston luontevaksi yhteistyökumppaniksi. Hyvänä esimerkkinä tästä tendenssistä voidaan mainita toisaalta OCR-editorin suunnitteleminen, mutta myös osallistuminen *digital humanities* -kentällä tapahtuvaan keskusteluun.

Jatkohankkeen toteutuessa yhteistyötä tiivistetään edelleen osallistumalla yhdessä tieteelliseen keskusteluun, konferensseihin, kansalaistieteiden järjestämiseen.

Kansalliskirjastolla on perinteisesti ollut vahva kokemus yhteistyöstä venäläisten kirjastojen kanssa. Viime vuosina aikana yhteistyö on ollut tiiveintä ennen kaikkea Venäjän kansalliskirjaston kanssa, mutta myös Karjalan tasavallan kansalliskirjasto on ollut keskeinen kirjastoalan yhteistyökumppani. Ennestään vahvat suhteet ovat saaneet Sukukielten digitointiprojektissa konkreettisia muotoja ja tuloksia, jotka ovat olleet kehittämässä venäläisten kirjastojen toimintaa tuleville vuosille. Ne toimintatavat ja -käytänteet, joita pilotoinnin aikana voineet venäläisille kirjastoille esitellä ja perustella ovat osaltaan avaamassa venäläisissä kirjastoissa olevia digitoituja aineistoja laajemmalle käyttäjäkunnalle. Olemme pilotoinnin aikana puhuneet venäläisille kirjastoille aineiston avoimesta saavutettavuudesta ja suositelleet useissa tapauksissa digitoitien saattamista avoimeen verkkokäyttöön. Puutteita on myös standardisoinnissa, sillä digitoitulle aineistolle ei ole useinkaan asetettu yhteisiä laatumäärittäjänsä, mikä on ollut esteenä aineiston hyödyntämiselle mm. kielentutkimuksen parissa. Samassa yhteydessä on pyritty myös lobbaamaan OCR-tekstin tuomista osaksi digitointiprosesseja, mikä ei aina ole ollut käytäntönä edes suurissa valtiollisissa kirjastoissa. Tässä suhteessa perustelut ja keskustelut aineistojen avoimesta verkkokäytöstä ovat mahdollistaneet asenneilmaston muutosta. Kansalliskirjastolla ja Sukukielten digitointiprojektilla on ollut selkä rooli eri kokonaisuuksia yhdistävänä organisaationa, jonka osaaminen digitointiin, digitaalisten aineistojen esittämiseen ja kokemus kansainvälisestä yhteistyöstä avaa maailmaa myös venäläisille toimijoille.

Asenneilmasto on muuttunut myös tekijänoikeudenselvityksen myötä. Sukukielten digitointihankkeessa tehty tekijänoikeuksienselvitys tehtiin tässä muodossaan ensi kertaa. Venäläisissä kirjastoissa on vallinnut skeptisyyttä tekijänoikeuksien selvittämistä kohtaan ylipäänsä, mutta hankkeessa voitiin osoittaa, että harjoitettu menettelytapa on mahdollinen. Pilottihankkeen aikana ilmeni, että tekijänoikeuksienselvityksen suorittanut järjestö, National Library Resource, saanee laissa määritellyn aseman venäläisen kirjastoalan lisenssiorganisaationa, mutta Venäjän duuma ei ole vielä päättänyt lakimuutoksesta. On siis mahdollista, että jatkossa kirja-alan tekijänoikeuksia selvitetään laajemminkin Sukukielten digitointiprojektissa suoritettulla tavalla. (Ks. myös National Library Resourcen johtajan kirjoitus selvityksestä hankkeen blogissa: <https://blogs.helsinki.fi/fennougrica/2013/05/20/gordian-knot-of-copyrights/>)

Tekijänoikeuksien selvityksen ja asenneilmaston tuulettamisen seurauksena Kansalliskirjaston asema luotettavana ja vastuullisena toimijana venäläisten kirjastojen silmissä on vahvistunut pilotoinnin aikana. Tämä saavutettu maine ja luottamus on hyödyksi myös tulevaisuuden projektin kannalta, sillä hankkeessa on luotu edellytyksiä laajemmallekin kirjastoalan yhteistyölle.

Projektin aikana on pyritty kasvattamaan myös Kansalliskirjaston omaa osaamista. Periaatteena on ollut, ettei hanke kuluttaisi Kansalliskirjaston resursseja, vaan myös kirjasto hyötyisi osaamisen kasvusta, eli osaavan henkilökunnan hyödyntäminen hankkeessa täytyy tuottaa jotain uutta Kansalliskirjastollekin. Pilottivaiheessa käytettiin hyväksi ennen kaikkea Slaavilaisen kirjaston (Slavica) ja kirjastoverkkopalveluiden osaamisvarantoja. Vakituisen henkilökunnan työpanosta siirrettiin Sukukielten digitointiprojektin myönnöltä työpanosta vastaava korvaus niin kirjastoverkkopalveluille kuin Slaavilaiselle kirjastollekin. Slaavilaisessa kirjastossa kompensatorahalla työskenteli ylimääräinen kirjastosihtööri ja kirjastoverkkopalveluissa voitiin palkata yksi tietojärjestelmäasiantuntija lähes puoleksi vuodeksi.

Kuva. Kansalliskirjaston intranet.

Onnistujan nimi: Sukukielten digitointiprojekti & Jussi-Pekka

Hakkarainen (18.6.2013)

Hankkeiden ja perustoiminnan nivominen yhteen ei aina ole helppoa. Hankkeilla on omat tavoitteensa, aikataulunsa ja rahoituksensa. Perustoiminta taas pyörii omaa rytmiään, usein varsin niukoin resurssein. Hankkeet eivät kuitenkaan ole muusta toiminnasta irrallisia ja kerrallisia juttuja, vaan niistä pitää aina jäädä perustoiminnan laariin jotain pysyvää.

Sukukielten digitointihankeen ja Slavicin toiminnan sovittaminen yhteen on sujunut hyvin, kummankin hyödyksi. Hankkeessa on tarvittu Slavicassa olevaa kuvailuosaamista ja venäjän kielen taitoa: Marina Ivanova on luetteloanut digitoidut kirjat ja kääntänyt käyttöliittymän venäjäksi. Samalla digitaalisen kirjaston osaaminen on kasvanut. Kompensaatioksi Marinan työstä Slavicassa on voitu palkata Erna Vuori luetteloimaan käsittelemättä jääneitä lahjoituksia. Näin asiakkaille saadaan uutta aineistoa esille tilanteessa, jossa hankintarahat ovat puolittuneet. Onnistumisen taustalla on hyvin tehty hankesuunnittelu.

9.3 Tulokset suhteessa Koneen Säätiön Kieliohjelmaan

Pilotoinnin onnistumista on syytä arvioida myös Koneen Säätiön Kieliohjelmassa esitettyjen tavoitteiden ja pyrkimysten valossa. Koneen Säätiön kieliohjelman päätavoitteena on edistää pienten suomalais-ugrialaisten kielten, suomen sekä Suomen vähemmistökielten dokumentointia. Lisäksi aineistojen saatavuuden, saavutettavuuden ja käytettävyyden parantamisen osalta

Kieliohjelman tavoitteena on saattaa sekä vanhoja että uusia kieliaineistoja tiedeyhteisön ja muun yhteiskunnan avoimeen käyttöön. Erikseen Kieliohjelmassa mainitaan, että ”Pietarissa sijaitsevassa Venäjän kansalliskirjastossa on runsaasti suomalais-ugrilaisilla pienillä kielillä kirjoitettuja aineistoja, kuten Neuvostoliiton alkuvuosina tehtyjä oppikirjoja, valistuskirjallisuutta, kaunokirjallisuutta, lehtiä ja aikakauskirjoja. Kieliohjelman tavoitteena on saada esimerkiksi näitä aineistoja avoimeen käyttöön.”

Sukukielten digitointiprojekti on pystynyt vastaamaan näihin Kieliohjelmassa esitettyihin tavoitteisiin ja haasteisiin kiitettävällä tavalla. Pilottihankkeen aikana digitoitu aineisto on saatettu avoimeen kansalaiskäyttöön Fenno-Ugrica -kokoelmassa ja näin on tuotettu uusia kieliaineistoja sekä tutkijayhteisön, että yhteiskunnan avoimeen käyttöön.

Saavutettavuutta ja käytettävyyttä on pyritty käyttöön asettamisen yhteydessä huomioimaan mahdollisimman laajasti. Fenno-Ugrican käyttöliittymä on käännetty myös venäjäksi ja julkaisuarkistoa on pyritty optimoimaan venäläisten hakukoneita varten, jolloin aineiston käytettävyys ja saavutettavuus on parantunut. Lisäksi aineiston kuvailuun, sen metadataan, on kiinnitetty erityistä huomiota.

Sukukielten digitointiprojektissa on pyritty vastaamaan myös sirpaleisen tiedon yhteen saattamiseen ja auttamaan niin tutkijoita kuin muitakin käyttäjiä löytämään sopivia kieliaineistoja. Viime aikoina suomalais-ugrilaisten aineistojen digitointi Venäjän federaation alueella toimivissa kansalliskirjastoissa on kasvanut merkittävästi viimeisten vuosien aikana. Monet venäläiset kirjastot ovat muodostaneet omia digitaalisten aineistojen kokoelmiaan ja antaneet yleisölle joko avoimen tai rajatun mahdollisuuden käyttää aineistoa. On ajateltu, että sekä uusien että vanhojen suomalais-ugrilaisten aineistojen digitointi ja aineistojen käyttöön saattaminen tukee sukukielten puhujien kielen oppimista, sillä näiden kielten voidaan ajatella edustavan omien yhteisöjensä, kulttuuriensa ja perinteidensä tärkeää voimavaraa ja maailmankuvaa, ja kulttuuri, perinteet ja maailmankatsomus välittyvät parhaiten äidinkielen vapaassa kielenkäytössä.

Samaan aikaan kuitenkin tiedonjakaminen ja -harmonisointi ovat jääneet digitointi-innon jalokoihin ja venäläisten kansalliskirjastojen kanssa käytyjen keskustelujen myötä on tullut selväksi, ettei tällä hetkellä kukaan kontrolloi kuinka paljon ja mitä suomalais-ugrilaisia aineistoja Venäjän federaation alueella digitoidaan – tämä vaikeuttaa aineistojen saavutettavuutta, digi-

tointiprojektien järjestelmällistä suunnittelua ja voi johtaa päällekkäiseen työhön, jolloin resursseja käytetty tehottomasti.

