

Tutkimusaineiston arkistoinnista ja arvonmäärittämisestä

Arkistoinnilla tarkoitetaan tutkimusaineiston luovuttamista arkistoon, jolla on Kansallisarkiston antama seulontapäätös. Arkistoinnissa vastuu aineistosta siirtyy arkistolle ja arkistointi mahdollistaa myös henkilötietoja sisältävän tutkimusaineiston säilymisen ja uusiokäytön alkuperäisen tutkimustarkoituksen päätyttyä. Tämä arkistointi tapahtuu Jyväskylän yliopistossa yleisen edun perusteella. Sen mukainen arkistointitarkoitus osoitetaan Kansallisarkiston arvonmäärittämis- ja seulontapolitiikan mukaan tehdyillä seulontapäätöksillä. Jyväskylän yliopistolla on Kansallisarkiston seulontapäätös¹, jonka mukaan kaikkien merkittävien tutkimusten tausta-aineistot tulee säilyttää pysyvästi. Näitä tausta-aineistoja, joita arkistolaki ja tietosuojasääntely koskevat ovat esim. lähdekeräelmät, havainto-, mittaus-, koe- ja haastatteluaineistot, laboratoriopäiväkirjat ja muistiinpanot sekä tutkimusaineiston dokumentaatio.

Näin ollen, kun tutkimusaineisto halutaan arkistoida, on sille tehtävä arvonmäärittäminen. Siinä määritetään, mitkä tutkimusaineistot arkistoidaan (eli säilytetään pysyvästi) ja mitä aineistoista säilytetään määräajan. Arkistoinnin ohella tutkimusaineistojen kohdalla on mahdollista julkaista (henkilötietoja sisältämätön) aineisto avoimesti saataville sen vaikuttavuuden lisäämiseksi. Aineiston julkaisemisen soveltuvassa tietoarkistossa voi tehdä riippumatta siitä, arkistoidaanko aineisto vai ei.

Mahdollisuuksien mukaan alustava arvonmäärittäminen tai ennakkoseulonta on tehtävä jo aineistonhallinnan suunnitteluvaiheessa, jotta aineistoa kootessa osataan huomioida arkistointitarkoituksesta syntyvät informointivelvoitteet sekä metatietojen että dokumentaation laatiminen ynnä arkistointiin soveltuvien tiedostomuotojen valinta. Tässä yhteydessä on myös tärkeää arvioida, koskeeko arkistointi aineistoa henkilötietoineen, vai voiko aineiston anonymisoida. Arkistolaki lähtee siitä, että arkistoidaessa tutkimukseen tai muuhun kulttuuriperintöarvoon perustuen, koskee arkistointivelvoite myös näihin asiakirjoihin sisältyviä henkilötietoja.

Lopullinen arvonmäärittäminen tehdään, kun aineisto halutaan tarjota luovutettavaksi Jyväskylän yliopiston arkistoon. Aineistosta vastaava tutkija [hankkeen tapauksessa vastuullinen johtaja/principal investigator (PI)] tekee esityksen arvonmäärittämisestä näiden ohjeiden perusteella, jonka jälkeen Avoimen tiedon keskus ottaa kantaa arvonmäärittämiseen ja tekee seulontapäätöksen.

Arvonmäärittämisestä neljä ulottuvuutta²

Arvonmäärittämisessä tulee suorittaa Kansallisarkiston arvonmäärittämisohjeiden mukaan 1) tehtäväanalyysi, 2) informaatioarvon analyysi, 3) säilytysmuodon analyysi ja 4) kustannustekijöiden analyysi. Näitä eri ulottuvuuksia tulee arvioida yhdessä ja suhteessa toisiinsa. Näiden lisäksi on arvioitava säilyttämisen esteitä.

Tehtäväanalyysi

Tehtäväanalyysissä arvioidaan sen tehtävän merkittävyyttä, jossa arkistoitava aineisto on syntynyt. Tutkimus sinänsä on lähtökohtaisesti merkittävä tehtävä ja yliopiston substanssitehtävään (tutkimus) suoraan liittyvänä tutkimusaineisto on arvokasta.

¹ Arkistolaitoksen päätös Jyväskylän yliopiston asiakirjojen pysyvästä säilytyksestä, 14.4.2010, AL/8216/07.01.01.03.01/2008.

