

lina Hepolehto, Mari Elisa Kuusniemi,  
Soile Manninen ja Susanna Nykyri

# RDA:ssa edistetään aineistohallintaa

Datamaailman sillanrakentajana toimiva Research Data Alliance (RDA) juhlisti kymmenettä toimintavuottaan maaliskuussa 2023 Göteborgissa; samassa kaupungissa järjestettiin ensimmäinen RDA-kokoontuminen vuonna 2013. Mitä kymmenessä vuodessa on tapahtunut ja mitä on tulossa seuraavien vuosien aikana?

**R**esearch Data Alliancen (RDA) kokoontumisia eli plenaryja järjestetään kaksi kertaa vuodessa, joten 21.–23. 3. 2023 Göteborgissa järjestettiin jo 20. RDA-kokoontuminen. Osallistujia oli 62 eri maasta, joista paikalla oli lähemmäs 500 ja etänä 230. RDA-sihteeristön lisäksi hybridi-konferenssin paikallisina järjestäjinä olivat Chalmersin teknillinen yliopisto, Göteborgin yliopisto ja Ruotsin kansallinen datapalvelu (Svensk nationell datatjänst, SND).

Avajaispäivän ensimmäisissä istunnoissa neuvottiin muun muassa göteborgilaisessa kevätpukeutumisessa (esim. pilkkihaalari) ennen kuin päästiin muistelemaan menneitä. RDA:ta perustamassa olleet John Wood ja Ross Wilkinsonin kuvailivat, miten kymmenen vuotta sitten ilmassa oli intoa, uupumusta ja jännitystä. Yhteinen kysymys oli, miten data saataisiin tutkijoiden ”ongelmasta” kansainväliseksi hyödyksi. Näkökulman muutos vaati vaikuttamista päättäjiin ja rahoittajiin, joten RDA-konkarit oppivat olemaan jatkuvassa hissihuuvalmiudessa. Linjaustekstit piti muotoilla ymmärrettäviksi, eikä niissä saanut olla tavoitteita kymmentä enempää, mielellään vähemmän.

Toimintaympäristö on muuttunut kymmenessä vuodessa paljon: ensimmäisessä RDA:n kokoontumisessa osallistujia oli 250, ja tällä hetkellä järjestössä on lähemmäs 14 000 jäsentä. Näin suurella joukolla on

20. RDA-kokoontuminen järjestettiin Lindholmenin konferenssikeskuksessa.  
Kuva: lina Hepolehto, CC BY 4.0



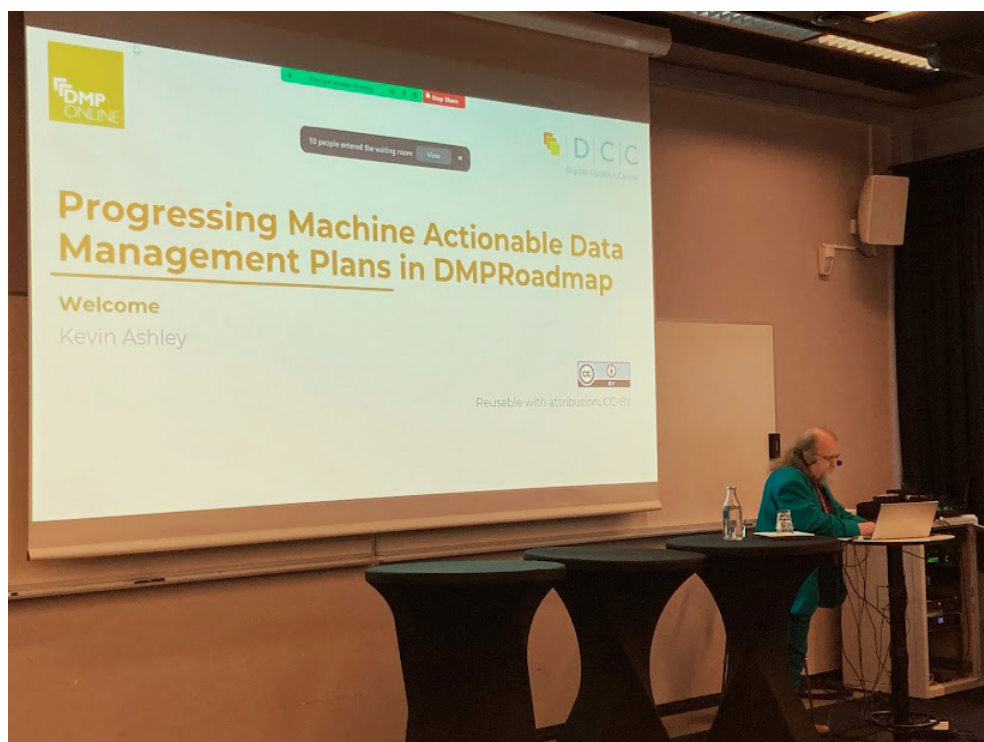
mahdollista ratkaista jopa globaaleja ongelmia. Paikalla oli runsaasti myös tutkimusrahoittajien edustajia, eli lobbaus ei ole mennyt hukkaan.

