

This is a self-archived version of the original publication.

The self-archived version is a publisher's pdf of the original publication.

To cite this, use the original publication:

**Lehto, A. & Suikkanen, E. 2021. Ammattikorkeakoulun TKI-aineistot haltuun datakatalogilla. Kreodi 3/21: Avoin tki & oppiminen. URN: NBN:fi-fe2021060634287.**

Link to the original publication: [URL](#)

All material supplied via Turku UAS self-archived publications collection in Theseus repository is protected by copyright laws. Use of all or part of any of the repository collections is permitted only for personal non-commercial, research or educational purposes in digital and print form. You must obtain permission for any other use.

# Ammattikorkeakoulun TKI-aineistot haltuun datakatalogilla

06.06.2021



KUVA: UNSPLASH, CC0.

Kuten monissa ammattikorkeakouluissa, Turun ammattikorkeakoulun tutkimus-, kehitys- ja innovaatiotoiminnassa (TKI) syntyvien aineistojen hallinnan haasteet ovat moninaiset. Vaikka kaikki liittyikin kaikkeen eikä datanhallinnan prosesseja ole mahdollista kehittää tyhjiössä, moniulotteisen kokonaisuuden hahmottamiseksi on hyödyllistä jakaa se ajatuksellisesti kahteen pääosaan: varsinaisen TKI-aineiston hallintaan ja sitä kuvaavan metadatan hallintaan. Jälkimmäisen nostaminen prioriteetiksi sisäisessä kehittämistyössä avaa uudenlaisia näkökulmia siihen, mitä TKI-aineisto on ja mitä sillä kannattaisi tehdä.

Datanhallinnan kehittämistä tukipalveluiden avulla

Nykyisen ja varsinkin tulevan toimintaympäristömme asettamien vaatimusten täyttävän sisäisen infrastruktuurin ja tukipalveluprosessien luominen kysyy ennakointia. Näitä vaatimuksia ovat mm. avoimen tieteen vahvistuva paradigma (Open Science), tietosuojasäätelyn vaihtuvat tulkinnat sekä lakisääteisten tehtävien edistäminen kaikin tavoin vastuullisella ja kestäväällä tavalla.

Datanhallinta on Turun ammattikorkeakoulussa lähivuosien yksi keskeisimmistä kehittämiskohteista, mitä tulee TKI-lyhenteen alla tehtävään työhön. Aineistohallinnan viitekehys on niin laaja ja monisyinen, että merkittävien kehittämistoimien aikaansaaminen vaatii pitkäkestoista ja jatkuvaa panostusta sekä oman organisaatiomme rajojen ylittävää yhteistyötä. Turun ammattikorkeakoulun lähtökohta erityisesti datanhallintaan on rakentaa riittävän vahvoja tukipalveluita pitkällä tähtäimellä.

## Pienistä puroista datavirtaan

Ammattikorkeakoulujen viitekehyksessä tutkimusaineistolla tarkoitetaan TKI-toiminnassa syntyviä ja prosessoituja aineistoja. Tyypillisesti tällaisia aineistoja ovat esimerkiksi mittaustulokset, tilastot, kuvat, äänitallenteet, haastattelujen litteroinnit, videot, datakoosteet, kyselyaineistot ja kenttätyöhavainnot. Erityisesti tukipalvelujen näkökulmasta on hyvä nähdä myös itse ”TKI-aineistot” mahdollisimman laajana terminä, sisältäen käytännössä myös palvelutoiminnan ja koulutuksen pää- tai sivutuotteena syntyvän datan. Tällä lähestymistavalla pystytään fokuoitumaan haluttuihin aihepiireihin joustavammin eri näkökulmista ja samalla yhteisöjemme jäsenien kynnys osallistua eri toimintoihin madaltuu.

