

Miksi en avaisi? Dataosaaminen on korkeakoulun kulttuurimuutos

31.8.2023

Ammattikorkeakouluissa on tehty hyvää työtä aineistonhallinnan toimintamallien rakentamisessa ja vakiinnuttamisessa. Paljon on vielä kesken. Mitkä ovat hyviä lähtökohtia korkeakoulun dataosaamisen vahvistamiseen ja datan avaamiseen? Toukokuussa ammattikorkeakoulujen asiantuntijat pohtivat tulevaisuuden datamaisemaa amk-päivien työpajassa; tässä artikkelissa tiivistetään pajan tulokset.

Laureassa toukokuun alussa AMK-päivien yhteydessä järjestettiin innopaja, jonka teema oli *TKI: avointa vai suljettua*. Se oli suosittu työpaja, jossa ammattikorkeakoulujen ja sidosryhmien asiantuntijat ideoivat avoimen tutkimus-, kehittämis- ja innovaatiotoiminnan (TKI) kehittämistä neljän osa-alueen näkökulmasta: näkökulmat olivat 1) avoin julkaiseminen, 2) avoin oppiminen, 3) tutkimusympäristöt ja 4) tutkimusdatat ja -tulokset.

Työpajassa havaittiin, että datan avaaminen oli osallistujille jo varsin tuttu asia ja sen yhteiskunnalliset hyödyt, kuten innovaatioprosessin nopeutuminen ja läpinäkyvyys, tunnettiin hyvin. Tästä huolimatta ammattikorkeakouluissa avattuja datasettejä on kuitenkin vielä hyvin vähän tarjolla. Työpajan osallistujat keskustelivat dataosaamisen edistämisestä ammattikorkeakouluissa oman asiantuntijuutensa kautta.

Datan avaaminen integriteetin varmistaja

Keskustelua tutkimusten toistettavuuteen ja luotettavuuteen liittyvistä kysymyksistä on käyty viime vuosina aktiivisesti monella alalla. Kun aiempia tutkimuksia on yritetty toistaa, on huomattu, ettei se aina onnistu. On herännyt laajempi kysymys tutkimuksen luotettavuudesta ja etiikasta. Tutkimusdatan avaaminen artikkelin julkaisun yhteydessä on yksi mahdollinen ratkaisu nykyiseen replikaatiokriisiin. Tätä edellyttävät jo jotkut

tiedejulkaisut, ja tulevaisuudessa yhä usemmat. Data avulla myös artikkelin review-prosessissa pystytään paremmin arvioimaan lähetetyn käsikirjoituksen luotettavuutta.

Jos data avataan, periaatteessa kuka tahansa voi tarkastella, millä tavoin tuloksiin on tultu. Tutkimusdatan avaaminen toimii hyvän tieteellisen käytännön ja tutkimusetiikan pohjana. Avaaminen parantaa tutkimuksen läpinäkyvyyttä ja ehkäisee mahdollisuutta tiedevilppiin. Avaamisen pohjana on hyvä aineistonhallinta, jonka korkeakoulu mahdollistaa oman tutkimusinfransa ja palveluidensa kautta.

Datan avaamisen tilanne

Ammattikorkeakoulut ovat tehneet hyvää työtä aineistonhallinnan toimintamallien luomisessa viime vuosina. Ohjaavia dokumentteja ja palveluja on kehitetty yhdessä kunkin ammattikorkean omaan tarpeeseen. Edistyminen havaittiin Avoimen tieteen ja tutkimuksen viimeisimmässä vuoden 2022 seurannassa, jossa aineistojen ja infrastruktuurien osalta ammattikorkeakouluista 78 % sijoittui ylimmille arvioinnin tasoille neljä ja viisi, keskiarvon ollessa neljä (arviointi tasoilla 1–5). Huomattavaa kuitenkin on, ettei arviointiin vaikuttanut avattujen datasettien tai metadatan määrä vaan huomiota kiinnitettiin lähinnä tarjolla olevaan koulutukseen ja tukipalvelujen saatavuuteen.

