

# Tutkimushallinnon tietovirrat ja tutkimustietovaranto -johtoryhmä

Kokousmuistio

Paikka: Opetus- ja kulttuuriministeriö, kh. Jukola, Meritullinkatu 1, Helsinki

Aika: 30.10.2017 klo 13–15

Osallistujat

- Tiedeasiantuntija Erja Heikkinen (ryhmän puheenjohtaja), opetus- ja kulttuuriministeriö
- Vararehtori Kaisa Miettinen, Jyväskylän yliopisto
- Rehtori Petri Raivo, Karelia-ammattikorkeakoulu, ARENE-edustaja
- Tietotekniikkapäällikkö Minna Harjuniemi, Helsingin yliopisto, FUCIO-edustaja
- IT-päällikkö Jaakko Riihimäki, Haaga-Helia ammattikorkeakoulu, AAPA-edustaja
- Tietoasiantuntija Aija Kaitera, Helsingin yliopisto, TUHA-verkoston edustaja
- ~~Kirjaston johtaja Jarmo Saari, Itä-Suomen yliopisto, Kirjastojen neuvoston edustaja ESTE, varahenkilö kirjastonjohtaja Ulla Nygren, Turun yliopisto ESTE~~
- Kehittämispäällikkö Riitta Meretoja, Helsingin ja Uudenmaan sairaanhoitopiiri
- Pääjohtaja Heikki Mannila, Suomen Akatemia
- ~~Rahoitusjohtaja Jari Toivo, Innovaatiorahoituskeskus Tekes ESTE~~
- Pääjohtaja Juhani Damski, Ilmatieteen laitos ESTE varahenkilö LYNET-päällikkö Sanna Marttinen, LUKE
- Neuvotteleva virkamies Saara Leppinen, sosiaali- ja terveysministeriö
- ~~Pääsuunnittelija Kai Husso, työ- ja elinkeinoministeriö ESTE~~
- ~~Professori Pekka Appelqvist, Puolustusministeriö ESTE~~
- Professori Virpi Tuunainen, Aalto yliopisto, Professoriliiton edustaja
- Informaatikko Maija Paavolainen, Helsingin yliopisto
- Toimitusjohtaja Päivi Tikka, Säätiöiden ja rahastojen neuvottelukunta
- Johtaja Kristiina Hormia-Poutanen, Kansalliskirjasto
- Professori Anne Mäntynen, Jyväskylän yliopisto, Tiedekustantajien edustaja
- ~~Puheenjohtaja Keijo Hämäläinen, Tieteellisten seurain valtuuskunta ESTE~~
- Johtaja Antti Mäki, CSC - Tieteen tietotekniikan keskus
- Ylitarkastaja Jukka Haapamäki, opetus- ja kulttuuriministeriö
- Kehityspäällikkö Hanna-Mari Puuska (ryhmän sihteeri), CSC - Tieteen tietotekniikan keskus

Asialista

## 1. Kokouksen avaaminen ja asialistan hyväksyminen

Puheenjohtaja avasi kokouksen klo 13. Hyväksyttiin asialista.

## 2. Edellisen kokouksen pöytäkirjan hyväksyminen

Hyväksyttiin edellisen kokouksen pöytäkirja ([Liite 1](#)).

## 3. Tilannekatsaus

Hanna-Mari Puuska esitteli ([Liite 2](#))

Kommentoitiin, että yliopistojen lisäksi pitäisi muistaa myös tutkimuslaitokset, jotka ovat osittain jäljessä tietojärjestelmäkehityksessä. Olisi hyvä voida hyödyntää jo yliopistoissa tehtyä työtä ja olemassa olevia järjestelmiä, ettei kaikkien tarvitsisi tehdä kaikkea itse.

**Päätettiin, että tietovarantoa koskevia käyttötapauksia edelleen tarkennetaan ja konkretisoidaan.**

Kuvauksissa tulisi käydä esille aktiivinen toimija, joka käyttää tietoa johonkin tarkoitukseen. Yliopisto voisi näin ollen ennakoida, mitä tietoa on käytettävissä ja varautua tarvittaviin tietojärjestelmämuutoksiin.

