

Paikka Väinämöinen, Meritullinkatu 10

Aika: To 23.8.2018 klo 12-15

Osallistujat:

- Ylitarkastaja Jukka Haapamäki (ryhmän puheenjohtaja), opetus- ja kulttuuriministeriö
- Opetusneuvos Matti Kajaste, opetus- ja kulttuuriministeriö
- Tutkimuksen kehittämispäällikkö, Timo Taskinen Jyväskylän yliopisto
- Tutkimuspäällikkö Seliina Päällysaho, Seinäjoen ammattikorkeakoulu
- Kehittämispäällikkö Minna Liikala, Terveyden ja hyvinvoinnin laitos
- Ylilääkäri Ari Lindqvist, Helsingin ja Uudenmaan sairaanhoitopiiri
- Tutkimuksen tietoasiantuntija Yrjö Lappalainen, Tampereen teknillinen yliopisto
- Palvelupäällikkö Anne Sunikka, Aalto-yliopisto
- Ennakointipäällikkö Soile Ollila, Business Finland
- Johtaja Marja Kylämä, Suomen Akatemia
- Tiedejohtaja Kalle Korhonen, Koneen Säätiö
- Tietojärjestelmäasiantuntija Jyrki Ilva, Kansalliskirjasto
- Suunnittelija Janne Pölönen, Tieteellisten seurain valtuuskunta
- Kehittämispäällikkö Ari Leppälähti, Tilastokeskus
- Kehityspäällikkö Hanna-Mari Puuska (ryhmän sihteeri), CSC - Tieteen tietotekniikan keskus

Vierailija:

- Projektipäällikkö Tiina Strengell, CSC - Tieteen tietotekniikan keskus
- Projektipäällikkö Anssi Kainulainen, CSC - Tieteen tietotekniikan keskus

Asialista

1. Kokouksen avaus ja asialistan hyväksyminen

- Kokous avattiin kello 12.05 ja asialista hyväksyttiin.

2. Edellisen kokouksen pöytäkirjan ja muistion hyväksyminen [Liite 1](#)

- Edellisen kokouksen muistio hyväksyttiin muutoksitta.

3. Tilannekatsaus [Liite 2](#)

a. Lähdetietovarantojen tilannekatsaus

i. VIRTA –julkaisutietopalvelu

Hanna-Mari Puuska esitteli VIRTA-julkaisupalvelun tilanteen liitteen 2 mukaisesti. VIRTAan on parhaillaan toteutuksessa integraatio

OpenAIREen, jota varten tehdään CERIF-tietomallin mukainen rajapinta.

Kysyttiin, mitä tietoja OpenAIREen on tarkoitus viedä vai päättääkö jokainen organisaatio viettävät tiedot erikseen. Lisäksi kysyttiin sitä, mitä OpenAIRE-versiota käytetään.

Kerrottiin, että prosessia mietitään tarkemmin VIRTAYhteys henkilöiden kanssa. 29.8. pidettävässä Kota –seminaarissa on myös hyvä keskustella näistä asioista.

Kerrottiin myös, että korkeakoulujen toimittamat JUFO-tasot säilytetään jatkossa VIRTAY-julkaisutietopalvelussa. Niitä verrataan VIRRAssa tunnistettaviin JUFO-tasoihin ja virheet näkyvät virheraporteilla ja extra-Vipusessa.

ii. Infrastruktuuriryhmän jatko

Tiina Strengell esitteli infrastruktuuriryhmän tilanteen liitteen 2 mukaisesti.

iii. Hanketietovarannon tilanne

Tiina Strengell esitteli hanketietovarannon kehityksen tilanteen liitteen 2 mukaisesti.

Keskusteltiin hanketietovarannon tietosuojaan liittyen. Päätöksiä tietosuojasta ei ole tulossa vielä tänä vuonna. Ennen tuotantoon menoa täytyy joka tapauksessa huolehtia, että mm. suostumukset on pyydetty.

iv. Käynnistyvä selvitys tutkijan omien tietojen luvittamisesta

Hanna-Mari Puuska kertoi tutkijan omien tietojen luvittamiseen liittyvästä selvityksestä. Selvityksen tekeminen aloitetaan syyskuussa Harri Hongon (W2C) toimesta. Selvityksen on tarkoitus valmistua joulukuussa.

b. Hyödyntämisen suunnittelu

Tiina Strengell esitteli tutkimustietovarannon hyödyntämiseen liittyviä suunnitelmia.

