

2016-01-27

Julkaisuarkistopalveluiden asiantuntijaryhmä, 27.1.2016

- [Julkaisuarkistopalveluiden asiantuntijaryhmä, 27.1.2016](#)
 - 1. Kokouksen avaus
 - 2. TAJUA-projekti ja muut Kansalliskirjastossa toteutettavat ATT-projektit
 - [Avoin tiede ja tutkimus -hanke](#)
 - [Tajua-hanke](#)
 - [Avoimen tieteen koulutukset ATT:n puitteissa](#)
 - [Kotimaiset tieteelliset lehdet avoimiksi ja vaikuttamaan](#)
 - [Keskustelua ATT- ja TAJUA-hankkeesta](#)
 - 3. Kansallinen pitkäaikaissäilytysjärjestelmä
 - [Keskustelua PAS:iin liittyen](#)
 - 4. Pimeä arkistointi ja opinnäytteiden säilytys
 - 5. Muut asiat
 - 6. Seuraava kokous

Aika: 27.1.2016 klo 10.00-12.00

Paikka: Fabianian kokoushuone, Yliopistonkatu 1, 2. krs

Puheenjohtaja: Mikko Pennanen (Turun yliopisto)

Sihteeri: Samu Viita (Kansalliskirjasto)

Läsnä: Mikko Pennanen (Turun yliopisto), Lassi Lager (Kansalliskirjasto), Jyrki Ilva (Kansalliskirjasto), Esa-Pekka Keskitalo (Kansalliskirjasto), Minna Liikala (THL), Samu Viita (Kansalliskirjasto)

Etänä: Mikko Saunamäki (Luonnonvarakeskus), Minna Marjamaa (AMKIT konsortio), Teemu Makkonen (Jyväskylän ammattikorkeakoulu), Toni Raja-Hanhela (Lapin yliopisto), Tiina Tolonen (AMKIT konsortio), Iiris Tahvanainen (Lappeenrannan teknillinen yliopisto)

Poissa: Päivi Rosenström (Kansalliskirjasto), Joonas Parviainen (Maanpuolustuskorkeakoulu)

1. Kokouksen avaus

2. TAJUA-projekti ja muut Kansalliskirjastossa toteutettavat ATT-projektit

Avoin tiede ja tutkimus -hanke

- Esa-Pekka Keskitalo esitteli Avoin tiede ja tutkimus -hanketta (ATT) ja erityisesti sen rahoittamaa TAJUA-projektia. Projektilla on myös julkaisuarkistoihin ja niiden kansalliseen yhteistyöhön liittyviä tavoitteita, ks. [hankkeen esittely Kansalliskirjaston uutiskirjeessä 2/2015](#).
- Tällä hetkellä tieteellinen tieto on huonosti saatavilla. ATT-hankkeen tarkoituksena on avoimen tieteen ja avoimen yhteiskunnan edistäminen. Hankkeessa pyritään valitsemaan ja kehittämään kansallisella tasolla yhteiset tieteen prosessit ja välineet, sekä saattamaan tutkimusdata, julkaisut ja tieteellinen viestintä kaikkien tiedonhaluisten saataville.
- Avoin tiede pitää sisällään avoimen tieteellisen yhteistyön, avoimen saatavuuden, avoimen datan, avoimen lähdekoodin ja avoimen laboratoriotekniikan. Jälkimmäisin tarkoittaa tutkimusprosessin avoimuutta, jossa muut tutkijat voivat kommentoida ja mahdollisesti osallistua tutkimusprosessiin jo sen varhaisessa vaiheessa.
- Erityisesti avoimen datan osalta Kansalliskirjastolla on keskeinen rooli sen tarjoamien tunnistepalvelujen vuoksi, tunnistet ovat elinehto datan tuottamisessa ja uudelleenkäytössä.
- Meneillään oleva ATT-hankekausi on vuosille 2014 - 2017. Tavoitteena on, että hankkeen tuloksena Suomi on avoimuuden johtavia maita.
- Hankkeen tarkoituksena on tuottaa uusia avauksia, mutta myös integroida jo olemassaolevia, avointa tiedettä edistäviä palveluja kuten tutkimusdatan tallennuspalvelu IDA:aa, yhteiskunnallisen tietoarhivon portaalia Ailaa, Kansalliskirjaston julkaisuarkistoja, Fintoa, datanhallinnan suunnittelutyökalua Tuulia ja tutkimus-PAS:ia.

