

Suunnitelmallisuus on datanhallinnan kulmakivi

Datanhallinnassa suunnitelmallisuus, eettisyys ja elinkaariajattelu tukevat laadukasta ja vastuullista datan hallintaa. Tutkimusdatanhallinta on olennainen osa vastuullista ja suunnitelmallista tiedon tuottamista.

TEKSTI: ANNE KARHAPÄÄ

Datan käytön ja hallinnan parissa työskentelee yhä kasvava joukko ammattilaisia erilaisissa rooleissa ja erilaisilla nimikkeillä, kuten data-analyytikko, data steward, data scientist ja data engineer. Yksi uusista data-ammateista akateemisessa maailmassa on tutkimusdata-asiantuntija, jonka tehtävänä on tukea tutkijoita tutkimusaineistojen hallinnassa ja avaamisessa. Viime vuosina korkeakoulut ovat panostaneet tutkimusdatanhallinnan kehittämiseen ja rakentaneet palveluja sen tueksi. Taustalla vaikuttavat paitsi tutkimusrahoittajien vaatimukset datanhallinnasta ja datan avoimuudesta, myös ymmärrys siitä, että hyvin hallittu data on edellytys vastuulliselle ja vaikuttavalle tutkimukselle.

Yliopistossa työskentelevänä tutkimusdatanhallinnan asiantuntijana saan usein kurkistaa tutkimustulosten taustalle – sinne, missä vasta suunnitellaan datan, kuten mittausten, havaintojen ja haastattelujen, keräämistä. Tässä vaiheessa voin auttaa tutkijoita pohtimaan, miten datanhallinta kannattaa toteuttaa tutkimuksen eri vaiheissa. On olennaista varmistaa, että dataa kerätään ja käsitellään oikein. Erityisen paljon huomioitavaa on silloin, kun data sisältää henkilötietoja tai muuta sensitiivistä aineistoa. Tutkimusdatanhallinta ei ole vain tekninen tukitoimi, vaan olennainen osa vastuullista ja suunnitelmallista tiedon tuottamista.

>>



*Data on hyödyllistä vain silloin,
kun se on käyttökelpoista.*

Datanhallinnan suunnittelu – hyödyllistä myös muille kuin tutkijoille

Tutkimusdatanhallinnassa keskeistä on kokonaiskuvan luominen aineiston elinkaaresta. Tutkijoiden tulee tutkimuksen alussa laatia aineistohallintasuunnitelma (data management plan, DMP), joka auttaa hahmottamaan datan elinkaaren ja sen eri vaiheissa tarvittavat toimenpiteet. Vaikka suunnitelma saatetaan aluksi kokea ylimääräiseksi työksi, moni tutkija on todennut sen kerran tehtyään, että suunnitelma on erittäin hyödyllinen työkalu. Se auttaa ennakoimaan ja huomioimaan olennaiset asiat sekä vähentää riskejä, jotka voisivat hankaloittaa tai jopa estää tutkimuksen toteutumisen. Siten se sujuvoittaa tutkimuksen tekemistä.

Tutkimusdatan aineistohallintasuunnitelmalle on kansallisessa yhteistyössä laadittu mallipohja¹, joka ohjaa suunnittelua rakenteellaan ja kysymyksillään. Se toimii sekä tutkijan työkaluna että hyödyllisenä dokumenttina tukipalveluiden ja hallinnon kanssa asioidessa. Aineistohallintasuunnitelmalla on keskeinen rooli akateemisessa maailmassa tutkimusdatan kanssa, mutta ajattelen että siihen on tiivistetty paljon hyödyllisiä asioita, joita voi soveltaa myös muun kuin vain tutkimusdatan hallinnassa. Esittelen tässä tuota mallia hyödyntäen keskeisiä näkökulmia datanhallinnan.

Datanhallinnan kuusi keskeistä osa-aluetta

1. Aineiston kuvaus ja laadunhallinta

Suunnittelu on hyvä aloittaa sillä, että selkiytetään, millaista dataa käytetään tai tuotetaan, ja miten sen laatua hallitaan. Esimerkiksi erityyppisten aineistojen kuvaaminen taulukkomuodossa voi selkeyttää omaa ajattelua ja tarvittavien toimien suunnittelua. Kun datan kanssa työskennellään, siirretään järjestelmästä toiseen tai tiedostomuodosta toiseen, on olennaista varmistaa, ettei siihen päädy virheitä. Esimerkiksi versionhallinnan avulla voidaan seurata datan ja siihen liittyvien dokumenttien muutoksia.

2. Juridiset ja eettiset näkökulmat

Dataan liittyy aina juridisia ja eettisiä seikkoja, jotka on huomioitava. On pohdittava mitä dataa voi kerätä ja miten. Data voi sisältää esimerkiksi henkilötietoja, jolloin on varmistettava, että niitä käsitellään oikein ja tarvittavat seikat huomioiden. On tärkeää myös määritellä datan omistajuus, käyttöoikeudet ja vastuut sekä tehdä etukäteen tarvittavat sopimukset näihin liittyen.