Pilotointiin liittyen kysyttiin tutkimusaineistojen hyödyntämisestä. Tutkimusaineistot haravoidaan ensin Etsimeen, josta ne ladataan myöhemmin tutkimustietovarantoon.

Tutkimustietovarannon kehitys ja roadmap

Tiina Strengell esitteli tutkimustietovarannon hyödyntämiseen liittyviä suunnitelmia.

Tämän yhteydessä kysyttiin aktiviteettitietojen toimittamisesta. Siitä on keskusteltu VIRTAn julkaisutietopalvelun ja julkaisutyöryhmän yhteydessä.

Kysyttiin myös, mitä roadmapilla näkyvät VIRRAn lataukset tarkoittavat ja vaikuttavatko ne siihen aikatauluun, joilla tiedot on vietävä VIRTaan, jotta ne saa tutkimustietovarantoon asti. Roadmapilla näkyvät lataukset tarkoittavat julkaisutietojen metatietojen vientiä tutkimustietovarantoon, jonka jälkeen niitä voidaan yhdistää muihin tutkimukseen liittyviin tietoihin, kuten hanketietoihin. Latauksia tehdään toistuvasti, joten myös VIRTaan myöhemmin lisättäviä tietoja tullaan lataamaan tutkimustietovarantoon.

4. FAIRdata-palvelut [Liite 3](#)

Anssi Kainulainen esitteli Fairdata-palveluja liitteen 3 mukaisesti. Palvelut on toteutettu CSC:llä ja Kansalliskirjastolla, ja ne koostuvat seuraavista osista: IDA, FairData-PAS, Etsin, Qvain ja METAX.

Palveluista IDA, jonne aineistoja tallennetaan ja säilytetään, on tuotannossa. Osa migraatioista on vielä kesken, mutta uudet projektit voivat ottaa uuden IDA:n käyttöön. Uudessa IDAssa palvelut pääosin samana ja se sisältää ainoastaan teknisiä muutoksia, jotka eivät vaikuta käyttöön.

Kysyttiin, kuka siirtää metatiedot palveluihin. Tutkijat eivät itse siirrä tietojaan, vaan ne tulevat keskitetysti organisaatioittain. Myös sitä kysyttiin, saavatko muut kuin yliopistojen tutkijat siirrettyä aineistonsa palveluun. Tutkijan täytyy kuulua johonkin organisaatioon, jonka kautta tiedot siirretään. Lisäksi ajatuksena on se, että aineistosta huolehditaan myös sen jälkeen, kun tutkija lopettaa työskentelyn tietyssä organisaatiossa.

Käytiin keskustelua myös ammattikorkeakoulujen yhteistyökumppanina toimivien yritysten aineistojen siirtämisestä. Käytännössä tähän liittyvät asiat on sovittava organisaatiokohtaisesti. Myös yritys voi saada tunnukset, mutta mm. huolehtiminen tunnusten jaosta ja tietojen siirtämisestä on aina vastuorganisaation vastuulla.

Kysyttiin myös, kuka voi saada IDA:an tunnukset. Tällä hetkellä rajauksena on se, että tunnuksen saajan täytyy tehdä töitä suomalaisessa projektissa.

Viittaukset tutkimusaineistojen ja muiden tutkimukseen liittyvien tietojen välillä ovat tärkeitä, jotta eri tyyppiset aineistot löytyvät. Tulevaisuudessa tutkimustietovarannon kautta saadaan lisää linkityksiä mm. julkaisuihin.

Erityisesti Qvaimen osalta kysyttiin aikataulusta. Qvaimen aikataulu on Kansalliskirjaston käsissä, koska he kehittävät palvelua. Tarkoitus on kuitenkin julkaista Qvain tämän vuoden aikana. Julkaisun yhteydessä vanhan ja uuden Etsimen välillä on tehtävä migraatio.