## Data, yhteistyö ja rahoitus

**Y**ksi tärkeimmistä toimijoista tutkimusdatanhallinnan kentällä on [European Open Science Cloud \(EOSC\)](#), jolle pyhitettiin konferenssissa oma tuokionsa. EOSC on yhdistys, jonka tavoitteena on luoda ”Web of FAIR Data and Services for science in Europe”. Tukholman yliopiston ylikirjastonhoitajalta Wilhelm Widmarkilta kuulumme, että Ruotsissa useimmat suuret yliopistot ovat EOSC-yhdistyksen jäseniä. Hyvällä koordinaatiolla ruotsalaisten ääni kuuluu EOSC:ssa koko ajan suurempana. Suomalaisia yliopistojäseniä on

toistaiseksi vähän, ensimmäiset ovat vasta hankkineet tarkkailijajäsenyyksiä. Suomen yliopistoja EOSC:ssa edustaa CSC – Tieteen tietotekniikan keskus Oy, vaikka se ei olekaan varsinaisesti yliopisto.

Yhdysvalloissa avoin tiede ja data ovat kuin nukkuva jättiläinen. Erilaisia linjauksia on tehty, mutta laajempi käytännön toteuttaminen esim. EOSCin tapaan ei ole onnistunut. Yhtenä RDA:n haasteena on pidetty Eurooppa-keksisyyttä, joten tämänkertaisessa koontumisessa kiinnitettiin erityisesti huomiota Latinalaiseen Amerikkaan, Aasiaan ja Afrikkaan. Latinalaisessa Amerikassa onkin käynnissä varsinainen tutkimusdata-buumi. Koska julkista rahoitusta ei juuri ole, palveluiden kehittäminen tapahtuu yksityisellä rahoituksella. Yhteisten rahoittajien puu-



DMP-työpajan aiheena oli DMP:t ja koneluettavuus. Tilaisuuden vetäjänä oli Kevin Ashley DCC:ltä. Kuva: Soile Manninen, CC BY 4.0

te kuitenkin vaikeuttaa yhteistyötä. Keskeinen uusi avaus RDA:ssa olikin yritys yhteistyön edistäminen, ja siitä saatiin esimerkkejä paneelikeskustelussa, jossa oli mukana keskustelijoita Volvota, Ikealta, Astra Zenecalta ja Amazonilta. Yritykset ostavat dataa käyttöönsä, ja koska data on rahanarvoista tavaraa, siitä pidetään myös huolta. Tämä kannattaa muistaa myös dataa tuottavissa tutkimushankkeissa.

