

VIRTA-ESISELVITYS, FUNIDATAN KOMMENTIT

Alla eriteltyinä eri teemoihin Funidatan huomioita nykyisestä VIRTA-opintotietopalvelusta, joita toivoisimme, että uudistuksen esiselvityksessä voitaisiin huomioida.

1. Hallintamalli, jolla varmistetaan datan parempi laatu

VIRTA-data luodaan hyvin hoidetun skeeman pohjalta, jolla taataan, että sisääntuleva data on samassa muodossa riippumatta lähteestä. Tämä ei kuitenkaan varmista, että tieto oli semanttisesti samanlaista vaan osin selkeän ja valvotun ohjeistuksen puuttuessa dataa voidaan toimittaa monella eri tavalla oman prosessin mukaan.

Tämä epäilemättä aiheuttaa ongelmia datan tulkintavaiheessa, kun asioita on mahdollista olla kirjattuna opintotietojärjestelmissä vaihtelevilla tavoilla. Tätä voidaan toki paikata tietovarastossa tapahtuvissa transformaatioissa, mutta selkeämmän datamallin avulla voitaisiin vähentää tulkintaepäselvyyksiä.

Samoista syistä olisi toivottavaa, että VIRTA edelleen toimittaisi tiedot kansallisille toimijoille (kuten YTHS), jotta datan kierrättämiseltä useamman eri toimijan ja tietovaraston läpi vältyttäisiin. Ne ovat omiaan lisäämään datan tulkinassa ilmeneviä epäselvyyksiä. Useat korkeakoulujen rahoitusmittarit ovat pohjautuneet tälle datalle, joten olisi myös tärkeää, että se on yhteismitallista ja mahdollisimman selkeää.

2. Inkrementaalinen toimintamalli tiedon toimittamisessa

Nykyisen opintotietopalvelun tapa ottaa vastaan tietoa ei palvele nykyaikaisia tietojärjestelmiä. Koko tietosisällön pakkaaminen tiedostoihin joka yö on laskennallisesti haastavaa monimutkaisista operatiivisista tietovarastoista, varsinkin datamäärien lisääntyessä vuosi vuodelta. Sisu toteuttaa inkrementaalisen laskennan VIRTA-datalle, jossa öisin muodostetaan vain muuttuneet VIRTA-tietueet ja olisi järkevää, että datan voisi toimittaa myös opintotietopalveluun tällä tavoin.

Toisena haasteena on datan lataaminen palveluun, jossa tulevaisuudessa datamäärien vielä kasvaessa aika loppuu kesken korkeakoulukohtaisista ikkunoista. Tiedostoja voidaan myös toimittaa vain ikkunoissa ja vain koko tietomassan ollessa valmiina, joka ei mahdollista muutoksia lyhyillä aikaikkunoilla vaan vaatii pitkät iteraatiosykliä.

3. Parempi palvelukerros tietovaraston lisäksi

Nykyinen VIRTAtietopalvelu toteuttaa käytännössä tietovarannon lisäksi melko vähän palveluita esimerkiksi tiedon tuottajan suuntaan. Funidata ei itse toimita XML-tiedostoja opintotietopalveluun, mutta ongelmien selvitystä tehdään osin meidän toimestamme.

Tiedostopohjainen siirtoformaatti on haastava, koska palautteen latauksen onnistumisesta saa vasta kun koko tietomassa on siirretty ja yritetty lukea. Rajapintaratkaisulla pystyttäisiin reagoimaan myös mahdollisiin virheisiin nopeammin. Tiedostoja voidaan toki hyödyntää myös toisiokäyttöön, mutta toivoisimme rajapintojen olevan ensisijainen vaihtoehto.