Tajua-hanke

- Tajua-hankkeen tehtävänä ATT-hankkeessa on tieteen tuotosten, kuten tutkimusdatan ja tieteellisten julkaisujen pysyvä avoin saatavuus. Tuotokset tulee siis pitää avoimena ja nopeasti saatavilla nyt ja tulevaisuudessa luotettavalla menetelmällä.
 - Hankekausi on ajalle 9/2015 - 3/2017

- Hankkeella on ollut alussa käynnistysvaikeuksia: Vaikka hanke käynnistyi syyskuussa, Helsingin yliopiston rekrytointikiellon takia hankkeessa on ollut aluksi vain yksi työntekijä, nyt saadaan kaksi uutta henkilöä, joiden rekrytointi on käynnissä.
- Paljon yhtymäkohtia muihin ATT-hankkeisiin, Tieto käyttöön -hankkeisiin ja muualle.
 - ATT-työryhmät toimivat hankkeen katselmoijina.
- Tajua-hankkeessa kehitetään ja yhtenäistetään rinnakkaistallennuksen infraa kehittämällä Kansalliskirjaston nykyisiä julkaisuarkistoja. Luodaan malli mahdollisimman helposti käyttöönotettavasta ja edullisesta Dorian tapaisesta, mutta "karvalakki" -palvelusta, johon pienten ja vähän julkaisuja tuottavien organisaatioiden on helppo ja edullista liittyä. Tähän palveluun ei ole saatavilla asiakaskohtaista räätälöintiä, vaan syöttölomakkeet, ulkoasu ja käytänteet ovat yhtenäiset kaikille. Jos asiakaskohtaista räätälöintiä tarvitaan, silloin käytetään Doriaa tai muita vastaavia julkaisuarkistopalveluja.
 - Tutkitaan Dspace-Cris laajennoksen kelvollisuutta julkaisurekisteritarpeisiin.
 - Tutkitaan myös mahdollisuutta tarjota "orvoille" tutkijoille julkaisuarkistopalveluita, joilla ei ole muutoin suoraa pääsyä sellaiseen oman organisaationsa kautta.
 - Pyritään kehittämään yhteisiä avoimen julkaisemisen käytäntöjä, jotka ovat nyt jääneet puolittiehen. Projekti kehittää yhteistyöverkoston, jossa tehdään tarpeelliset kansalliset määrittymiset ja suositukset julkaisuarkistoille:
 - Eri aineistotyyppien kuvailuohjeet
 - Ohjeet julkaisulisenssien valintaan, käyttöön ja merkitsemiseen
 - OpenAire yhteensopivuusohjeet
 - Ohjeet siitä, miten julkaisut, tutkimusdata ja metodit viittaavat toisiinsa
 - Ohjeet tilastointiin ja raportointiin
- Kehitetään avoimen julkaisemisen tuen malleja:
 - Kirjoittajamaksut: Kuka maksaa ja kuinka paljon?
 - Tehdään selvityksiä ja arvioita mahdollisuuksien rajoissa
 - Esitetään vaihtoehtoja rahoitusmalleiksi
 - Esitetään suosituksia sopimusehdoiksi
 - Lisätään tietoisuutta avoimesta julkaisemisesta
 - Tiedonkeruu
 - Avoimen julkaisemisen tulevaisuuskuvien tarkastelu
- Tutkimusdatan kuvailutyökalu:
 - Tutkijalle suunnattu helppo työkalu, jonka suunnittelu kytketään osaksi tutkimus-PAS:in suunnittelutyötä, erityisesti sen paketoitupalvelun osalta
 - Työ koordinoitava Tuulin, IDA:n yms. kanssa (IDA:n kanssa tulossa suuria uudistuksia, koska järjestemälusta kilpailutetaan)
- Tunniste ja nimitietopalvelu
 - Asterin viitetietojen julkaiseminen
 - ISNI-tunnusten käyttöönotto ja niiden hallinnan prosessit
 - URN-palveluiden kehittäminen, mm. tuki sille, että URN voi viitata useampaan vaihtoehtoiseen URL:ään, mahdollisesti myös Tutkimus-PAS:issa olevaan, jolloin pysyvyyden takaaminen olisi eri luokkaa kuin nykyisellä infralla
- Tieteellisten julkaisujen PAS-valmiuden parantaminen
 - Digitaalisten aineistojen saatavuus saattaa lakata yllättäen, miten saatavuutta voi varmistaa? Avoimissa aineistoissa puuttu rahan tuoma vipuvoima.
 - Pyritään varmistamaan, että kotimaassa kiinnitetään pitkäaikaissaatavuuteen huomiota.