3. Dokumentointi ja metadata

Data on hyödyllistä vain silloin, kun se on käyttökelpoista. Siinä ratkaisevassa roolissa on sitä kuvaileva tieto eli metadata, jonka tulisi kertoa kaikki olennainen datasta niin, että se on ensinnäkin

löydettävissä ja sitten myös käytettävissä. Se edellyttää sitä, että data on organisotu, nimetty ja kuvailtu selkeästi ja johdonmukaisesti. Selkeät käytännöt tiedostojen nimeämisessä, kansiorakenteessa ja versioinnissa helpottavat datan käyttöä, myös datan kerääjälle itselleen. Vaikka kuvittelisi myöhemmin muistavansa kaiken

olennaisen, niin ettei sitä tarvitse kirjoittaa ylös, on varmaa että kun datan pariin palaa pitkän tauon jälkeen, mystisesti nimetyistä tiedostoista on hankalaa päästä kartalle mikä onkaan oikea versio, tai jos Excel sarakkeissa ei kerrota mistä sarakkeen numeroissa on kyse, niin ne muuttuvat pelkäksi

satunnaisten numeroiden listaksi. Varsinkin jos datan kanssa työskentelee useita henkilöitä, on tärkeää, että luodaan yhteiset käytänteet ja ohjeet datan dokumentoimiseksi.

4. Säilytys ja varmuuskopiointi

On olennaista suunnitella missä dataa säilytetään ja miten sen varmuuskopiointista huolehditaan. Organisaatioiden tarjoamat palvelut mahdollistavat turvallisen säilytyksen ja pääsynhallinnan, niin että vain henkilöt, joilla on oikeus käsitellä dataa pääsevät siihen. Jos data sisältää erityisen sensitiivistä sisältöä, kuten terveystietoja, se voi edellyttää erillisiä suojaustoimia, jotka on syytä suunnitella etukäteen. Varmuuskopiointi on kriittistä. Laitteita rikkoutuu ja dataa tuhoutuu aina silloin tällöin. Silloin vain eri sijainnissa oleva varmuuskopiointi voi pelastaa datan ja sen tuottamiseen käytetyn työajan. On syytä olla tietoinen, onko käytettävässä säilytysratkaisussa automaattinen varmuuskopiointi vai tulisiko se tehdä itse manuaalisesti. Siksi on tärkeää, että säilytyspalvelut on kuvattu niin, että niiden käyttäjät ymmärtävät helposti tarvittavat tekniset tiedot.

5. Datan elinkaaren loppupää

On tärkeää kiinnittää huomiota myös datan elinkaaren loppupäähän – eli suunnitella, miten pitkään dataa säilytetään ja miten tarpeeton data poistetaan. Esimerkiksi henkilötietoja saa säilyttää vain niin kauan kuin se on perusteltua ja tarpeellista siihen tarkoitukseen, jota varten ne on kerätty. Datan käsittely ja säilytys vievät resursseja. Tuotamme erilaista dataa koko ajan kiihtyvää vauhtia, ja sen säilyttäminen kuluttaa energiaa. On hyvä pohtia, miksi dataa

säilytetään, milloin se on perusteltua, ja jos dataa säilytetään pidempään, miten huolehditaan sen käyttökelpoisuudesta. Kaikki data ei ole niin arvokasta, että sitä voitaisiin säilyttää pysyvästi.

Tutkimusdatan avoimuuden edistäminen on yksi keskeinen tavoite korkeakoulujen datanhallinnassa. Datan halutaan olevan mahdollisimman avoimesti myös muiden käytettävissä. Datan avaaminen muiden käyttöön ja avoimen datan hyödyntäminen voi säästää huomattavasti resursseja. Tässäkin suunnitelmallisuus on tärkeää: on osattava ennakoita, millaisia toimia tarvitaan, jotta dataa voidaan avata muiden käyttöön eettiset ja juridiset seikat huomioiden. Esimerkiksi tutkittavia on tie-

Dataa pitää jalostaa, että siitä saadaan lisäarvoa.

Tutkimusdatanhallinnassa keskeistä on kokonaiskuvan luominen aineiston elinkaaresta.

On olennaista varmistaa, että dataa kerätään ja käsitellään oikein.

dotettava jo aineistoa kerätessä, jos se aiotaan antaa avoimesti jatkokäyttöön. Jos datalla on arvoa uudelleen käytön kannalta, tulee huolehtia, että se myös säilyy pitkään. Esimerkiksi tiedostomuodon valinnalla voi vaikuttaa siihen saako tiedoston auki, vaikka järjestelmät muuttuisivat.