Etsin on hakukone, jolla aineistot löytyvät. Sieltä löytyvät tiedot haravoidaan tai syötetään aineistojen metatietojen Qvaimen kanssa. Se kuka omistaa aineiston, määrittelee siihen liittyvät ehdot.

Fairdata-PASin ideana se, että aineiston täytyy säilyä ymmärrettävänä ja käytettävänä. PASiin siirretään aineisto, joka on merkittävä organisaatiolle tai kansallisesti. Tällä hetkellä tallennettuja aineistoja on kolme.

Tähän liittyen kysyttiin aineistojen valinnasta käytännössä: Aallossa esim. tähtitieteen dataa, joka heidän mielestään tärkeää. Kuka päättää valinnasta käytännöstä? CSC ei päättää valinnoista, vaan se on organisaation vastuulla. Kriteerit on mietitty valmiiksi ja oletuksena kerätään jonkinlainen ryhmä aineistojen valintoja varten. Käytännössä kapasiteetti 10 teratavua per organisaatio. Mikäli tarpeita osoittautuisi olevan enemmän, mietitään asiaa uudestaan. Huomiona esitettiin myös, että Fairdata-PAS säilöo ainoastaan tutkimusaineistoa koskevat datat, ei esim. kulttuuriperintöön liittyvää aineistoa.

METAX on ns. liima Fairdata-palveluiden välillä. Se huolehtii mm. linkityksistä, sanastoista, tunnisteista ja versioinnista. METAXin API on julkisesti katsottavissa.

METAXin volyymistä kysyttäessä kerrottiin, että se sisältää pari-kolmesataa projektia, joiden koko vaihtelee todella paljon. Kaikki pyritään sitomaan aineistokuvauksiin.

HUS:lla on keskusteltu tutkimusaineistojen tallentamisesta erityisesti biolääketieteellisen tiedon osalta. Heidän keräämä aineisto on käytännössä big dataa ja se tuo eteen uusia kysymyksiä. Tämän osalta kysyttiin vaikutuksia aineistojen tallentamiseen Fairdata-palveluihin. Fairdata-palveluihin tallennettavat aineistot ovat rakenteisessa muodossa. Mikäli aineisto ei ole rakenteista dataa, käyttötarve on jonkin verran erilainen. Esimerkiksi iltatieteenlaitoksella on erillinen tietokanta, joka ottaa ja analysoi aineiston. Tämän jälkeen Fairdata-palveluihin voidaan siirtää vain olennaiset tiedot. Näidenkin osalta on otettava huomioon se, ettei palveluihin voi viedä arkaluontoista aineistoa.

Tähän liittyen käytiin keskustelua myös siitä, miten tällaisissa tapauksissa suhtaudutaan tieteen avoimuuteen. Biolääketieteessä vaatimuksena julkaisukanavan toimesta on, että julkaisun takana oleva aineisto pitäisi tuottaa julkisena. Ongelmana on kuitenkin mm. henkilötietojen käsittely. Tämän osalta on todettu, että koska kaikki tieto ei voi olla avointa, aineistot julkaistaan mahdollisemman avoimena. Lisäksi mietitään erikseen muun muassa luvitukset, lisenssit, käyttöluvut.

Keskusteltiin myös USA:n ja EU:n tietosuojasetuksen problematiikasta. Tietojen luovutus näiden välillä on vaikeaa eikä ole tällä hetkellä lainkaan tutkijoiden hallinnassa. Eurooppalainen lainsäädäntö on hyvin erilainen kuin USAssa. Myös henkilötietojen käsittely on iso ongelma joka puolella. Tästä nousi esimerkkinä myös mm. kielipankki. Biolääketieteessä tästä voisi tulla suuri ongelma. Fairdata-palveluihin kerättävä metatieto sisältää kuitenkin vain tekijätiedot, mutta ei muita henkilötietoja. Henkilötiedot datan sisällä ovat kuitenkin tutkijoiden huolehdittavia asioita. Fairdata-palvelujen toimesta tarjotaan ohjeistusta, mutta valvomiseen ei käytännössä ole resursseja.