Avoimen tieteen koulutukset ATT:n puitteissa

- Kansalliskirjasto koordinoi ATT-hankkeen puitteissa järjestettäviä avoimen tieteen koulutuksia, ks. [koulutusten esittely ATT-hankkeen sivuilla](#). Koulutusten tarkoituksena on lisätä tieteen ja tutkimuksen avoimuutta ja siihen tarvittavan osaamisen vahvistamista. Tarkoitus on koota tieteen ja tutkimuksen parissa toimivilta asiantuntijoilta koulutuspaketti, jota voi muokata edelleen erityyppisiin koulutustilaisuuksiin.

Kotimaiset tieteelliset lehdet avoimiksi ja vaikuttamaan

- Tieteellisten seurain valtuuskunnan kanssa toteutettava yhteisprojekti *Kotimaiset tieteelliset lehdet avoimiksi ja vaikuttamaan* on alkanut (ks. [projektia esittelevä artikkeli Tietolinja-lehdessä](#) ja [projektin kotisivu](#)). Projektissa kehitetään TSV:n tieteellisille lehdille tarjoamaa teknistä infrastruktuuria ja OA-lehdille soveltuvaa rahoitusmallia. Rahoitusmallia suunnittelevan tietoasiantuntijan rekrytointi on parhaillaan käynnissä (ks. [hakuilmoitus](#)).
 - Kuten Tajua-projektissa, alussa rekrytointiongelmia. Suunnittelija tulee sijoittumaan Kansalliskirjastoon.
 - Open Journal Systems -alustaa (OJS) tarjotaan palveluna, sitä ajanmukaistetaan ja kehitetään eteenpäin.
 - Kotimaiset lehdet ottavat DOI-tunnisteet käyttöön.
 - Selvitetään myös Orcidin käyttöönottoa
 - Rajapintojen osalta selvitetty sitä, kuinka OJS:stä voisi tallentaa dataa julkaisuarkistoon.
 - Rahoitusmalli on myös yksi tärkeä selvitettävä asia, jonka osalta työ on vasta alkamassa.

Keskustelua ATT- ja TAJUA-hankkeesta

- Esa-Pekka Keskitalo totesi, että kirjastojen on laajennettava osaamistaan julkaisujen lisäksi tutkimusdatan puolelle.
- Esa-Pekka mainitsi myös laskelmasta, jonka mukaan nykyisin Suomessa maksettavilla lisenssimaksuilla pystyttäisiin rahoittamaan julkaisujen OA-julkaisemisen kulut.
- Kokousväen keskuudesta tuli useammaltakin taholta kysymyksiä ja toiveita yhtenäisistä suosituksista Dublin Coren käytön osalta. Esimerkiksi Theseuksessa asia on ajankohtainen, kun rinnakkaistallenteita varten suunnitellaan metadatta. Myös Julkarissa on meneillään metadatan paranteluprosessi, jossa pyritään ottamaan huomioon omien tarpeiden lisäksi julkaisutiedonkeruun ja OpenAire-yhteensopivuuden vaatimukset. Hankkeelta on tarkoitus tulla yhtenäisiä suosituksia metadatan osalta.
- Jyrki Ilva otti esiin kysymyksen päällekkäisyydestä: TAJUA-hankkeen puitteissa perustetaan yhteinen työryhmä pohtimaan julkaisukysymyksiä ja siinä on muitakin kuin KK:n asiakkaita. Tulisi pohtia, miten mahdolliset päällekkäisyydet julkaisarkistojen asiantuntijaryhmän kanssa voitaisiin välttää. Minna Liikala ehdotti yhtä yhteistä työryhmää. Ilva mainitsi, että tämä olisi toivottavaa, mutta toisaalta asiantuntijaryhmässä keskitytään enemmän KK:n ylläpitämiin käytännön asioihin, TAJUA-ryhmä keskittyy yleisempiin asioihin. Keskitalo ehdotti, että voitaisiin supistaa asiakaskohtaiset asiat puoleen tuntiin ja sitten kaikille yhteiset asiat.
- Iris Tahvanainen kysyi, vastaisiko OJS -järjestelmä Lappeenrannan teknillisen yliopiston tarpeisiin, sillä tutkijat ovat kysyneet, voisiko esim. konferenssijulkaisut julkaista LutPubissa ulkomaisen lehden sijaan. Ilvan mukaan voisi vastata, mutta ei olisi kuitenkaan välttämättä konferenssijulkaisujen kannalta optimaalinen palvelu, sillä se on optimoitu ensisijaisesti tieteellisille lehdille. TSV:n palvelu on lisäksi suunnattu lähinnä TSV:n jäsenseuroille, joille se on ilmainen.

