

Tutkimustietovarannon työpaja tutkimuslaitoksille 12.2.2019

Aika: Tiistai 12.2.2019 klo 13-16

Paikka: CSC, Life Science Center, 1. krs (Keilaranta 14, Espoo).

Tutkimuslaitoksilla ei velvollisuutta siirtää tietoja tutkimustietovarantoon, mutta sen kautta tiedot saa hyödynnettäviksi laajasti (esim. rahoittajien palveluissa), tutkimustietovaranto tuo näkyvyyttä ja tutkimuslaitokset voivat hyödyntää rahoittajien ja muiden tutkimusorganisaatioiden tietoja.

Esittelypisteillä esiin nousseet asiat:

Tutkimustietovarannon portaali

Tutkimustietovarannon portaali on avoin kanava tutkimustietovarantoon kootun tiedon hyödyntämiseksi. Portaalissa on tarkoitus mahdollistaa suomalaisen tutkimukseen liittyvien tietojen haku eri näkökulmista ja esitellä tutkimuksen kokonaiskuvaa. Esittelypisteellä tuli ilmi, että portaali oli osalle tuttu. Portaalia selailtiin mielenkiinnolla ja odotetaan mitä tulee jatkossa. Kevään tilaisuuksissa on tarkoitus äänestää nimeä portaalille.

Tutkimusaineistojen metatiedot – Fairdata –palvelut

Tutkimusaineistojen metatietoja kootaan Fairdata –palveluihin kuuluvaan METAXiin. Tutkimusaineistojen metatietojen osalta keskusteltiin pysyvistä tunnisteista ja tietovirroista. Aineistoista kiinnostuneet voivat liittyä TSV:n vetämiin avoimen tieteen ryhmiin <https://avointiede.fi/fi/asiantuntijaryhmat> . Ryhmässä keskustelua mm. dataviittauksista ja tulossa työpaja tunnisteista.

Tutkimusaineistojen säilytys – Fairdata –palvelut

Fairdata-palveluihin kuuluu myös tutkimusdata säilytyspalvelu IDA. Tutkimuslaitokset ovat saaneet kesällä tunnukset ja voivat ottaa IDA:n käyttöön. Myös testitunnuksia jaettu ja pidetään voimassa ainakin seuraavan viikon ajan. Mikäli aiheesta on kysyttävää, voi olla yhteydessä asiakaspalvelu@csc.fi. Keskustelua käytiin myös käyttöpolitiikasta ja todettiin, että IDA-tilan oikeudet jakaa organisaation vastuuhenkilö. IDA:an tallennetaan tiedostot, jotka voi siirtää pysyväksi.

Tutkimusinfrastruktuurien tietopankki

Tutkimusinfrastruktuurit tarkoittavat kansallisesti merkittäviä tutkimusvälineitä, laitteistoja, aineistoja ja palveluita, jotka mahdollistavat tutkimus- ja kehitystyön. Keskusteltiin, mitä infrastruktuuri käytännössä tarkoittaa eli millaisia infrastruktuureja tietopankkiin kerätään. Nykyisessä infrapankissa on listattu Suomen Akatemian tiekartalla olevat infrastruktuurit, jotka ovat laajoja kokonaisuuksia, mutta jatkossa on tarkoitus kerätä infrastruktuurien tietoja laajemmin myös mm. tutkimuslaitoksista. Käytännössä organisaatioilla on päätösvalta sen suhteen, millaiset infrastruktuurit ja palvelut tutkimustietovarannossa näytetään. Jotta tietojen toimittaminen olisi mahdollista ja mahdollisimman helppoa, palvelun tietomallia on muokattu yksinkertaisemmaksi yhteistyössä TUHA:n infrastruktuuriryhmän kanssa, jossa on osallistujia etupäässä yliopistoista, mutta myös tutkimuslaitokset ovat tervetulleita: <https://wiki.eduuni.fi/x/Vpz4Ag> .