Tiukasti sisäisten toimijoiden hyppysissä pysyvät aineistot mahdollistaisivat vaikkapa opinnäytetöiden systemaattisemman tekemisen ja teettämisen tietyn aihepiirin hankkeiden ympärillä tai opiskelijaryhmien kerryttämän datan käyttämisen palvelutoiminnan kehittämiseen siten, että sillä voisi olla merkitystä tutkimusryhmätasolla. Näin ollen irrallisin sykäyksin tyypillisesti kertyvä hankepohjainen aineisto voisi saada rinnalleen muistakin lähteistä syntyvää aineistoa, joka puolestaan tukisi myös tulevaa hanketoimintaa.

Haasteelliseksi ammattikorkeakoulujen kestävä datanhallinnan tekee jo lähtökohtaisesti se, että ammattikorkeakouluissa tuotetaan tyypillisesti paljon big datan vastakohtaa eli small dataa – kyselyn tulokset siellä, hometesti tuolla. TKI-aineistoja on usein vaikea tunnistaa arvokkaaksi, koska ne tuntuvat yhtäaikaaisesti sekä irrallisilta että kontekstisidonnaisilta. Kun samaan aikaan seuraava projekti on jo tippunut työpöydälle, datanhallintaan panostaminen laajasti yli rahoittajan asettamien minimivaatimusten saattaa tuntua lähes utopialta.

Mutta jos vaikkapa Turun ammattikorkeakoulussa tiedettäisiin kattavasti, mitä dataa meillä on ja missä, sen avulla olisi helpompaa tuottaa uutta dataa. Edellisen projektin data muuntuisi osaksi vakuuttavaa uutta hankehakemusta, palvelutoimintana synnytetty aikasarja akateemisesti kiinnostavaa artikkelia. Opinnäytetyön yhteydessä generoitunut data olisi ponnistuslautana seuraavalle valmistumista vailla olevalle, yhden opiskelijaryhmän kyselyaineisto auttaisi toisen porukan lentoon muuten tylsässä toimeksiannossa. Jos sitä kohti aktiivisesti pyrittäisiin, jo

muutamassa vuodessa pienistä datavirroista voisi syntyä aidosti toimintaa tukevaa ja ohjaavaa materiaalia, jonka päälle voisi rakentaa uutta toimintaa uudella tavalla, Horisontti Eurooppa -ohjelmasta ja ensimmäisen vuoden projektipajoihin.

On selvää, että aineistojen avaaminen avoimen tieteen periaatteiden mukaisesti on osa väistämätöntä toimintakulttuurin muutosta niissäkin korkeakouluissa, joissa asiaan ei ole vielä käytännössä kiinnitetty riittävää huomiota. Laadukkaat aineistot eivät avaa itse itseään. Vaatii paljon työtä ja monta askelta ”kuratorintipolulla”, ennen kuin jokin tietty data on saatu avattavaan muotoonsa. Tämä aiheuttaa suurta painetta organisaatiokohtaisille tukipalveluille, koska yksittäisen toimijan ei voida olettaa hallitsevan koko datanhallinnan kokonaisuutta.

Toisaalta aineiston avaaminen on vain yksi asianmukaisen aineistohallinnan mahdollistamista lopputulemista. Siksikin on hyödyllistä tarkastella datanhallintaa myös oman organisaation kehittymistä edistävänä toimintana ja etsiä synergiaetuja samalla kuin datan avaamisen volyyymia kasvatetaan. Sitä kautta voitaisiin lisätä aineistointensiivisyyttä esimerkiksi erilaisissa laboratorioympäristöissä, millä olisi positiivinen vaikutus sekä hanke- että palvelutoimintaan koulutuksen monipuolistamisesta puhumattakaan. Tukipalveluiden näkökulmasta realistinen tavoite kohti aineistointensiivisyyttä voisi olla vaikkapa se, että aineistojen hallinta toimisi kokonaisuutena vähintään yhtä paljon oman organisaation mahdollistamana kuin siitä huolimatta. Yhteiset työkalut ja yhtenäiset prosessit auttavat järjestämään pientä datasilppua vaikuttaviksi kokonaisuuksiksi.

