

VIRTA-julkaisuyhteyshenkilöt - kokous 9.3.2016

Aika: Keskiviikko 9.3.2016 klo 13-14.30

ACP: <https://connect.funet.fi/virta-julkaisu>

[Nauhoite 9.3.2016](#)

Niille, joille ajankohta ei sopinut, järjestettiin kokous samansisältöisenä pe 4.3.2016, [ks. nauhoite 4.3.2016](#)

Osallistujat:

Ajankohtaiset asiat

1. Pilotoinnin sujuminen

Muistiinpanot 4.3.2016

- Viisi pilottia lähettänyt dataa. Lähes 8 000 julkaisua ladattu tietovarantoon. Ladataan tällä hetkellä kerran päivässä sisään. Jatkossa tiheämmin.

Pilottien tilannekatsaukset:

Turun yliopisto: Siirtotiedostoihin liittyviä virheitä ISSN:issä, jotka todennäköisesti hoituvat automaattisella korjauksella. XML-tiedosto muodostetaan tutkimustietojärjestelmästä tietovaraston raportointiminäisyyksien avulla. Ei erityisiä ongelmia VIRTA-tiedonsiirrossa.

Lahden ammattikorkeakoulu: XML-tiedosto tehty CSV-XML-muuntimella. Työkalu ollut hyödyllinen tarkistuksissa. Tavoitteena saada 2016 julkaisuista neljä kertaa vuodessa siirto tietovarantoon.

Kysymys: Tulevatko seurantaraportit sitten extra-vipusen puolelle?

Vastaus: Kyllä. Tarkoitus toteuttaa kesän-syksyn aikana.

Kysymys: Kysymys virheraportista: saako julkaisuun painetun virheellisen ISBN:n ilmoittaa (esim. tarkistusnumero väärin) vai pitäisikö ISBN muuttaa julkaisutietoihin oikeaksi? Eli riittääkö ISBN:n oikea muoto vai pitääkö numeroittenkin olla oikein?

Vastaus: Toistaiseksi tietovarannossa pystytään tarkistamaan vain, onko tunnus oikeanmuotoinen.

Kysymys: CSV-XML:ssä ei kenttää julkaisun tilalle? Nyt oletuksena julkaisun tila on 0, kun siirretään tietovarantoon.

Vastaus: Voidaan lisätä sarake julkaisun tilalle työkaluun. Ennen sitä muokataan käsin tietovarantoon.

Muistiinpanot 9.3.2016

- Lähes kaikki pilotit toimittaneet ainakin testidataa. Tietovarantoon siirretty noin 11 000 kappaletta.
- Piloteille lähetetty kahdenlaisia raportteja jokaisella latauskerralla:
 - Tilanneraportti sisältää tiedot julkaisun identifioivasta tunnuksesta (organisaation oma sekä tietovarannossa generoitu), organisaatitunnuksen, julkaisun nimen sekä luonti- ja muokkauspäivämäärän. Lisäksi julkaisun tila kertoo, omko julkaisu hylätty, puutteellinen tai hyväksyttävä OKM-tiedonkeruuseen. Tilanneraportti toimitetaan joka latauksella.
 - Virheraportti sisältää julkaisun identifioivasta tunnuksesta (organisaation oma sekä tietovarannossa generoitu), organisaatitunnuksen, julkaisun nimen sekä virhekoodin ja julkaisun tilan. Raportti toimitetaan vain, jos tiedoissa on virheitä tai puutteita.
- C2-typpejä koskevaan ohjeistukseen jäänyt virhe
 - Pakollisuuksissa oli jäänyt huomioimatta toimitetut lehtien erikoisnumerot, joilla voi olla ISSN, mutta ei ISBN-tunnusta
 - Jatkossa pitää olla jompi kumpi
 - Asia korjataan 9.3. aikana sekä KOTA-alustalle, tietovarantoon että CSV-XML-työkaluun.
- CSV-XML-työkaluun tehty muutamia pieniä muutoksia:
 - Lisäksi CSV-työkaluun lisätty julkaisun tila siten, että oletuksena jokaisen julkaisun tila on 2. Arvon voi muuttaa vielä XML-tiedostoon. Julkaisun tila -koodia ei ole vielä mukana mallitiedostossa.
 - Muutettu myös pakollisuuksia tutkimuslaitosten ja sairaaloiden kohdalla.
- Julkaisutiedonkeruu
 - Deadline 14.3.
 - [ks. VIRTA ja julkaisutiedonkeruu](#)

- OKM-tiedonkeruussa huomioidaan ko. päivään mennessä tietovarantoon siirretyt julkaisut, joiden tilakoodi on 2.

Kysymys: CSV/XML-muuntotyökalu listaa virheet työkalussa tai lokitiedossa. Olisiko mahdollista saada tuottamaan tietoa taulukkomuodossa?

Vastaus: Voitaisiin jatkossa tehdä csv-muotoisena. Huomioidaan työkalun myöhemmässä versiossa.

Kysymys: Pitääkö elementti poistaa XML:ää, jos ei siihen tuoda tietoa (esim. julkaisumaa)

Vastaus: Julkaisumaa-koodi elementin voi jättää pois, mutta jos sen antaa, se ei voi olla tyhjä, vaan on annettava validi maakoodi.

