

Mikko Lappalainen, Juha Hulkkonen, Juho Inkinen,  
Aleksi Kallio, Markus Koskela, Mona Lehtinen,  
Mats Sjöberg, Osma Suominen ja Laxmana Yetukuri

# Automaattisen sisällönkuvailun ohjelmiston rakentaminen – case Annif

Sisällönkuvailun automatisointiratkaisut ovat puhuttaneet kirjastomaailmassa viime vuosina, ja erilaisia kokeiluja on tehty niin Suomessa kuin maailmallakin. Kansalliskirjastossa kehitetty automaattisen sisällönkuvailun Annif-työkalu on herättänyt paljon mielenkiintoa monissa organisaatioissa ja kokemukset ensimmäisistä käyttöönotoista ovat olleet lupaavia. Mitä kehitysvaihtoehtoja Annifia rakennettaessa on tehty, ja minkälaisia haasteita kuvailun automatisointiin ylipäätään liittyy?

Tässä artikkelissa esitellään tekstiaineistojen automaattisen sisällönkuvailun ohjelmiston rakentamisen haasteita ja ratkaisuja erityisesti KAM-sektorin (kirjastot, arkistot ja museot) organisaatioiden näkökulmasta. Esimerkkinä käytetään Kansalliskirjastossa kehitettävää automaattisen sisällönkuvailun työkalua Annifia. Artikkelin pohjautuu selvitykseen, *Sisällönkuvailun automatisoinnin haasteita ja ratkaisuja kulttuuriperintöorganisaatioissa* (Hulkkonen ym. 2021), joka julkaistiin alkuvuodesta 2021 osana CSC:n, Kansalliskirjaston ja Kansallisarkiston yhteistä High-Performance Digitisation -hanketta.

## 1. Sisällönkuvailun automatisoinnin lähtökohdat

**1.1 Manuaalinen, puoliautomaattinen ja automaattinen kuvailu**  
Sisällönkuvailun tarkoituksena on ilmaista kuvailtavan aineiston (esim.

kirja, artikkeli, verkkosivu jne.) aiheaineiston löytämisen helpottamiseksi. Kirjastoissa ja muissa kulttuuriperintöalan organisaatioissa sisällönkuvailua on tehty pitkään ihmistyönä valmiita asiansastoa ja/tai muuta kontrolloitua sanastoa hyödyntäen, ja sisällönkuvailijoista on muodostunut oma ammattiryhmänsä ainakin isoimmissa kirjastoissa. Verkkomaailmassa sisällönkuvailua tehdään usein sivutyönä toisen pääasiallisen työn ohessa, kuten esimerkiksi toimittaja sisällönkuvailijaksi kirjoittamansa artikkelin osana toimitustyötä. Tämä käyttäjäryhmien moninaisuus onkin hyvä huomioida automaattisen kuvailun ohjelmistojen ja palveluita suunniteltaessa, erityisesti silloin kun ajatuksena on tuottaa monissa erilaisissa ja erikokoisissa organisaatioissa hyödynnettäväksi soveltuva yleiskäyttöinen ohjelmisto.

Puoliautomaattisessa sisällönkuvailussa jokin järjestelmä antaa halutulle tekstille asiansaehdotuksia, jotka käydään läpi ihmistyönä. Tarkoituksena on

sujuvoittaa kuvailutyötä näillä pohjaehdotuksilla; ihmiselle jää kuitenkin päävastuu kuvailusta.

Täysin automaattisessa kuvailussa ohjelmiston/algoritmin tuottama kuvailu etenee sellaisenaan ilman ihmistarkistusta järjestelmän tietokantaan. Tässä menetelmässä sudenkuoppana on se, ettei asiansanoituksen laatu välttämättä vastaa ihmistyönä tehtyä. Positiivisena puolena kuitenkin voidaan nähdä se, että näin saatetaan saada edes jollain tasolla kuvailtua aineistoa, joka muuten jäisi kokonaan kuvailematta. Kuvailutyötä automatisoivilla ratkaisuilla voidaan myös rikastaa olemassa olevaa metadataa.