## Metadataprosessi datanhallinnan ytimessä

Aineistopohjainen toiminnanohjaus ei ole mahdollista ammattikorkeakoulun tai edes tutkimusryhmien tasolla, jos käytettävissä ei ole tietoa olemassa olevista aineistoista ja niiden hyödyntämisen reunaehdoista. Varsinaisen aineiston ja sen metadatan hallinta ovat toisiaan täydentäviä, pakollisia osia datanhallinnan kokonaisuutta. Mutta entä jos metadata nähtäisiin yksittäisessä organisaatiossa prioriteettina?

Oletuksena tässä näkökulmassa on se, että itse TKI-aineisto on jo kertynyt tai kertymässä johonkin. Kehittymättömässä datanhallinnan toimintakulttuurissa on selvää, että metadataosio on heikommissa kantimissa kuin itse dataosio, koska varsinaisella datalla on enemmän välitöntä itseisarvoa. Siksi vaikuttavimmat muutokset datanhallintaa koskettavaan toimintakulttuuriin on mahdollista tehdä juuri metadatan suhteen. Tämänkaltaisen priorisoinnin taustalla on ajatus TKI-aineiston arvosta itsenäisenä toiminnan tuloksena, ei vain kertakäyttöisenä raaka-aineena vaikkapa prosessin, palvelun, toimintatavan, tuotteen tai opintosuorituksen synnyttämiseksi.

Oli näkökulma sitten kehittämislähtöinen tai ei, TKI-aineistojen tehokkaan hallinnan ytimessä on mahdollisimman aukoton metadataprosessi, jota seuraamalla voidaan ylläpitää kunkin aineiston laatu riittävällä tasolla koko aineiston elinkaaren ajan sen synnyttämisen suunnittelusta aina sen tuhoamiseen. Metadatasta huolehtiminen luo osaltaan myös tiettyä suunnitelmallisuutta, joka sujuvoittaa tutkimusdatan arkistoinnista ja avaamista sekä mahdollistaa aineistohallinnan potentiaalsiin ongelmiin valmistautumisen ajoissa. Tieto siitä, mitä on olemassa, on jo puoli

voittoa, eivätkä täydellisetkään tallennusratkaisut auta, jos niissä säilöttyä dataa ei kyetä löytämään ja/tai hyödyntämään. Tällöin puhumme tarpeesta metadataan, joka mahdollistaa varsinaisen aineiston luvallisen käytön (hallinnollinen metadata), sen luonteen ymmärtämisen (kuvaileva metadata) sekä sen käytännön hyödyntämisen (rakenteellinen metadata). Kaikista kolmesta tulokulmasta rikastettu metadata on helppo avata käyttöön ulkopuolisille tahoille – ja se on oman organisaation kannalta potentiaalisesti hyvinkin hyödyllistä, avattiin sitä lopulta tai ei.

## Datakatalogi metadataluettelona

Ammattikorkeakoulukontekstissa datakatalogin yksinkertainen ajatus on luoda listaus kaikesta edellä mainituista aineistoista, ensimmäisen vuoden opiskelijoiden ryhmätyönä laatimasta kyselyaineistosta Horisontti-hankkeen tutkimustuloksiin. Datakatalogi ei siis ensisijaisesti arvota varsinaisia aineistoja vaan antaa mahdollisuuden niiden paikantamiselle ja näkyväksi tekemiselle. Käyttäjien ei myöskään tarvitse arvottaa omia aineistojaan itse, koska metatietojen syöttäminen järjestelmään ei sitä vaadi. Vaikka toiminnanohjaus sinänsä tarvitsee laadukasta tietoa sekä organisaatio- että tutkimusryhmätasolta, datakatalogin käyttöä ohjaavat kriteeristöt ovat toissijaisia.

