

Kokousmuistio

Tutkimushallinnon tietovirrat ja tutkimustietovaranto -ohjausryhmän 1/2017 kokous

Paikka: OKM, Meritullinkatu 10, Kultakabinetti

Aika: 9.5.2017 klo 12–15

Osallistujat

Ylitarkastaja Jukka Haapamäki (ryhmän puheenjohtaja), opetus- ja kulttuuriministeriö

Tutkimuksen kehittämisspäällikkö, Timo Taskinen Jyväskylän yliopisto

Tutkimuspäällikkö Sellinä Päälyssaho, Seinäjoen ammattikorkeakoulu

Kehittämisspäällikkö Minna Liikala, Terveyden ja hyvinvoinnin laitos

Yliääkäri Ari Lindqvist, Helsingin ja Uudenmaan sairaanhoitopiiri

Tutkimuksen tietoasiantuntija Yrjö Lappalainen, Tampereen teknillinen yliopisto

Asiantuntija Anne Sunikka, Aalto-yliopisto

Yksikön johtaja Marja Kylämä, Suomen Akatemia

Ennakointipäällikkö Soile Ollila, Innovaatorahoituskeskus Tekes

Tiedeasiamies Kalle Korhonen, Koneen Säätiö

Tietojärjestelmäasiantuntija Jyrki Ilva, Kansalliskirjasto

Suunnittelija Janne Pölönen, Tieteellisten seurain valtuuskunta

Kehittämisspäällikkö Ari Leppälahti, Tilastokeskus

Kehityspäällikkö Hanna-Mari Puuska (ryhmän sihteeri), CSC - Tieteen tietotekniikan keskus

Vierailijat

Kehitysjohtaja, Avoin tiede ja tutkimus -hankkeen pääsihteeri Pirjo-Leena Forsström, CSC - Tieteen tietotekniikan keskus

Vanhempi myyntikoordinaattori, sihteeri Leena Jukka, CSC - Tieteen tietotekniikan keskus

1. Kokouksen avaaminen, osallistujien esittäytyminen ja työjärjestyksen päättäminen

- Puheenjohtaja avasi kokouksen klo 12.05.
- Työjärjestys hyväksyttiin muutoksitta.

2. Kokouskäytännöt

- Ryhmän varapuheenjohtajaksi hyväksyttiin opetusneuvos Matti Kajaste, opetus- ja kulttuuriministeriö.
- Kokousmateriaalit julkaistaan ryhmän wiki-sivulla <https://wiki.eduuni.fi/x/4YFTAg>.
- Kokousmateriaaleja saa jakaa verkostoille, jotta saadaan verkostojen kommentit ja näkemykset esille.
- Jäsenet voivat nimetä varahenkilön, ja erityisesti verkostojen edustajille toivotaan varahenkilöitä
- Varahenkilöt ilmoitetaan Leena Jukalla (leena.jukka@csc.fi), heidät lisätään ryhmän sähköpostilistalla ja he saavat myös tiedon kokousmateriaaleista.
- Mahdolliset matkalaskut lähetetään Leena Jukalle.
- Kokouskutsut lähetetään kaksi ja materiaalit viikkoa ennen kokousta (materiaalien jakaminen on suotavaa)
- Kokouspöytäkirjat ovat kaksiosaisia. Päätösaasiat kirjataan pöytäkirjaksi ja keskusteluasiat julkaistaan muistioina. Molemmat julkaistaan wiki-sivulla, materiaaleja ei lähetetä sähköpostissa.
- Etäosallistuminen ei ole mahdollista, sillä kokouksia on harvoin ja tavoitteena saada keskustelua. Kokousaikataulu pyritään järjestämään niin, että kauempaakin tulevat ehtivät hyvin.

3. Asettamiskirje: ryhmän asettaminen ja tausta. Jukka Haapamäki esitteli.

- TUHA-verkoston viitearkkitehtuuryö kytketään ATT:n viitearkkitehtuuryöhön. TUHA on korkeakoulujen ja tutkimuslaitosten organisoima tutkimuksen tuen ja hallinnon verkosto. OKM rahoittaa verkostoa antamalla CSC:n koordinaatiopalvelut verkoston ja joidenkin sen ryhmien käyttöön.
- Onko prioriteetteja? Kyseessä on vähän eritasoisia asioita, joten kaikkia viedään eteenpäin samaan aikaan.
- Tutkimustietovarannosta on oma projektisuunnitelma.

4. Tutkimustietovarantojen tilannekatsaus

4a. Tutkimustietovaranto ja tutkimustietovirrat. Hanna-Mari Puuska esitteli.