Opetus- ja kulttuuriministeriön vuonna 2020 käyttöönotettu kansallinen tutkimustietovaranto Tiedejatutkimus.fi -palvelu nostaa esille ammattikorkeakoulujen tuloksia ja avattuja datasettejä. Palvelusta voidaan havaita, että ammattikorkeakouluilla on vielä varsin vähän avattuja aineistoja ja avattua datan metadataa olemassa.

Ammattikorkeakoulujen tuloksiksi on tähän mennessä kirjautunut yhteensä 77 aineistoa 24 ammattikorkeakoulusta eri vuosilta. Tilanne on parempi kuin vielä pari vuotta sitten, sillä suurin osa avatusta aineistosta on vuosilta 2022 tai 2023 (54/77 aineistoa). Huomattavaa on, että samaan lukuun tutkimustietovarannossa listautuvat niin avatut aineistot kuin avattu metadata, mikä tarkoittaa, että avattuja datasettejä on luvusta vain osa.

Onkin hyvä kysyä, jääkö aineistojen avaaminen tulevaisuudessakin vain strategiatasolle ja pysyvätkö Euroopan komission tiedepolitiikat vain julistuksina ilman jalkautusta. Toukokuisessa työpajassa esitettiin hyviä ajatuksia korkeakoulujen aineistonhallinnan edistämisestä.

Millaista on ammattikorkeakoulujen aineisto? Small data

Ammattikorkeakoulujen TKI-hankkeet ja rahoitukset ovat usein pieniä ja tutkimusdataa kerätään yksittäiseen konkreettiseen kehittämistehtävään. Tämä tuottaa sirpaleista dataa. Keskeinen kysymys tulevaisuudessa on, kuinka tehdä datan keruusta ammattimaista sillä tavoin, että data ylittäisi pienet yksittäiset hankkeet ja tuottaisi laajempia kokonaisuuksia sillisalaatin sijaan.

Toisaalta pientä ja sirpaleista dataa yhdistämällä voi tuottaa yllättäviä uusia avauksia. Tästä on erityisesti julkisen sektorin datan käytöstä useita esimerkkejä: kun avataan dataa kenen tahansa käytettäväksi,

mahdollistetaan uusien ideoiden ja tuotteiden kehittäminen.

Datan arvonmääritys tulee tärkeäksi tulevaisuudessa, kun datan avaaminen on käynnistynyt laajemmassa mittakaavassa. Kun tällä hetkellä dataa on avattu vielä varsin vähän, on tärkeintä edelleen edistää datan avaamisen käynnistämistä eikä rajata sitä, mitä kannattaa avata.

Avaaminen ei kuitenkaan saa olla ainoa datan tavoite. Tärkeää on tunnistaa, voisiko ammattikorkeakouluissa tuotettu data sisältää kaupallistamismahdollisuuksia tai muunlaista lisäarvoa esimerkiksi yritysten liiketoimintaan. Tällöin aineisto tulee suojata ja käyttää ensin itse, ja vasta sen jälkeen avata laajemmin yhteiskunnan käyttöön. Osa rahoitusinstrumenteista edellyttää avoimuutta, joten datan avaaminen saattaa olla rahoituksen edellytys.

Big data mahdollisuutena

Oman aineiston keräämisen sijasta ammattikorkeakoulujen kannattaa jatkossa laajemmin hyödyntää jo olemassa olevaa laaja dataa. Muun muassa Terveyden ja hyvinvoinnin laitos ja sairaalat keräävät paljon aineistoa, jota tutkijat tai vaikka opinnäytetöiden tekijät voivat käyttää.