Erilaisten dokumenttien sisältöä kuvaillaan monista eri lähtökohdista. Varsinainen asiansanoitus, koneellisesti tai ihmisvoimin, tehdään käyttötarkoitukseen sopivan kuvailusanaston perusteella, esimerkkinä vaikkapa Yleinen suomalainen ontologia yso tai sen edeltäjä Yleinen suomalainen asiansanasto ysa. Tällä pyritään varmistamaan kuvailun yhteismitallisuutta ja aineiston löydettävyyttä.

Aineistoa voidaan myös kuvailla vapaammin – ja astetta subjektiivisemmin – käyttämättä kontrolloitua sanastoa tai laventamalla sanaston käyttöön liittyviä ohjeistuksia. Tällöin kuitenkin puhutaan ennemmin vaikkapa avainsanojen määrittämisestä tai aineiston ”taggaamisesta”.

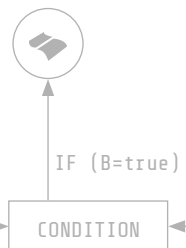
## 1.2 Sisällönkuvailun laadun mittaus

Sisällönkuvailun laadun mittaaminen ei ole suoraviivainen tehtävä. On vai-

keaa määrittää ehdottomat kriteerit laadukkaalle kuvailulle tai löytää jokin vertailukohta, joihin muiden samasta tekstistä tehtyjen kuvailujen oikeellisuutta voitaisiin verrata. Kuvailijoiden välillä, ja usein myös saman kuvailijan eri kerroilla tuottamien kuvailujen välillä, on harvoin kovin suurta yhdenmukaisuutta. Empiirisissä kokeissa eri ihmisten tuottamien kuvailujen yhdenmukaisuus on ollut 30–50 % (Medelyan 2009).

Laadun mittaamisessa ja arvioinnissa voidaan hyödyntää esimerkiksi kuvailijajoukon keskinäistä samanmielisyyttä. Mitä enemmän yksittäinen kuvailu sisältää samoja asiansanoja kuin muut samasta tekstistä tehdyt kuvailut, sitä laadukkaampana tätä kuvailua voidaan pitää (ks. Rolling 1981). Tässä kuvailijaksi voidaan laskea algoritmitkin. Voidaan esimerkiksi laskea, kuinka monta samaa asiansanaa algoritmit ovat ehdottaneet dokumentille ja mitä ihmisten valitsemia asiansanoja jäi puuttumaan. Näin voidaan laskea tiedonhaun arvioinnista tuttuihin saantiin ja tarkkuuteen perustuvia tunnuslukuja, esim. F1-tunnusluku on saannin ja tarkkuuden harmoninen keskiarvo. Joskus arvioinnissa on myös otettava huomioon ehdotettujen asiansanojen järjestys, etenkin tilanteessa jossa oletetaan, että ensimmäiset asiansanat ovat tärkeimpiä. Mittarina esim. NDCG (Normalized Discounted Cumulative Gain) huomioi tämän.

Toinen tapa lähestyä sisällönkuvailun laadun arviointia on kysyä ihmisten mielipidettä ehdotetusta kuvailusta (ks. esim. Golub ym. 2016). Asiantuntijat





(tai maallikot) voivat arvioida tekstile annettujen asiasanojen laatua ja hyödyllisyyttä jollain asteikolla (esim. 1–5 tai hyödyllinen/kelvollinen/väärä). Vastaaajilta voidaan kysyä lisäinformaatiota ja perusteluja avovastauksissa tulosten tulkinnan avuksi.