- Julkaisujen metadataan ei liity mitään tietoja siitä, missä hankkeessa tai millä laitteistolla ne on tehty. Mikäli on mahdollista yhdistää nämä ilman yliopistojen lisätyötä, niin yliopistot ovat hankkeen kannalla. Tietojen yhdistäminen ei lisää raportointia. Esimerkiksi Suomen Akatemian tapauksessa, kun tutkija raportoi hankkeesta, hän saa hankkeen aikana syntyneiden julkaisujen tiedot VIRT-julkaisutietopalvelusta suoraan raporttiin. Akatemiasta puolestaan tulee VIRTaan hankenumero.
- Infratietopankissa ei ole tietoa julkaisusta, mutta sieltä voidaan poimia infrojen tiedot raportteihin.

- Tavoitteena on, että kaikki tekijätiedot ovat avointa dataa, vaikka sisältävätkin henkilötietoja, siis tekijöiden tietoja. Tavoitteena on käsittelyperusteen antava lainsäädäntö, joka mahdollistaisi tämän: esim. jonkin toimijan lakisääteinen tehtävä.
- Joku joutuu kuitenkin tuottamaan metatiedot ja tämä työ vaikuttaa suurelta. Onko tarkoitus, että tässä vaiheessa pohditaan sitä tutkijaa tai muuta inhimillistä panosta, joka sitoutuu tähän metatiedon tuottamiseen? Onko teknologista ratkaisua siihen, miten metatietoa tuotetaan? Vastauksena on, että tutkijat tuottavat tällä hetkellä metatietoa ja monta kertaa. Jos ajatellaan esim. ORCID-kehitystä, niin metadataa kertyy muualtakin kuin tutkijan itsensä tallentamana. Tässäkkin ryhmässä voidaan käydä tätä asiaa koskevaa keskustelua, jotta tämä kokonaisuus selkiytyisi.
- Jos esim. tutkija on lähettänyt käsikirjoituksen lehteen, niin ORCID auttaisi metadatojen siirtämisessä.
- Tavoitteena avoimet rajapinnat, mutta toistaiseksi tilanne on sellainen, että jos joku haluaa esimerkiksi VIRTA-julkaisutietopalvelun tietoja käyttää, on täytynyt pyytää lupa.
- Tekes Match -kehitystyö, jossa kehitetty tukea yritysten ja tutkijoiden yhteistyöhön: sopivassa vaiheessa projektipäällikkö voisi tulla esittelemään asiaa ohjausryhmälle.
- Akatemian ja yliopistojen yhteistyönä on kehitetty Etsi expertti -palvelu.
- Julkaisutietovarannon päälle on Turun yliopiston RUSEn tutkimushankkeessa tehty altmetriikka-sovellus, jossa julkaisuille voidaan laskea keskeisiä altmetriikkaindikaattoreita.
- Ohjausryhmä on hyvä foorumi keskustella siitä, kannattaako tehdä kansallisia ratkaisuja vai ostavatko yliopistot palvelun muualta.
- Tekes Match - etsii semanttisia älykkäitä yhteyksiä, AMK-puolella huomattu, että heidän tiedon käyttäjänsä on usein pk-yritys, jolla ei tiedonhakukokemusta.
- Ovatko populaarit julkaisuja mukana? Nyt Virrassa on OKM:n keräämät ammatilliselle ja suurelle yleisölle tarkoitetut julkaisut. Ministeriön määritelmän mukaisesti mukana ovat ministeriön tiedonkeruuta varten keräämät julkaisut.
- Konkreettiset hyödyt -kalvoa pidettiin hyvänä ja levityskelpoisena. Siihen voi lisätä hyötyjä sitä mukaa, kun niitä tulee esiin ja oman kohderyhmän näkökulmaa silmällä pitäen.
- Milloin jostain tulee hanke tai mikä on tutkimusinfra? Onko kyseessä hanke silloin, kun se saa rahoitusta vai kun siitä rupeaa tulemaan tuotoksia? Tarkempaa tietoa ja konkretiaa esiin, niin asiaa on helpompi viedä eteenpäin omassa organisaatiossa. Pienestä lähdetään liikkeelle ja sitten lisätään.
- Onko sisältö rajattu vain suomalaisten organisaatioiden tutkijoihin? Entä jos tutkija on toiminut ulkomailla? Ohjausryhmässä voidaan jatkossa pohtia tutkimustietovarannon rajauksia tältä kannalta.
- Tietojen vieminen järjestelmään: osittain automaattisesti, Web of Science:sta ja Scopuksesta saadaan julkaisun tieto tuotua suoraan, mutta tutkijat voivat myös itse tallentaa tietoja. Joissain kirjastoissa tehdään tätä työtä.
- Esimerkiksi Tampereen teknillisestä yliopistosta tiedot päivittyvät VIRTAAan kerran vuorokaudessa. Virheiden korjausta varten TTY:llä on rakennettu oma tietokanta ja siihen virheen tarkastuksia. Jos löytyy virhe, siitä tulee samantien ilmoitus. Tietokanta skannataan monta kertaa päivässä ja virheet korjataan ennen kuin ne lähtevät VIRTAAan .
- Yliopistojen tietojärjestelmät voisivat olla hyvä lähtökohta, mutta tutkijoiden pitäisi itse voida päättää omista tiedoistaan, mitä he haluavat näyttää ja mitä piilottaa
- Entä tiedon luotettavuus? Kun tutkijat luovat itse, tieto ei ole välttämättä luotettavaa. Jos tietoja tulee muista järjestelmistä, miten tutkija voi järjestellä näitä tietoja tai entä jos tutkija haluaa salata jotain tietoja? Vaikuttaako tämä ja miten yliopiston rahoitukseen?

