

Ohjausryhmä: Tutkimushallinnon tietovirrat ja tutkimustietovaranto

Aika: ma 27.5.2019 klo 12-15

Paikka: nh. Väinämöinen, OKM (Meritullinkatu 10)

Kutsutut:

- Opetusneuvos Jukka Haapamäki (puheenjohtaja), opetus- ja kulttuuriministeriö
- ~~Opetusneuvos Matti Kajaste, opetus- ja kulttuuriministeriö ESTE~~
- ~~Tutkimuksen kehittämispäällikkö, Timo Taskinen Jyväskylän yliopisto ESTE~~
- suunnittelupäällikkö Pirjo Rauhala, Tampereen yliopisto (OK)
- ~~Tutkimuspäällikkö Seliina Päälysaaho, Seinäjoen ammattikorkeakoulu ESTE~~
- erityisasiantuntija Susanna Näreaho, Metropolia ESTE
- Kehittämispäällikkö Minna Liikala, Terveyden ja hyvinvoinnin laitos
- ~~Yliääkäri Ari Lindqvist, Helsingin ja Uudenmaan sairaanhoitopiiri~~
- Tutkimuksen tietoasiantuntija Yrjö Lappalainen, Tampereen teknillinen yliopisto
- Palvelupäällikkö Anne Sunikka, Aalto-yliopisto
- Yksikön johtaja Marja Kylämä, Suomen Akatemia
- Ennakointipäällikkö Soile Ollila, Innovaatorahoituskeskus Business Finland
- ~~Tiedejohtaja Kalle Korhonen, Koneen Säätiö ESTE~~
- Tietojärjestelmäasiantuntija Jyrki Ilva, Kansalliskirjasto
- Suunnittelupäällikkö Janne Pölönen, Tieteellisten seurain valtuuskunta
- Kehittämispäällikkö Ari Leppälahti, Tilastokeskus
- Kehityspäällikkö Hanna-Mari Puuska (sihteeri), CSC - Tieteen tietotekniikan keskus

Vierailijat

- Yliaktuaari Mervi Härkönen, Tilastokeskus
- Tietoasiantuntija Timo Vilen, Kansalliskirjasto
- Ohjelmapäällikkö Tuomo Suortti, Business Finland
- Johtava tiedeasiantuntija Jussi Alho, Suomen Akatemia
- Tietojärjestelmäasiantuntija Osmo Suominen, Kansalliskirjasto
- Projektipäällikkö Tiina Strengell, CSC - Tieteen tietotekniikan keskus

Asialista

1. Kokouksen avaus ja työjärjestyksen hyväksyminen

Puheenjohtaja avasi kokouksen klo 12. Hyväksyttiin työjärjestys.

2. Edellisen kokouksen pöytäkirja [Liite 1](#)

Hyväksyttiin edellisen kokouksen pöytäkirja.

3. Tilannekatsaus (tiedoksi) [Liite 2](#)

- Kysyttiin, milloin tutkimustietovarannon portaalia esitellään tarkemmin. Esitellään portaalin demoa seuraavassa kokouksessa.

4. Viestintäsuunnitelma (Hanna-Mari Puuska, CSC) [Liite 3](#)

- Nimikilpailuun liittyen kysyttiin, onko parhaalle ehdotukselle palkintoa? Ei olla tässä vaiheessa suunniteltu annettavan varsinaista palkintoa.
- Kotimaisten kielten keskusta konsultoidaan nimen valinnan yhteydessä.
- Portaalin lanseeraus tapahtuu näillä näkymin tammi-helmikuu 2020
 - Ensimmäisessä vaiheessa portaalissa näytetään julkaisuja, rahoituspäätöksiä, infrastruktuureja, uutisia, mutta ei esim. linkitettyjä tietoja
 - Uusi ministeri voi vaikuttaa myös siihen, aletaanko erillislakia edistää ja jos, niillä aikataululla.
 - Rahoitustietojen osalta Koneen säätio, SA, BF olisi mahdollista saada mukaan ensimmäiseen versioon.
 - Sitä mukaa, kun saadaan portaalia valmiiksi, voidaan viestiä enemmän. SA:n julkaisutietojen haku toimii hyvänä käytännön esimerkkinä ja auttaa viestinnässä.