3. Kansallinen pitkäaikaissäilytysjärjestelmä

- Esa-Pekka Keskitalo kertoi kansallista pitkäaikaissäilytysjärjestelmää (PAS) rakentavasta hankkeesta.
 - ATT-hankkeessa puhutaan pitkäaikaissäilytyksen sijaan pitkäaikaissaatavuudesta. Nämä termit tarkoittavat kuitenkin samaa.
 - Kansallisen PAS:in tarkoituksena on se, että tutkimustulokset ja aineistot ovat aidosti käytössä. Käytännössä niiden sotkeminen on kuitenkin helppoa, vika usein lopulta ihmisissä, kuten tiedon hallinnoinnissa ja organisoinnissa, ei niinkään tekniikassa, vaikka sitä usein syytetäänkin.
 - Haasteita pitkäaikaissaatavuudelle: Teknologian muutokset, hajanaisen heterogeenisen datan säilytys, käyttöoikeusasiat ja suojaukset, katastrofit kuten tulipalo yms.
 - Dataa voi olla tallessa, mutta sen merkitys voi olla vaikeaa selvittää, jos esim. konteksti on hukassa. (esim. vanha taulukko, jossa sarakkeiden nimet puuttuvat).
 - Tällä hetkellä esimerkiksi pelkästään eläköityminen on valtava ongelma: Pilottien myötä tehdyssä kartoituksessa on selvinnyt, että usein entisiin tai jo eläkkeellä oleviin työntekijöihin otetaan yhteyttä ja kysellään erinäisten tärkeiden tietojen perään.
 - PAS:in ajatuksena tiedon säilytys ja saatavuus yli teknologian, ihmisten ja organisaatioiden muutosten.
 - Säilytys tapahtuu bittitasolla ja aineistoa huolletaan varmuuskopioimalla, eheystarkastuksilla ja teknologiaa sekä formaatteja uusimalla.
 - Teknisen säilyttämisen lisäksi huomiota kiinnitetään myös käytettävyyden säilyttämiseen, kuten käyttöoikeuksien ja sisällön ymmärrettävyyden säilyttämiseen, johon kuuluu mm. kontekstuaalisen tiedon säilytys.
 - KDK-PAS -palvelua on rakennettu pitkään, Kansalliskirjasto otti sen vuodenvaihteessa käyttöön verkkoarkiston osalta.
 - Sekä KDK-PAS, että tutkimus-PAS käyttävät samaa pitkäaikaisen saatavuuden järjestelmää.
 - KDK-puolella on isoja toimijoita, kuten kirjastot, arkistot ja museot, Tutkimus-PAS:issa sen sijaan ei: Tutkimusryhmiä yms...
 - Rootitus: Datavastuu organisaatiolla, käytänteitä täytyy viedä siihen suuntaan, että vastuun siirto tutkijalta yms. organisaatiolle siirtyy jouhevasti ja on itsestäänselvä osa prosessia.
 - Pilotteja Tutkimus-PAS:in osalta tehty viime vuonna tutkimusryhmien kanssa. Kaikki onnistuivat, mutta tutkijoille koitui todella paljon lisätyötä esim. sen osalta, miten data esitetään standardimuodossa.
 - Pyritään viemään hyvää datanhallintakäytäntöä käynnissä olevaan tutkimukseen keskipitkän säilytyksen osalta (ei vielä PAS-tasolla, mutta bittitason tallennuksilla). Tämä vaatii hieman kuratointia pitkäaikaissäilytyskelpoiseksi muunnettaessa. Ideana tarjota eritasoisia tallennusvaihtoehtoja.
 - Vaatimusmäärittelyvaihe menossa vielä, tällä hetkellä tehdään teknistä ja käyttötapaussuunnittelua.
 - Tänä vuonna pilotoidaan enemmän hallinnollisia prosesseja. Erytystapauksena tänä vuonna genomidatan pilotti, jossa suoritetaan pidemmälle viety testi, miten käytännössä asia hoidetaan.