Nykyisestä pilottivaiheessa olevasta palvelusta (<https://avaa.tdata.fi/web/infrat>) luovutaan, kun infrapankki liitetään tutkimustietovarannon yhteyteen.

Tutkijoiden profiili- ja aktiviteettitiedot ja suostumuksenhallinta

Julkaisujen, aineistojen ja infrojen lisäksi tutkimustietovarantoon on tarkoitus koota tutkijoista profiili- ja aktiviteettitietoja (esim. tutkimusalat, nimen kirjoitusasu, affiliaatiot, asiantuntijatehtävät), joita tutkija on tallentanut itsestään esim. ORCID-palveluun tai oman kotiorganisaation järjestelmään. Tutkimuslaitoksista osa kerää tällaisia tietoja, mutta tiedonkeruu toteutettu hyvin eri tavoin. Jotta tutkijoiden omat tiedot voidaan hakea ORCIDista tai omasta organisaatiosta sekä näyttää tutkimustietovarannon portaalissa ja välittää hyödyntäjälle, vaaditaan tutkijalta suostumusta.

Kommentoitiin mm. tutkijan motivointia – miksi tutkija näkisi vaivaa suostumuksen antamiseen? Tutkijalle suurimpana hyötynä tässä on näkyvyyden edistäminen kansallisella tasolla sekä tietojen välittäminen organisaatioiden välillä esim. rahoitusta hakiessa tai organisaatiota vaihtaessa (yliopistot, tutkimuslaitokset rahoittajat). Tutkijat itse todennäköisesti hakevat muita tutkijoita muista palveluista, kuten ResearchGate:sta. Tarkoituksena on, että kansainvälinen ORCID-tunniste olisi avain tutkijan tunnistamiseen ja tietojen yhdistämiseen. Aikataulun suhteen kerrottiin, että tänä vuonna pilotoidaan ja suunnitellaan tutkijoiden tietojen yhdistämistä, mutta ensi vuonna aloitetaan varsinainen toteutus.

Hanketietovaranto

Hanketietovaranto kokoaa rahoittajien rahoituspäätöstiedot, joita aletaan kerätä kevään 2019 aikana Esittelypisteellä kysyttiin, miten hanketiedot näkyvät tutkimuslaitoksissa ja miten tietoja saa hyödynnettäväksi? Sen jälkeen, kun tietoja saadaan kerättyä tutkimusrahoittajilta, tietoja on mahdollista saada ensin esim. CSV -tiedostona, mutta suunnitteilla on myös API –rajapinta ja portaali, joiden kautta tietoja voi hakea. Hanketietovarannon osalta kysyttiin myös, kuinka kattavasti tietoa saadaan ja mikä hyöty sen suhteen on. Ensi vaiheessa mukana ovat Akatemia, Business Finland sekä muutamia yksityisiä säätiöitä. Tutkimuslaitoksia kiinnostavat myös kiinnostuneita EU-rahoitus ja ministeriön rahoittamat hankkeet.

Tietosuojasiat

Tiivistelmä tietosuojasi-asioista on agendan yhteydessä (<https://wiki.eduuni.fi/pages/viewpage.action?pageId=83535778>). Esittelypisteellä keskusteltiin käsittelyperusteista. Tarvitaanko näitä, koska nämä tiedot ovat saatavissa muualtakin? Rahoitusten osalta tarvitaan suostumus/sopimus, kuten myös tutkijoiden tietojen osalta. Käsittelyperusteet arvioitava käyttötarkoitusten mukaan: julkaiseminen portaalissa, edelleen luovuttaminen, ministeriön hyödyntäminen ohjauksessa ja tilastoinnissa. Yhteisrekisterinpitäjinä toimivat tietoja luovuttavat tahot. Käsittelyperusteet eivät päde linkitetyille tiedoille, koska ne muodostavat oman tietokokonaisuuden, joten yhdistelyn osalta harkitaan erillislakia.

OKM toimii rekisterinpitäjänä, joten tietojen toimittajat tekevät sopimukset OKM:n kanssa. CSC tietojen käsittelijänä.