Kysymys: olisiko muuntotyökalusta mahdollista saada jossakin vaiheessa komentoriviskripti? Helpottaisi paljon xml-tiedoston muodostamista, eikä kaikkien tarvitsisi välttämättä rakentaa omaa ratkaisua.

Vastaus: Voisi tehdä version, joka vie virheet lokiin. Nykyinen on tehty pienempiä korjauksia ja tarkistusta varten. Jotenkin virheet pitää ilmoittaa, eli ei voi muodostaa xml:ää ei-validista aineistosta. Jonkun on siinä oltava välissä korjaamassa. Voidaan ottaa kehitykseen versio, joka on komentorivipohjainen.

Kysymys: JUULissa on lehtiartikkeleita, joissa on vain lehden nimi ja julkaisu vuosi. Voisiko tarkistuksiin lisätä, että jotain muutakin tietoa pitää olla (joko numero, sivut, päivämäärä ..)?

Vastaus: Voidaan mahdollisesti ottaa puutelistaukseen. OKM-tiedonkeruun osalta täytyy pohtia erikseen.

Pilottien tilannekatsaukset:

Oulun yliopisto: CSV/XML-työkalu ollut tarpeellinen. Etukäteistarkistukset helpottaneet työtä. Alunperin oli suunniteltu tehtävän eri tavoin. Siirto sujui yllättävän helposti. Jatkossa tavoitteena siirtyminen automaattiseen tiedonsiirtoon. Virheraportit hyödyllisiä, edellytti kuitenkin organisaatiokohtaisen tunnuksen poimimista ja tarkistamista omassa järjestelmästä ja korjaamista XML:ään.

Tampereen yliopisto: Kokonaisuutena mennyt hyvin. Virheitä yllättävän vähän. Luodaan 2 tiedostoa, tietovarastosta suoraan csv-tiedosto sisäiseen käyttöön. Käytetään csv-xml-työkalua sen tutkimiseen. XML tuotetaan SQL-serverialustalla erikseen. Tietoja tulee joka yö.

Tampereen teknillinen yliopisto: Pure-järjestelmä viime vuonna otettu käyttöön. Nyt ensi kertaa käytössä raportoinnissa. Poimitaan automaattisesti tietoja ja lisäksi toistaiseksi käytetty CSV/XML-työkalua tiedoston muodostamiseen. TTY:llä koodattu "poimuri", jolla Puren syötteestä saadaan poimittua vaadittavat tiedot. Seuraavassa vaiheessa tuotetaan VIRTAAan XML suoraan omalta palvelimelta. Tiedot ajastettu siirtymään kerran vuorokaudessa.

HAMK: Nyt tehdään käsin, mutta jatkossa varmaan tavoitellaan jatkuvaa siirtoa, eli 1/yö. Ei käytössä mitään maksullista valmista ohjelmaa. On tehty omaa amk-kontekstiin sopivaa ratkaisua.

Työterveyslaitos: Tietojen kokoaminen käynnissä. Julkaisutietojärjestelmästä toistaiseksi puuttuu tiedonsiirrossa vaadittavia tietoja. Vuoden 2016 julkaisuja syötetään uuteen järjestelmään. Järjestelmästä saadaan suoraan XML-aineistoa.

FIMEA: Vuoden 2015 tietoja parhaillaan syötetään. Järjestelmässä ei kaikkia kenttiä, joten tietoja joudutaan täydentämään. Julkaisuja melko vähän ja tehdään paljon käsityötä.

Seuraavaksi

- Yhteisjulkaisujen ja julkaisufoorimitasojen tunnistaminen VIRT-julkaisutietopalvelusta
 - Julkaisutyypin ristiintaisuuDET tulevat korjattaviksi organisaatioille keVäällä samoin kuin viime vuonna
 - Pilotit voivat tehdä korjaukset suoraan VIRTaan.
 - Julkaisufoorumi-tasot vahvistetaan kesäkuussa.
- Tiedot viedään JUULiin kesäkuun loppupuolella.
- Tavoitteena, että virhe- ja tilanneraportit syksyllä luettavissa Extra-Vipusen web-palvelussa käyttöliittymän kautta

Kysymys: Aikataulukysymys: milloin yhteisjulkaisujen tyyppiristiriidat ym XML:n lähetyksen jälkeen havaitut korjaukset on jätettävä?

Vastaus: Maalis-huhtikuun vaihteessa tiedot OKM:ään. OKM tekee pohjaehdotuksen julkaisutyypikorjauksista sekä sisästen duplikaattien poistamisesta, jonka korkeakoulut ovat joko hyväksyneet tai hylänneet. Tänä vuonna mennään ainakin vielä näin, jatkossa voidaan miettiä vaihtoehtoisia toimintatapoja. Korjauspyynnöt arviolta huhtikuun puolessa välissä eteenpäin organisaatioille. Organisaatioilla noin pari viikkoa aikaa korjata.

Tiedoksi

- ORCID-seminaari konsortiojäsenyysasioista järjestetään pe 22.4.2016

Seuraavat kokoukset

Pe 1.4. klo 13-14.30

Pe 29.4. klo 13-14.30

Pe 3.6. klo 13-14.30