Samaa ongelmaa kuvastaa aineiston sirpaleinen saavutettavuus. Loppukäyttäjien näkökulmasta monien käyttöliittymien käyttö ja niiden hallinta vaatii aikaa ja tietoteknistä osaamista, mutta myös valtakielen hallitsemista. Ei ole olemassa järjestelmää, minkä avulla kaikkia digitoituja suomalais-ugrilaisia aineistoja voitaisiin etsiä, selata ja tarvittaessa jopa käyttää.

Edistääkseen Koneen Säätiön Kieliohjelmassa kuvattuja pyrkimyksiä Kansalliskirjasto toteutti Fenno-Ugrica -kokoelman lisäksi Opetus- ja kulttuuriministeriöltä saadulla avustuksella yhteistyöhankkeen, jonka tavoitteena oli luoda yhteinen ja avoin tietojärjestelmäinfrastruktuuri eri kirjastoissa digitoituille suomalais-ugrilaisille kieliaineistoille. Uralica-portaali avattiin elokuussa 2013 ja siihen linkitetty digitoituja aineistoja Karjalan tasavallan kansalliskirjaston (Petroskoi) ja Udmurtian tasavallan kansalliskirjaston (Iževsk), Göttingenin valtion ja yliopiston kirjaston (*Niedersächsisches Staats- und Universitätsbibliothek*) ja Viron kielen instituutin (*Eesti Keele Instituut*, Tallinna) kokoelmista.

10. SUOSITUKSET JATKOHANKKEEN TOTEUTTAMISTA VARTEN

Seuraavassa esitellään sellaisia suorituksia mahdollisen jatkohankkeen toteuttamista ajatellen, joiden huomioonottaminen olisi tarkoituksenmukaista hankkeen hyvälle toteuttamiselle.

Jatkohankkeen aikana on tarkoitus digitoida ja saattaa käyttöön lähes 1100 monografia- ja 51 sanomalehtinimekettä. Monografiasivuja suunnitelman mukaisesti kertyy noin 88 300 ja sanomalehtisivuja noin 72 500. Digitoitavat aineistot on valikoitu yhteistyössä kotimaisen tutkijakunnan kanssa ja sen on katsottu palvelevan niin kotimaista kuin ulkomaista fennougristiikan alan tutkimusta. Hankkeen toteutuessa aineistot muodostaisivat maailman suurimman uralilaisien kielten resurssin, joka toisi tutkijakunnan ulottuville sellaisia kieliaineistoja, joihin aikaisemmin heillä ei ole ollut mahdollista tutustua ja johon kaikilla käyttäjillä on asuinpaikasta riippumatta avoin pääsy. Aineisto saatetaan sekä tutkijoiden hyödynnettäväksi että avoimeen kansalaiskäyttöön Kansalliskirjaston ylläpitämässä Fenno-Ugrica -kokoelmassa (fennougrica.kansalliskirjasto.fi)

Jatkohankkeen aineistojen kohdalla on pyritty painottamaan ennen kaikkea hyvin vähän digitoituja sekä vaikeasti saavutettavia alueellisia ja periferisiä julkaisuja. Sisällön puolesta keskuksen

ulkopuolella sijaitsevien alueiden kieli (paikallislehdet) on keskuksen kieltä mielenkiintoisempaa, koska periferiassa voi ilmetä joko keskuksen kielestä poikkeavaa murrevariaatiota tai konservatiivisuutta myös kirjoitetussa kielessä. Paikallisten sanomalehtien digitointia puoltaa myös pyrkimys edistää aineiston saavutettavuutta: paikallismateriaaliin keskittymällä tuodaan tutkijoiden ulottuville aikaisemmin huonosti saatavaa materiaalia ja digitoida sellaisia sanomalehtiä, joiden asema venäläisten kirjastojen digitointisuunnitelmissa on joko hyvin marginaalinen tai jopa poissaoleva.

10.1 Toteutustapa

Jatkohankkeen osalta toteutustapa noudattelee pääsääntöisesti pilottihankkeen aikana luotuja ja testattuja käytänteitä muutamien pienin poikkeuksin. Aineisto digitoidaan pääsääntöisesti Venäjän Kansalliskirjaston (Pietari) kokoelmista ja aineistoja koskevia tekijänoikeuksia selvitetään yhteistyössä National Library Resource (Moskova) kanssa. Aineisto luovutetaan vain TIFF-muotoisena Kansalliskirjastolle jatkokäsittelyä varten. Jatkokäsittely suoritetaan joko Kansalliskirjaston Digitointi-, mikrofilmaus- ja konservointiyksikössä Mikkelissä tai yhteistyössä tutkijoiden kanssa. Tavoitteena on rakenteellista kuvamuotoista aineistoa, eli siihen lisätään OCR-tunnistus ja palstoitus. Rakenteellistettu aineisto siirretään PDF-muotoisena osaksi digitaalista Fenno-Ugrica -kokoelmaa, missä aineisto myös kuvaillaan. Tutkijat voivat hyödyntää palstoitettua aineistoa rikastaessaan materiaalia OCR-editorin avulla. OCR-editorilla korjattu kielimateriaali luovutetaan edelleen FIN-CLARINin ylläpitämään Kielipankkiin ja se on siten myös muun tutkijayhteisön käytettävissä.

Sukukielten digitointiprojektia johtaisi Suomen Kansalliskirjasto ja se vastaa yhteistyöstä ja sen koordinoinnista kansainvälisten ja kotimaisten toimijoiden kanssa. Keskeisimpiä ulkomaisia toimijoita ovat Venäjän Kansalliskirjasto Pietarissa, jonka kokoelmista suurin osa Sukukielten digitointiprojektin aineistoista digitoidaan. Koska jo aineistojen digitointiin vaaditaan tekijänoikeuksien haltijoiden lupa, niin digitoitavien aineistojen tekijänoikeuksia selvittää molempien kansalliskirjastojen toimeksiannosta moskovalainen tekijänoikeusjärjestö National Library Resource. Toimijoiden välinen työnjako noudattelee pilottivaiheen aikaisia käytäntöjä ja määrittelee toimijoiden vastuualueita: Venäjällä selvitetään paikalliset tekijänoikeudet ja tutkijoiden valitsevat aineistot digitoidaan venäläisistä kokoelmista, mutta aineisto saatetaan käyttöön Suomessa. Tämä tuotantomalli on kansainvälisesti ainutlaatuinen ja avaa uusia yhteistyömahdollisuuksia Venäjän ja lännen välillä niin kirjastojenvälisellä kuin humanistisen tutkimuksen kentällä yleensä.

10.2 Jälkikäsitteily

Yksi jatkohankkeen tavoitteista koskee Venäjän federaation alueella tehtävää digitointia ja sen laatua. Yhteistyön avulla suomalais-ugrialaisten aineistojen digitoinnille voitaisiin luoda yhdessä määritellyt laatuksiteerit, sillä yhtenäinen, standardinmukainen digitointijälki ja kuvailutieto mahdollistavat myös aineiston pitkäaikaissäilytyksen ja turvaa sen säilyvyyden tuleville sukupolville.

Sukukielten digitointiprojektilla on kiinnekohtia myös kieliteknologiseen tutkimukseen, sillä projektin yhdeksi päämääräksi voidaan laajasti ajatella digitaalisten kirjasto- ja arkistoaineistojen käyttötapojen ja käytettävyyden parantaminen. Projektissa edistetään suomalais-ugrilaisen aineiston käyttöön saattamisen lisäksi menetelmiä, joilla digitoitua raakadataa voidaan jalostaa entistä käyttökelpoisemmiksi aineistoiksi ja joilla tällaista aineistoa voidaan hyödyntää. Näillä menetelmillä tarkoitetaan digitoidun aineiston OCR-tunnistuksen lisäämistä, tunnistetun tekstimassan palstoittamista sekä ennen kaikkea kielenkorjaukseen tarkoitettun OCR-editorin jatkokehittämistä, jonka avulla voidaan korjata digitoinnin ja OCR-tunnistuksen yhteydessä jääneitä virheitä. Pilottivaiheen aikana kehittämä OCR-editorin edelleen kehittäminen tukisi siis myös digitointialan kokonaisvaltaista kehittämistä

Projektin pilottivaiheen kuluessa käytettiin hyvin paljon työtä Pietarista saatujen PDF-dokumenttien analysoinnissa ja ALTO XML-tiedostojen tuottamisessa. Jatkossa on pyrittävä saamaan digitoitu aineisto mieluiten digitoijan tuottamana ALTO XML:nä. Formaatti on digitoinnissa käytettävien ohjelmistojen yleisesti tukema.

Pilottiprojekti on ollut verraten erillinen osa Kansalliskirjaston kehitystoiminnassa. Siinä ei ole pystytty täysin käyttämään hyväksi kirjaston asiantuntijoita esim. käytettävyys- ja graafisen suunnittelun, laitteistopalvelujen alalla - ja yleensä tukena ja ajatusten kaukupohjana suunnittelun ja toteutuksen aikana. Projektin jatkuessa työn järjestelyyn ja johtamiseen on näiltä osin löydettävä toisenlaisia ratkaisuja. Työssä voisi myös paremmin käyttää hyväksi Kansalliskirjaston tehtävienhallinnan (JIRA) ja muita välineitä.

OCR-editori lienee tällaisenaan ainoa lajissaan, mutta tähän mennessä siihen kehitetyt toiminnallisuudet eivät vastaa vielä tutkijakunnan tarpeita kokonaisuudessaan, joten jatkohankkeessa tulisi tätä työkalua edelleen kehittää yhteistyössä niin tutkijoiden kuin eri hankkeiden kanssa jatkohankkeen aikana. OCR-editorin kehitystyö tehtäisiin Kansalliskirjaston kirjastoverkkopalveluissa ja kehitystyöhön tulisi palata jatkohankkeen varoista toinen tietojärjestelmäasiantunti-

ja, jonka työtehtävät liittyisivät editorin jatkokehittämiseen yhdessä jo käytettävissä olevan resurssin kanssa.