² Kansallisarkisto, Arvonmäärittämis- ja seulontapolitiikka. Versio 1.6. 16.12.2020.

Tutkimuksen merkittävyyttä suhteessa muihin tutkimuksiin joudutaan arvioimaan erityisesti, mikäli muissa analyysialueissa nousee esiin arkistointia vastustavia tekijöitä. Tällöin arvioidaan, onko tutkimusaineistolla jo sellaisenaan tiedehistoriallista arvoa ja tähän liittyen, kuinka merkittäviin tutkimuksiin ja tutkimuskysymyksiin aineistoa on käytetty. Kysymyksiä voivat olla esimerkiksi, onko tutkimuksessa käytetty menetelmä tai teoria tai sen tuottama tieto kuinka merkittävässä roolissa tieteellisessä keskustelussa? Aineistolla voi olla myös historiallista ja kulttuurista arvoa sinänsä. Samoin voidaan arvioida, kuinka suuri aineiston tarjoama otanta on tai onko se osa jotain suurempaa aikasarjaa. Merkittävää on myös, valaiseeko aineisto poikkeavaa yhteiskunnallisen kehityksen muutosta, murrosaikaa tai käännekohtaa tai dokumentoi merkittävää ilmiötä tai tapahtumaa. Yksi tehtävän merkittävyyttä lisäävä tekijä on myös, mikäli aineisto tukee yleisemmin yliopiston tehtävää. Nämä tekijät voivat puoltaa aineiston arkistointia myös silloin, kun muista analyyseistä ei tätä tukea löydy.

Informaatioarvon analyysi

Informaationarvon analyysissa tarkastellaan, onko tutkimusaineiston tieto kuinka ainutkertaista eli sisältyykö se jo muualla saatavilla olevaan tai arkistoituun aineistoon. Ainutkertaisuuden suhteen on syytä myös arvioida, onko aineistoa mahdollista tuottaa uudelleen vai onko käsillä sen ainoa mahdollinen ilmentymä. Ainutlaatuista tilannetta tietyllä ajanhetkellä kuvaava aineisto on lähtökohtaisesti ainutkertaista verrattuna koeasetelmiin ja mittauksiin, jotka voidaan sellaisenaan toistaa ja koota sama aineisto uudelleen.

Ainutkertaisuuden lisäksi informaatioarvoa tuo myös tietojen kattavuus alueellisesti tai kohteensa puolesta. Tähän liittyy myös kysymys siitä, onko data jo kokonaisuudessaan analysoitu.

Tässä kyse on myös datan laadusta. Tällöin arvioidaan, onko tieto helposti käytettävää ja yhdisteltävissä muihin tietoihin. Laadukkaan datan tunnusmerkkejä on myös, että se on oikeaa (vastaa todellisuutta), on tarkkaa (kuva kohdetta) ja on johdonmukaista (yhdenmukaista ja itsessään johdonmukaista). Tähän vaikuttaa muun muassa tiedon rakenne, muoto sekä ennen kaikkea aineiston dokumentaatio – eli onko aineiston mukana riittävä kuvaus sen sisällöstä, jotta toinen tieteen tekijä kykenee aineistoa käyttämään ymmärtäen sen rajoitteet ja mahdollisuudet sekä sen, miten aineisto on kerätty/tuotettu. Arvokas data on siis aina kattavasti kuvailtu, jolloin data ja sen muutokset voidaan jäljittää ja datan alkuperä on tunnettua. Mikäli alalla on standardeja aineiston tai sen metadatan esittämiseen, niitä seurataan.

Informaatioarvoltaan arvokkaan aineiston yleinen löydettävyyteen liittyvä metadata on hyvin kuvailtu ja auttaa ymmärtämään ja löytämään datan. Arkistoinnin tapauksessa tämän metadatan laadun varmistavat yliopiston asiantuntijat, joten sitä erikseen arvonmäärittämisessä arvioida.

