

VIRTA-julkaisuyhteyshenkilöt - kokous 12.2.2016

Aika: Perjantai 12.2.2016 klo 13-14.30

ACP: <https://webinar.nordu.net/virta-julkaisu/>

Nauhoite: <https://webinar.nordu.net/p8rh8d9brs3/>

Osallistujat: Anneli Korhonen HAMK, Eeva-Maria Suojärvi Xamk, Fredrik Finnberg CSC, Hanna-Mari Puuska CSC, Heile Tofferi VSSHP, Heli Peltola LAMK, Ilkka Nokelainen LUT, Jarkko Seppä OY, Jukka Haapamäki OKM, Jukka Rantasaari TY, Jussi Kärki SAMK, Kati Sajaniemi TTL, Kirsi Varpa TAMK, Kristian Karlström Arcada, Liisa Hallikainen LaY, Maria von Hertzen Arcada, Miika Nurminen JY, Mikko Kuru TaY, Mikko Saunamäki LUKE, Minna Karvonen SeAMK, Mira Juppi FIMEA, Mira Tiilikainen OY, Neea Salo OAMK, Olli-Pekka Kaurahalmes TY, Raimo Körkö OY, Riitta Kokko, Raine Rapo CSC, Riitta Ruuska JAMK, Sirpa Hangasmäki Tay, Sirpa Riikkinen LUT, Tarja Laakkonen Laurea, Tommi Lehtokangas TaY, Tua Hindersson-Söderholm Hanken, Vaasan yliopiston ryhmä, Virpi Miettinen OY, Yrjö Lappalainen TTY

Ajankohtaiset asiat

1. Tilannekatsaus

- Yhteys on avattu, pilotit voivat siirtää tiedostoja
- Ensimmäisten tiedostojen kanssa edetään manuaalisesti, sen jälkeen kytketään automaattiset virheraportit päälle.
- Sähköpostiviesti piloteilta, kun tiedostoja siirretty.
- Muutamalta pilotilta puuttuvat vielä tiedot yhteyden avaamista varten (IP-osoite ja SSH-avaimet)
- Päästään kunnolla testaamaan, kun saadaan oikeaa aineistoa

Kysymys: Olisiko mahdollista saada suoraan virheraportit csv -muodossa samalle palvelimelle, jonne tiedostot siirretään, jolloin niitä voisi automaattisesti sieltä lukea - ei pelkästään sähköpostilla.

Vastaus: Ei vielä suunniteltu, otetaan jatkokehityksessä pohdittavaksi.

Kysymys: Onko sähköpostilla lähetettävä koosteraportti, vai identifioidaanko siinä myös julkaisut, joissa virheitä on?

Vastaus: Identifioidaan. Raportissa näkyy organisaatiokohtainen julkaisutunnus ja virheen kuvaus. Aikaisemmin yhteyshenkilökokouksessa sovittu, että raportilla lisäksi julkaisun nimi tai osa siitä. Tarkoitus tuottaa myös kontrolliraportti, mitkä julkaisut poimittu tiedonkeruuseen samaan tapaan kuin KOTASTA nykyisin.

2. Skeemamuutokset

- XML-skeemaan tulossa pieniä muutoksia. Jos tehty dataa nykyisen skeeman mukaan, muutokset eivät edellytä toimenpiteitä. Vanhalla versiollakin voi tuoda tietoja.
- Korjataan seuraavat skeemassa havaitut virheet/puutteet:
 - Ilmoitusvuosi vapaaehtoiseksi (generoidaan VIRTA-julkaisutietopalvelussa)
 - Artikkelinumero vapaamuotoiseksi kentäksi (max 50 merkkiä)
 - POLAMK lisään organisaatiokoodistoon
 - Julkaisun tila -koodistoon lisätään koodi -1 (=tietoa ei viedä tietovarantoon)

Kysymys: Onko tietokanta toteutettu samoin kuin opintotietopalvelussa, että se tyhjennetään ennen kuin uutta data ladataan?

Vastaus: Ei. Tässä voi tuoda tietoja julkaisu kerrallaan. Jos tuodaan tieto samalla organisaatiokohtaisella julkaisutunnuksella, vanha tieto poistetaan ja uusi tuodaan tilalle. Julkaisun tila koodilla =-1" tieto poistetaan kokonaan tietovarannosta.

Kysymys: Eli tiedonsiirto on inkrementaalien, että voi lisätä vähän kerrallaan?

Vastaus: Kyllä.