5. Tutkijan omat tiedot -käyttötapauksia (Hanna-Mari Puuska, CSC) [Liite 4](#)

- Hankkeen johtoryhmä päättänyt, että suostumuksenhallintaa aletaan toteuttaa ORCID –tunnisteeseen pohjautuen.
 - Pohditaan lainsäädännön lisäksi myös sitä, miltä omien tietojen näkyvyyttä tutkija näkökulmasta.
 - Peruskäyttötapauksien lisäksi löytyy useita erityistapauksia, jotka täytyy miettiä erikseen.
 - Kysyttiin myös, miten toimittaisiin, jos tietoja tulisi monesta eri lähteestä. Ainakin alkuvaiheessa julkaisut ym tiedot tulisivat kotiorganisaatioista, mutta tutkijan omat ”profiilitiedot” ORCIDista.
 - Tiedon hyödyntäminen: Kuinka automaattiseksi tiedon hyödyntäminen saadaan? Tuleeko olemaan esim. tutkijan julkaisuluettelo suoraan vai vaatiiko manuaalisia toimia?
- Erityistapauksia:
 - Tutkija vienyt ORCID –profiiliin muita kuin kotiorganisaatiossa olevia tietoja, esim. julkaisuja. Pitäisikö mahdollistaa myös näiden tietojen tuonti? Haasteena on ainakin muutoksen hallinta. Tieto voi olla ristiriitaista esim. julkaisutiedonkeruun kanssa.
 - Kerrottiin, että jos organisaatio on konsortion jäsen, se voi siirtää tietoja ORCIDiin ja takaisin
 - Onko tarkempaa tunnistautumista kuin ORCID-kirjautuminen? Missä tutkimusorganisaatioiden tekemä validointi tapahtuu?
 - Myös tiedottamissuunnitelman osalta pitäisi ottaa huomioon se, että portaalin tieto ei kata kaikkea.

- Käytännön esimerkkinä 33% uusista professoreista on ulkomailta ja heillä on Suomen ulkopuolisia julkaisuja.
 - 30v historia näyttää erilaiselta – vanhat julkaisut ja ulkomaiset julkaisut?
 - Portaalin osalta linjaus voisi olla, että vanhoja ja ulkomaisia tietoja ei näytetä, mutta ne löytyvät kannasta.
- Julkaisut ORCIDissa ja CRIS:ssä? Ongelmia voi tulla duplikaattien kanssa. ORCIDiin voi lisätä esim. julkaisuja myös manuaalisesti ja tämä altistaa ongelmille. Suostumuksenhallinnassa voisi olla valinta mitä portaalissa näytetään.
- Vanhat ja ulkomaiset julkaisut voi tuoda nytkin VIRTaan. Tutkija voi haluta, että myös vanhat julkaisut näkyvät. Voisiko saada näkymään siten, että mainittaisiin esim. "external"?
- Käyttötapaus 2: Tutkija antaa suostumuksen siirtoon ORCIDista, mutta tietoja ei voi yhdistää julkaisuihin, koska niissä ei ole tunnistetta.
- Käyttötapaus 3: Tutkija antaa suostumuksen siirtoon ORCIDista ja myös yliopistosta, mutta tiedot eroavat toisistaan.
 - Tutkija voisi valita tiedon ensisijaisuuden.
 - Käyttötapausten ottaminen huomioon tekee toteutuksen huomattavasti työläemmäksi. Jos lähteenä on vain ORCID, on helpompaa toteuttaa.
 - Ovatko historiatiedot näkyvissä? Jos vaihtaa organisaatiota, niin täytyy käydä muuttamassa omat tiedot.
 - ORCIDissa on näkyvissä myös historia.
- Henkilötietojen lähteet olisivat ORCID ja henkilö itse. Tietojen tuonti kotiorganisaatioista olisi helpoin toteuttaa niin, että ORCID toimisi tunnisteena.
- Keskusteltiin mahdollisuudesta olla kaksi tunnistetta. Mikäli ORCID -profiileja olisi useampi, yhdistettäisiin tutkijalle vain toiseen profiiliin liitettyjä tietoja.
 - Esim. Puressa tutkijalla ei voi olla monta ORCIDia.
 - Kuinka yleinen tapaus on, että on useampi ORCID? Yksittäisiä tapauksia, esim. kun on vuosia sitten tehnyt ja unohtanut.
 - Suostumuksenhallinnassa voisi ohjeistaa tunnusten yhdistämiseen.
 - Isompi ongelma on se, että tutkijat saataisiin perustamaan ORCID.
- Onko ok, että tunnistautuminen perustuu pelkästään ORCIDiin?
 - Tietosuojavaltuutetut tavattu 23.5. ja kysytty asiasta. Tietojen luovuttamisessa ORCIDin perusteella ei nostettu esille ongelmia.
 - Huolehdittava, että korkeakoulut eivät ottaisi kaikkea riskiä itselleen (vapautukseen lausekkeet, ettei synny taloudellisia riskejä) eikä myöskään tutkija.
 - Mietitty myös Haka –tunnistautumista. Tällöin saataisiin myös tietoja, jotka eivät ole kytketty tutkijan ORCID-tunnisteeseen.
 - Asiaan voi vaikuttaa myös erillislaki, joka odottaa uuden hallituksen muodostamista.
 - ORCID–tiedonsiirto on selkeä, mutta jos haetaan tietoja yliopistosta, niin on selvitettävä tarkemmin.
 - Olisi hyvä luoda ohjeistus tunnisteiden yhdistämiseen, jos tutkijalla on useampi ORCID.
- Käyttötapaus 5: Turha, jos ei ORCID integraatiota. Prosessi voi olla iso ja tehdä toteutuksen sekavaksi.