Ensimmäisessä metadataprosessin kehittämisen vaiheessa on tärkeää saada tietoa siitä, mitä dataa on yleensä olemassa. Toisin sanoen olisi olennaista pitää huolta, että katalogia voitaisiin täydentää ja selailta joustavasti useista eri käyttäjäryhmistä käsin. Perusjako näille sisäisille muokaus-, katselu-, haku- ja raporttioikeuksille voisi olla vaikkapa:

1. toiminnanohjaus
2. toiminta
3. tukitoiminta
4. yhteisö.

Datakatalogi toimisi siis ammattikorkeakoulun TKI-aineistojen kuvailevien tietojen tallennuspaikkana. Katalogista näkisi, mitä aineistoja ja niihin liittyvää metadataa yhteisön jäsenet ovat tuottaneet. Toisin sanoen katalogista voisi tarkistaa mm. omistajuuteen, käyttöoikeuksiin ja tutkimusmenetelmiin liittyviä yksityiskohtia, joiden puute usein tulee datan jatkohyödyntämisen tielle. Eri käyttäjäryhmillä olisi tarpeen mukaan joustava pääsy toistensa metadatoihin, samoin kuin tilanteen mukaan myös organisaation ulkopuolisilla käyttäjillä (esimerkiksi hankepartnerit).

Metadatan tuotanto tällaiseen datakatalogiin olisi äärimmäisen matalan kynnyksen toimintaa, mutta sitä silti tuettaisiin julkaisutiedonkeruuseen rinnastettavissa olevalla keskitetyllä tukipalvelulla. Pääajatuksena metadatan syöttämisessä järjestelmään olisi rikastaa metadataa toiminnan aikana samalla, kun varsinaista aineistoa kerrytetään ja prosessoidaan. Kunkin aineiston metadatatietue on tärkeä saada luotua välittömästi datan kerryttämisen alkaessa tai parhaimmillaan jo ennen sitä, jotta metadatan vaiheittainen rikastaminen on mahdollista. Yksi

toiminnanohjauksellinen keino tämän varmistamiseksi olisi antaa tallennustila aineistolle vasta sen jälkeen, kun kyseessä olevan aineiston perustason metadata on syötetty datakatalogiin.

Datakatalogissa eri käyttäjäryhmien käytön mahdollistaminen ja heidän erilaiset näkymänsä metadataan ovat keskiössä. Tämä koskee erityisesti myös opiskelijoita, joiden koulutukseen olisi hyvin tärkeää saada sisällytettyä entistä kattavammin perustason osaamista aineistojen vaalimisesta. Tällä olisi samalla myönteinen vaikutus kunkin ammattikorkeakoulun datanhallinnan kulttuuriin. Kuten todettua, ei ole mitään estettä sille, että datakatalogiratkaisussa hallittaisiin myös muuta kuin TKI-projekteissa tuotettuja aineistoja, kuten vaikkapa opinnäytetöissä ja palvelutoiminnassa syntyneitä dataa, miksei jopa itse tuotettuja oppimateriaaleja. Näin hyödynnettynä datakatalogi tarjoaisi organisaation eri tasoille kattavan toiminnanohjauksellisen kokonaiskuvan eri aineistoista, mutta myös mahdollisuuden meritoitumiseen sekä yksilöille ja ryhmille. Jokainen datakatalogista julkaistu metatietue tulisikin pystyä linkittämään eri TKI-toiminnan perusyksiköihin kuten projektiin, tutkimusryhmään, infrastruktuuriin ja/tai henkilöön.