#### 4. Tietosuoja-asiat: Henkilötietojen käsittely tutkimustietovarannossa

Hanna-Mari Puuska esitteli ([Liite 3, selvitys](#))

Kysyttiin, miksi nimen ja affiliaatitietojen käsittely on ongelmallista, kun ne ovat jo valmiiksi julkisia kaikissa julkaisuissa. Vastattiin, että Helsingin yliopistossa tietosuojavastaava on ottanut asiaan kantaa. Yliopiston tutkimustietojärjestelmä erottelee enemmän kuin tekijätiedot viitetietokannoissa, sillä tutkimustietojärjestelmän kautta tulevat henkilötiedot on tunnistettu oikeiksi henkilöiksi. OpenAI:en tapauksessa tutkija on antanut erikseen luvan, että tiedot voidaan tallentaa, kun taas VIRTAn julkaisutietopalvelussa tiedot ovat ilman henkilön varsinaista suostumusta. Toisaalta vaikka henkilö antaisi luvan, teknisen käyttöyhteyden avaaminen ei mahdollista nykyisen julkisuuslain puitteissa.

Todettiin, että asiaa on vaikeaa saada ratkaistua kansallisesti ja pitäisi saada samat käytännöt koko EU:n alueella. Tiukin mahdollinen tulkinta ei ole paras kokonaisuuden kannalta. **Päätettiin, että tutkimustietovarannon ohjausryhmässä seurataan myös muiden maiden lainsäädäntökehitystä.** Esimerkiksi Ruotsissa on tekeillä hyvää tieteellistä käytäntöä koskeva lakiluonnos.

Mainittiin, että korkeakouluissa tietosuoja-asetuksen parissa on jumpattu paljon. Jos tietosuoja-asetusta tiukkaan tulkitaan, pitää jatkokäyttö olla ennalta määrätty ja tiedonluovutukset pitäisi pystyä näkemään etukäteen. Olisi hyvä, että tulkinnat tiedon käytössä eivät olisi liian tiukkoja.

Todettiin, että laki ei anna tulkintoja, vaan ne jäävät oikeuskäytännöissä tutkittaviksi. Tämä on hankala tapa testata sitä, mikä on sallittua. Henkilötietokysymykset koskevat koko tutkimustietovarantoa, mutta tulkinnat ovat vielä auki.

Kerrottiin, että korkeakoulujen tietosuojavastaavat tapaavat säännöllisesti, seuraavan kerran tällä viikolla. Tarkoituksena on löytää yhteisiä käytäntöjä. Tapaamiset on jaettu teemakohtaisiin ryhmiin. T&K-puolella ovat mukana tietosuoja-asioihin perehtyneet juristit Jarkko Reittu (HY) & Maria Reh binder (Aalto).

**Päätettiin, että tutkimustietovarantoasioissa tehdään yhteistyötä korkeakoulujen tietosuojavastaavien verkoston kanssa.**

Kommentoitiin, että myös tietojen elinkaaren osalta tarvittaisiin tulkintaa - esim. oikeus tulla unohdetuksi. Pitäisi saada ääneen lausuttu linjaus, että "ei ole oikeutta tulla unohdetuksi" ja sitten oikeusistuin aikanaan asiasta tarvittaessa tulkitsee.

Mainittiin, että julkaisujen lisäksi myös esimerkiksi infrastruktuurien kanssa on törmätty henkilötietokysymyksiin, kun yliopistoissa on kerätty osana kuvaavaa dataa myös yhteyshenkilöiden tietoja.

Kysyttiin, onko ajatuksena nivoa yksityinen puoli, erityisesti rahoittajasäätiöt, tiiviimmin tähän järjestelmään. Todettiin, että tietojen luovutus ei julkisilla organisaatioilla ole sen yksiselitteisempää muuten kuin lakisääteisen käytön osalta. Esimerkiksi yliopistot ja tutkimuslaitokset ovat tällä hetkellä rekisterinpitäjiä julkaisutietopalvelussa.