### 1.3 Automaattisen sisällönkuvailun työkalu Annif

Sisällönkuvailun avuksi on kehitetty työkaluja, joista yksi on tässä artikkelissa käsiteltävä Annif (Suominen 2019). Annifin toiminta perustuu erilaisiin kieliteknologiaa ja koneoppimista hyödyntäviin työkaluihin ja algoritmeihin, joita voi käyttää erikseen tai yhdessä toistensa kanssa. Annif on avointa lähdekoodia, ja kenen tahansa hyödynnettävissä. Annif on myös muokattavissa tietyn käyttötapauksen tarpeita vastaavaksi. Annifilla on rajapintapalvelu, joka mahdollistaa sen integroinnin käyttäjien omiin järjestelmiin.

Artikkelin kirjoittamishetkellä Annifia on hyödynnetty eniten suomen-, ruotsin- ja englanninkieliseen asiasanoitukseen Yleisen suomalaisen

ontologian yso:n perusteella. Annif on kuitenkin periaatteessa kieliriippumaton, sikäli kun taustatyökaluista löytyy halutun kielen tuki. Annif on ainakin toistaiseksi ollut käytössä pääasiassa puoliautomasoiduissa sisällönkuvailuprosesseissa. Annifista ja sen taustamenetelmistä voi lukea lisää englanniksi GitHubin wikisivulta ([github.com/NatLibFi/Annif/wiki](https://github.com/NatLibFi/Annif/wiki)) tai osoitteesta [annif.org](http://annif.org). Annifilla on myös suomenkielinen asiakaswiki ([www.kiwi.fi/display/Finto/Annif](http://www.kiwi.fi/display/Finto/Annif)).

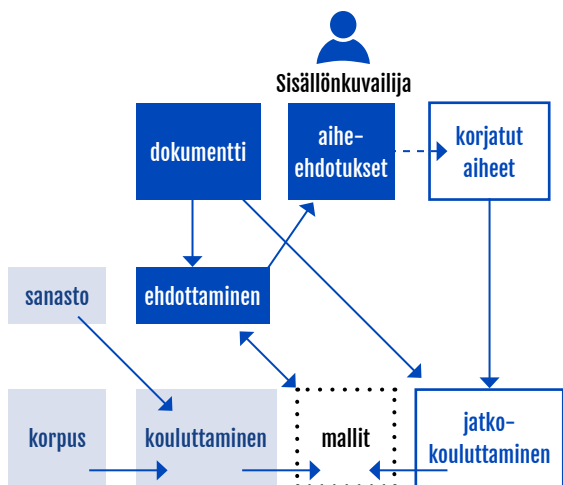
## 2. Automaattisen kuvailun ohjelmiston keskeiset toiminnallisuudet

Automaattisen kuvailun työkalut tarjoavat useita eri toimintoja. Tässä luvussa on keskitytty erityisesti Annif-työkalun tarjoamiin toiminnallisuuksiin. Muissa vastaavissa työkaluissa toiminnallisuudet voivat olla hiukan erilaisia. Automaattisen sisällönkuvailun ohjelmiston keskeiset toiminnallisuudet ja prosessit on esitetty kuvassa 1.

### 2.1 Aihe-ehdotusten tuottaminen

Sisällönkuvailijat ovat puoliautomaattisen sisällönkuvailun järjestelmän varsinaisia käyttäjiä. Sisällönkuvailijan kannalta järjestelmän keskeisin toiminnallisuus on järjestelmän sille syötetyn dokumentin pohjalta tuottamat asiasana- tai luokkaehdotukset.