- Jos tutkija ei anna suostumusta? Onko tietty ryhmä tutkijoita, joka ei anna suostumusta tai jotka tarvitsisivat suostumuksen?
 - Esimerkkeinä tietoturvan tutkijat, jotka eivät ole halukkaita näyttämään tietoja. Näyttäisivät julkaisutiedot, mutta ei muita tietoja.
- Kommentoitiin, että portaalia ei voi markkinoida kattavimpana lähteenä, koska ei ole lainsäädäntöä, joka pakottaa tutkijan tuomaan tietoja.
 - Olisi ainakin hyvä paikka perustiedon etsimiseen.
 - Perustiedot tutkimuksesta ovat laajemmin näkyvillä, tutkijan profiilitiedot ovat enemmänkin tukemassa ja vain pieni osa tutkimustietovarannon sisältöä.
 - Tutkijan tiedoilla kaksi käyttötarkoitusta: tutkijan markkinointi mm. yrityksille ja perustietojen, kuten julkaisujen näyttäminen.
 - Tiedon hyödyntäjänä myös rahoittajat. Erityisesti yksityiset säätiöt haluavat muuta kuin julkaisutietoja. Ajatuksena on se, että rahoittajaa helpotetaan, jos he tarvitsevat tietoonsa esim. CV:n tai affiliaation, joista ei saa tietoa muualta. Affiliaation tuonti ei ole tosin ihan yksinkertaista toteuttaa.
- *Aikataulu: Syksyllä suunnitellaan toteutusta myös teknisesti, ensimmäisenä toteutetaan ORCID –integraatio.*

6. OKM:n yhteisrekisterinpitäjäyysopimukset rahoittajien ja tutkimusorganisaatioiden kanssa (Jukka Haapamäki, OKM) [Liite 5](#)

- Tutkimustietovarantoon toivotaan liittyvän 40 korkeakoulua ja toista sataa rahoittajaa. Ei järjestetä erillisiä neuvotteluja kaikkien kanssa vaan muokataan kaksi erilaista pohjaa siten, että sopimus voitaisiin hyväksyä organisaatioissa suoraan.
- Nyt rahoittajien osalta pohja on kunnossa ja vaatii sopimusten teon. Seuraavaksi käydään SA:n ja BF:n kanssa läpi sopimuskentekoprosessi läpi.
- Yhteisrekisterinpitäjäyysopimuksen sijaan voitaisiin hoitaa asia erillislailla. Tehdään kuitenkin sopimus, koska lain säätäminen ei onnistu puolen vuoden sisällä.
- Korkeakoulujen kanssa asia edennyt hitaammin.
- Korkeakoulujen tietosuojavastaavat tuoneet esille tieteen vapauden perusteena julkaista tietoja. Asia hyvin tulkinnanvarainen. Lähtökohtaisesti julkaisut ovat julkisia ja tieteelliseen toimintaan liittyy julkisuus.
- Mikäli akateemista vapautta käytettäisiin perusteena, täytyisi silti ottaa huomioon, että tiedoista muodostuu henkilörekisteri.
- Iso osa tulkinnoista koskee myös kaikkia Euroopan maita. Onko muissa maissa tarkennettu ja miten muissa maissa tulkittu? Esimerkkinä näistä OpenAIRE, jossa käsittelyperusteita ei ole määritelty.
- *Rahoittajien osalta sopimukset ovat kunnossa, korkeakoulujen osalta ne vielä tarkentuvat.*
- *Selvitystä muihin maihin liittyen edistetään*