Keskeinen osa informaatioarvon analyysia on datan laadun lisäksi sen käyttötärpeiden arviointi. Tällöin tulee arvioida, soveltuuko aineisto muuhun kuin juuri siihen tutkimukseen, jossa se on tuotettu. Tärkeää on osata silloin ajatella myös omaa alaa laeemmin. Käyttötärpeet voivat tulla yllättävistäkin suunnista tai syntyä pitkän ajan kuluessa. Tutkimuksen lisäksi aineistolle voi olla myös muita ulkopuolisia käyttötärpeitä (yhteiskunnan, talouden, ympäristön ja kansalaisten tiedon, sivistyksen ja itseymmärryksen tarpeet), jotka on syytä arviota tehdessä huomioida. Tutkimusdata voi akateemisen impaktinsa lisäksi tuottaa yhteiskunnallista tai taloudellista hyötyä.

Uudelleenkäytettävyyys liittyy myös aineistoon liittyvien oikeuksien arviointi. Onko uudelleenkäyttö juridisesti mahdollista tai onko sille muita sopimuksellisia esteitä.

Tietomuodon analyysi

Tietomuodon analyysissä arvioidaan, onko tieto sellaisessa muodossa, että se on hyödynnettävissä. Tällöin keskitytään digitaalisten aineistojen kohdalla arvioimaan, ovatko käytetyt tiedostomuodot sellaisia, että ne kestävät aikaa. Toisin sanoen onko niitä oletettavasti mahdollista avata vielä kymmenien vuosien kuluttua.³

Tietomuoto liittyy myös edellä arvioituun uudelleenkäytettävyyteen. Onko aineistossa käytetty sellaisia tiedostomuotoja, että niitä voi ylipäättään avata yleisesti saatavilla olevilla ohjelmistoilla. Tällöin avoimen lähdekoodin ohjelmistoille annetaan arvioinnissa suurempi painoarvo. Erityisen kalliit tai vaikeasti saatavilla olevat ohjelmistot taas voivat muodostaa riskin tietomuodon käytettävyydellä ja siten arkistoinnille.

Mikäli aineistolla on myös analoginen ilmentymä, on syytä arvioida, onko sillä säilytys- ja käyttöarvoa digitaaliseen ilmentymään verrattuna. Mikäli aineisto on ainoastaan analogisessa muodossa, on arvioitava, onko se muodossa, jossa sen käyttäminen muihin tarkoituksiin ja siten arkistointi on mielekästä ja mahdollista.

Kustannustekijöiden analyysi

Kustannustekijöiden analyysissä arvioidaan ensisijaisesti sitä, kuinka kallista arkistointi on. Yleisesti sanoen mitä kooltaan suurempi aineisto, sitä kalliimpaa sen säilyttäminen on. Tässä keskeisin mittari on digitaalisten aineistojen kohdalla aineiston koko tavuina. Kooltaan megatavuihin tai muutamiin gigatavuihin jäävän aineiston säilyttäminen on sangen triviaalia, mutta satojen gigojen tai teratavujen kokoisten aineistojen kohdalla tämä vaikuttaa jo arvonmäärittäykseen.

Suurten aineistojen tapauksessa on erityisen tärkeää peilata kustannuksia aineiston ainutkertaisuuteen. Mikäli aineisto on mahdollista tuottaa uudelleen, arvioidaan mikä on sen kustannus verrattuna arkistoinnin kustannuksiin.

Oma arviointikohteensa voi olla myös aineiston tuottamisen hinta sellaisenaan. Kalliilla tuotettu aineisto on tärkeää arkistoida, mikäli muut arkistoinnin edellytykset antavat lainkaan osviittaa arkistoinnin puoltamisesta.

Säilyttämisen ehdottomat esteet

Arvonmäärittäyksen yhteydessä on tutkittava myös mahdollisia säilyttämisen esteitä. Henkilötietoja sisältävän aineiston tapauksessa tulee tutkia, onko tutkittavien informoinnit hoidettu niin, että ne mahdollistavat arkistoinnin (toisin sanoen tutkittavia on tiedotettu, että aineisto tullaan arkistoimaan). Mikäli näin ei olla toimittu, voidaan vielä tarkastella mahdollisuutta uudelleeninformointiin. Tähän arviointiin vaikuttaa myös, onko tiedot kerätty alun perin rekisteröidyltä itseltään vai muualta. Jos tietoja ei ole kerätty rekisteröidyltä, on laki joustavampi.