### Työryhmien tapaamiset parasta antia

**S**uurin työ tehdään RDA-kokouksetusten välillä erilaisissa työryhmissä (working groups, WG) ja yhteistyöryhmissä (interest groups, IG), joita on tällä hetkellä yli 100. Mo-

nissa ryhmätapaamisissa painottui enemmän näkemysten vaihto, mutta muutamissa pääsi konkreettisesti tekemäänkin. Koska kaikissa paikoissa ei voi olla yhtä aikaa, ryhmien kokouksumisiin voi palata RDA:n [Youtube-kanavalla](#), jossa on julkaistu 70 videota.

[Sensitiivisen datan yhteistyöryhmässä](#) keskusteltiin datasta eri tieteenalojen näkökulmasta, ja mukana on yleensä jokin ei-sensitiivinen tieteenala. Astronomiassa datan jakamista on tehty kauan, ja mottona onkin: ”As open as possible, closed only if necessary”. Laitteet ovat kalliita, joten yhteistyön tekeminen on fiksua. Yhteentoimivuutta tukevat myös yhteiset standardit. Muutoin yhteiskuntatieteisiin ja elämäntieteisiin painottuneessa keskustelussa muistutettiin myös, että ei-sensitiivinen data voi ajan kuluessa muuttua sensitiiviseksi, mutta tästä puhutaan liian vähän.

Tekoäly herätti monenlaisia näemyksiä. [Artificial Intelligence and Data Visitations -työryhmän \(AIDV\)](#) tapaamisessa kuultiin alatyöryhmien fellow- tutkijoilta tila-



Göteborgin kaupunginkirjaston vihki käyttöön Astrid Lindgren huhtikuussa 1967.

Kuva: Soile Manninen CC BY 4.0

tut esitykset, jotka tukivat ryhmän asiantuntemusta ja tavoitteita. Datavierailuissa data ei siirry itse palvelusta mihinkään, mutta tekoälyn päästäminen käsiksi dataan on riski yksilön oikeuksille ja itsemääräämisoikeudelle, eli henkilöiltä tarvitaan nimenomainen suostumus tekoälyn käyttöön heitä koskevan datan käsittelyssä.

[RDA:n datakoulutusryhmässä \(Education and Training on Handling Research Data, ETHRD\)](#) työn alla on määritellä aineistohallinnan koulutusmateriaaleille vähimmäistason kuvailutiedot. Tavoitena on materiaalien uudelleenkäytettävyys ja yhdisteltävyys, jolloin niitä voidaan hyödyntää erilaisissa konteksteissa, eli ne olisivat käytettävissä kansalaisille, opiskelijoille ja tutkijoille yli maa- ja tieteenalajojen erilaiset taustat huomioiden. Kuulostaakin tutulta jo tutkimusdatamaailmasta. Konferenssin jälkeen julkaistut suositukset [löytyvät Zenodosta](#).

Toinen ETHRD-ryhmän aihe oli aineistohallintakoulutusten pelillistäminen. Luennolla esiteltiin mielenkiintoinen listaus, miten pelaamisen yhdistäminen koulutukseen edistää syväoppimista. Pelillisuus muun muassa motivoi, korjaa mahdollisia harhakäsityksiä keskustelujen kautta, kertaa opittua, sekä auttaa hyödyntämään opittua käytännössä, jolloin uusi asia siirtyy lyhytaikaisesta muistista pitkäaikaiseen muistiin. Sessiossa testattiin leikkillisesti sanaselityspeliä, joka kansainvälisesti tunnetaan Taboo-nimellä. Pienryhmissä jaettiin tietoa aineiston-

## Luennolla esiteltiin mielenkiintoinen listaus, miten pelaamisen yhdistäminen koulutukseen edistää syväoppimista.

hallintaan ja avoimeen tieteeseen liittyvistä peleistä, joista on [koottu kirjasto Zenodoon ja Figshare-data-arkistoon nimellä YUFERING open science games](#).