Ammattikorkeakouluissa tehdään jo nyt jonkin verran tutkimusta isoista sosiaali- ja terveysalan rekisteriaineistoista sekä tilastoaineistoista, joiden pohjalta voidaan rakentaa pitkittäistutkimusta. Näiden aineistojen käyttö perustuu tutkimusetiikan ja tietosuojalainsäädännön noudattamiseen ja niiden käyttöön tarvitaan aina lupa; lupia myöntävät Findata ja Tilastokeskus. Findata on sosiaali- ja terveysalan tietolupaviranomainen, jonka toiminta perustuu toisiolakiin eli lakiin sosiaali- ja terveystietojen toissijaisesta käytöstä (2019/552).

Korkeakouluissa on myös kehitetty toimintamalleja oman pitkittäisdatan keräämiseen. Tässä keskeistä on rakentaa pitkäjänteisesti toimintamalleja ja tutkimuslinjoja. Mallin toiminta perustuu aineiston kerääjien ja käyttäjien sitoutumiseen. Näin datasta saadaan aidosti hyödyntämiskelpoista.

Opiskelijat ja yritykset dataosaajiksi

Dataosaaminen on työelämään perustaito eikä se koske vain hanketyötä: korkeakoulujen tärkeä tehtävä on kouluttaa opiskelijoita datanhallinnan ja tietosuojan perusasioiden osaajiksi. Opinnäytetyön aineistonhallintasuunnitelman teko on tässä konkreettinen oppimisprosessi. On arvioitu, että osaaminen valuu korkeakouluista yrityksiin noin viidessä vuodessa opiskelijoiden siirtyessä työelämään. Tämä tuo laajalla rintamalla uusia mahdollisuuksia työpaikoille.

TKI-hankkeiden integrointi opetukseen voi jatkossa tapahtua datan analyysivaiheessa. Opiskelijat tarkastelevat hankkeen tai yrityskumppaneiden dataa osana opintojaksoa ja tuottavat siitä uusia avauksia. Toisaalta avattua dataa voi jatkossa käyttää opinnäytetyössä niin, ettei opiskelijan tarvitse kerätä aineistoa itse. Opiskelija ei ole tutkija, joten opinnäytetyön tekeminen ja oppimisen tavoitteet tulee suhteuttaa tähän.

Ammattikorkeakouluilla on hyvä tilaisuus ottaa rooli työelämän toimijoiden kouluttamisessa

ammattikorkeakoulujen oman ja muiden toimijoiden aineistojen hyödyntämisessä. Lisäksi ammattikorkeakoulut voivat edistää erityisesti yritysten osaamista oman datansa avaamiseen järkevästi niin, että suojataan tuotekehitystä ja yrityssalaisuuksia, mutta samalla luodaan läpinäkyvyyttä.

Miksi en avaisi? Kulttuurimuutos

Yksi datan avaamisen este on se, että oma asenteemme avointa dataa kohtaan ei tue sen avaamista eikä käyttämistä. Tutkijat ja hanketoimijat mieltävät usein oman aineistonsa vähäpätöiseksi ja ajattelevat, ettei pieni aineisto tuota hyötyä tai meriittiä kenellekään. Huomaamatta jää, että avattu data voi osoittautua hyödylliseksi esimerkiksi opinnäytetyön tekijälle tai työelämälle.

Tarvitsemme uudenlaista ajattelua laajasti korkeakouluissa ja yhteiskunnassa. Tavoite on, että data avataan aina kun mahdollista. Jos dataa ei voi avata esimerkiksi yrityssalaisuuksien tai muun sensitiivisen luonteen takia, yleensä kuitenkin metadata voidaan avata. Pelkän metadatan avaaminenkin edistää datojen käyttöä ja tuottaa tuloksia ammattikorkeakoululle ja tutkijalle itselleen.

Toinen avaamisen este on se, että aineistonhallintaa ei ole hoidettu kunnolla. Aineistonhallinta tulee aloittaa hyvissä ajoin ennen datan keräämistä. Aineistonhallinnan suunnittelu, tutkittavien informointi, anonymisointi sekä sopimukset hankekumppaneiden kesken ovat perusta datan avaamiselle. Kun aineistonhallinta on kunnossa, data voidaan huoletta avata.