10.3 Tutkijayhteistyö ja kansalaistieteet

Jatkohankkeessa digitoitavaksi esitetty aineisto on valittu yhdessä tutkijakunnan kanssa. Tuottamalla tämän aineiston Kansalliskirjasto vahvistaa omaa tutkijayhteistyötään kielentutkimuksen partnerina, mutta vahvistaa myös yhteistyötä venäläisen kirjastokentän kanssa. Sukukielten digitointiprojektin jatkohankkeen keskeisiä partnereita ovat erityisesti Koneen Säätiön Kieliohjelman apurahalla työskentelevät tutkijat ja tutkijayhteisöt. Jatkohankkeen kannalta keskeisessä asemassa ovat seuraavat tutkimushankkeet ja -projektit:

- Turun yliopisto, Volgan alueen kielten tutkimusyksikkö: ”Marin kielen sanaston kehitys 1920- ja 1930-luvuilla”
- Helsingin yliopisto, Jack Rueter ja työryhmä: ” Morfologisten jäsentimien luominen suomalais-ugrilaisille vähemmistökielille”
- Helsingin yliopisto, Riho Grünthal ja työryhmä: ”Itämerensuomalaisten kielten muutos monikielisessä ympäristössä”
- Helsingin yliopisto, Suomalais-ugrilaiset kielet ja internet -hanke.
- Kotimaisten kielten tutkimuskeskuksen vanhan kirjasuomen sanakirjat
- Kirjallisuuspankki-hanke
- FIN-CLARIN
- HFST (Helsinki Finite-State Transducer Technology)
- Tromssan yliopisto, Giellatekno

Kotimaisen tiedeyhteistyön on tarkoitus olla pelkkää aineiston tuottamista laajempaa ja ylittää tieteiden välisiä rajoja. Ylirajaista yhteistyötä jatketaan jatkohankkeessa OCR-editorin tiimoilta, jota on kehitetty pilottihankkeessa vuoden 2012 lopulta lähtien. Editoria hyödynnetään myös eräissä muissa Koneen Säätiön rahoittamissa hankkeissa, kuten Kotimaisten kielten tutkimuskeskuksen vanhan kirjasuomen sanakirjan 3. osan loppuunsaattamiseen tähtäävässä hankkeessa. Tarkoituksena on, että OCR-korjattu vanhan kirjasuomen aineisto liitetään Kotuksen tekstikorpuskokoelmaan ja tuodaan muiden korpusten tavoin julkisesti saataville.

Kirjallisuuspankki on Suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitoksen, Kansalliskirjaston ja Helsingin yliopiston kirjaston yhteishanke. Se on kattava ja lähdekriittinen kokonaisuus suomen- ja ruotsinkielistä kaunokirjallisuutta sekä siihen liittyvää tutkimustietoa, kulttuurihistoriallinen resurssi koko maan tutkimukselle, koulutukselle

ja kansansivistystyölle. Kirjallisuuspankin aineisto on niin ikään kaikkien vapaasti käytettävissä internetin välityksellä. Kirjallisuuspankissa julkaistavien digitoitujen klassikkotekstien stilisointi on tarkoitus suorittaa OCR-editoria hyödyntäen.

Pilottihankkeessa on rakennettu kansainvälistä yhteistyöverkostoa ja jatkohankkeen aikana tulisi yhteistyötä syventää ainakin seuraavien toimijoiden kanssa projektin eri osa-alueilla:

- Venäjän kansalliskirjasto, Pietari
- Venäjän Valtiollinen kirjasto, Moskova
- National Library Resource, Moskova
- Udmurtian tasavallan kansalliskirjasto, Iževsk
- Karjalan tasavallan kansalliskirjasto, Petroskoi
- Viron kielen instituutti, Tallinna
- Tarton yliopisto, Tartto
- Valtion- ja yliopistonkirjasto, Göttingen
- Max Planck-Institut für Wissenschaftsgeschichte, Berliini
- Suomen Moskovan-suurlähetystö
- Hanti-Mansian tasavallan kansalliskirjasto, Hanti-Mansijsk
- Komin tasavallan kansalliskirjasto, Syktyvkar
- Marin tasavallan kansalliskirjasto, Joškar-Ola
- Mordovian tasavallan kansalliskirjasto, Saransk

Sukukielten digitointiprojektin jatkohankkeen yksi keskeinen tavoite on luoda kansalaistieteiden verkosto Fenno-Ugricassa olevien aineistojen ja jatkohankkeen kumppanuuksien ympärille. Tarkoituksena on kerätä kansalaistieteilijöiden joukko, joka pystyvät toimimaan paikallisissa olosuhteissa talkoistamisen ohjaajina niin oppilaitoksissa kuin vaikkapa alueellisissa kirjastoissa. Tällä tavoin Sukukielten digitointiprojekti pyrki jatkohankkeen aikana talkoistamaan Fenno-Ugricassa olevan aineiston käsitellen sitä OCR-editorilla.

Sopivien kansalaistieteilijöiden kouluttamista Venäjän federaation alueella pyritään mahdollistamaan Suomalais-venäläisen kulttuurifoorumin (<http://www.kultforum.org>) kautta. Vuonna 2014 suomalaiset partnerit saavat ehdottaa yhteistyöhankkeita ja Kansalliskirjasto pyrkii mahdollistamaan kansalaistieteiden levittämisen luomalla tarvittavat partneriverkostot niin paikallisten kirjastojen kuin koulujenkin kanssa. Sopivien partnerien paikantamisessa työskennellään yhteistyössä tutkijakunnan kanssa.

11. LIITTEET

Liite 1 – projektisuunnitelma

Tiivistelmä

Tässä projektisuunnitelmassa esitetään tuotantovälineet ja –prosessi, organisaatio (**Liite 1**) ja kustannusarvio (**Liite 2**) suomalais-ugrialaisten aineistojen (**Liitteet 3, 4 ja 5**) digitoinnin pilottihankkeelle vuosille 2012–2013. Pilottihanke on osa Koneen Säätiössä valmisteltua kieliohjelman hankekokonaisuutta ja siihen viitataan Koneen Säätiön Kieliohjelmassa 2012–2016 otsikolla *Pietarin aineistojen digitoinnin pilottihanke (2012–2013)*, tästedes tässä yhteydessä vain *pilottihanke*.

Pilottihanketta varten digitoitava aineisto julkaistaan Kansalliskirjaston ylläpitämässä ja avoimeen lähdekoodiin pohjautuvassa julkaisujärjestelmässä, jossa esitettävä digitaalinen aineisto saatetaan avoimeen kansalaisyhteisöön sekä taataan tiedeyhteisölle mahdollisuus näiden kieliaineistojen hyödyntämiseen tutkimuksessa. Lisäksi aineiston käyttöön asettamisen yhteydessä luodaan tarkoituksenmukaiset välineet ja edellytykset talkoistamispalveluille.

Selvityksessä ehdotetaan työnjaoksi, että pilottihankkeen toteuttamisesta ja koordinoinnista vastaa erikseen nimitettävä projektipäällikkö, joka sijoittuu Kansalliskirjastoon ja raportoi toiminnasta suoraan pilottihankkeen ohjausryhmälle. Projektipäällikkö vastaa Venäjän Kansalliskirjastossa (Pietari) tehtävän digitoinnin laatumääritysten noudattamisesta. Hän valvoo aineiston käyttöön asettamista Kansalliskirjaston julkaisujärjestelmässä ja aineistoa koskevien tekijänoikeussäädösten noudattamista sekä koordinoi Kansalliskirjastossa suoritettavia työvaiheita. Lisäksi pilottihankkeen viestintä kuuluu projektipäällikön työtehtäviin yhteistyössä Kansalliskirjaston viestintäyksikön kanssa. Julkaisujärjestelmän ylläpidosta ja teknisestä kehittämisestä huolehtii Kansalliskirjaston Kirjastoverkkopalvelut ja kuvailu- ja formaattityöstä Kansalliskirjaston Tutkimuskirjastopalvelut. Kaikki projektista syntyvät kustannukset, mukaan lukien palkkakulut, katetaan Koneen Säätiöltä anottavalla apurahalla. Kansalliskirjasto huolehtii työntekijöiden työvälineistä, työpisteistä ja muusta välittömästä infrastruktuurista.

Selvityksen perusteella Koneen säätiöltä anotaan 356 040 euron suuruinen apuraha pilottihankkeen toteuttamiseksi aikavälillä heinäkuu 2012–joulukuu 2013. Pilottihankkeen avulla kehite-

tään toimintatapaa ja teknistä toteutusta, joita on tarkoitus hyödyntää muiden aineistojen suhteen jatkohankkeessa vuosina 2013–2016.

Osapuolet ja organisaatio

Kansalliskirjasto tuottaa Koneen Säätiön kieliohjelmassa tarvittavan tutkimusaineiston ja infrastruktuurin Koneen Säätiöltä haettavalla apurahalla. Kansalliskirjasto vastaa myös yhteistyöstä Venäjän Kansalliskirjaston (Pietari) ja Helsingin yliopiston kirjaston kanssa. Kansainvälistä työnjakoa ohjaa Kansalliskirjaston ja Venäjän Kansalliskirjaston välinen yhteistyösopimus, jonka perusteella pilottihanketta varten neuvotellaan erikseen keväällä 2012 kirjastojen välinen sopimus, missä määritellään hankkeen puitteissa toteutettavat kummankin osapuolen vastuut ja velvollisuudet ehtoineen. Aineistojen tekijänoikeuksien selvittämisestä ja korvaamisesta sovietaan erikseen Kansalliskirjaston ja venäläisen kirjastoalan yhteisen tekijänoikeuksia selvittävän toimiston National Library Resourcen (NLR, venäjäksi Национальный библиотечный ресурс, НБР) välillä.

Pilottihanketta varten luodaan ohjausryhmä. Ohjausryhmä koostuu kotimaisten osapuolten ja tieteen tekijöiden edustajista siten, että myös kirjastoalan erityisosaaminen on ohjausryhmässä edustettuna. Kansalliskirjasto apurahan vastaanottajana vastaa pilottihankkeen raportoinnista ja tulosten esittelemisestä Koneen Säätiölle.

