

# PAS-pilotit toivat uutta tietoa palvelun kehittämiseen

Fakta Julkaistu 14.04.2016 klo 14:42 Kirjoittaja Tommi Kuttilainen

ATT-hankkeen Tutkimuksen pitkäaikaissaatavuuspalvelu (Tutkimuksen PAS-palvelu) toteutti vuoden 2015 aikana kolme pilottia, joiden tarkoituksena oli selvittää millaisia vaatimuksia valitun aineiston säilyttäminen asettaisi Tutkimuksen PAS-palvelulle ja parantaa pilottiin osallistuvien tutkimusorganisaatioiden teknisiä valmiuksia PAS-palvelun käyttöön. Pitkäaikaissaatavuuden edellytyksenä on pitkäaikaissäilytys, ja näissä piloteissa keskityttiin ensisijaisesti aineistojen ymmärrettävyyden säilytyksen kannalta olennaisten kysymysten ratkaisemiseen.

Pilotteja toteutettiin yhteensä kolme kappaletta:

- Aalto-yliopiston Brain & Mind Laboratory (BML), aineisto Elokuvan käyttö aivotutkimuksessa (8 gigatavua)
- Jyväskylän yliopiston Kiihdytinlaboratorio, aineisto 250-Nobeliumin hajoamisspektroskopia (200 gigatavua)
- Turun yliopiston Avaruustutkimuslaboratorio (ATL), aineisto ERNE-instrumentin vuohavainnot 1996–2014 energioissa 1.6-200 MeV/n (22 gigatavua).

Tutkimusaineiston elinkaari ajatellaan yleisesti laajana kokonaisuutena, joka pitää sisällään mm. tutkimussuunnitelman laatimisen, aineiston analyysin, tulosten julkaisun ja aineiston pitkäaikaissäilytyksen. Pilottien käytännön kokemukset osoittavat, että pitkäaikaissäilyttämisen onnistuminen edellyttää asian huomioimista koko aineiston elinkaaren ajan.



## Johtopäätöksiä

Aineistojen säilyttämiskelpoisuus oli määritelty merkittävässä määrin jo ennen varsinaisen tutkimuksen alkamista tutkimuksen suunnitteluvaiheessa. Käytännössä tämä tarkoitti esimerkiksi mittaus suunnitelmaa (Kiihdytinlaboratorio), eettisen toimikunnan lupapyyntöä (BML) tai aineiston jakelusta tehtyä sopimusta (ATL). Kaikissa näissä tapauksissa jo tutkimussuunnitelmassa tehdyt valinnat vaikuttivat siihen, kuinka aineistoa voidaan uudelleen käyttää alkuperäisen tutkimuksen jälkeen. Aineisto voi olla täysin avointa (ATL), sensitiivistä (BML) tai täysin luvanvaraista (Kiihdytinlaboratorio). Erityisesti aineistojen jakelu kolmansille osapuolille (saatavuus) edellyttää käyttöoikeuksien täsmällistä määrittelyä. Tutkimuksen PAS-palvelun tulisi tukea tutkimusaineistojen käyttöoikeuksien ja rajoitusten kuvaamista.

Aineistojen käsittely parantaa aineiston hyödynnettävyyttä tulevaisuudessa. Tutkimuksen PAS-palvelun tulisi mahdollistaa tutkimusaineiston säilyttäminen sen käsittelyn kaikissa vaiheissa.

Tutkimustulosten tieteelliset julkaisut voivat olla merkittäviä myös aineiston jatkokäyttöä ajatellen. Esimerkiksi aineiston tai siihen liittyvän mittalaitteen keskeinen dokumentaatio (ATL) voi olla tieteellisessä julkaisussa. Kustantajan kanssa tehty julkaisusopimus voi estää tai rajoittaa julkaisujen säilyttämistä. Julkaisun alkuperäisen käsikirjoituksen tekijänoikeudet ovat kuitenkin tutkijalla itsellään, joten niiden säilyttäminen mahdollisimman aikaisessa vaiheessa tulisi mahdollistaa.

Pilottien aineistot olivat laskennallisia ja niihin liittyi jokin mittalaite, jolla ne olivat tuotettu. Tämä tarkoitti sitä, että aineisto oli alkuperäisesti mittalaitteen tuottamassa ”custom-formaatissa”. Ymmärrettävyyden säilyttäminen on haastavaa ja jotta se ylipäättänsä olisi mahdollista, täytyy formaatille ja sen dokumentaatiolla luoda kriteeristö. Tutkimuksen PAS-palvelussa tulisi olla ennalta määritetyt kriteerit tutkija- tai laitekohtaisten tiedostomuotojen vastaanottamiselle.

Aineistojen tiedostomuotoon voi liittyä myös pakkausalgoritmi (BML). Säilyttämisen kannalta pakkausalgoritmit ovat ongelmallisia, koska koko aineiston ymmärrettävyys riippuu tästä algoritmista. Tästä syystä Tutkimuksen PAS-palvelun tulisi tapauskohtaisesti tukea myös pakattuja tiedostomuotoja. Tutkimuksen PAS-palvelun tulisi myös tukea yleisesti käytettyjä tekstimuotoisen tiedon esitystapoja sekä yleisesti käytettyjä teksti-, kuva- ja videotiedostomuotoja.

Tutkimusaineistoihin liittyvät metatiedot selvitettiin ja arvioitiin näiden merkitys säilytyksen kannalta. Tutkimuksen PAS-palvelun tulee tukea sekä tutkimusaineistojen yhteisiä kuvailevia metatietoja että tieteenalakohtaisia kuvailevia metatietoja.

Aineiston paketointi teknisesti ei ollut vaativaa ja [KDK:n](#) (Kansallinen digitaalinen kirjasto) paketointimääritykset osoittautuivat toimiviksi myös tutkimusaineistoille. Tutkimuksen PAS-palvelun tulisi tarjota yleinen paketointikomponentti, joka mahdollistaisi siirtopakettien luomisen automatisoinnin.