## Oman ja yhteisen tasapainoilua

Kansallisten aineistohallintaan liittyvien palveluiden maksimaalinen hyödyntäminen on perusteltua viimeistään silloin, jos ne vastaavat suoraan oman organisaation tarpeisiin. Korkeakoulun näkyvyyden ja uusien yhteistyöavausten syntyminen kannalta on olennaista, että näihin avoimesti saavutettaviin palveluihin pystyttäisiin viemään metadataa mahdollisimman kattavasti. Jokaisella organisaatiolla lienee kuitenkin tarve hallita myös sellaista dataa, jota ei ole välttämättä mitenkään perusteltua avata laajemmin – mutta tälle datalle ei ole järkevää luoda omaa työkalua.

Oli työkalu mikä tahansa, sen loppukäyttäjien heterogeenisyys luo haasteensa metadatan tuotantoon. Haaste voidaan nähdä positiivisessakin valossa, sillä olisihan hienoa, jos TKI-aineistojen metadatan tuottaisi aidosti ammattikorkeakoulun koko yhteisö opiskelijoista tutkijoihin ja opettajista tukipalveluihin. Tämä toisaalta asettaa datakatalogin käytettävyydelle ja muille ominaisuuksille suuria odotuksia ja toiveita, joihin toistaiseksi kansalliset palvelut eivät ole pystyneet kattavasti vastaamaan. Näitä ovat esimerkiksi:

- organisaation muita tietojärjestelmiä muistuttava (kerta)kirjautumisjärjestelmä ja tutulta tuntuva käyttökokemus
- käyttäjäryhmien joustava määrittely
- käyttöoikeuksien jako vaivattomana tukipalveluna
- tukipalveluiden pääsy keskeneräisiin metatietoihin pääkäyttäjinä.

Qvaimen tietomalli on hyvä lähtökohta metadatan kerryttämisen tavoitetasoksi datakatalogissa ja kyseisen tietomallin mukailu helpottaisi metadatan julkaisemista niissä tapauksissa, joissa siihen päädytään. Qvain ei kuitenkaan ole järjestelmänä kaiken metadatanhallinnan tarpeen tyydyttävä yksittäisen organisaation lähtökohdista, sillä haasteiden eteen ajaututaan esimerkiksi eri käyttäjäryhmien vaatimuksien ja avaamattomaksi tarkoitetun metadatan suhteen.

Uudenlainen datakatalogiratkaisu, oli se sitten organisaatiokohtainen tai kansallinen, voisi lisätä joustavuutta sekä loppukäyttäjän että tukipalvelujen näkökulmasta ja muuttaa toimintakulttuuria dataa vaalivampaa suuntaan myös opiskelijoiden kohdalla.

Datakatalogi selkeyttäisi aineistonhallinnan kannalta opinnäytetyöprosessia, projektioppimisympäristöjen toimintaa ja palveluprosesseja sekä mahdollistaisi eri toimintojen tuottamien tausta-aineistojen jatkoehdyntämisen tulevissa opinnäytetöissä ja asiakasrajapinnassa. Siten opiskelijat pystyisivät tukeutumaan vahvemmin toistensa tekemään työhön, samalla kuin uudet toimintatavat luovat uudenlaista konkreettista tarvetta ja positiivista painetta oman työn tuloksista huolehtimiselle. Samalla koulutusalojen toimintaa ohjaavat henkilöt kykenisivät paremmin fokusoimaan opiskelijatyötä tutkimusryhmien pitkän aikavälin tarpeisiin ja tiettyihin ydinteemoihin seurantatutkimusten ja jopa aikasarjojen synnyttämisen kautta. Näin myös koulustoitinnalla sisällä aineistoihin liittyvät tavoitteet olisivat linjassa projektitoiminnan aineistointensivisempien lähtökohtien kanssa.