Ehdottaminen voi tapahtua rajapintapalvelun kautta. Tällöin sisällönkuvailija käyttää omaa kuvailujärjestelmäänsä, joka on integroi-



Kuva 1. Automaattisen sisällönkuvailun ohjelmiston keskeiset toiminnallisuudet ja prosessit

tu automaattisen sisällönkuvailun järjestelmään. Toisessa vaihtoehdossa, joka ei vaadi järjestelmäintegraatiota, sisällönkuvailija syöttää tekstin suoraan automaattisen sisällönkuvailun järjestelmään esimerkiksi verkkolomakkeen kautta ja saa ehdotukset nähtäville kopioitavaksi varsinaisen kuvailujärjestelmään. Tässäkin tapauksessa verkkolomake tukeutuu automaattisen sisällönkuvailun järjestelmän rajapintapalveluun. Annifiin perustuva verkkolomake löytyy esimerkiksi Finto AI -palvelusta (kuva 2.).

Järjestelmän tuottamissa ehdotuksissa on toisinaan aiheita, jotka ovat dokumentin kannalta toissijaisia tai virheellisiä. Toisaalta jokin aivan keskeinen aihe voi puuttua ehdotuksesta. Tällöin sisällönkuvailija voi korjata kuvailuehdotusta poistamalla epärelevantteja aiheita ja lisäämällä uusia ennen kuin kuvailu tallennetaan kuvailujärjestelmään. Samalla kuvailujärjestelmä voi lähettää sekä alkuperäisen dokumentin että korjatun kuvailun automaattisen sisällönkuvailun järjestelmään, joka niiden perusteella päivittää mallia, korjaa toimintaansa ja kehittää ehdotuksia paremmiksi.

## 2.2 Koneoppimismallien kouluttaminen

**T**ekstin sisällönkuvailuun tarkoitettujen koneoppimismallit ovat käytännössä aina sidoksissa tiettyyn sanastoon sekä luonnolliseen kieleen. Käytettävän sanaston ja kielen valinta tapahtuu luomalla ”projekti”, jossa määritellään joukko asetuksia, kuten käytettävä sanasto, kieli, algo-

ritmi ja mahdolliset hyperparametrit eli koneoppimismallien rakenteeseen liittyvät asetukset.

Kun projekti on määritetty, se täytyy yleensä vielä kouluttaa esimerkiksi dokumenteilla. Kouluttamisessa luodaan projektin määritysten mukainen koneoppimismalli ja opetetaan se esimerkkidokumenttien avulla ennustamaan aiheita dokumenteille. Kouluttamisen jälkeen projekti on käyttövalmis.

Kouluttaminen tapahtuu syöttämällä koneoppimismallille korpus, joka koostuu joukosta etukäteen kuvailtuja dokumentteja. Kouluttamisen lopputulokseen vaikuttavat myös käytettävät esikäsittelytavat (esim. sanojen perusmuotoistaminen eli lemmatisointi) sekä koneoppimismallin hyperparametrit. Kouluttaminen saattaa olla raskasta ja vaatia paljon laskentaresursseja ja/tai RAM-muistia. Neuroverkkomallien kouluttamista nopeutetaan usein grafiikkasuorittimilla (GPU, graphics processing unit) suoritettavalla laskennalla, mutta siitä huolimatta mallin koulutus saattaa viedä tunteja tai jopa useita päiviä.

The screenshot shows the Finto AI web interface. At the top, there is a navigation bar with the Finto AI logo, a search icon, and language options (pääsivellä, svenska, English). Below the navigation bar, there is a text input field labeled "Kuvailtava teksti" containing a snippet of text from a journal article. To the right of the input field, there is a "Palaute" button and a "Tietoa" button. Below the input field, there is a "Sisällönkuvailu" section with a dropdown menu for "Sanasto ja tekstin kieli" set to "YSO suomi (2021-09-23)". Below this, there is a section for "Ehdotusten enimmäismäärä" set to "15" and "20". There is also a section for "Aihe-ehdotusten kieli" set to "sama kuin tekstin kieli". At the bottom, there is a section for "Ehdotetut aiheet" with a "Kopioi" button and a list of suggested topics: "aikakauslehdet", "kirjastoala", "järjestöt", and "kirjastot".