### Kun DMP ei riitä

**A**ineistohallinnan suunniteluun liittyvät tapaamiset olivat suosittuja, ja niille olisi ollut kysyntää enemmänkin. Ennen varsinaista konferenssia osallistuimme [Digital Curation Centren \(DCC\) järjestämään työpajaan](#), jonka aiheena oli aineistohallintasuunnitelmien (DMP) koneluettavuus. Esitysten aiheina olivat muun muassa erilaiset pysyvät tunnisteet (persistent identifier, PID) sekä niiden käyttö järjestelmien välisissä integraatioissa. Pyöreän pöydän keskusteluissa osa ryhmistä pureutui tarkemmin teknisiin haasteisiin, mutta olennainen ylätason kysymys on, mitä koneluettavuuden eteen kannattaa tehdä, jotta siitä hyötyisi mahdollisimman moni.

RDA:n eri ryhmissä nostettiin esille ohjelmistoille tarkoitettut software management planit (SMP), ja erillisen SMP:n tarpeesta keskusteltiin myös suomalaisten osallistujien kesken. Nykyisessä kansallisessa DMP-pohjassa on

väljyyttä käsitellä myös ohjelmistojä, mutta mikään ei estä organisaatioita tarvittaessa kehittämästä omia SMP-pohjia. Työpajojen keskusteluissa tuli esiin myös vielä laajempi suunnitelma, nimeltään tutkimustuotossuunnitelma (Research Output Plans), kattaen kaikki erilaiset tutkimustuotokset: datan, koodit, ohjelmistot, tietokannat, menetelmät, koodistot, ym. Tällainen suunnitelma oli käytössä esim. NASALLa. Suomessa DMP-konsortio on aloittamassa nykyisen kansallisen aineistonhallintasuunnitelma-pohjan päivittämistä. Päivitystyössä on tarkoitus ottaa huomioon tuore [Tutkimusmenetelmien ja -infrastruktuurien avoin saatavuus -osalinjaus](#), jossa suositellaan mm. menetelmien lisäämistä aineistonhallintasuunnitelmaan.

### Kirjastojen rooli UNESCO:n avoimen tieteen suosituksen toteuttamisessa

**K**irjastojen rooli tutkimusaineistoasioissa on RDA:ssa hyvin ja laajasti tunnistettu. [Libraries for research data -ryhmän](#) kokoontumiseen osallistuu perinteisesti paljon väkeä ja keskustelu on vilkasta. Session alussa tapahtui puheenjohtajiston kapulanvaihtoa: Uusina toivotettiin tervetulleiksi Jennifer Gu (Hong Kong University of Science & Technology Library), Susanna Nykyri (Tampereen yliopisto) ja Tim Dennis (University of California, Los Angeles (UCLA) Library). Pitkän kauden tehneitä pestinsä jättäviä puheenjohtajia Chris Erd-

mann (Michael J. Fox Foundation, New York) ja Birgit Schmidt (University of Goettingen, State and University Library) kiitettiin hienosta kaudestaan.

Ryhmän tapaamisen erityisteenä oli kirjastojen rooli UNESCO:n avoimen tieteen suosituksen toteuttamisessa. Tähän oli pyydytty esitykset useasta eri näkökulmasta.

Iryna Kuchma (EIFL), joka toimi myös session moderaattorina, esitteli seuraavia:

- [UNESCO Open Science Toolkit](#)
- [UNESCO:n avoimen tieteen yhteisön tuotokset \(Zenodo\)](#)
- [Checklist for open access publishers on implementing the UNESCO Recommendation on Open Science \(OASPA:n kanssa yhteistyössä\)](#)
- [Checklist for universities on implementing the UNESCO Recommendation on Open Science \(LIBSENSEN kanssa yhteiskehitetty\).](#)

Seuraavaksi Cécile Swiatek Cassafieres (Library Director, Université Paris Nanterre) kertoi UNESCO:n avoimen tieteen suosituksen laadinnasta ja suosituksen vaikutuksesta erityisesti LIBERin työn kautta.

Myös tässä esityksessä nostettiin esille konkreettisia tuotoksia ja työkaluja, joista on hyötyä laajasti kirjastoissa ympäri maailman, kuten [avoimen tieteen osaamisen sisältöjä kristallisoivasta visualisoinnista](#).