7. Apurahatietojen keruu Tilastokeskuksessa (Ari Leppälahti ja Mervi Härkönen, Tilastokeskus) [Liite 6](#)

- Tähän asti tietoja on kerätty työsuhteessa olevilta. Tästä eteenpäin koskee myös ulkopuolisia.
- Paljonko ostettuihin palveluihin liittyy tutkimukseen liittyviä palveluja? Aletaan kysellä yliopistoilta.
- Mikä on tutkimukseen liittyvä palvelu? Esim. koodaajalta ostettu palvelu. Apurahatutkija on myös sisäinen T&K -henkilö, mikäli sopimus. Tieteellistä apurahaa saava opiskelija on ulkopuolinen.
- Kuinka suuria summia apurahatutkijat? Kuvattu liitteen sivulla 5.
 - Ei täysin selkeitä rajoja.
 - OECD –koulutustilastojen osalta kerätään myös vastaavia tietoja. Siellä iso kysymys liittyy jatko-opiskelijoihin, lasketaan joissain maissa työntekijöiksi, joissain enemmän opiskelutyypistä (mm. USA).
 - Jos saa apurahaa, mutta on jatko-opiskelija? Onko jatko-opiskelija järkevä termi? Myös maisteriopiskelijat menevät näihin ryhmiin. He kuitenkin harvoin saavat apurahaa, joten tippuvat siten ulkopuolelle.
 - Kuinka paljon annetaan apurahoja niillä, jotka eivät ole työsuhteessa? OECD:n koulutustilastojen puolella ei ole nostettu esille, mutta taustat ovat varmaankin samat.
- Apurahaa saaneiden ilmoitetut apurahat
 - Mietitty mikä on T&K -työtä.
 - Ei täysin selkeitä rajoja: esimerkiksi henkilöistä osa on töissä yrityksissä yo:n maksamana ja vain 28 yliopistolla? Yrityksen vai yliopiston T&K:ta?
- Keskustelua:
 - Minkä vuoden tiedoista alkaen tulee tiedonkeruukohde? Ei vielä päätetty.
 - Milloin yliopistojen pitäisi alkaa kerätä tietoja? EU:n tilastoasetusta ei vielä hyväksytty. Todennäköisesti keväällä kyselyyn ostettuja T&K -palveluja. Nyt yrityskysely pilotoitavana ja jos pilotointi onnistuu, laajentuu ensi keväälle.
 - Onko työpaikkatieto vuoden lopun tieto? Jos on ollut 2016 aikana työssä. Vai käytetäänkö vuoden lopun tietoa? Työnantaja voi vaihdella vuoden aikana.
 - Liittyykö tutkijasopimukseen, resurssisopimukseen tai vierailijasopimukseen? Tehdään niille, jotka eivät ole opiskelijoita tai työsuhteisia. Tarkoittaako samaa joukkoa? Tätä pystytään raportoimaan ja on myös yliopiston intresseissä seurata mm. hlötyö määrää.
 - Täytyykö tunnistaa alaluokat vai riittääkö raportointi könttänä? Nyt lähdetty rekistereistä, joten jos yliopistolla on jo seuranta, niin olisi hyvä laajentaa kyselyä. Ainakin Tampereella tiedot omassa rekisterissä.
 - OKM:n vierailijat ilmoitetaan vain summina, ei hlöpohjaisia.
 - Tilastokeskus vierailee yliopistoissa ja selvittää raportoinnin tilannetta käytännössä.