Keskustelua PAS:iin liittyen

- Jyrki Ilva tiedusteli, miten julkaisuarkistot tulevat näkymään PAS:issa. Esa-Pekka Keskitalo vastasi, että lähtökohtaisesti Kansalliskirjasto-vetoisesti, tavoitteena on, että säilytettävä materiaali olisi käytettävissä muuallakin kuin vapaakappaletyöasemilla. (vrt. suunnitelma kytkeä PAS-aineisto osaksi URN-resolvointia, jolloin pysyvä tunniste viitaisi PAS-versioon, vaikka alkuperäinen dokumentti olisikin poistettu tai hävinnyt).
- Pohdittiin sitä, menevätkö julkaisuarkistojen aineistot KDK- vai tutkimus-PAS:iin. Meneekö osa aineistoista aineistotyyppiin ja taustaorganisaation perusteella toiseen ja osa toiseen?

- Kysyttiin myös, onko kyseessä ilmaisupalvelu. Keskitalon mukaan hankerahaa on haussa, puhutaan kuitenkin 2020 jälkeisestä ajasta.
- Tutkimus-PAS:in osalta joudutaan myös miettimään tilausmekanismia. Tapahtuu organisaation kautta, koska asiaan liittyy lupa-asioita. Vapaakappaleiden ja PAS:in liittymäkohdassa vielä avoimia ja vaikeita asioita.

4. Pimeä arkistointi ja opinnäytteiden säilytys

- Ammattikorkeakoulujen kanssa on alustavasti keskusteltu mahdollisuudesta rakentaa Theseuksen yhteyteen erillinen ns. pimeä arkisto, joka palvelisi mm. opinnäytteiden säilytystä, etenkin niiden töiden osalta joita ei julkaista avoimesti verkossa. Palveluun tallennettu aineisto olisi käytettävissä rajatusti esim. kunkin organisaation kirjastossa. Asiaa selvittävän työryhmän työ on vielä alkuvaiheessa; ensimmäisiä selvitettäviä asioita ovat mahdollisen uuden palvelun konsepti ja tällaiseen palveluun kohdistuvan kiinnostuksen laajuus AMK-sektorilla.
- Jyrki Ilva mainitsi, että aiheesta on pidetty yksi suunnittelupalaveri AMK:ien kanssa. Ensimmäinen selvitettävä asia on konsepti ja käyttötapaukset järjestelmälle, toinen se, kuinka laajaa kiinnostusta pimeän arkistoinnin palveluun liittyy.
- Haasteina mm. salaisten tai osittain salaisten opinnäytteiden näkyvyys: onko niitä ylipäättään mielekästä tallentaa järjestelmään?
- Valmistelu vielä hyvin alkuvaiheessa, tarkoitus vielä pitää pari kokousta ennen kevään Theseus-seminaaria, jossa asiasta keskustellaan laajemmin.

5. Muut asiat

- Julkaisuarkistopalveluiden **uusi asiakaswiki** avattu. Wikin on tarkoituksena toimia yhteisenä ohje- ja viestintäpaikkana Kansalliskirjaston julkaisuarkistoihin liittyvissä asioissa.
- LUT:in OpenAire-setti valmis. Pohdinnassa vielä, pitääkö EU-projekti-id olla pakollinen tieto vai ei, OpenAiressa ristiriitaista ohjetta tämän osalta. Tällä hetkellä tieto on vapaaehtoinen.
 - LUT:in LutPub-aineisto on Doriassa ja sille on oma OpenAire-setti: http://www.doria.fi/oai/lutpub?verb=ListRecords&metadataPrefix=oai_dc
 - Pilotin myötä keskitetyille arkistoille on siis mahdollista tehdä oma OpenAire-setti, eikä siis esim. Doriassa ole rajoitetta sen suhteen, että olisi vain yhteinen setti kaikille Doriaa käyttäville organisaatioille.
 - OpenAireen liittyminen vaatii aina asiakaskohtaista suunnittelua metadatakenttien ja syöttöomakkeiden osalta. Haasteita lisäävät OKM:n tiedonkeruuvaatimukset, mahdollisesti pitkälle viety räätälöinti organisaation metadatan osalta ja muut yhteensopivuudet ja suositukset. Vaarana on, että samoja tietoja joutuu toistamaan hieman eri muodossa. Myös totuttuihin arvoihin, esim. type-kentän osalta tarvitaan todennäköisesti muutoksia, jotta vastaa OpenAiren vaatimuksia.
- Theseus-seminaari 26.4. Helsingissä

6. Seuraava kokous

- 5.4. 2016 klo 10 alkaen, paikka tiedotetaan myöhemmin.