Kuva 2. Finto AI -palvelun Annifiin perustuva verkkolomake, johon käyttäjä voi syöttää haluamansa tekstin automaattisesti sisällönkuvailtavaksi.

Projektien sisältämät koneoppimismallit voidaan kouluttaa uudelleen, jos siihen on riittävästi syitä. Muuttunut sanasto, parempi koulutusaineisto tai halu kokeilla uusia hyperparametreja voivat toimia perusteena. Sanastoa päivitettäessä haasteena on sanastoon lisättyjen uusien käsitteiden tunnistaminen. Useimmat koneoppimismallit pystyvät tunnistamaan vain niitä aiheita, jotka esiintyvät koulutusdatassa, joten ne eivät pysty ehdottamaan uudessa sanastoversiossa lisättyjä käsitteitä.

### 2.3 Automaattinen laadun arviointi

Kun projekti on koulutettu, sen tuotamien ehdotusten laatua voidaan arvioida tilastollisesti jo kuvailtuja dokumentteja vasten. Arvioinnin ajatuksena on tuottaa joukolle dokumentteja (testikorpus) aihe-ehdotukset ja verrata algoritmin antamia ehdotuksia aiemmin ihmistyönä tehtyihin kuvailuihin. On tärkeä varmistaa, että testikorpukseen sisältyviä dokumentteja ei ole käytetty projektin kouluttamisessa eli ne ovat algoritmin kannalta uusia, koska muuten algoritmi voi muistaa yksityiskohtia koulutusdokumenteista ja antaa siksi epärealistisen hyviä ehdotuksia.

Ehdotusten laatua arvioidaan tyyppillisesti samankaltaisuusmittareilla kuten saanti, tarkkuus, F1-score ja NDCG. Jos arvioitu laatu ei tyydytä, voi ylläpitäjä kokeilla vaihtaa projektin asetuksia (esimerkiksi hyperparametrit) ja uudelleenkoulutuksen jälkeen arvioida laatua uudelleen, kunnes saavutettu laatu on riittävän hyvä.

---

## 3. Käytettävien algoritmien valinta

---

### 3.1 Tekstin luokittelu sisällönkuvailussa

Koneoppimisen näkökulmasta automaattiseen sisällönkuvailuun on useita erilaisia lähestymistapoja ja siihen voidaan soveltaa monia eri algoritmeja. Yleisesti ottaen sisällönkuvailua voidaan pitää moniluokkaisena tekstin luokittelutehtävänä, jossa jokaiselle syötetekstille poimitaan sitä parhaiten kuvaavat asiasanat käytössä olevasta sanastosta.

Sovellettaessa tekstin luokittelua sisällönkuvailuun, jokaista sanaston asiasanaa vastaa erillinen luokka eli asiasana-avaruudessa on yhtä monta ulottuvuutta kuin sanastossa on asiasanoja. Automaattinen sisällönkuvailu poikkeaa tavanomaisista koneoppimisen luokittelutehtävistä luokkien lukumäärän osalta. Koneoppimisessa on tyyppillistä, että luokkia on suhteellisen vähän, yleensä korkeintaan joitakin kymmeniä tai satoja. Sisällönkuvailussa käytetyissä sanastoissa asiasanojen lukumäärä voi sen sijaan olla kymmeniätuhansia tai jopa enemmän. Esimerkiksi Wikipediassa käytetään yli miljoonaa kategoriaa sivujen luokitteluun. Tekstin luokittelua suureen määrään eri luokkia on viime vuosina tutkittu suhteellisen paljon. Siihen on kehitetty myös useita uusia menetelmiä, sillä sille on löydetty useita taloudellisesti merkittäviä sovelluksia.

Useat käytössä olevat algoritmit palauttavat myös asiasanoja vastaavat hyvyysarvot eli arviot asiasanojen rele-

vanssista kyseessä olevalle syötetekstil-  
le. Näitä hyvyysarvoja voidaan käyttää  
asiasanojen rajoittamiseen esimerkiksi  
siten, että ainoastaan tiettyä kynnysar-  
voa korkeampia hyvyysarvoja vastaa-  
vat asiasanat otetaan huomioon, tai  
asiasanojen järjestämiseen relevanssin  
mukaiseen järjestykseen.