Esteiden yhteydessä on myös, ovatko aineiston oikeudet ja omistajuus niin selvät, että se voidaan arkistoida. Toisin sanoen onko arkistointia esittävällä taholla riittävä päätösvalta aineistoon sen arkistoinnista. Tämä on erityisen tärkeää useiden osapuolten hankkeissa, joissa yksi osapuoli ei yleensä voi omin päin päättää arkistoinnista ja arkistoon luovuttamisesta.

Aineistoa kootessa on voitu myös solmia sopimuksia, jotka saattavat kieltää arkistoinnin. On saatettu sopia tietojen tuhoamisesta heti tutkimushankkeen päätyttyä. Sopimuksissa on mahdollisesti myös kielletty käyttö muihin tarkoituksiin, joka puolestaan on huomioitava informaatioarvon analyysissä. Samoin esimerkiksi erilaisten sovellusten käyttöehdot (esimerkiksi sosiaalisen median osalta)

³ Säilytys- ja siirtokelpoiset tiedostomuodot, <https://digitalpreservation.fi/specifications/fileformats>

saattavat estää arkistoinnin. Mikäli aineiston käyttö on syystä tai toisesta rajattua myös tulevaisuudessa, tulee aineiston yhteydessä olla selkeä selvitys siitä, mihin perustuen käyttöä on rajoitettu ja mikä käyttö on mahdollista.

Henkilötietojen arkistoinnista⁴

Vaikka arkistointi yleisen edun nimissä mahdollistaa tutkimus aineiston arkistoinnin tunnistetiedoin, ei tämä ole missään nimessä automaatio. Tietojen minimointiperiaate on huomioitava arkistointiaessakin. Vaikka jollakin henkilötietoja sisältävällä tietoaaineistolla katsottaisiin olevan arvomäärityskriteerien perusteella arkistollista arvoa, ei se välttämättä aina tarkoita, että näiden tietojen käsitteleminen olisi myös ”ehdottoman välttämätöntä” arkistointikäsitteilyllä tavoiteltujen päämäärien saavuttamiseksi.

Näin ollen henkilötietojen minimoinnin periaatteella on erityinen merkitys arvomäärityksen kannalta, koska arvomäärityksellä osoitetaan oikeutus yleisen edun mukaiselle arkistointitarkoitukselle. Vaikka arvomäärityskriteerit ovat yleensä sisällöllisiä kriteerejä, on tietojen minimoinnin periaate taas muodollinen vaatimus, johon pitää pystyä vastaamaan. Tietosuoja-asetuksen sääntelyn näkökulmasta kyse on ennen kaikkea sen arvioimisesta, onko kyseisillä tiedoilla sellaista pysyvää yleistä etua koskevaa merkitystä, joka mahdollistaa niiden käsittelemisen arkistointitarkoituksessa. Tämä ei koske ainoastaan henkilötietojen arkistointia yleisesti vaan aineiston henkilötietojen laajuuden arviointia siten, että osa henkilötiedoista saatetaan todeta asianmukaisiksi, olennaisiksi ja tarpeellisiksi suhteessa arkistointitarkoituksiin ja osa taas ei.

Arvomäärityksen yhteydessä on siis arvioitava, ovatko tutkimusaineiston sisältämät henkilötiedot aineiston kulttuurihistoriallisen ja tutkimuksellisen arvon kannalta niin merkittävässä roolissa, että aineisto on perusteltua arkistoida tunnistetiedoin. Kun henkilötietojen hävittäminen alkuperäisen käsittelyn päätyttyä on pääsääntö, tulee niiden säilyttämiselle pystyä esittämään erityisiä perusteluja. Tässä arvomääritysprosessin ulottuvuuksien kautta arvon osoittaminen on merkittävässä roolissa, mutta niitä on peilattava erikseen myös aineiston henkilötietoja sisältävään osaan ja henkilötietojen luonteeseen ja merkitykseen aineistokokonaisuudessa. Näin arvomäärityksellä voidaan osoittaa, että arkistointipäätös kytkeytyy henkilötietojen käsittelyn oikeasuhtaisuuden ja tarpeellisuuden vaatimuksiin.