Avoimia kysymyksiä:

Olisiko tieto muiden tutkimuslaitosten tekemisestä hyödyllinen? Voisi olla hyödyllistä erityisesti niille, jotka tekevät päätöksiä. Olisi hyvä tietää muiden tekemisistä, jotta saataisiin asioita tehokkaammin edistettyä. Tästä voisi olla hyötyä ihan konkreettisestikin, mikäli eri organisaatioilla on samanlaisia järjestelmiä. Sanna Marttinen ottaa asian esille tutkimusjohtajien tapaamisessa.

Kuinka paljon halutaan tietoa takautuvasti tutkimusaineistojen näkökulmasta? Ei ole tehty rajausta. Käyttötarkoitus määrittää ja tullut esiin tarvetta myös historiatiedoille. Hanketietojen osalta historiatietojen keräys ongelma, kun se vaatii suostumusta.

Kysyttiin kustannuksista: henkilöresurssit ja järjestelmien kehittämisen kustannukset? Olisi hyvä selvittää tarkemmin, jos puhutaan johdolle. Käytännössä pitäisi tulla "samalla vaivalla". Moni asia kerätään joka

tapauksessa. Koettiin, että VIRTta on ollut luonteva jatko, vaikka on teettänyt jonkin verran ylimääräistä työtä. Yhdeksi hankaluudeksi koettiin se, että kokonaisuus ei kuulu organisaatioissa yhdelle henkilölle vaan vastuu jakautuu. Tämän asian voisi nostaa esiin tutkimusjohtajille.

Tällä hetkellä eniten keskustelua on käyty aineistojen osalta, koska niiden tietojen on tulevaisuudessa oltava esillä.

Kysyttiin CSC:n toteuttamien Fairdata-palvelujen osalta sitä, onko vastaavia palveluja ja onko näitä synkattu toisiinsa. On esim. EUDAT-palvelu, ISAACUS ja kansainvälisiä palveluja. Kysymys on avoimena, miten tietovirtoja saataisiin kehitettyä niin, että tietoja voitaisiin yhdistää. Nyt Fairdata on tutkimustietovarannon ensisijaisena lähteenä, mutta lähteitä voi tulevaisuudessa olla muitakin.

Tutkimustietovarannon aikataulussa aineistojen metatietojen lisääminen on juuri sen vuoksi myöhemmin, koska kenttä ei niin selkeä kuin muut osa-alueet.

Tutkimusdatan osalta – projektin data vai seurannan yms data? Rajanveto ei ole selkeä. Osa tutkimusaineistoista on myös tutkimusinfrastruktuureja, esim. terveystiedot koko väestöstä.

SYKE:n tutkimusaineistostrategiassa puhuttu tutkimus ja asiantuntijatyöstä. Esim. FinELiBin aineistot kattavat sekä tutkimus- että asiantuntijatyön.

Jatko ja tulevia tapahtumia:

Tutkimustietovarannon pilotointiin otetaan mielellään osallistujia – esimerkiksi hankkeiden hyödyntämisen, infrojen ja aineistojen siirtämisen osalta. Sovittiin, että toteutetaan myöhemmin kysely halukkuudesta osallistua tutkimustietovarannon kehittämiseen ja hyödyntämiseen.

TUHA-seminaarissa 9.4. teemana ovat tutkimusaineistot. Agenda julkaistaan lähiaikoina ja laitetaan tietoa myös osallistujille.

EuroCRIS –tapaaminen 27.-29.5. esittelee tutkimustiedon eri osa-alueita. <https://www.eurocris.org/>
Työpajassa ilmoitetusta poiketen tapahtuma on ilmainen euroCRISin jäsenorganisaatioiden jäsenille (muille 60e)

Kerrottiin tiedoksi, että Sanna Marttisen kautta voi tutkimuslaitoksista välittää viestiä tutkimustietovarantohankkeen johto- ja ohjausryhmälle.