Mitä tulee laajemmin toiminnanohjaukseen, datakatalogin erilaiset näkymät ja raporttimahdollisuudet voisivat tarjota organisaatioiden eri tasoille uutta tietoa tehdyn työn sisällöistä sekä mahdollisuuden vaikuttaa merkittävästi siihen, mihin aineistoihin ammattikorkeakoulu voi saada omistajuuden. Tätä kautta datan jatkoehdyntämisen potentiaali kasvaa huomattavasti. Kuten todettua, laadukas organisaationsisäinen aineistonhallinta mahdollistaa myös tutkimuksellisesti merkittävien aikasarjojen kumuloitumisen ja liiketoiminnallisesti kiinnostavan aineiston jakamisen eteenpäin suoraan yrityksille. Monet kaupungit ovat jo aloittaneet avoimen datan jakamisen, mutta yrityssectorilla on vaikuttanut olevan joskus hieman haasteita löytää heille aidosti hyödyllistä materiaalia käytettävässä muodossa. Ammattikorkeakoulun ainutlaatuinen asema korkeakoulutuksen ja yritystoiminnan polttopisteessä voisi kuitenkin mahdollistaa tutkimusaineistojen kerryttämisen ja jakamisen entistä vaikuttavammalla tavalla. Vaikka katalogi olisikin tarkoitettu ensisijaisesti sisäiseen käyttöön, sieltä pystyttäisiin joustavasti avaamaan metadatan kansallisiin ja kansainvälisiin järjestelmiin. Pidemmälle rikastetut oman organisaation ulkopuolisia kiinnostavat metadatat tulisikin pystyä avaamaan nappia painamalla vaikkapa kansallisessa datapalvelussa pysyvällä tunnisteella varustettuna.

Oman organisaation synergiahyödyn huomioiminen aineistojen avaamisen mahdollistamisen yhteydessä on joka tapauksessa sijoitus tulevaisuuteen. Keskitetysti kertyvän aineiston ja metadatan avulla johdolle syntyisi toiminnanohjauksellisesti merkittävä kokonaiskuva siitä, mitä teemoja omassa organisaatiossa on mahdollista viedä tulevaisuudessa tehokkaammin eteenpäin. Sama pätee pienimuotoisemmin myös tutkimusryhmien tasolla. Koko yhteisön jäsenien tekemistä ja työn tuloksia olisi silloin helpompi liittää yhteen lintuperspektiivistä, mikä johtaisi entistä aineistointensivisempien ja vaikuttavampien teemakokonaisuuksien ja sitä kautta toimintakulttuurin kehittymiseen.

Edellä kuvattu datakatalogiratkaisun tavoitetila vaatii käytettävyydeltään laadukkaan sähköisen työkalun lisäksi organisaatiokohtaista tukipalvelua metadatanhallintaan. Itse järjestelmän teknisen ratkaisun luonne on Turun ammattikorkeakoulun ja varmasti monen muunkin

organisaation kannalta auki. Pääsemme metadatanhallinnan työkalujen, tukipalvelujen ja osaamisen kannalta eteenpäin sitten hitaasti tai nopeasti, on hyödyllistä tunnistaa metadatan mahdollisuudet yhtä lailla sekä oman toiminnan kehittämisen välineenä kuin datan avaamisen näkökulmasta. Pitkässä juoksussa datanhallinnassa vahvat ammattikorkeakoulut ovat myös avoimen tieteen kannalta hyödyllisempiä kuin yksittäistapauksina avatut aineistopakettit.

---

## Kirjoittajat



Anttoni Lehto

erityisasiantuntija, TKI

Turun ammattikorkeakoulu

[Kirjoittajan muut artikkelit](#) >



Eija Suikkanen

Vastaava tietoasiantuntija

Turun ammattikorkeakoulu

[Kirjoittajan muut artikkelit](#) >

## Artikkelin tiedot

Kirjoittaja: Anttoni Lehto, Eija Suikkanen

Número: 3/2021 Avoin tki & oppiminen

URN: <http://urn.fi/URN:NBN:fi-fe2021060634287>

## Lisenssit





Tämä teos on lisensoitu [Creative Commons Nimeä 4.0 Kansainvälinen -lisenssillä](https://creativecommons.org/licenses/by/4.0/).