- Tällä hetkellä Research Gatesta tai academia.edu:sta ei voi tuoda tietoa.
- Mikä on suhde research.fi -portaaliin? Tuleva korvaisi nykyisen.
- Puhutaan kansallisesta tiedon tuottamisesta, mutta tieto on kansainvälistä ja esimerkiksi biolääketieteen tutkijoilla tieto on kv. tietokannoissa. Tutkimus on globaalia, mutta esim. osaajia ja yrityksiä haetaan kansallisesti. Kansallisella palvelulla on paitsi ohjauksellinen myös näyteikkunallinen rooli. Se näyttää, mitä Suomessa tehdään. Varsinaiseen tutkimuksen tekemiseen tällä on rajallinen vaikutus. Esim. Koneen Säätiö ja Akatemia rahoittavat tutkimusta, jossa osa porukasta on ulkomailla. Suomalaiset tutkimusorganisaatiot ja rahoittavat ovat hyvä rajaus.
- Finnan metatiedot on lisensoitu CC0-lisenssillä. Täytyy vain tietää lisensoiva taho, mutta kun se on selvillä, niin tämä on hyvä tapa.
- Tuntuu hassulta, että erikseen täytyy pyytää tietojen käyttämiseen lupa, kuten Turun yliopiston RUSEn ja Suomen Akatemian tapauksessa aiemmin.
- Joskus julkaisija vaatii, että artikkelin käsikirjoituksen lisäksi pitää submitata analyysin pohjana oleva tutkimusaineisto. Tämä on ehkä enemmän ATT-asiaa, mutta metatiedot ovat myös tämän hankkeen asiaa.
- Jos tutkija antaa luvan laittaa nimensä yhteen asiaan ja sitä siirretään sitten eteenpäin, niin tarvitaanko uusi lupa tietojen siirrolle? Lakimiehet ovat tarkkoja siitä, että lupa pitäisi olla. Tämä voisi olla esimerkiksi yliopistoilla tietosuojaselosteessa. Rekisteröity voi antaa suostumuksensa vain tiettyyn käyttötarkoitukseen.
- Mitä tietojärjestelmiin pitää korjata riippuu siitä, minkälainen kansallinen lainsäädäntö on asetuksen pohjalta. OKM Immo Aakkula on OM:n Tatti-ryhmässä, joten asia menee monella rintamalla eteenpäin.

4b. Julkaisutietovaranto. Hanna-Mari Puuska esitteli.

- Talousraporttien erilaisuus puhuttanut jo aiemminkin rahoittajia. Toivottavaa olisi, jos Virrasta saisi edes jonkun osan, joka olisi yhteinen rahoittajille. Erilaisuutta rahoittajien kesken on paljon: raportointia tehdään joko vuosittain tai vasta hankkeen päätyttyä, rahoitettavat asiakkaatkin ovat erilaisia.
- Jos tiedossa palvelun tarjoajia, jotka haluavat integroitua Virtaan, voi olla yhteydessä CSC:hen. Tietojen käyttö vaatii kuitenkin lupaprosessin.
- Koska ei ole tutkijaa identifioivia tunnisteita, nimikaimojen kanssa voi olla ongelma, sillä Virrasta tulee molempien julkaisut. ORCID voisi olla ratkaisu.
- Tekesillä ja Suomen Akatemialla on yhteinen aineistohallintasuunnitelma. Tiedot ovat DMPTuuli-aineistohallintatyökalussa, mutta mietittävä miten se liittyy kokonaisuuteen. Suunnitelmassa kuvattuja aineistoja ei ole vielä olemassa. Tutkijat käyttävät samaa aineistohallintasuunnitelmaa usein.
- ATT- ja tutkimustietovaranto-hankkeen liittymäpinnat ja vastuunjako kirjoitettava auki. Tätä varten on järjestetty kokous ma 15.5.2017, johon osallistuu ATT:n ja tutkimustietovaranto-hankkeen edustajia.