Ohjausryhmän on tarkoitus tukea hanketta varten palkattavaa projektipäällikköä, joka raportoi ohjausryhmälle projektin edistymisestä. Projektipäällikön sijoittuminen Kansalliskirjaston organisaation sisälle antaa hänelle parhaat mahdollisuudet toteuttaa pilottihanketta sekä valvoa hankkeen työvaiheita ja edistymistä. Hankkeen alkuvaiheessa, alkaen heinäkuusta 2012 projektipäällikkö vastaa pilottihankkeen yleisestä suunnittelusta, budjetoinnista, sopimusten solmimisesta hankkeen eri osapuolien ja muiden tarvittavien tahojen kanssa sekä alkuvaiheen tietojärjestelmäkoulutuksesta ja yhteistyökäytänteiden sopimisesta niin kotimaisten kuin ulkomaisten osapuolien kanssa. Myöhempisiin projektipäällikön tehtäviin kuuluvat digitoitavan aineiston laadunvalvonta, työprosessien suunnittelu, käyttöön asettamisen valvonta, talkoistamispalveluiden käyttäjähallinta sekä tekijänoikeuksien valvonta. Projektipäällikkö vastaa myös yhteistyöstä tieteen tekijöiden kanssa. Projektin viestintä ja sen järjestäminen yhteistyössä Kansalliskirjaston viestinnän kanssa kuuluvat niin ikään projektipäällikön tehtäviin. Laskutuksiin ja rahaliikenteeseen liittyen projektipäällikkö työskentelee yhteistyössä Kansalliskirjaston taloushallinnon kanssa.

Projektipäällikön vastuiden ja työtehtävien lisäksi hankkeen eri vaiheissa tarvitaan erilaista kirjastoalan teknistä osaamista, jollaisen osaamisen avulla julkaisujärjestelmien ylläpitoa ja kehittämistä voidaan toteuttaa. Lisäksi pilottihankkeessa tarvitaan kirjastoalan kuvailu- ja metadatakokemusta, erityisesti kohdistuen hankkeen poikkeaviin metadatakäytänteisiin ja formaatteihin. Projektipäällikön arvioitu henkilötyövuosikertymä pilottihankkeen aikana (heinäkuu 2012-joulukuu 2013) on 8 henkilötyökuukautta (htkk), teknisestä ylläpidosta ja kehittämisestä vastaavan henkilön 6 htck (syyskuu 2012-joulukuu 2013) ja kuvailu- ja formaattiasiantuntijan 4 htck (syyskuu 2012-joulukuu 2013). Organisaatio- ja prosessikaavio esitetty **liitteessä 1**.

Aineisto

Koneen säätöön edustajat ovat syksyllä 2011 määritelleet pilottihankkeessa digitoitavan suomalais-ugrilaisen aineiston, joka kuuluu Venäjän kansalliskirjaston (Pietari) kokoelmiin. Tämä tutkijoiden valitsema suomalais-ugrilaisilla kielillä julkaistu aineisto digitoidaan Pietarissa Venäjän Kansalliskirjastossa ja toimitetaan Kansalliskirjastolle digitaalisessa muodossa. Valikoitu aineisto sisältää noin 7000 sivua inkeröisten- ja vepsänkielisiä julkaisuja, yhteensä 66 monografiaa, jotka ovat pääsääntöisesti oppikirjoja ja sanakirjoja. Monografioiden lisäksi digitoidaan lähes 20 000 sivua marilaisia ja mordvalaisia sanomalehtiä, jotka on julkaistu pääsääntöisesti 1920- ja 1930-luvuilla. (Ks. aineiston tarkempi kuvaus **liitteissä 3, 4 ja 5.**)

Pilottihankkeen aikana digitoitavan aineiston teknisiä vaatimuksia on määritelty yhdessä Venäjän Kansalliskirjaston kanssa. Koska aineistoja ei voida luovuttaa Suomeen digitoitavaksi, digitointi suoritetaan Venäjän Kansalliskirjastossa ja heidän digitoimansa aineistot toimitetaan edelleen Kansalliskirjastolle. Monografiat digitoidaan harmaasävyskannattuna 300 dpi tarkkuuteen, kun taas sanomalehtien resoluutioksi on haluttu määritellä 400 dpi. Primääriskannaus suoritetaan TIFF-kuvaformaattiin ja skannattujen kuvien yhteyteen lisätään automaattisen tekstin tunnistamiseen (OCR, optical character recognition) vaadittavat elementit. OCR-digitoitu teksti liitetään kuvatiedostoon ja nämä elementit liitetään yhdeksi tiedostoksi, PDF-formaattiin, joka on myös aineiston esitysformaatti julkaisujärjestelmässä. Mahdollista myöhempää käsittelyä varten myös TIFF-muotoiset kuvat luovutetaan Kansalliskirjastolle.

Venäjän Kansalliskirjasto toimittaa myös digitoitavien aineistojen bibliografiset tiedot Kansalliskirjastossa (julkaisujärjestelmässä) tapahtuvaa aineistonkuvailua varten. Bibliografisia tietoja ei sellaisenaan voida hyödyntää julkaisujärjestelmässä, sillä venäläiset kirjastot käyttävät

aineistojen kuvailuformaattina RusMarcia, kun taas pilottihankkeessa käytettävä digitaalisten aineistojen kuvailuformaatti on DublinCore, joka on digitaalisten julkaisujen kuvailuun kehitetty metadataformaatti. Luetteloinnin yhteydessä bibliografista metadataa muokataan vastaamaan DublinCoren kuvailumääritelmiä digitaalisesta aineistosta. Pilottihankkeessa digitoitavaa aineistoa ei luetteloita paikallistietokantoihin, vaan aineiston saavuttavuutta edistetään mahdollistamalla sen löydettävyys Kansallisen Digitaalisen Kirjaston (KDK) ja mahdollisesti Euroopan käyttöliittymissä, niiltä osin kuin se on tekijänoikeuksien puitteissa mahdollista.

Järjestelmät ja rajapinnat

Pilottihankkeessa digitoitu aineisto julkaistaan Kansalliskirjaston ylläpitämässä julkaisujärjestelmässä, joka perustuu avoimen lähdekoodin DSpace-ohjelmistoon. Pilottihanketta varten perustetaan muista Kansalliskirjaston digitaalisista kokoelmista erillinen DSpace-instanssi (julkaisujärjestelmäympäristö), jonka avulla voidaan vastata niin tutkijakunnan kuin kansalaistieteen tarpeisiin. Rääätälöidyn DSpace-instanssin myötä kehitetään kansalaistieteen työkaluja sekä luoda paremmat edellytykset hanketta varten kehitettäville toiminnallisuuksille. Eräs pilottihankkeen aikana toteutettava kansalaistieteen työkalu on ocr-tekstieditori, jonka avulla julkaistun aineiston ocr-tekstiä voidaan korjata ja muokata julkaisujärjestelmässä. Työkalun kehittämisessä Kansalliskirjasto ja tutkijat toimivat yhteistyössä.

Pilottihankkeen rääätälöidyn DSpace-instanssin avulla edistetään digitoidun aineiston saavutettavuutta ja käyttöä myös ulkomailla, niin kansainvälisen tiedeyhteisön keskuudessa kuin Venäjällä asuvien suomensukuisten kielten puhujien keskuudessa. DSpace-instanssi toteutetaan myös venäjänkielisenä ja sen käyttöliittymä voidaan mahdollisesti julkaista myös muilla kielillä. Viestinnän ja tunnettuuden edistämiseksi pilottihankkeen DSpace-instanssille luodaan rääätälöity ja tunnistettava ulkoasu. Näin toteutettu julkaisujärjestelmä tukee Koneen Säätiön kieliohjelmassa esitettyjä tavoitteita avoimuuden ja vuorovaikutuksen kulttuurin edistämisestä.

Aineiston esittäminen DSpace-julkaisujärjestelmässä edistää myös aineiston löydettävyyttä. Järjestelmän oma hakutoiminto tukee sekä tiedonhakua julkaisujen kuvailutiedoista että kokotekstihakua julkaisujen sisällöstä. Järjestelmään tallennetuille julkaisuille annetaan pysyvät URN-osoitteet, joiden avulla aineisto on löydettävissä, vaikka julkaisujen sijainti joskus muuttuisi. Pilottihankkeen aineisto saatetaan avoimeen käyttöön myös tulevan Kansallisen digitaalisen kirjaston (KDK) asiakasliittymän ja vastaavien kansainvälisten hakupalveluiden kautta. Aineistot ovat löydettävissä myös Googlella ja muilla yleisillä hakukoneilla.

Tekijänoikeudet

Pilottihankkeessa on otettava huomioon sekä suomalaiset että venäläiset tekijänoikeuslainsäädännöt, koska digitointi tehdään Venäjällä ja aineisto asetetaan käyttöön Suomessa. Koska suurin osa digitoitavasta aineistosta kuuluu tekijänoikeuslainsäädännön piiriin, niin aineiston digitointia ja luovuttamista koskevia edellytyksiä on edelleen selvitettävä yhteistyössä Venäjän Kansalliskirjaston ja National Library Resourcen kanssa.

Projektisuunnitelman valmistelussa tekijänoikeuskysymyksiin ole saatu selkeitä ja yksiselitteisiä vastauksia, joiden perusteella pilottihanke voidaan toteuttaa. Pilottihankkeen käynnistysvaiheen kannalta on keskeistä, että tekijänoikeuskäytänteet selvitetään ja niihin kohdistuvat todelliset kustannukset määritellään yhteistyössä venäläisten tahojen kanssa. Tämän vuoksi kustannuslaskelmassa on tekijänoikeuksista koostuvat kustannukset esitetty tämänhetkisenä arviona.

Viestintä

Pilottihankkeen viestintä kuuluu projektipäällikön työtehtäviin yhteistyössä Kansalliskirjaston viestintäyksikön kanssa. Pilottihankkeelle luodaan oma hankesivu Kansalliskirjaston verkkosivuille. Tämän sivukokonaisuuden avulla tiedotetaan hankkeen tapahtumista tarvittavilla kielillä, vähintään suomeksi, ruotsiksi, englanniksi ja venäjäksi sekä mahdollisuuksien rajoissa myös muilla kielillä. Hankkeen käytettävissä ovat myös muut Kansalliskirjaston yleiset viestintäkanavat, kuten asiakaslehdet ja sosiaalisen median viestimet, joiden avulla voidaan kohdentaa viestintää ja tiedottamista erilaisille käyttäjäryhmille. Pilottihankkeen projektipäällikkö vastaa viestinnällisistä sisällöistä ja tietojen ajantasaisuudesta.