8. APC-maksutietojen keruu (Timo Vilen, Kansalliskirjasto) [Liite 7](#)

- Vuoden mittainen projekti OKM:n rahoittamana, toteuttajana FinELib kansalliskirjastossa

- Edelleen hämärä käsitys, kuinka paljon kirjoittajamaksuja maksetaan kokonaisuudessaan.
- Open APC Finland – vastaava projekti Ruotsissa
- Nykytila: heikko tietämys kaikesta, nykytilaa kartoitetaan
 - Myöhemmin pilotointi, jossa 5 osallistujaa eri sektoreilta, eri järjestelmien käyttäjiä (SoleCris, Pure..). Tuloksena arviot mm. vaikutuksista ja maksuista
- Tulevaisuudessa julkaisutiedoille oma kenttä maksuille Virtaan. Pyritään myös jakamaan tietoja avoimesti.
- Erilaisia tarpeita, mm. rahoittajilta (julkaisujen avoimuus)
- Kysyttiin, että jos jatkossa suurin osa lehdistä on avoimia (lukea ja julkaista), mitä summaa käytännössä haetaan ja todistetaan? Jos yliopisto maksaa lukemisen ja julkaisemisen könttäsummana?
 - Pitkäaikainen skenaario on se, että lisenssimaksuista päästään eroon. Tähän on kuitenkin matkaa, koska nyt ostetaan aikaa vain omiin julkaisuihin.
 - Onko odotettavissa, että lisenssimaksujen hävitessä päästään halvemmalla?
 - Nyt on olemassa paljon kustannuksia, joista ei tiedetä.
 - Jo nyt rahat riittäisivät siihen, että kaikki avattaisiin, mutta siihen on kuitenkin muilta osin matkaa.
- Eri tahojen intressien yhteen nivominen on hankalaa. Tiedetään ongelmista ja niitä ilmenee lisää.
 - Pyritään yksinkertaistamaan.
 - Mikä on syy siihen, että APC -maksutietoja kerättäisiin julkaisutietojen yhteydessä? Liittyy kuitenkin talouteen ja tieto on kirjanpidossa.
 - Julkaisutiedonkeruun yhteydessä kerääminen on ajateltu helpottavan seurantaa.
 - Mitä metatietoja on tarkoitus kerätä?
 - DOI, APC -maksu, vuosi, hybridi/ei-hybridi
 - Päämäärä on liittää kustantajiin ja julkaisuihin
 - Tiedot eivät selviä kirjanpidosta eikä kirjastosta
 - Tietojen pitäisi myös kulkeutua ja olla saatavilla
 - Kuka raportoi? Ainakin Puressa on kenttä, mutta ei maksukenttää.
 - Tutkijoille lisätyö
 - Hyvä, että yritetään selvittää myös se osuus kokonaiskustannuksesta, että koitetaan välttää rinnakkaiskustannukset.
 - Tähän ei menisi tutkija raportoimaan, vaan tieto tulisi kirjastosta.
 - OpenAPC on esimerkki, mitkä tiedot kiinnostavat.
- Iso työ, kun joudutaan tekemään yhteistyötä talouden kanssa ja yhdistämään Puressa.
 - Riittäisikö pelkästään kustantajittain? Onko olemassa esim. lehtikohtaisia tilejä?
 - Maksun maksamisen metatiedoista ei löydy niin yksityiskohtaista tietoa, että voisi liittää kustantajaan.
 - Turku hakee APC –maksut taloudelta ja pallo heitetään kirjastolle.
 - Käytännössä yhdistetään maksu (summa) ja maksun syy (artikkeli) - Yhdenmukaistetut tiliöintitiedot – laskussa pitäisi olla tieto.