### 3.2 Algoritmityyppit

Automaattisessa sisällönkuvailussa käy-  
tettyjä algoritmityyppiejä ovat esimer-  
kiksi:

- leksikaaliset algoritmit
- tiedonhakualgoritmit
- luokkakohittaiset luokittelualgoritmit
- puualgoritmit
- tiivistysalgoritmit
- hierarkkiset luokittelualgoritmit
- neuroverkot.

Eri algoritmityyppit lähestyvät teh-  
tävää eri näkökulmasta. Esimerkiki-  
si tiedonhakualgoritmien tapauksessa  
sisällönkuvailutehtävä voidaan tulkita  
tiedonhakuongelmana, jossa kuvail-  
tava syöteteksti vastaa hakukyselyä ja  
jokaista sanaston asiasanaa vastaa si-  
tä varten laadittu asiasanaa kuvaileva  
dokumentti tai joukko dokumentteja.  
Puualgoritmit puolestaan perustuvat  
hajota ja hallitse -periaatteeseen, jossa  
alkuperäinen ongelma jaetaan yksin-  
kertaisemmiksi osaongelmiksi ja joiden  
tuloksista alkuperäisen ongelman rat-  
kaisu lopuksi koostetaan.

Annif hyödyntää tekstin kirjoit-  
tushetkellä pääasiassa leksikaalisia ja  
puualgoritmeja, mutta kokeiluja eri  
algoritmityyppien toimivuudesta teh-  
dään jatkuvasti.

Algoritmityyppien rajat eivät ole  
tarkasti määriteltyjä, ja sisällönkuvai-  
lussa käytettävät menetelmät saatta-  
vatkin sisältää osia tai ominaisuuksia  
useista algoritmityypeistä. Algoritme-  
ja voidaan myös yhdistää, ja parhaat  
tulokset saavutetaankin usein erilais-  
ten algoritmien tuloksia yhdistelevillä  
(ensemble) malleilla. Annif tukee yk-  
sittäisten algoritmien käytön lisäksi  
yhdisteleviä malleja.

### 4. Yhteenveto

Sisällönkuvailua tehdään aineiston  
löydettävyyden parantamiseksi, eli  
tiedonhakihoita varten. Kuvailun au-  
tomatisoinnilla pyritään helpottamaan  
ja sujuvoittamaan kuvailijan työtä ja  
saattamaan kuvailun piiriin aineistoja,  
jotka muuten saattaisivat jäädä kuvai-  
lematta. KAM-sektorilla kuvailutyötä  
tekevät esimerkiksi organisaatioiden  
tiedonhallinnan ammattilaiset, mutta  
kuvailua tehdään paljon myös ns. oman  
toimen ohella. Automaattisen kuvailun  
järjestelmiä rakennettaessa yhtenä haas-  
teena onkin järjestelmää hyödyntävien  
kuvailijoiden heterogeeninen tausta, ja  
sitä kautta erilaiset tarpeet.

Kuvailua automatisoitaessa on tär-  
keää seurata ja mitata kuvailun laatua.  
Sisällönkuvailun laadun mittausta ei ole  
kuitenkaan suoraviivainen tehtävä, kos-  
ka absoluuttisen oikeaa kuvailua on  
usein mahdotonta määritellä. Useissa  
tapauksissa voikin olla tarpeen hyö-  
dyntää erilaisia laatumittareita saman-  
aikaisesti.