Kesän ja syksyn aikana pilottihanketta on tarkoitus esitellä ja saatetaan tietoisuuteen myös erilaisissa kotimaisissa (esim. Kirjastoverkkopäivät) ja kansainvälisissä tapahtumissa (esim. IFLA 2012) yhteistyössä niin Koneen Säätiön kuin muiden toimijoiden kanssa. Pilottihanketta esitellään myös yksityiskohtaisemmin syksyllä järjestettävässä kick off- tapahtumassa, jonka tarkoituksena on viestiä niin tutkijayhteisölle kuin suurelle yleisölle.

Kustannukset

Pilottihankkeen kustannukset vuosille 2012–2013 muodostavat Koneen Säätiöltä haettavan apurahan suuruuden. Seuraavassa on esitelty ja eritelty niitä kustannuseriä, jotka pilottihankkeesta muodostuvat. Pilottihankkeen kustannukset esitetään tässä laskemina, jotka perustuvat

selvitystyön aikana tehtyihin arvioihin. Kaikkia kustannuseriä, kuten tekijänoikeuksista koituvia kustannuksia, ei voida esittää kovin yksityiskohtaisesti tässä yhteydessä. Todellisista kustannuksista saadaan parempi ja yksityiskohtaisempi käsitys pilottihankkeen kuluessa. Yhteenvedo kuluista esitetty **liitteessä 2**.

Aineiston digitoinnista koituvat kustannukset

Venäjän Kansalliskirjasto on pyynnöstä arvioinut valitun aineiston digitoinnista aiheutuvia kuluja ja on esittänyt arvionaan, että teoslistauksen mukaisesti monografioiden (7000 sivua) skannaus maksaisi noin 120 000 ruplaa (noin 3 500 euroa). Tähän laskelmaan kuuluvat mukaan bibliografiset palvelut, eli bibliografisten tietojen luovuttaminen Pilottihankkeessa hyödynnettäväksi. Sanomalehtien skannauksesta aiheutuvat kulut ovat korkeammat, sillä aineisto on usein heikkolaatuista ja vaatii erikoisosaamista. Venäjän Kansalliskirjaston on arvioinut, että lasketun 20 000 sivun digitoimisen kustannukset olisivat 100 ruplaa per sivu (yhteensä 2 000 000 ruplaa) ja bibliografiset kustannukset mukaan lukien sanomalehtien digitoinnista aiheutuvat kulut ovat yhteensä 2 200 000 ruplaa (noin 60 000 euroa). Tekstitunnistuksen lisäämisestä ja TIFF-formaatin lähettämisestä aiheutuvia kuluja ei ole pystytty tässä yhteydessä tarkoin arvioimaan, mutta Venäjän Kansalliskirjastolta ostettavien palveluiden kustannukset, mukaan lukien arvonnäköalaverot (100 000 euroa + ALV 23%), on arvioitu yhteensä **123 000 euron** suuruisiksi.

Tekijänoikeudet

Selvitystyön aikana on selvitetty niin aineistojen tekijänoikeuksia yleensä kuin niiden digitaalisesta käyttöön saattamisesta koituvia tekijänoikeuspalkkioiden suuruutta. Näiltä osin on toimitettu yhteistyössä Venäjän Kansalliskirjaston kanssa, joka on pyytänyt selvityksen National Library Resourcelta Moskovasta tällaisesta työstä aiheutuvista kustannuksista. Toistaiseksi ei ole tiedossa National Library Resourcen työstä aiheutuvien kustannusten ja mahdollisten tekijänoikeuspalkkioiden kustannukset, joten tämä kustannuserä on jouduttu arvioimaan ilman tarkempia laskelmia kaikkinaensa 20 000 euron (+ ALV 23%) eli yhteensä **24 600 euron** suuruiseksi.

Laitteistot ja järjestelmät

DSPACE-julkaisujärjestelmän osuus haettavasta apurahasta muodostuu sen perustoiminnallisuuksista, joka sisältää asiakaskäyttöliittymän, tallennustilan, DSpace:n omat syöttölomakeet, kokoelman näkymän räätälöinnin ja virkailijakäyttöliittymän sekä käyttäjäkoulutusta järjestelmän parissa työskenteleville – tämän osuuden laskennallinen hinta on 13 000 euroa. Tutkijoiden käyttöön räätälöidyn kielenkorjaustyökalun kehittämiskustannukset ovat noin kolmasosa

(7000 euroa) järjestelmiin kohdistuvista kuluista, joten laitteistojen ja järjestelmien osalta kustannukset ajalle syyskuu 2012-joulukuu 2013 ovat yhteensä **20000 euroa**.

Palkkakustannukset

Pilottihankkeen ajalla työvoimatarve vaihtelee eri työtehtävien ja työtehtävien kestojen osalta siten, että pilottihankkeeseen palkattava projektipäällikön olisi aloitettava viestintään, sopimukseen ja tekijänoikeuksiin liittyvät työt parissa heinäkuussa 2012. Muut projektiin osallistuvat työntekijät aloittaisivat pilottihankkeessa vaiheittain syyskuusta 2012 alkaen. Työ kaikissa tehtävissä on osa-aikaista ja se muodostaa yhteensä 18 henkilötyökuukautta (htkk) ajalle heinäkuu 2012 – joulukuu 2013, kuitenkin seuraavasti jaettuna:

- 1) Projektipäällikkö, 8 htkk, vaatitaso 10 = 39 000 euroa
- 2) Tekninen asiantuntija, 6 htkk, vaatitaso 9 = 31 000 euroa
- 3) Kuvailu- ja formaattiasiantuntija, 4 htkk, vaatitaso 9 = 21 000 euroa

Lisäksi palkkakustannuksiin on huomioitu yleiskorotus (1,2 prosenttia) ja mahdollinen järjestyserä (0,7 prosenttia) vuodelle 2013, jolloin palkkakustannukset olisivat koko pilottihankkeen ajalta yhteensä **92 800 euroa**

Muut kulut ja palkkiot

Pilottihankkeen aikana maksetaan muita kuluja yhteensä 35 000 euroa (+ ALV 23%) eli yhteensä **43 050 euroa**. Kulut koostuvat osallistumisista hanketta esitteleviin seminaareihin, hankkeen toteuttamisen kannalta välttämättömiin matkoihin Venäjälle, mahdollisiin sopimusten ja viestintämateriaalien käänköspalkkioihin ja hankkeen aikana annettavaan koulutukseen. Koulutukseen varattu osuus apurahasta kohdennetaan Kansalliskirjaston ja Venäjän Kansalliskirjaston yhteistyöhön digitoitikäytänteiden kehittämiseksi. Tarkoituksena on, että Pilottihankkeen aikana voidaan jakaa osaamista digitoinnin kehittämisen osalta, ennen kaikkea liittyen aineiston esittämisen ja metadatan kannalta tärkeiden METS/Alto-tiedostojen tekemiseen, joita Venäjän Kansalliskirjasto ei tällä hetkellä pysty tarjoamaan. Koulutuksen tavoitteena on että Venäjän Kansalliskirjasto pystyisi tarjoamaan digitoituja aineistoja hyödyntäen METS/Alto-tiedostoina viimeistään jatkohankkeen aikana. Tällöin voidaan varmistaa digitoitujen aineiston pitkäaikaissäilytys sekä kehittää kansalaistieteen palveluita edelleen jatkohanketta silmällä pitäen.

Jatkohanke

Pilottihankkeen aikana (heinäkuu 2012-joulukuu 2013) kehitetään toimintatapaa ja teknistä toteutusta, joita on tarkoitus hyödyntää muiden kieliohjelmaan liittyvien aineistojen digitoinnissa jatkohankkeessa vuosina 2014–2016. Tarkoituksena on luoda myös uusia edellytyksiä kansainväliselle yhteistyölle ja laajentaa mahdollisuuksien mukaan yhteistyötä kattamaan myös muita suomalais-ugrilaisia kieliaineistoja hallinnoivia ja niitä digitoineita muistiorganisaatioita Venäjällä ja Suomessa. Yhteistyön kehittäminen pilottihankkeen aikana luo paremmat edellytykset jatkohankkeen toteuttamiselle, erityisesti digitaalisten aineistojen metadataformaattien (METS/Alto-paketit ja RusMarc-konversio) suhteen.

Liite 2 - selvitys tekijänoikeuksista

<p style="text-align: center;">ЮРИДИЧЕСКОЕ ЗАКЛЮЧЕНИЕ о правомерности оцифровки произведений на языках финно-угорской группы в рамках международного проекта «Языковая программа Фонда Коне на 2012- 2016 годы»</p> <p>г. Москва марта 2013 года</p> <p style="text-align: right;">«26»</p> <p>Настоящее заключение составлено по поручению Автономной некоммерческой организации «Национальный библиотечный ресурс» на основе действующего законодательства Российской Федерации.</p> <p>Предметом рассмотрения в настоящем заключении являются вопросы получения разрешения правообладателей на оцифровку произведений на языках финно-угорской группы в рамках международного проекта «Языковая программа Фонда Коне на 2012-2016 годы».</p> <p style="text-align: center;">I. Суть проекта.</p> <p>Автономная некоммерческая организация «Национальный библиотечный ресурс» (г. Москва) (далее – «Национальный библиотечный ресурс») совместно с Национальной библиотекой Финляндии (г. Хельсинки) (далее – «Национальная библиотека Финляндии») и Федеральным государственным бюджетным учреждением «Российская национальная библиотека» (г. Санкт-Петербург) (далее – «Российская национальная библиотека») участвует в международном проекте «Языковая программа Фонда Коне на 2012-2016 годы», которая направлена на сохранение культурного наследия малых народов путем оцифровки печатных изданий на языках финно-угорской группы (далее – «Издания»), т.е. воспроизведения и хранения указанных изданий в электронной (цифровой) форме.</p> <p style="text-align: center;">II. Правовые основы оцифровки. Интеллектуальные права на произведения науки, литературы и искусства.</p> <p>Нормативно-правовой основой национальных программ по внедрению информационно-коммуникационных технологий в сферу культуры,</p>	<p style="text-align: center;">LEGAL OPINION on the Legality of the Works in Finno-Ugric Languages Digitization within the Framework of the International Project “Kone Foundation Language Program for 2012-2016”</p> <p>Moscow March 26, 2013</p> <p>This legal opinion, commissioned by the Autonomous Non-Profit Organization National Library Resource, is based upon the Law of the Russian Federation.</p> <p>The subject matter of this opinion are the issues to be solved in obtaining the authorization of the copyright holders to digitize the works in Finno-Ugric languages within the framework of the international project “Kone Foundation Language Program for 2012-2016”.</p> <p style="text-align: center;">I. Essence of the Project</p> <p>Autonomous Non-Profit Organization “National Library Resource” (Moscow) (hereinafter referred to as the “National Library Resource”) together with the National Library of Finland (Helsinki) (hereinafter referred to as the “National Library of Finland”) and the Federal State Budgetary Institution “Russian National Library” (Saint Petersburg) (hereinafter referred to as the “Russian National Library”) participate in the international project “Kone Foundation Language Program for 2012-2016”. The project is aimed at the preservation of small-numbered peoples’ cultural heritage by way of digitizing the printed publications in Finno-Ugric languages (hereinafter referred to as the “Publications”), i.e. reproduction and storage of the mentioned publications in electronic (digital) format.</p> <p style="text-align: center;">II. Legal Framework for Digitizing. Intellectual Rights to Scientific, Literary and Artistic Works</p> <p>Constitution of the Russian Federation and the Fundamentals of the Russian Federation Law on Culture No. 3612-1 of October 9, 1992 (here-</p>
---	--

предусматривающие, в частности, создание и развитие электронных информационных ресурсов библиотек, в том числе представленных в сети «Интернет» (см. Постановление Правительства РФ от 3 марта 2012 г. № 186 «О федеральной целевой программе «Культура России (2012-2018 годы)»), являются, в первую очередь, Конституция РФ и Основы законодательства Российской Федерации о культуре от 9 октября 1992 г. № 3612-1 (далее – «**Основы законодательства о культуре**»).