9. Tutkimustiedon tekstinlouhinta

- Tutkimustietovarannon idea on koota paljon tietoa ja tietoja olisi hyvä hyödyntää. Sen pohjalta pyritään tuomaan ajatuksia, miten tekstinlouhintaa on tehty ja miten sitä pystyttäisiin hyödyntämään.
- **TekesMatch / Business Finland Connect (Tuomo Suortti, Business Finland)**
 - Käytiin läpi Business Finland Connectin demoa. Tämän avulla oli toteutettu asiakasdatan klusterointia.
 - Hakuja suomeksi ja englanniksi
 - Semanttinen haku, joka hakee kaiken, joka liittyy hakusanaan. Käytännössä hakee ja klusteroi. Pystyy muodostamaan näkymän temaattiseen avaruuteen ja klusteroimaan sen.
 - Millaisia teemoja muodostuu? Esim. syöpä ja siihen liittyvät, vaikkei välttämättä sanaa mainittu.
 - Kuvaa myös metatietoja, kuten käsittelijät (hankeryhmän henkilöt)
 - Yritykset etsivät osajia – yritykset näkevät osajia, tutkijat voivat hakea yrityksiä.
 - Kehitystä jatketaan kehittämällä tekoälyä seuraavalle tasolle:
 - klusteroidaan osaamisia esim. startup -yrityksen hakemus
 - Käytännön esimerkkinä mielenkiintoisten kontaktien yhdistäminen asiakkaille.
 - Patenttitietokanta – ketkä toimijat tekevät saman tyyppistä, ja sitä kautta tietoa vastaavista patenteista. Tämän toteutus alkaa kesän aikana.
 - Toiveena, että olisi kollegoiden, asiakkaiden käytössä
 - Yrityksen yhdistäminen kontekstiin vaatii luvan. Ajatus, että rahoitusprosessin kylkeen tulisi kysely tähän liittyen.
 - SA:lta potentiaalisia toimijoita, joiden kanssa voisi esim. keskustella osarahoituksesta.
 - EU:lta yhteistyö ja sen toimijoiden etsiminen
 - Hyödyntämismahdollisuudet Tutkimustietovarannon portaalissa
 - Tieto siitä, mikä on ajankohtaista tutkimusmaailmassa? Mitä haetaan eniten? Mihin rahaa käytetään eniten?
 - Ennakoinnissa kokeiltu, miten tietty termi näkyy hakemuksissa.
 - Esim. droneja kehittävien startupien rahoitus on kasvanut ja miten sen organisaatiotyyppi vaihtuu startupista laajempaan.
 - Voisi olla mahdollisuus hyödyntää tutkimusrahoituksen hankkeiden kuvauksia ja julkaisujen abstrakteja.
 - Täytyisi rakentaa jonkinlainen hakumoottori eli ei löydy suoraa pluginia, mutta periaatteessa mahdollista.
- **Tiedonlouhinta Akatemian rahoitushakemuksissa (Jussi Alho, Suomen Akatemia)**
 - [Liite 9](#)
 - Nyt analysoitu mm. mitkä organisaatiot ja tiedekunnat saavat rahoitusta ilmastotutkimukseen liittyen.

- TTV:ssa hyödyntäminen
 - Nousevat aiheet potentiaalinen käyttökohde
 - Analyysiä koskien mm. teemoja (tutkimusalat + teemat) ja ilmiöitä (avainsana + avainilmaisupohjaisesti)
 - Voisi hyödyntää samoihin tilanteisiin kuin edellistä.
- **Annif-aiheentunnistustyökalu (Osma Suominen, Kansalliskirjasto) Liite 10**
 - Kehittäminen on lähtenyt kirjastojen yms tarpeista
 - erilaisia ja erityyppisiä lähteitä
 - Annifin avulla toteutetaan asiasanotusta ja luokitusta. Tuottaa sisällönkuvausta tekstin perusteella hyödyntäen kieliteknologiaa ja koneoppimista.
 - Idea lähtenyt Finnan tiedon hyödyntämisestä (Finna – anniF)
 - Finnasta haetaan, Annif ehdottaa aiheita
 - Hyödynnetty myös JYX:ssa, jossa ehdottaa avainsanoja opinnäytteille. Analysoi koko gradun, ei ainoastaan abstraktia.
 - Mobiilisovellukset (m.annif.org – toimii ”linssinä”)
 - Tehty käsityönä (mm. finto)
 - Avointa lähdekoodia, GitHubissa
 - Annif.org – lomake, jolla voi testata
 - rajapintapalveluja käytettävissä

Käydään myöhemmin läpi, miten tutkimustietovarannossa voidaan hyödyntää.

10. Seuraava kokous

- 21.8.2019

11. Kokouksen päättäminen

- Puheenjohtaja päätti kokouksen noin 15.05