#### 5. OKM:n tiedonkeruu ja tietojen käyttö, Tiede-Vipunen

Jukka Haapamäki esitteli ([Liite 4](#))

Todettiin, että lyhyessä ajassa kehitys indikaattoritietojen keruista tiedon laajempiin käyttömahdollisuuksiin on ollut huimaa. Korkeakouluilla on velvollisuus toimittaa OKM:lle kaikki sen pyytämät tiedot. Tietoja voidaan kuitenkin käyttää aiempaa paremmin hyödyksi myös muussa kuin OKM:n omassa toiminnassa.

Kerrottiin, että Akatemian Tieteen tila –tarkastelu on muuttumassa reaaliaikaisemmaksi. Myös Vipusen palvelumuotoilu on käynnistymässä tiedon hyödynnettävyyden parantamiseksi. Tietopohjaa on aiempaa paremmin tarjolla myös tieteen tutkimukselle. Komissio 9. po valmistelu. enemmän riskiä tutkimusrahoituksessa, tulevaisuuden lupaukset.

Komissio on painottanut 9. puiteohjelman valmistelussa, että kansallisesti pitäisi olla valmiimpi riskeihin tutkimuksen rahoittamisessa: miten tietoa hyödynnetään, jos ei enää rahoiteta ainoastaan aiemman menestyksen pohjalta.

## 6. Tutkimustietovarannon portaalin kehittäminen: benchmarking

Hanna-Mari Puuska esitteli ([Liite 5](#)).

Todettiin, että tutkimustietovarantohankkeella on OECD:n DSIP-hankkeessa esimerkkitapauksena urauurtava rooli, koska muissa paikoissa ei ole edetty samoin kuin tässä on tarkoitus.

Mainittiin, että portaaleja on kehitetty useita myös kirjastopuolella ja KDK- ja ATT-arkkitehtuurityö on käynnissä suunnittelun ja toteutuksen tueksi. Kansalliskirjastossa on panostettu erityisesti Finnan kehitykseen (kirjastojen, arkistojen ja museoiden aineistot). Kysyttiin, onko pohdittu näiden olemassa olevien portaalien käytettävyyden arvioimista ja käyttämisvaihtoehtoja. Todettiin, että yhteyksiä pohditaan myös OKM:n sisällä. Toteutus- tai teknologiaratkaisuja ei ole vielä tehty, eikä asiaa ole käsitelty vielä ohjausryhmässä. Nyt ollaan alkuvaiheessa ja haetaan vielä, mitä portaalilla tehdään. Olemassa olevat ratkaisut otetaan huomioon.

Kommentoitiin, että portaalin suunnittelussa ei voida lähteä tekniikka edellä, sillä käyttäjäkokemus on ensimmäinen asia. Yleensä tehdään konseptisuunnitelma, jonka tekemisessä on mukana suunnittelun ammattilaisia ja käyttäjäryhmiä ja sen jälkeen vaatimusmäärittely ja arkkitehtuuru suunnitelma. **Päätettiin, että vaatimusmäärittelyä ja arkkitehtuuria käsitellään myös ohjaus- ja johtoryhmissä.**

Todettiin, että asiassa on kaksi näkökulmaa: yleinen tietosisältö ja käyttäjäkokemus. Nyt ei saisi tehdä valintoja, joita huomataan myöhemmin huonoiksi. Tietosisältö tulisi huomioida niin, ettei kohta huomata, että jokin tarpeellinen tieto jäi ottamatta mukaan.

Mainittiin, että tutkimustietovaranto ei ole pelkästään rakennettava portaaliksi, vaan sen alla tehdään myös tietovarantoa. Tietoarkkitehtuuri alla mahdollistaa tietojen käyttämisen lähdejärjestelmistä, esim. Akatemia hyödyntää tietoja omassa hankeraportoinnissaan. Pohditaan myös, onko portaaleita yksi vai useita eri näkökulmista.