Часть 2 ст. 44 Конституции РФ гласит: **«Каждый имеет право на участие в культурной жизни и пользование учреждениями культуры, на доступ к культурным ценностям»**. При этом ч. 3 ст. 44 закрепляет положение о том, что **каждый обязан заботиться о сохранении исторического и культурного наследия, беречь памятники истории и культуры**.

В соответствии со ст. 6 Основ законодательства о культуре **Российская Федерация признает равное достоинство культур, равные права и свободы в области культуры всех проживающих в ней народов и иных этнических общностей, способствует созданию равных условий для сохранения и развития этих культур, обеспечивает и укрепляет целостность российской культуры посредством законодательного регулирования федеральной государственной культурной политики и федеральных государственных программ сохранения и развития культуры**.

Ст. 12 Основ законодательства о культуре закрепляет положение о том, что **каждый человек имеет право на приобщение к культурным ценностям**, на доступ к государственным библиотечным, музейным, архивным фондам, иным собраниям во всех областях культурной деятельности. Ограничения доступности культурных ценностей по соображениям секретности или особого режима пользования устанавливаются законодательством Российской Федерации.

Таким образом, ограничения доступности культурных ценностей могут быть связаны только с соображениями секретности или особого режима пользования. Иные ограничения доступности культурных ценностей вводиться не могут.

Согласно ст. 20 Основ законодательства о культуре народы и иные этнические общности в Российской Федерации имеют **право на сохранение и развитие своей культурно-**

inafter referred to as the “**Fundamentals of Law on Culture**”) serve the primary source of regulation for the national programs aimed at implementing the achievements of information and communication technologies in the field of culture. These programs account for the creation and development of the electronic information resources in libraries, including those present in the Internet (see Resolution of the Government of the Russian Federation No. 186 of March 3, 2012 “On the Federal Target Program “The Culture of Russia (2012-2018)”).

Part 2 of Article 44 of the Constitution of the Russian Federation stipulates the following: **“Everyone shall have the right to participate in cultural life and use cultural establishments and to access to items of cultural value”**. With that, Part 3 of Article 44 contains the provision that **everyone shall be obliged to care for the preservation of cultural and historical heritage and protect monuments of history and culture**.

In accordance with Article 6 of the Fundamentals of Law on Culture, **the Russian Federation recognizes equal merit of culture, equal rights and freedoms in the sphere of culture of all peoples living within the territory of the Russian Federation and other ethnic groups, contributes to the creation of equal conditions for preservation and development of these cultures, provides for and supports the integrity of Russian culture by way of legislative regulation of federal state cultural policy and federal state programs of preservation and development of culture**.

Pursuant to Article 12 of the Fundamentals of Law on Culture, **each person is entitled to join in the cultural values**, to access the state libraries, museums, archives, other collection in all fields of cultural activities. Limitations of cultural values availability on the grounds of secrecy or special mode of use are established by the laws of the Russian Federation.

Thus, limitations of cultural values’ availability may be connected only with the grounds of secrecy or special use mode. Other limitations of access to cultural values shall not be introduced.

In accordance with Article 20 of the Fundamentals of Law on Culture, peoples and other ethnic groups in the Russian Federation have the

национальной самобытности, защиту, восстановление и сохранение исконной культурно-исторической среды обитания.

В соответствии с Соглашением между Правительством Российской Федерации и Правительством Финляндской Республики о сотрудничестве в области культуры, образования и научно-исследовательской деятельности (Хельсинки, 11 июля 1992 г.), а именно абзацу второму ст. 2, высокие договаривающиеся Стороны будут оказывать поддержку тесным контактам в области архивного дела и **содействовать дальнейшему развитию связей, существующих между библиотеками и музеями обеих стран.** Ст. 3 предусматривает, что **особое внимание будет уделяться сотрудничеству в охране культурного наследия,** обмену опытом, связанным с охраной, реставрацией и использованием памятников истории и культуры.

Работы по оцифровке произведений науки, литературы и искусства на территории Российской Федерации должны проводиться с учетом требований, предъявляемых к защите прав интеллектуальной собственности, предусмотренных Гражданским кодексом РФ.

На результаты интеллектуальной деятельности и приравненные к ним средства индивидуализации (результаты интеллектуальной деятельности и средства индивидуализации) в силу ст. 1226 ГК РФ признаются **интеллектуальные права.** Они включают **исключительное право, являющееся имущественным правом, а в случаях, предусмотренных ГК РФ, также личные неимущественные права и иные права** (право следования, право доступа и другие).

Согласно п. 1 ст. 1255 ГК РФ интеллектуальные права на произведения науки, литературы и искусства являются авторскими правами. При этом п. 2 ст. 1255 ГК РФ закрепляет за автором произведения следующие права:

- 1) исключительное право на произведение;
- 2) право авторства, т.е. право признаваться автором произведения (п. 1 ст. 1265 ГК РФ);
- 3) право автора на имя, т.е. право использовать или разрешать использование произведения под своим именем, под вымышленным именем (псевдонимом) или без указания имени, то есть анонимно (п. 1 ст.

right to preservation and development of their cultural and national identity, protection, restoration and preservation of genuine cultural and historical living environment.

According to the Agreement between the Government of the Russian Federation and the Government of the Republic of Finland on Cooperation in the Field of Culture, Education and Scientific Research (Helsinki, July 11, 1992), namely Part 2 of the Article 2, the High Contracting Parties will provide assistance to close contacts in the field of archiving and **contribute to further development of connections existing between the libraries and museums of both countries.** Article 3 envisages **special attention to be paid to cooperation in cultural heritage preservation,** sharing experience related to protection, restoration and the use of historical and cultural monuments.

Activities to digitize the scientific, literary and artistic works in the territory of the Russian Federation shall be conducted with regard to the requirements to intellectual property rights protection, as prescribed by the Civil Code of the Russian Federation.

The results of intellectual activity and the means of individualization equivalent to them (results of intellectual activities and individualization means) are deemed **intellectual rights** by virtue of article 1226 of the Civil Code of the Russian Federation. They include **exclusive right constituting the property right, and, in cases provided by the Civil Code of the Russian Federation, also personal non-property rights and other rights** (resale royalty right, access right, etc.).

In accordance with Part 1 of Article 1255 of the Civil Code of the Russian Federation, the intellectual rights to scientific, literary and artistic works are copyright rights. With that, Part 2 of Article 1255 of the Civil Code of the Russian Federation allocates to the author of a work the following rights:

- 1) exclusive right to the work;
- 2) copyright, i.e. the right to be recognized as the author of the work (Part 1 of Article 1265 of the Civil Code of the Russian Federation);
- 3) author's right to the name, i.e. the right to use or permit the use of the work under his or

1265 ГК РФ);

4) право на неприкосновенность произведения (ст. 1266 ГК РФ);

5) право на обнародование произведения (ст. 1268 ГК РФ).

Смысл и функция исключительного права заключается в том, что, согласно п. 1 ст. 1229 ГК РФ **гражданин или юридическое лицо, обладающие исключительным правом на результат интеллектуальной деятельности или на средство индивидуализации (правообладатель), вправе использовать такой результат или такое средство по своему усмотрению любым не противоречащим закону способом.** Правообладатель может распоряжаться исключительным правом на результат интеллектуальной деятельности или на средство индивидуализации (статья 1233), если ГК РФ не предусмотрено иное.

Правообладатель может по своему усмотрению разрешать или запрещать другим лицам использование результата интеллектуальной деятельности или средства индивидуализации. Отсутствие запрета не считается согласием (разрешением).

Другие лица не могут использовать соответствующие результат интеллектуальной деятельности или средство индивидуализации без согласия правообладателя, за исключением случаев, предусмотренных ГК РФ.

Вместе с тем ГК РФ предусматривает целый ряд случаев свободного использования произведений без согласия автора или иного правообладателя и без выплаты вознаграждения, но с обязательным указанием имени автора, произведение которого используется, и источника заимствования (ст. 1273-1280 ГК РФ). Однако рассматриваемый проект Фонда Коне не подпадает под особый режим использования произведений, предусмотренных вышеуказанными статьями, в том числе под специальные изъятия в интересах библиотек и архивов.

Так, в соответствии с п. 2 ст. 1274 ГК РФ в случае, когда библиотека предоставляет экземпляры произведений, правомерно введенные в гражданский оборот, во временное безвозмездное пользование, такое пользование допускается без согласия автора или иного правообладателя и без выплаты вознаграждения. При этом выраженные в цифровой форме экземпляры произведений, предоставляемые библиотеками во временное

her name, under a fictitious name (a pseudonym) or without name indication, i.e. anonymously (Part 1 of Article 1265 of the Civil Code of the Russian Federation);

4) right to the integrity of the work (Article 1266 of the Civil Code of the Russian Federation);

5) right to make the work public (Article 1268 of the Civil Code of the Russian Federation).