4c. Hanketietovaranto. Hanna-Mari Puuska esitteli.

- ESR- ja rakennerahastot ovat amk:ien päärahoittajia, mutta hanketietovarannon suunnittelua varten tehty kysely ei lähtenyt heille. Katsottiin, että ne [maakuntaliitot ja ELY-keskukset] ovat toimijoina niin erilaisia rahoittajina, ettei niitä otettu tähän kyselyyn mukaan. Asiaa tulisi kuitenkin selvittää yhteistyössä amk:ien kanssa.
- Rahoittajaryhmän kokous pidetään 8.6., ja silloin käydään kyselyn tulokset läpi sekä mietitään jatkoa. Mikäli jäsenille tulee mieleen kehittämissuhteita tai kommentteja, niin mielellään otetaan vastaan.

4d. Tutkimusaineistojen metatietovaranto. Pirjo-Leena Forsström esitteli.

- Tutkimusaineistojen metatietovaranto eli Metax toteutetaan avoimen lähdekoodin ratkaisulla. Ytimenä on tietokantamoottori, jonka päälle tehdään erilaisia kyselyalgoritmeja.
- Tutkimusaineistojen tietomalli on kansainvälisesti yhteensopiva. Tietomallityöryhmän vetäjä on Pauli Assinen Helsingin yliopiston kirjastosta.
- Tietomallit kerätään IOW-yhteentoimivuusvälineistöön iow.csc.fi, jonne tulee myöhemmin tutkimustietovarannonkin tietomalli. Tutkimusaineistojen osalta on huomattava, että on paljon tietomalleja eri tieteenaloilla. IOW:n avulla pystytään rakentamaan yhteentoimivuutta.

4e. Tutkimusinfrastruktuurien tietopankki. Pirjo-Leena Forsström esitteli.

- Tutkimusinfrastruktuurien tietopankki infrat.avointiede.fi.
- Infrastruktuurien viittauskäytäntöjä ei maailmalla ole, vaan Suomi on ensimmäisten joukossa.
- Infojen tunnisteet olisi hyvä saada näillä infoilla tuotettujen aineistojen ja julkaisujen metatietoihin.
- Käsitteenä infrastruktuuri on vielä kehittymässä, mutta lähdetty liikkeelle Suomen Akatemian infoista, koska ne ovat tässä kontekstissa selkeästi määriteltyjä.
- TUHA-verkoston tietoinfrastruktuurityöryhmä on kehittänyt yhteistä luokittelua.
- Tietosuoja sääntely tulee vaikuttamaan jatkossa. Nyt esim. Suomen Akatemia on rekisterinpitäjä infrastruktuuritietopankissa, mutta jos korkeakoulut tulevat mukaan, heistäkin tulee rekisterinpitäjiä.

5. Kansainväliset verkostot

5a. Katsaus kansainväliseen yhteistyöhön tutkimustiedon hallinnan alueella. Hanna-Mari Puuska esitteli.

5b. Eurooppalaisen julkaisutietopalvelun pilotointi. Hanna-Mari Puuska esitteli.

6. Tiedoksi

6a. Kansainvälinen benchmark-selvitys <https://wiki.eduuni.fi/x/sC9iAg>. Selvitystä päivitetään hankkeen kuluessa yhteisvoimin.

6b. Johtoryhmän kokousajat 12.6.2017 ja 30.10.2017

6c. Tutkimustietovarantoseminaari järjestetään 30.10.2017

7. Muita asioita

- Ei muita asioita.

8. Seuraavat kokoukset

- Syksyllä järjestetään kolme kolmentunnin pituista kokousta.
- Kokousajankohdista lähetetään jäsenille doodle-äänestys.

Päätöslista

2. Kokouskäytännöt

- Varapuheenjohtajaksi hyväksyttiin opetusneuvos Matti Kajaste, opetus- ja kulttuuriministeriö.
- Jäsenet voivat nimetä itselleen varahenkilöt.
- Kaikki materiaalit julkaistaan ryhmän wiki-sivulla, eikä niitä lähetetä postitse.
- Etäosallistuminen ei ole mahdollista.

4b. Julkaisutietovaranto

- ATT- ja Tutkimustietovarantohankkeiden liittymäpinnat on kirjoitettava auki ja vastuunjako tuotava esille.

8. Seuraavat kokoukset

- Syksyllä pidetään kolme kolmen tunnin kokousta. Doodle-äänestys ajankohdista avataan.