6. Roolit ja vastuut

Datanhallinnan roolit ja vastuut on hyvä selkiyttää jo alussa. On myös tärkeää arvioida tarvittavat resurssit, kuten työaika, välineet ja palvelut, kuten tarvitaanko esimerkiksi erityisen paljon työaikaa datan käsittelyyn tai hankkia jotakin tiettyjä välineitä tai palveluja.

Tutkimusdatanhallinnasta arvokasta tietoa tietalojen ammattilaisille

Sanonta ”data on uusi öljy” on jo aikansa elänyt, mutta sitä voi tarkastella myös siltä kannalta, että mahdollisuuksien lisäksi dataan liittyy riskejä, jotka kenties realisoituvat vasta paljon myöhemmin. Hallitsemattomasti kertyvä data voi aiheuttaa tietotulvaa, vaikeuttaen olennaisen tiedon löytämistä. Lisäksi se kasvattaa tietoturvariskejä, erityisesti silloin, jos arkaluontoista tietoa säilytetään huolimattomasti. Organisaatioiden toiminnassa syntyy paljon dataa. Jos se hajaantuu eri paikkoihin hallitsemattomasti,

Tutkimusdatan avoimuuden edistäminen on yksi keskeinen tavoite korkeakoulujen datanhallinnassa.

sitä ei voida hyödyntää. Dataa pitää jalostaa, että siitä saadaan lisäarvoa esimerkiksi päätöksenteon tueksi. Datan käsittelyssä on kuitenkin tärkeää huomioida eettisyys ja vastuullisuus: on arvioitava, mitä tietoa

voidaan kerätä, miten sitä voidaan käyttää ja mitä datanhallinnan toimia tarvitaan.

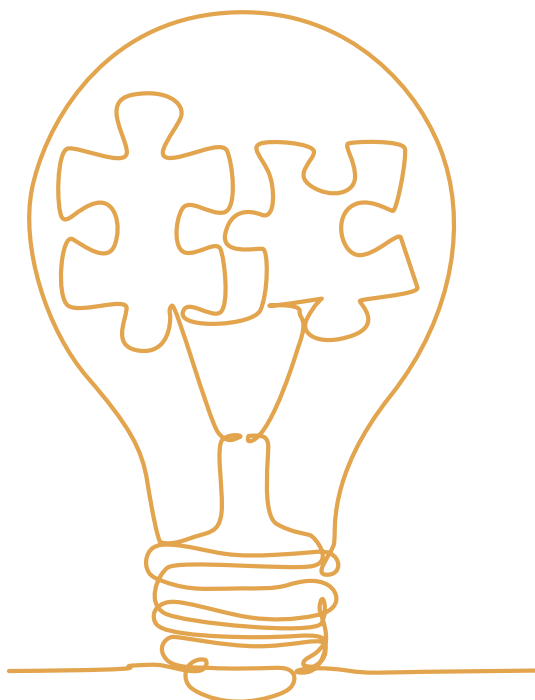
Tutkimusdatanhallinta tarjoaa arvokkaita oppeja kaikille tietalojen ammattilaisille. Datanhallinnassa suunnitelmallisuus, eettisyys, ja elinkaariajattelu tukevat laadukasta ja vastuullista datan hallintaa – riippumatta siitä, missä kontekstissa dataa tuotetaan ja käytetään. ■

Kirjoittaja työskentelee tietoasiantuntijana Itä-Suomen yliopiston kirjastossa ja työskentelee tutkimusaineistohallinnan parissa yliopiston Datatuessa. Hän on myös aikuiskasvatustieteen väitöskirjatutkija Jyväskylän yliopistossa.

1 <https://zenodo.org/records/5242629>



Hallitsemattomasti kertyvä data voi aiheuttaa tietotulvaa, vaikeuttaen olennaisen tiedon löytämistä.



Tekoälyn ja median tutkimukseen lahjoitettiin professuuri

Helsingin Sanomain Säätiö, Media-alan tutkimussäätiö ja C. V. Åkerlundin mediasäätiö lahjoittavat professuurin yhteisrahoituksella tekoälyn ja median tutkimuksen ja opetuksen vahvistamiseksi. Tavoitteena on vastata nopeasti niihin haasteisiin ja mahdollisuuksiin, joita generatiivinen tekoäly tuo mukanaan journalismiin, mediakulttuuriin ja demokratiaan. Tampereen yliopiston ITC-tiedekunnassa yhdistyy teknologinen ja viestintätieteellinen osaaminen, mikä tarjoaa professuurille ainutlaatuisen ympäristön.

Lahjoitusprofessuuri on kolmivuotinen, minkä jälkeen rahoitus siirtyy pysyväksi tehtäväksi Tampereen yliopiston viestintätieteiden yksikköön.

Lähde: Helsingin Sanomain Säätiö, tiedote 3.10.2025