Essence and reason of the exclusive right is expressed in Part 1 of the Article 1229 of the Civil Code of the Russian Federation proclaiming that, **a citizen or legal entity having exclusive right to the result of intellectual activities or the means of individualization (the copyright holder) is entitled to use such result or such means at his or her own discretion in the manner which is not contradictory to the law.** The copyright holder may dispose of the exclusive right to the result of intellectual activity or means of individualization (article 1233), unless otherwise provided by the Civil Code of the Russian Federation.

The copyright holder may, at his or her discretion, permit or prohibit other persons to use the intellectual activities result or individualization means. Lack of ban is not an authorization (permit to use).

Other persons cannot use the results of intellectual activity or means of individualization without the consent of the copyright holder except in special cases provided for by the Civil Code of the Russian Federation.

Alongside with that, the Civil Code of the Russian Federation provides a number of cases of free use of writings without the author's or another copyright holder's consent and without the payment of the remuneration but with the obligatory indication of the name of the author whose work is used as well as the source of borrowing (Articles 1273-1280 of the Civil Code of the Russian Federation). However, the regarded Kone Foundation project is not subject to special mode of the use of works, as provided for by the above mentioned articles, including special exceptions for the benefit of libraries and archives.

Thus, in accordance with Part 2 of Article 1274 of the Civil Code of the Russian Federation, when the copies of writings, that have been legally commissioned to commercial civil turn over, are provided for temporary free use by libraries, such use is admitted on exceptional basis without the author's or another copyright holder's consent and without the payment of the remuneration. That is why the digital copies of the

безвозмездное пользование, в том числе в порядке взаимного использования библиотечных ресурсов, могут предоставляться только в помещениях библиотек при условии исключения возможности создать копии этих произведений в цифровой форме.

Данное исключение в интересах библиотек и архивов не подлежит применению, так как Договор между Национальной Библиотекой Финляндии и Национальным Библиотечным Ресурсом о передаче прав на использование финно-угорских материалов от «___» _____ 2012 года предусматривает предоставление Национальной библиотеке Финляндии права осуществлять публичный показ на серверах своей информационной системы оцифрованных произведений и материалов без какого-либо территориального или временного ограничения данного права.

Следовательно, для правомерной оцифровки и дальнейшего использования Изданий необходимо получить согласие их правообладателей.

В соответствии с п. 3 ст. 1228 Гражданского кодекса РФ исключительное право на результат интеллектуальной деятельности, созданный творческим трудом, первоначально возникает у его автора. Это право может быть передано автором другому лицу по договору, а также может перейти к другим лицам по иным основаниям, установленным законом. Кроме того, исключительное право на произведение может принадлежать работодателю, если произведение было создано в пределах, установленных для работника (автора) трудовых обязанностей.

Закон закрепляет **срок действия исключительного права на произведение.** Согласно п. 1 ст. 1281 ГК РФ исключительное право на произведение действует, по общему правилу, в течение всей жизни автора и семидесяти лет, считая с 1 января года, следующего за годом смерти автора. Исключительное право на произведение, созданное в соавторстве, действует в течение всей жизни автора, пережившего других соавторов, и семидесяти лет, считая с 1 января года, следующего за годом его смерти. В соответствии с п. 5 ст. 1281 ГК РФ, если автор работал во время Великой Отечественной войны или участвовал в ней, срок действия исключительного права, установленный настоящей статьей, увеличивается на четыре

works provided by libraries for temporary free use, including the mutual joint use of library resources, may be accessed to only in the library premises provided that the possibility to make digital copies of these works is excluded.

This exception that benefits libraries and archives is not applicable to the given case for the agreement between the National Library of Finland and the National Library Resource for transfer of rights to use Finno-Ugric materials dated _____, 2012 envisages granting to the National Library of Finland the right to demonstrate digitized works and materials to public via its information system servers without any restriction of this right in time or territory.

Thus, for digitizing the Publications and further use of the digitized works in compliance with Law the libraries are expected to obtain the consent of the respective copyright holders.

In accordance with Part 3 of Article 1228 of the Civil Code of the Russian Federation, exclusive right to the result of intellectual activity created in the course of creative work is attributed to its author. This right may be transferred by the author to another person, as well as may be transferred to other persons on other basis established by Law. Apart from that, exclusive right to a work may belong to the employer, if the work was created as the work for hire prescribed by employee's (author's) labor contract.

The law limits **the term of the exclusive right to the work.** In accordance with Part 1 of Article 1281 of the Civil Code of the Russian Federation, the exclusive right to writing is valid, as a common rule, during the life time of the author and seventy years starting from January 1 of the year following the year of the author's death. Exclusive right to the writing created in joint authorship is valid during the life span of the surviving co-author and seventy years after his/her death. In accordance with Part 5 of Article 1281 of the Civil Code of the Russian Federation, if the authors worked during the Great Patriotic War or participated in it, the term of exclusive right established by this article is prolonged for 4 years.

года.

Статья 1282 ГК РФ закрепляет положение о том, что по истечении срока действия исключительного права произведение науки, литературы или искусства, как обнародованное, так и необнародованное, переходит в общественное достояние (п. 1). Произведение, перешедшее в общественное достояние, может свободно использоваться любым лицом без чьего-либо согласия или разрешения и без выплаты авторского вознаграждения. При этом охраняются авторство, имя автора и неприкосновенность произведения (п. 2).

Необходимо также отметить, что произведения, не имеющие конкретных авторов, считаются произведениями народного творчества (фольклором) и, в силу подп. 3 п. 6 ст. 1259 ГК РФ, не являются объектами авторских прав.

Ответственность за «очистку» (получение) авторских прав на оцифровку Изданий в соответствии с Договором с Национальной библиотекой Финляндии принял на себя Национальный библиотечный ресурс. Получение согласия правообладателей осложнилось тем, что публикация ряда Изданий была осуществлена в 1913-1947 гг., и значительное число периодических Изданий, а также издательств к настоящему моменту прекратили свое существование.

III. Действия Национального библиотечного ресурса, связанные с «очисткой» авторских прав на Издания.

В целях соблюдения требований законодательства Российской Федерации о культуре, об авторском праве и руководствуясь принципами добросовестности, разумности и справедливости, Национальный библиотечный ресурс осуществил целый ряд действий по проверке наличия или отсутствия обладателей исключительных прав на включенные в проект Фонда Коне Издания или их части.

В рамках этих действий Национальный библиотечный ресурс обратился с запросами:

- в Федеральную миграционную службу РФ;
- в Федеральную службу безопасности РФ;
- в Общероссийскую общественную организацию «Российское Авторское Общество»;
- в Общероссийское общественное движение «Ассоциация финно-угорских народов Российской Федерации»;
- в Санкт-Петербургское государственное

Article 1282 of the Civil Code of the Russian Federation has the provision that upon expiration of the term of exclusive right scientific, literary and artistic work, both published and unpublished, is transferred to public domain (Part 1). The work transferred to public domain may be used freely by any person without any consent or authorization and without payment of royalty fee. With that, the authorship, author's name and integrity of the work are retained (Part 2).

It is also necessary to note that the writings having no certain authors are considered folk art (folklore), and by virtue of Part 1 of Article 1259 (par. 3) of the Civil Code of the Russian Federation they are not copyright items.

Responsibility for “clearance” (obtaining) of copyright for the digitizing of the Publications in accordance with the Agreement with the National Library of Finland is assumed by the National Library Resource. Obtaining consent of copyright holders was complicated by the fact that a number of the Publications were published in 1913-1947, and significant amount of periodic Publications, as well as publishing houses have ceased to exist so far.

III. Activities of the National Library Resource, which are related to the “clearance” of copyright to the Publications

For the purpose of compliance with the Russian laws concerning culture, copyright and being governed by the principles of fair practice, reasonableness and justice, the National Library Resource conducted a number of activities aimed at the check of presence or lack of exclusive rights holders to the Publications or their parts included in the Kone Foundation project.

Within the frames of these activities the National Library Resource made enquiries to:

- the Federal Migration Service of the Russian Federation;
- the Federal Security Service of the Russian Federation;
- the All-Russian Public Organization “Russian Authors’ Society”;
- the All-Russian Social Movement Association of Finno-Ugric Peoples of the Russian Federation;
- the Saint Petersburg State Government

казенное учреждение «Центральный государственный архив Санкт-Петербурга»;

– в государственное казенное учреждение Республики Карелия «Национальный архив Республики Карелия»;

– в государственное казенное архивное учреждение «Центральный государственный архив Республики Мордовия»;

– в государственное казенное учреждение «Государственный архив Республики Марий Эл».

Из вышеуказанных организаций были получены ответы, из которых следовало, что данные организации не располагают сведениями о наличии или отсутствии обладателей исключительных прав на произведения из перечня, приведенного Национальным библиотечным ресурсом в соответствующих запросах.

Кроме того, Национальный библиотечный ресурс разместил объявления о поиске правообладателей в «Литературной газете» (№ 1 (6399), 16-22 января 2012 г., общий тираж 136 000 экз.) и в газете «Марий Эл» (12 января 2013 года, тираж 6 200 экз.).

Несмотря на установленный Национальным библиотечным ресурсом разумный и достаточный срок подачи заявлений о наличии исключительных прав на Издания – 30 (тридцать) рабочих дней с момента опубликования объявления, – каких-либо заявлений получено не было.

Таким образом, Национальный библиотечный ресурс выполнил требования законодательства об авторском праве и осуществил все разумные действия по поиску обладателей исключительных прав на Издания.

Выводы.

1. Национальный библиотечный ресурс проявил необходимую добросовестность и должную осмотрительность в деле поиска правообладателей Изданий. В адрес НБР не поступило заявлений от физических и юридических лиц о принадлежности им прав на Издания, или их части.

2. Проведенная работа дает Национальному библиотечному ресурсу все основания полагать, что физических и юридических лиц, которые должны образом могли бы заявить о своих притязаниях на авторские права на произведения, вошедшие в

Agency Saint Petersburg Central State Archives;

– the State Government Agency of the Republic of Karelia National Archives of the Republic of Karelia;

– the State Government Archive Institution Central State Archives of the Republic of Mor-dovia;

– the State Government Agency State Archives of the Republic of Mari El.

The organizations mentioned above sent the answers from which it was possible to make a conclusion that these organizations had no information on the presence or absence of exclusive rights holders to the works from the list submitted by the National Library Resource in corresponding enquiries.