Cassafieres korosti eri osaamisalueiden tunnistamisen sekä tiedon, taitojen ja kompetenssien kehän hahmottamista, kun halutaan kehittää muiden

tai omaa avoimen tieteen osaamista. Esityksessä korostui kirjastossa toimivien asiantuntijoiden syvä arvostus ja merkittävä rooli avoimen tieteen edistämisessä.

Organisatorisempaa näkökulmaa UNESCO:n avoimen tieteen suosituksen puolestaan oli pyydetty Susanna Nykyriltä. Hän toi alkajaisiksi esille Suomessa tehtävän tiiviin kansallisen yhteistyön ja sen merkityksen. Vaikka kansallinen avoimen tieteen linjauksia kehittämistyö on hyvin aktiivista, Nykyri totesi UNESCO:n suosituksen sparraneen sekä kansallista että yliopistotason kehittämistyötä.

Aiempiä kansallisia linjauksia korostuneempaan asemaan ovat nousseet avoimen tieteen ratkaisujen taloudellisen kestävyuden merkitys, laajempi eri toimijoita ja kansalaisia yhdistävä yhteistyön lisääminen sekä kansalaistieteeseen panostaminen. Myös avoimen tieteen elimellinen yhteys hyviin tieteellisiin käytänteisiin on korostunut.

Suomessa tutkimusorganisaatiot eivät pyri ottamaan huomioon ei-kaupallisten toimijoiden ja palveluiden integroimista toiminnansuunnitteluunsa ja hankintoihinsa, kuten suositus tähdentää.

### **Arviointimalleja tulisi kehittää**

Nykyri korosti avoimen tieteen edistämisessä arvioinnin uudistamisen merkitystä ja sitä, miten COARA-yhteistyöllä pyritään aiempaa linjauksuutta konkreettisemmin uudistusta edistämään. Suomessa arvioidaan tutkimusorganisaatioiden avoimen tie-

## **Arviointimalleja tulee kehittää niin, että ne aiempaa paremmin rohkaisevat yhteiskehittämiseen eivätkä keskinäiseen kilpailuun.**

teen edistymistä myös kansallisella tasolla. Yleisössä mielenkiintoa herätti erityisesti AVOTT-seurantamalli ja DMP-konsortiomalli yhteiskehittämisen muotona. Nykyri korosti, että meidän tulee kehittää arviointimalleja niin, että ne aiempaa paremmin rohkaisevat yhteiskehittämiseen eivätkä keskinäiseen kilpailuun.

Paneelikeskustelua edeltävän osaamisen kehittämiseen keskittyvän alustuksen tarjosi Tim Dennis. Hän kertoi kirjasto- ja informaatioalan ammattilaisille tehdystä avoimen tieteen curriculum-projektista, jossa kehitettiin opetus sisältöjä vertaisarvioinnin menetelmin informaatioalan ammattilaisille. Sisällöt käsittelevät laajasti avoimen tieteen menetelmien ja digitaalisten prosessien eri tulokulmia. UCLA jatkaa tätä kehittämistyötä ja pyrkii osaltaan varmistamaan, että kirjasto- ja informaatioalan ammattilaisilla on tarvittavat taidot ja tiedot edistää avointa tiedettä osana tutkimuksen linkaarta.

Esityksiä seurasi paneelikeskustelu, jota juonsivat Iryna Kuchma ja Christopher Erdmann. Panelisteina olivat Susanna Nykyri, Tim Dennis ja Cécile Swiatek Cassafieres.