Muita virheitä/puutteita skeemassa voi laittaa tiedoksi virta-julkaisut@postit.csc.fi

3. Tarkistusraporttien sisältö

- Julkaisujen virheilmoitukset:
 - https://docs.google.com/spreadsheets/d/1nRcKlHq1TF0h_IikxA5dCWNFPdd546LJH5oLChQjOhE/edit?pref=2&pli=1#gid=0
 - Taulukossa eroteltu tiedonsiirrossa pakolliset, vapaaehtoiset ja muut tiedot vuoden mukaan
 - Jos julkaisussa on virhe, se vaikuttaa "Julkaisun tila" -koodiin taulukon mukaisesti:
 - Taulukossa numero, esim. -1 viittaa siihen, mihin tilaan julkaisu joutuu, jos siinä havaitaan virhe
 - Osa tarkistetaan XML-tasolla, eli XML ei validoidu, jos tiedossa virhe
 - Osa tarkistetaan vasta tietokannassa, esim. ISSN, sillä monimutkaisimpia tarkistuksia ei pystytty toteuttamaan xsd-määrittelyissä

- Jos tiedonsiirrossa pakollinen tieto puuttuu, julkaisu menee tilaan "-1" eli sitä ei siirretä tietovarantoon
- Jos muu tieto puutteellinen, esim. DOI vääränmuotoinen, siitä tulee merkintä virheraportille, mutta hyväksytään silti OKM:n tiedonkeruuseen
- Organisaatio voi myös itse kirjata julkaisun tilaksi esim. 0, mikäli se ei halua julkaisun näkyvän JUULissa
- Julkaisun tila -koodin voi jättää myös tyhjäksi. Jos kenttä on tyhjä, tila-koodi täydennetään tietovarannossa sen mukaan, onko tiedoissa puutteita/virheitä

Kysymys: Lähetetäänkö julkaisun tila koodilla 1, jos ok ja koodilla -1 jos halutaan poistaa seuraavassa ajossa-

Vastaus: Kyllä (koodilla 2, jos OKM-kelpoinen). Julkaisuja ei tarkisteta ylöspäin: jos ilmoitettu 1:nä, ei viedä tilaan 2, vaikka menisi tarkistuksista läpi.

Kysymys: Onko julkaisumaa nyt pakollinen kenttä?

Vastaus: Ei, mutta tulee virhe, jos ei vastaa koodistoa. Jos julkaisumaaksi merkitty muuta kuin koodiston mukainen koodi tai tyhjä, ilmoitetaan virheestä.

Kysymys: Voitaasiinko julkaisun tila -koodia käyttää siihen, että sen avulla voitaisiin raportoida myös julkaisuja, jotka eivät täytä OKM:n tiedonkeruukriteereitä? Esim. lyhyet konferenssipaperit tai uudelleenjulkaisut. Näytetäänkö JUULissa vain OKM-kelpoiset julkaisut vai myös laajemmin julkaisuja, joita tutkijat haluavat raportoida. JY:ssä käytetty lisätietorastia, jossa merkitty, jos julkaisu ei OKM-julkaisutiedonkeruukelpoinen.

Vastaus (OKM): Nyt määritelty julkaisut, joiden pohjalta arvioidaan korkeakoulujen toimintaa. JUULin ei kannata ottaa uutta luokkaa "muu". Julkaisutietovarannon hyödyntämisen laajentaminen se, tutkijat voisivat näyttää omaa profiiliaan. Julkaisuja, joilla esim. CV-käyttöä, voitaisiin niiden osalta pyrkiä kehittämään ja laajentamaan julkaisutyypiluokitusta. Ennenkin mukaan julkaisutyypiluokitukseen kuin omaksi kaatoluokakseen.

Vastaus (CSC): Alunperin julkaisun tila -koodin "1" tarkoitus, että puutteellisia tietoja voidaan julkaista JUULissa, mutta tiedot vielä puutteellisia joidenkin OKM:n edellyttämien tietojen osalta.

Kysymys: Onko niin, että korkeakoulu ei koskaan lähetä tilakoodilla 2?

Vastaus: Voi lähettää koodilla 2, mutta jos tietoja havaitaan tietovarannossa puuttuvaksi, korjataan koodiin 1.

Kysymys: Vaaditaanko ISBN-numeroihin yhä väliviivat?

Vastaus: Voi olla väliviivat, välilyönnit tai yhtenä pötkönä, mutta ei voi vain olla joistakin kohdista pois. Täytyy olla joko kaikki väliviivat tai ei lainkaan. -> Selvennetään vaadittava muoto elementtien kuvaukseen.