On tärkeää tietää, mitä tiedolla tullaan tekemään ja mitä johtopäätöksiä tiedoista tullaan vetämään. Tällä on vaikutuksia siihen, minkälaisen detaljien taso on relevanttia.

Kommentoitiin, että jo kahden vuoden kuluttua asiat muuttuneet. Rakentamisen pitää lähteä siitä, että maailma muuttuu koko ajan. Toteutusta pitäisi tehdä paloittain, jotta voidaan arvioida jo tehtyjä ratkaisuja pala kerrallaan. Myös organisaatiot itse haluavat hyödyntää tietoja.

Kommentoitiin, että tiedolle tulee helposti uusia yllättäviä käyttötarkoituksia ja odottamatonta merkittävää liiketoimintaa. Esimerkiksi opintotietoja on Norjassa suunniteltu tietoja avattavan rekryfirmoille. Mainittiin, että kaupallinen näkökulma potentiaalisena hyötynä puuttuu tästä. Esimerkiksi sosiaali- ja terveystietojen avaamisessa on ollut hyvin keskeistä yksityisen intressin näkökulman huomioiminen ja siitä seuraavat vaikutukset toimintatavoille, hinnoittelulle ja valvonnalle.

Tutkimustietovarannossa ei tarvitse rajautua siihen, että kaiken pitäisi olla avointa, vaan osa tiedosta voi olla myös suostumukseen perustuvaa. Mainittiin, että kansalliseen palveluväylään ollaan tekemässä yleistä suostumuksenhallintaa. Todettiin, että tässä sovelletaan periaatetta ”niin avointa kuin mahdollista, niin suljettua kuin tarpeen”.

## **7. Datanhallinnan ja laskennan kehittämishjelma (DHL)**

Anu Nuutinen esitteli ([Liite 6](#)).

Mainittiin, että Euroopan tieteellisten kirjastojen verkoston Open Science Policy Platformissa ovat nousseet esille ns. pehmeät asiat, kuten EOSC:n tunnettuuden ja tietämyksen lisääminen sekä se, että osaamisen kehittämisen pitäisi lähteä jo peruskouluopetuksesta.

Euroopan tasolla eri pääosastojen tukemien aloitteiden ja toimintalinjojen keskinäinen koordinaatio on heikkoa. Lisäksi eurooppalaisten aloitteiden rahallinen skaala on erittäin suuri eikä rahoitus ole suinkaan selvä. Kansallisesti tavoitteena on tehdä sellainen ympäristö, että olemme edelleen kansainvälisesti varteenotettava kumppani. Tutkimuslaitosten mukaantulossa ei ole kyse vain siitä, että tutkimuslaitokset alkavat käyttämään CSC:n palveluita vaan eri sektoreiden on myös varauduttava omien kyvykkyyksiensä ja järjestelmiensä kehittämiseen, jotta ne voivat käyttää CSC:n palveluita.

Datanhallinnan ja laskennan ohjausryhmän tehtävissä oli muotoiltu kysymykseksi, miten eri hallinnonalojen tavoitteet tukevat kehittämishjelmaa. Pitäisi ajatella myös, miten kehittämishjelma tukee hallinnonalojen tavoitteita ja muotoilee yhteisiä tavoitteita. Ei ole syytä jäädä odottamaan EU-tason kehittämistä.

Kysyttiin, mitä varsinaisesti ollaan hankkimassa. Kerrottiin, että hankitaan kokonaisuutta, jossa laskentakapasiteetti yksi osa ja tarvittavat palvelut toinen osa. Edelliseen superkonehankintaan verrattuna painotus laskennan ja datanhallinnan välillä on muuttunut. Ei puhuta vain suurteholaskennasta, vaan kokonaisuus on paljon monimuotoisempi. Investointi tulee ensimmäistä kertaa CSC:n taseeseen valtion taseen sijaan.

## **8. Muut asiat**

Ei muita asioita.

## **9. Seuraavat kokoukset**

Järjestetään ensi vuonna kolme kokousta.

## **10. Kokouksen päättäminen**

Puheenjohtaja päätti kokouksen klo 14.55.