Besides, the National Library Resource placed advertisements on the search for copyright holders in “Literaturnaya Gazeta” (“The Literary Newspaper”) (No. 1 (6399), January 16-22, 2012, general circulation 136,000 copies) and “Mari El” newspaper (January 12, 2013, circulation of 6,200 copies).

Despite the established by the National Library Resource reasonable term of filing applications on the presence of exclusive rights to the Publications (thirty (30) business days from the moment of the advertisement publishing) no applications were obtained.

Thus, the National Library Resource complied with the law requirements concerning copyright and carried out all necessary activities to search for the holders of exclusive rights to the Publications.

Conclusions

1. The National Library Resource showed necessary fairness and due diligence in its work on search for copyright holders of the Publications. No applications were submitted to the NLR from individuals or legal entities on their rights to the Publications or their parts.

2. The conducted search gives the National Library Resource all reasons to declare that there are no individuals and legal entities that have claimed or could have justified claims for copyright to the works contained in the Publication. As a consequence, copyright to such writings may with high degree of probability be referred to ownerless property (escheat).

Издания, нет. Следовательно, авторские права на такие произведения с высокой степенью вероятности можно отнести к выморочному имуществу.

3. Особенности наследования выморочного имущества регулируются статьей 1151 ГК РФ «Наследование выморочного имущества», которая гласит:

(1). В случае если отсутствуют наследники как по закону, так и по завещанию, либо никто из наследников не имеет права наследовать или все наследники отстранены от наследования, либо никто из наследников не принял наследства, либо все наследники отказались от наследства и при этом никто из них не указал, что отказывается в пользу другого наследника, **имущество умершего считается выморочным.**

(2). Выморочное имущество, которое не является жилым помещением, переходит в порядке наследования по закону в собственность Российской Федерации (п. 2 в ред. Федерального закона от 29.11.2007 N 281-ФЗ).

Согласно Приказу Минюста РФ от 28.07.2003 N 180 в редакции от 16.02.2009 N 49 Свидетельство о праве на наследство по закону, удостоверяющее право на выморочное наследственное имущество, выдается нотариусом. В свидетельстве должно быть указано из чего состоит объект наследования. Согласно сложившейся практике такие свидетельства оформлялись только в отношении движимого имущества в виде транспортного средства или недвижимого имущества.

Для перехода в собственность Российской Федерации авторских прав выдачи свидетельства о праве наследования по закону не требуется.

4. Согласно части 2 статьи 1283 ГК РФ («Переход исключительного права на произведение по наследству»): *«В случаях, предусмотренных статьей 1151 настоящего Кодекса, входящее в состав наследства исключительное право на произведение прекращается, и произведение переходит в общественное достояние».*

Переход произведения в общественное достояние означает, что такое произведение в силу ст.1282 ГК РФ может свободно использоваться любым лицом без чье-либо согласия или разрешения и без выплаты

3. Peculiarities of succession to ownerless property are governed by article 1151 of the Civil Code of the Russian Federation (“Escheat”), which says:

(1). If there are no legal heirs and heirs by will or if neither of the heirs has a right to inherit or all heirs have been deprived of their right of inheritance (Article 1117) or neither of the heirs have accepted the inheritance or all the heirs refused their inheritance and neither of them has indicated that the inheritance is waived for the benefit of another heir (Article 1158) the decedent's estate shall be deemed escheat.

(2). Escheat property shall pass into the ownership of the Russian Federation by succession by operation of law (Part 2 as amended by Federal Law No. 281-FZ of November 29, 2007).

In accordance with the Order of the Ministry of Justice No. 180 as of July 28, 2003, as amended on February 16, 2009 No. 49, the Certificate of Inheritance Right by virtue of law, which certifies the right to ownerless inheritance property, is issued by the notary. In the certificate it shall be mentioned what the inheritance object consists of. In accordance with the established practice, such certificates were executed only with regard to movable property in the form of vehicles or immovable property.

For transfer of copyright to the ownership of the Russian Federation no legal requirements provide for the issue of certificate of the inheritance right.

4. In accordance with Part 2 of Article 1283 of the Civil Code of the Russian Federation (“Transfer of Exclusive right to the Work by Inheritance”): *“In cases indicated in Article 1151 of this Code the exclusive right to the work included in the structure of heritage is terminated, and the work transfers to public domain”.*

The transfer of the work to public domain means that such work by virtue of Article 1282 of the Civil Code of the Russian Federation may be used freely by any person without any consent or authorization and without payment of royalty fee. With that, the authorship, author's name and the integrity of the work are retained.

Thanks to the activity of the National Li-

авторского вознаграждения. При этом охраняются авторство, имя автора и неприкосновенность произведения.

Благодаря работе, проведенной Национальным библиотечным ресурсом, удалось документально обосновать факт принадлежности авторских прав на Издания к выморочному имуществу, в отношении которого действует порядок использования произведений, перешедших в общественное достояние.

Исследование, выполненное Национальным библиотечным ресурсом дает все основания считать, что действия по оцифровке произведений на языках финно-угорской группы в рамках международного проекта «Языковая программа Фонда Коне на 2012-2016 годы» осуществляется в полном соответствии с Конституцией РФ, Гражданским кодексом РФ, Основами законодательства Российской Федерации о культуре от 9 октября 1992 г. № 3612-1, Постановлением Правительства РФ от 3 марта 2012 г. № 186 «О федеральной целевой программе «Культура России (2012-2018 годы)» и другими нормативно-правовыми актами.

Директор
Центра правовой защиты
интеллектуальной собственности,
адвокат

В.Л. Энтин

**АКТ ВЫПОЛНЕННЫХ РАБОТ
к Договору между Национальной
библиотекой Финляндии и Национальным
библиотечным ресурсом о передаче прав на
использование финно-угорских материалов**

Стороны Договора: Национальный библиотечный ресурс (National Library Resource), находящийся в г. Москва (Российская Федерация), и Национальная библиотека Финляндии, находящаяся в Хельсинки (Республика Финляндия),

подтверждают настоящим Актом нижеследующее:

1. Национальный библиотечный ресурс осуществил все необходимые действия, диктуемые правилами добросовестности и должной осмотрительности, по очистке авторских прав на произведения и материалы на ижорском, вепском,

brary Resource, it managed to documentarily prove the fact that the copyright to the Publications belongs to ownerless property (escheat), with regard to which the procedure of the use of works which fell into public domain is implemented.

The in depth search conducted by the National Library Resource gives all reasons to believe that the activities on digitizing the works in Finno-Ugric languages within the framework of international project “Kone Foundation Language Program for 2012-2016” are in full compliance with the Constitution of the Russian Federation, the Civil Code of the Russian Federation, the Fundamentals of Law of the Russian Federation on Culture No. 3612-1 as of October 9, 1992, the Resolution of the Government of the Russian Federation No. 186 as of March 3, 2012 “On the Federal Target Program “Culture of Russia (2012-2018)” and other regulations.

Director of
The Centre for Intellectual Property
Legal Protection,
Attorney at law

V. L. Entin

**ACCEPTANCE CERTIFICATE
to the Agreement between the National Li-
brary of Finland and the National Library
Resource on transferring the rights to use the
Finno-Ugric Materials**

Parties to the Agreement: the National Library Resource located in Moscow (Russian Federation) and the National Library of Finland located in Helsinki (Republic of Finland)

herewith confirm the following:

1. The National Library Resource performed all necessary activities, as prescribed by the rules of fair practice and due diligence, to clear the copyright to the works and materials in Ingrian, Vepsian, Mari and Mordvinian, as mentioned in appendices to the Agreement.

2. In witness of the performance of necessary and sufficient work on “clearance” of the copyright to

<p>марийском и мордвинском языках, указанные в приложениях к Договору.</p> <p>2. В удостоверение выполнения необходимых и достаточных работ по «очистке» авторских прав на произведения и материалы, указанные в приложениях к Договору, Национальный библиотечный ресурс передает Национальной библиотеке Финляндии Приложение, содержащее копии запросов и объявлений в газетах о предоставлении информации об обладателях исключительных прав на отобранные Сторонами произведения и материалы, а также копии ответов на указанные запросы.</p> <p>3. Настоящий Акт является неотъемлемой частью Договора, действует наряду с ним и вступает в силу с момента подписания Сторонами настоящего Акта.</p> <p>4. Во всем остальном, что не урегулировано настоящим Актом, Стороны руководствуются условиями Договора.</p> <p>5. Настоящий Акт составлен в 2 (двух) экземплярах, по одному для каждой из Сторон.</p> <p>6. Приложение, указанное в пункте 2, на _____ листах прилагается.</p>	<p>the works and materials, as mentioned in the appendices to the Agreement, the National Library Resource transfers to the National Library of Finland the Appendix containing copies of enquiries and newspaper advertisements on submission of information on copyright holders of exclusive rights to the works and materials selected by the Parties, as well as the copies of answers to the mentioned enquiries.</p> <p>3. This Certificate constitutes an integral part of the Agreement, is valid together therewith and becomes effective upon signing this Certificate by the Parties.</p> <p>4. In all matters not governed by this Certificate, the Parties shall be governed by the terms and conditions of the Agreement.</p> <p>5. This Certificate is made in two (2) counterparts, one copy for each Party.</p> <p>6. The Appendix mentioned in clause 2, on _____ sheets is attached.</p>
<p style="text-align: center;">ПОДПИСИ СТОРОН</p> <p>Москва, «__» апреля 2013 г. От имени Национального библиотечного ресурса</p> <p>Исполнительный директор Олег Махно</p> <p>Хельсинки, «__» апреля 2013 г. От имени Национальной библиотеки Финляндии</p> <p>Директор Кай Экхольм</p>	<p style="text-align: center;">PARTIES' SIGNATURES</p> <p>Moscow, April __, 2013 On behalf of the National Library Resource</p> <p>Managing Director Oleg Makhno</p> <p>Helsinki, April __, 2013 On behalf of the National Library of Finland</p> <p>Director Kai Ekholm</p>

12. YHTEYSTIEDOT JA TIEDUSTELUT

Liisa Savolainen
johtaja
Kansalliskirjasto
Tutkimuskirjasto
PL 15 (Fabianinkatu 35)
00014 HELSINGIN YLIOPISTO
liisa.savolainen@helsinki.fi
09 191 22745

<http://www.kansalliskirjasto.fi/kokoelmatjapalvelut/digitaalisetkokoelmat/finnougric.html>