Tällä hetkellä tarkistusten käsittely julkaisun tila -koodin avulla on vasta ehdotus. Julkaisun tila -koodia voidaan tarvittaessa laajentaa myös muihin käyttötapauksiin. CSC tuottaa yhteyshenkilöille tarkemman kuvauksen prosessista seuraavalla viikolla ja ottaa seuraavaan yhteyshenkilökokoukseen paremmin keskusteluun.

4. Vanhojen julkaisutietojen korjaaminen

- Keskustelua prosessista vanhojen julkaisujen korjaamiseksi sekä lisäämiseksi:
 - Tarkoituksena, että julkaisuja voi tuoda myös historiasta
 - 2011-14 tiedot viety VIRT-julkaisutietopalveluun
 - Jos vanhoja tietoja haluaa korjata, korjaus tehdään organisaatiokohtaisen julkaisutunnuksen avulla. Kaikilla vanhoilla julkaisuilla tätä tunnusta ei kuitenkaan ole.
 - Yksi vaihtoehto on, että tietovarannossa generoidaan tunnukset julkaisuille, joilla sitä vielä ole ja raportoidaan ne takaisin organisaatioille
 - Voidaan myös tunnistaa duplikaattina, mutta näkyy tällöin organisaatiolle duplikaattivirheenä

Kommentti: Open access -tieto tms voidaan haluta täydentää jälkepäin. Vanhat tiedot eivät kuitenkaan sisällä kaikkia kenttiä viimeisimpien määritysten mukaan ja tietoa ei välttämättä edes saatavilla.

Vastaus: Kriteerit löyhemmät vanhoille julkaisuille, joita ei enää huomioida tiedonkeruissa. Tietyt perustiedot (julkaisun nimi, tekijät ja julkaisuvuosi jne.) pakollisia, mutta muuten ei edellytetä kaikkia tietoja.

Kysymys: Korkeakoulujen toimittamista tiedoista CSC:llä koostettuihin exceleihin on joka tapauksessa viety jonkinlaiset ID:t. Voisiko niitä viedä tietovarantoon ja niiden perusteella korjata? Onnistuisiko myös kertaluonteinen siirto, jolloin voisi viedä julkaisulle alkuperäisen lähdejärjestelmässä olevan tunnisteiden. Helpottaisi lähtöpäässä, jos saataisiin tietovarantoon alkuperäiset tunnisteet vanhoistakin julkaisuista.

Vastaus: Exceleiden ID-tiedot saatavilla ja voisi olla mahdollista, että niiden avulla luodaan organisaatiokohtaiset tunnisteet, mikäli niitä ei ole. Mietitään myös prosessi, jotta organisaatioiden omia tunnisteita voisi jälkikäteen toimittaa.

Kommentti: Vanhoihin julkaisutietoihin ei pitäisi enää lisätä tietoja, jos ne on jo tilastoitu ja hyväksytty, koska tiedonkeruuväittämukset on muuttuneet useampana vuonna

Vastaus: Korjauksia ei ole tarkoitus viedä enää tiedonkeruuseen poimittavaan aineistoon tai tilastoihin. On kuitenkin tullut palautetta, että JUULin tietoja pitäisi voida korjata - etenkin jos tietoja hyödynnetään laajemmin muissakin palveluissa.

CSC tekee tästäkin tarkemman ehdotuksen jatkokeskustelua varten.

Tiedoksi

- TUHA-seminaari 21.4.2016 klo 9-16 (HY, Minerva-tori)

- Teemana tutkimuksen arviointi
- Iltapäiväosuudessa esillä myös tutkijan sähköiset palvelut
- VIRT-julkaisutietopalvelun hallintamallin luonnos löytyy [täältä](#)
- [VIRT-julkaisutietopalvelu - kommentteja ja kysymyksiä](#) -sivustolle lisätty useimmin kysyttyjä kysymyksiä
- Seuraavassa kokouksessa käsiteltäväksi: edellisen vuoden julkaisujen tarkistukset ja huomioonottaminen tiedonkeruussa

Seuraavat kokoukset

Ke 9.3. klo 13-14.30 (siirretään hiihtolomien ja päällekkäisen ATT-seminaarin vuoksi). Niille, jotka eivät pääse esim. pohjoisen hiihtolomaviikon takia, järjestetään kokous myös aiemmin sovittuna ajankohtana pe 4.3.2016 klo 13-14.30.

Pe 1.4. klo 13-14.30

Pe 29.4. klo 13-14.30

Pe 3.6. klo 13-14.30