Fairdata-palveluihin liittyen kysyttiin myös sitä, kuuluuko palvelut European Open Science Cloudiin. Ajatuksellisesti ne voidaan sanoa kuuluvaksi, mutta käytännön tasolla asiaa ei ole vielä edistetty.

Tiedon julkisuuden tasoihin liittyen kerrottiin, että lausuntopalveluissa (VM) puhutaan paljon tiedon julkisuuden tasoista. Tästä löytyy mm. dokumentti osoitteesta: https://vm.fi/artikkeli/-/asset_publisher/julkisen-hallinnon-linjaukset-tiedon-sijainnista-ja-hallinnasta-lausunnoille, josta löytyy tietoa aiheesta.

Kysyttiin myös Fairdata-palveluiden toteutusalueista, sen kaupallisuudesta ja kaupallisuuden merkityksestä. Palveluissa on käytetty pääasiassa avoimen lähdekoodin tuotteita. Toisinaan kaupallisten tuotteiden tukeminen on perusteltua, mutta ajatuksena on se, että palvelu pyritään tekemään niin, että sen ylläpidosta pystytään huolehtimaan jatkossakin itse.

Anssi Kainulainen esitteli lopuksi demon nykyisistä palveluista.

5. Tutkimustietovarannon portaalin palvelumuotoilu [Liite 4](#) + [tiivistelmä](#)

Tiina Strengell esitteli portaalin palvelumuotoilun tilanteen liitteen 4 mukaisesti.

Kerrottiin, että yliopistoilla organisaatiokohtaiset portaalit ovat huolenaiheena johtuen siitä, että järjestelmien toimittajat tarjoavat jatkossa ainoastaan täysin valmiita ratkaisuja, joita ei voi. Esimerkiksi Aalto-yliopistolla Elsevierin tuki nykyiselle järjestelmälle on loppumassa kolmen vuoden päästä, joten siellä harkitaan oman portaalin tekemistä. Siellä on toivottu portaaliin myös uusia moderneja elementtejä, esim. videoita ja mahdollisimman vähän manuaalista päivytystä, mutta tästä ei ole vielä konkreettisia suunnitelmia.

Tutkimustietovarannon ja yliopistojen portaalien päivityksiin liittyen keskusteltiin siitä, että näiden tekemistä olisi hyvä jollakin tasolla koordinoita. Aiheeseen liittyen kommentoitiin myös, että on hyvä erottaa kansallinen näkymä ja se, mikä organisaation tarve. Myös tutkimusryhmiltä on tullut toiveita omista sivuista. Käytännössä täytyisi peruuttaa tutkimustietovarannon portaalista kehityksessä taaksepäin, jos nämä otettaisiin kokonaan huomioon, joten se ei ehkä ole järkevää, mutta koordinoitua olisi kuitenkin hyvä tehdä.

Portaaliin liittyen nostettiin esiin myös saavutettavuusdirektiivi, joka on otettava huomioon varhaisessa vaiheessa, koska se voi olla isokin kustannuserä. Tästä löytyy tietoa mm. osoitteesta saavutettavuus.fi. Jukka Haapamäki kertoi, että Vipunen –portaalista on tarkoitus tehdä saavutettavuuden arviointi loppuvuodesta.

Sovittiin, että konseptia käydään läpi yliopistojen viestinnän henkilöiden kanssa.

6. Henkilötietojen käsittelyperusteet tutkimustietovarannossa [Liite 5](#)

Jukka Haapamäki esitteli henkilötietojen käsittelyperusteisiin liittyviä asioita liitteen 5 mukaisesti. Aihetta mietitty jo siitä lähtien, kun tutkimustietovarantoa alettiin rakentaa ja alussa valmisteluissa oli mukana OKM:ssä aikaisemmin juristina toiminut Immo Aakkula. Kesän jälkeen asiaa on valmistellut Laura Karppinen, joka oli esittelemässä tietosuojaan liittyviä asioita tutkimustietovarannon johtoryhmässä. Tavoitteena on saada selville periaatteet, joiden pohjalta henkilötietoja käsitellään. Aihe on ollut hankala edistää, koska kansallinen lainsäädäntö, jolla EU:n tietosuoja-asetuksen toimeenpanoa tarkennetaan, on vielä käsittelyssä.