Panelisteilta kysyttiin mm. mitä



## Mikä on RDA

- Research Data Alliance (RDA) on vuonna 2013 perustettu kansainvälinen yhteisö, jonka tavoitteena on rakentaa avoimen datan jakamisen mahdollistavaa sosiaalista ja teknistä infrastruktuuria.
- Kuka tahansa, joka hyväksyy RDA:n toimintaperiaatteet, voi liittyä jäseneksi rekisteröitymällä [RDA:n verkkosivustolla](#). Yksittäisjäseniltä ei peritä jäsenmaksua. Myös organisaatiot voivat liittyä jäseniksi vuosimaksua vastaan. Tällä hetkellä jäseniä on yli 13 800.
- RDA:n toimintaa johtaa sihteeristö ja kolme hallintoelintä: neuvosto, tekninen neuvottelukunta (Technical Advisory Board, TAB) ja organisaatorinen neuvottelukunta (Organisational Advisory Board, OAB).
- Käytännön työstä vastaavat työryhmät (Working Groups, WG) ja erillaiset yhteistyöryhmät (Interest Groups, IG; Communities of Practice, CoP). Työryhmien toimikausi on yleensä 12 - 18 kuukautta, ja ne tuottavat raportteja, suosituksia ym. ratkaisuja, jotka mahdollistavat datan jakamista, vaihtamista ja yhteentoimivuutta. IG:t toimivat ilman aikarajoitusta. Tällä hetkellä toimivia ryhmiä on 107.



Tupa oli täysi, kun aiheena olivat DMP:t.  
Kuva: Soile Manninen, CC BY 4.0


”alhaalla roikkuvia hedelmiä” kirjastoilla on napata edistääkseen UNESCO:n avoimen tieteen suosituksen toteutumista ja mitä erityisesti suosittelisi tekemään.

Panelistit olivat varsin yksimielisiä ja suosittelivat luomaan tarvittavan verkoston – jos sitä ei jo ole – ja ylipäänsä panostamaan yhteiskehittämiseen. Yksin tehden ei yksittäinen kirjasto pääse pitkälle, siihen tarvitaan nimenomaan yhteistyötä. Avoimen julkaisemisen edistäminen on jo tuttua, mutta se ei riitä.

Avoimien tieteiden ymmärrettävä laajana asiana, mutta on hyvä valita kulloisetkin edistämisen painopisteet, jotta niihin voi kylliksi panostaa. Kirjastoilta edellytetään huomattavaa aktiivisuutta ja osaamisen kehittämistä, jotta voimme edistää UNESCO:n avoimen tieteen tavoitteita ja tukea yhteisöme jäseniä palveluillamme.

Yleisöstä kuultiin myös eriäviä kommentteja perusteluna nyt jo liian laajat osaamisvaatimukset. Panelistitkin tunnustivat haastavuuden ja korostivat erikoistumismahdollisuuksien vahvistamista ja organisaatiotasoa laajemman yhteistyön merkitystä, mikä toki on haastavaa toteuttaa erityisesti pienten organisaatioiden näkökulmasta.

## Irti aikavyöhykkeistä

RDA:n 20. kokoontuminen oli toteutukseltaan oikeasti hybridi, ja etäosallistujat kuuluivat esitykset ja keskustelut välillä paikalla olijoihin paremmin. Ihmisten tapaamisella on kuitenkin omat etunsa, esim. aikavyöhykkeet ovat konkreettinen haaste, joka yhteisellä konferenssilla halutaan poistaa edes hetkeksi. [Seuraava RDA Plenary järjestetään Salzburgissa](#), Itävallassa 23.–26. 10. 2023 osana kansainvälistä dataviikkoa. 

## Kirjoittajat:

IINA HEPOLEHTO

Helsingin yliopiston kirjasto  
iina.hepolehto@helsinki.fi  
orcid.org/0000-0003-4011-7740

MARI ELISA KUUSNIEMI

Helsingin yliopiston kirjasto  
mari-elisa.kuusniemi@helsinki.fi  
orcid.org/0000-0002-7675-287X

SOILE MANNINEN

Helsingin yliopiston kirjasto  
soile.manninen@helsinki.fi  
orcid.org/0000-0003-1009-1180

SUSANNA NYKYRI

Tampereen yliopisto  
susanna.nykyri@tuni.fi  
orcid.org/0000-0002-5018-5176