Aineistonhallinnan prosessi on koko korkeakoulun asia. Se tulee huomioida kaikissa toiminnoissa niin TKI-hankkeissa, sidosryhmien kanssa työskennellessä opinnäytetöissä, hallinnon kehittämisessä – yleensä kaikessa, missä syntyy dataa. Tätä mallia tulee laajentaa myös yhteistyötahoille ja ohjata heitä toimimaan vastuullisesti.

Datan avaamisen kulttuurin luomisessa keskeistä on tukeutua ja luottaa asiantuntijayhteisöjen voimaan. Kun yhteisöille annetaan mahdollisuus luoda omia toimintamalleja, syntyy omasta datasta ja tuloksesta viestimiselle aito konteksti. Näin rakennetaan asiantuntijaviestintää. Kun tutkijat ymmärtävät, että data on tutkimusryhmän arvokas tulos, sitä halutaan suojata, siitä halutaan julkaista ja sen avulla halutaan luoda liiketoimintaa.

Toimintamallien juurruttamisessa viestinnän merkitystä tulee nostaa. Metadatan ja kuvailun tekeminen luo aineistosta löydettävän. Kun tutkijat ja hanketoimijat alkavat itse viestiä tekemästään datasta, se tuottaa asiantuntijaviestintää toisille tutkijoille. Tässä data-agenttitoiminta tai muu vastaava toimintamalli, jossa korkeakoulun omat asiantuntijat neuvovat dataan liittyvissä asioissa, on avainasemassa.

Ammattikorkeakoulujen hyvät muutokset ovat yleensä perustuneet kunnolliseen pohjatyöhön ja varmoihin prosesseihin. Suuri laiva lähtee liikkeelle hitaasti, mutta kun se liikkuu, liike etenee vähitellen läpi yhteisön.

Minna Marjamaa
tietoasiantuntija
Laurea-ammattikorkeakoulu

Anne Kärki
yliopettaja
Satakunnan ammattikorkeakoulu

Maija Merimaa
suunnittelija
Laurea-ammattikorkeakoulu

Seliina Päällysaho
tutkimuspäällikkö
Seinäjoen ammattikorkeakoulu

Tämä artikkeli on kirjoitettu osana Seinäjoen ammattikorkeakoulussa käynnissä olevaa TKI-kumppanuusmallilla enemmän ja vaikuttavampaa TKI-toimintaa Etelä-Pohjanmaalle -hanketta, joka on Euroopan unionin osarahoittama. Hanketta toteutetaan ajalla 1.1.2023 – 30.7.2025.

Lähteet

Ala-Mantila, M., Hilska-Heinänen, K., Jauhiainen, I., Juppo, V., Kiviluoto, J., Koivula, H., … & Ylönen, I. (2022). Avoimen toimintakulttuurin palveluiden itsearviointityökalu. https://urn.fi/URN:NBN:fi:tuni-202208176465

Avoimen tieteen koordinaatio, Tieteellisten seurain valtuuskunta. (2022). Avoimen tieteen ja tutkimuksen seurantamalli – periaatteet ja toimintatavat. Vastuullisen tieteen julkaisusarja 7:2022. Tiedonjulkistamisen neuvottelukunta ja Tieteellisten seurain valtuuskunta. https://doi.org/10.23847/tsv.237

Lahtinen H., Päällysaho S. & Marjamaa M. (2022). AVOTT-seurannan tulokset erinomaisia: lähes 80 % amkeista tasoilla 4 tai 5 – entä jatkossa?. *Kreodi*, (4). http://urn.fi/URN:NBN:fi-fe2022121270881

Valtonen, J., & Hankonen, N. (2020). Mitä toistettavuuskriisistä pitäisi ajatella? Johdanto tutkijoiden haastatteluihin. *Psykologia*, *54*(6 (2019)), 433-459. http://psykologia.fi/uusin-numero/236-artikkelit-uusin-numero/keskustelua/842-mita-toistettavuuskriisista-pitaisi-ajatella-johdanto-tutkijoiden-haastatteluihin