Laura Karppisen esityksen mukaan OKM on rekisterinpitäjä, koska se on määritellyt koko tietovarantoprojektin henkilötietojen käsittelyn tarkoituksen ja keinot. CSC on henkilötietojen käsittelijä, joka toimii OKM:n lukuun ja toimeksiannosta. Tämän hetkisen käsityksen mukaan lisäksi tietovarannon tiedot alun perin keräävät ja ne tietovarantoon toimittavat tahot olisivat rekisterinpitäjiä, koska ne ovat edustettuina suoraan tai välillisesti tietovarannon ohjaus- ja johtoryhmässä ja näin osallistuvat käsittelyn tarkoituksien ja keinojen määrittelyyn. Yhteisrekisterinpitäjinä ne voivat myös käyttää tietoja itse, koska tutkimustietovarannon tarkoitus on yhtenevä alkuperäisten rekisterinpitäjien käyttötarkoitusten kanssa.

Varmimmalla pohjalla oltaisiin, jos TTV:stä säädettäisiin erikseen laki. Tutkimuksen metatietoihin liittyvät asiat yritettiin saada OKM:n toimesta mukaan kansalliseen tietosuojalainsäädäntöön, mutta oikeusministeriö linjasi, että tämän tyyppisistä yksityiskohdista pitäisi säätää muualla.

Käytännössä joko nykyinen tuleva lainsäädäntökehikko sallii tutkimustietovarannon sellaisenaan tai vaihtoehtoisesti mikäli halutaan pelata varman päälle, niin voidaan säätää oma laki, kuten ns. Koski-laissa opintotietojen osalta on tehty.

Kysyttiin, onko tilastointi edelleen käsittelijän omaa toimintaa. Tähän liittyen tulkintaa voi tehdä vielä hyvin vapaasti.

Kysyttiin myös, onko ollut tiedossa miksi OM ei halunnut lain muutosta. Ilmeisesti kysymys oli siitä, että OM ei halunnut kuvata käyttötarkoituksia liian tarkkaan.

Fairdatassa on tehty tulkinta, jonka mukaan käytetään henkilötietojen käsittelyperusteena merkittävää yleistä etua. Tilastointiin liittyen on myös oma tilastolaki, jota on hyödynnetty. Tärkeäksi asiaksi on noussut osoittamisvelvollisuus sen suhteen mitä tehdään ja miksi, joten nämä täytyy kuvata selkeästi myös tutkimustietovarannon osalta. On myös paljon erilaisia tapauksia, joita täytyy miettiä tarkemmin, kuten tutkijat työsuhteessa ja apurahoina, organisaatioiden tietoina.

Sovittiin, että odotetaan, että uusi laki tulee voimaan ja kysytään sen jälkeen OM:n näkemystä asiaan.

Yhteisrekisterin pitäjyyteen liittyen kysyttiin, voivatko jatkossa esimerkiksi Business Finland ja Suomen Akatemia hyödyntää toistensa tietoja ilman lupien kysymistä. Jukka Haapamäen vastasi, että ainakin maallikkotulkinnan mukaan näitä voisi hyödyntää ja se on tutkimustietovarannon ajatuskin.

On myös muistettava, että kaikki, jotka käsittelevät toistensa tietoja, tarvitsevat voimassaolevat sopimukset. OKM antaa myös ohjeistuksen tietojen käytöstä.

7. Muita asioita

NordForskilta ollaan hakemassa rahoitusta VIRRAn pohjoismaiseen yhteistyöhankkeeseen liittyen. Hankkeessa on tavoitteena toteuttaa pohjoismainen julkaisutietokanta, jota voisi laajentaa myöhemmin muihin Euroopan maihin.

Tutkimustietovarannon tiimillä on tarkoitus tutustua Norjan Cristin-järjestelmään lokakuun puolivälissä.

KOTA –seminaarissa 29.8. esitellään myös tutkimustietovarantoa.

8. Seuraava kokous to 11.10.2018

9. Kokouksen päättäminen

Kokous päätettiin klo 14.29