Annifin kaltaisen automaattisen  
sisällönkuvailun työkalun käyttöön-

otossa lähtökohtana on sopivan kuvailusanaston valinta. Käytettävien koneoppimismallien opettamiseen tarvitaan valitulla sanastolla ihmistyönä kuvailtuja esimerkkidokumentteja. Sisällönkuvailupalvelun suunnittelussa tulee lisäksi ottaa huomioon käytettävien mallien käyttömahdollisuudet erilaisille aineistoille, koulutusaineistojen koko ja ylläpidon tarve.

Automaattiseen sisällönkuvailun ohjelmisto voi hyödyntää monia erilaisia algoritmeja, joiden lähestymistavat poikkeavat toisistaan. Parhaiden tulosten saavuttamiseksi onkin usein tarpeen yhdistellä eri tyyppisiä algoritmeja. Lopulta Annifin kaltainen ohjelmisto tulee tarjota käyttäjien hyödynnettäväksi mahdollisimman helpolla tavalla, esimerkiksi rajapintaintegraationa olemassa olevaan kuvailujärjestelmään. ❖

### Lähteet

GOLUB, K., SOERTEL D., BUCHANAN G., TUDHOPE D., LYKKE M. & HIOM D., 2016. *A framework for evaluating automatic indexing or classification in the context of retrieval*.

Journal of the Association for Information Science and Technology, 67(1): 3-16.  
<https://doi.org/10.1002/asi.23600>

HULKKONEN, J., INKINEN J., KALLIO A., KOSKELA M., LAPPALAINEN M., LEHTINEN M., SJÖBERG M., SUOMINEN O. & YETUKURI L., 2021. *Sisällönkuvailun automatisoinnin haasteita ja ratkaisuja kulttuuriperintöorganisaatioissa*. <https://urn.fi/URN:ISBN:978-951-51-7233-4>

MEDELYAN, O., 2009. *Human-competitive automatic topic indexing* (Doctoral dissertation, The University of Waikato).  
<https://researchcommons.waikato.ac.nz/handle/10289/3513>

ROLLING, L., 1981. Indexing consistency, quality and efficiency. *Information Processing & Management*, 17(2), 69-76. [https://doi.org/10.1016/0306-4573\(81\)90028-5](https://doi.org/10.1016/0306-4573(81)90028-5)

SUOMINEN, O., 2019. Annif: DIY automated subject indexing using multiple algorithms. *LIBER Quarterly*, 29(1), pp.1-25. DOI: <http://doi.org/10.18352/lq.10285>

MIKKO LAPPALAINEN  
 Kansalliskirjasto  
[mikko.lappalainen@helsinki.fi](mailto:mikko.lappalainen@helsinki.fi)

JUHA HULKKONEN  
 CSC – Tieteen tietotekniikan keskus  
[juha.hulkkonen@csc.fi](mailto:juha.hulkkonen@csc.fi)

JUHO INKINEN  
 Kansalliskirjasto  
[juho.inkinen@helsinki.fi](mailto:juho.inkinen@helsinki.fi)

ALEKSI KALLIO  
 CSC – Tieteen tietotekniikan keskus  
[aleksi.kallio@csc.fi](mailto:aleksi.kallio@csc.fi)

MONA LEHTINEN  
 Kansalliskirjasto  
[mona.lehtinen@helsinki.fi](mailto:mona.lehtinen@helsinki.fi)

MARKUS KOSKELA  
 CSC – Tieteen tietotekniikan keskus  
[markus.koskela@csc.fi](mailto:markus.koskela@csc.fi)

MATS SJÖBERG  
 CSC – Tieteen tietotekniikan keskus  
[mats.sjoberg@csc.fi](mailto:mats.sjoberg@csc.fi)

OSMA SUOMINEN  
 Kansalliskirjasto  
[osma.suominen@helsinki.fi](mailto:osma.suominen@helsinki.fi)

LAXMANA YETUKURI  
 CSC – Tieteen tietotekniikan keskus  
[laxmana.yetukuri@csc.fi](mailto:laxmana.yetukuri@csc.fi)