Henkilötietoja sisältävää aineistoa arkistointiaessa on syytä siis määrittää, ovatko henkilötiedot arvokkaita ja onko ne minimoitu niin pitkälle kuin pystytään. On siis arvioitava, kuinka helposti yksilöiden on oltava tunnistettavissa, jotta aineistolla on vielä tutkimuksellista arvoa – ja toki arvioitava, tuoko tunnistettavuus jossain muussa tutkimuskontekstissa merkittävää lisäarvoa. Tärkeää onkin huomioida, ettei minimointi saa vaarantaa tietoaaineistojen autenttisuutta, luotettavuutta, eheyttä ja käytettävyyttä. On myös arvioitava, rajoittaako minimointi merkittävästi ja pysyvästi aineiston tulevaa tutkimuskäyttöä tai sitä, millaista tutkimusta sen perusteella voidaan tehdä. Erityisesti lääketieteellisessä ja yhteiskuntatieteellisessä tutkimuksessa tunnisteiden perusteella tapahtuva tietoaaineistojen yhdistely on erityisen merkittävässä roolissa.

Käytännössä on siis arvioitava, voidaanko tutkimustarkoitus tulevaisuudessa toteuttaa anonymisoidulla tai muilla tavoin minimoidulla aineistolla. Vaaditun minimoinnin voidaan katsoa tapahtuvan myös arkistoinnilla aineisto esimerkiksi pseudonyyminä tai vain epäsuorin tunnistein, jolloin arkiston on myös jatkokäytön yhteydessä helppo huolehtia siitä, että uusiokäytössä tutkijalle

⁴ Lönnblad, Riikka, ”EU:n yleisen tietosuoja-asetuksen vaikutuksen arkistoinnin edellytyksiin”. 2025. https://kansallisarkisto.fi/documents/141232930/153230445/GDPR-arkistointi_selvitys_2025_final.pdf , viitattu 15.5.

välitetään vain sellaisia tietoja, jotka ovat tutkimuksen kannalta tarpeellisia ja esimerkiksi suora koodisto voidaan tarvittaessa jättää välittämättä. Näin ollen jo tehtyjä minimointitoimia voidaan hyödyntää arkistoitaessa ja lukea arvonmäärittämisessä aineiston eduksi.

Henkilötietojen minimointia ja arkistointia arvonmäärittämisestä kautta arvioitaessa on muistettava, että sitä tehdään erityisesti tutkimusaineistojen osalta tutkimuskontekstissa tai tiedehistoriallisen arvon kautta. Kantaa ei siis tarvitse välttämättä ottaa kulttuurisiin, yhteiskunnallisiin, hallinnollisiin tai oikeudellisiin tiedontarpeisiin tulevaisuudessa, vaan voidaan keskittyä arvioimaan sitä tieteen kontekstissa. Tässä arvonmäärittämisessä riittää, että arkistoitavien tietoaineistojen arkistollinen arvo (tutkimuksen tai tiedehistorian kautta) perustellaan riittävän huolellisesti ja yksityiskohtaisesti tunnistuen asianmukaiset, olennaiset ja tarpeelliset arkistoitavat tiedot ja rajaavan muut tiedot arkistoinnin ulkopuolelle. Mikäli kulttuurisiin tai yhteiskunnallisiin arvoihin (erityisesti tiedot, joilla on historiallista, kulttuurista, tutkimuksellista tai kulttuuriperintöön liittyvää merkitystä) kuitenkin osataan ja voidaan ottaa kantaa, ovat ne ehdottomasti ansioksi arvonmäärittämisestä tehtäessä ja vahvistavat ”pysyvän yleisen edun” perustelua.

Tässä yhteydessä yleisesti tietojen käsittelyn oikeasuhtaisuus toteutuu, kun alkuperäiset tiedotkin on kerätty tutkimustarkoitukseen, eikä niiden käyttö uusiin tutkimuksiin merkittävästi loukkaa yksilön oikeuksia. Henkilötietojen kohdalla oikeasuhtaisuutta arvioitaessa on kuitenkin pohdittava myös tietojen sisältöä. Jos tiedot ovat erityisen arkaluonteisia tai mahdollistavat rekisteröityä koskevan yksityiskohtaisen profiilin luomisen tai yksityiskohtaisten päätelmien tekemisen henkilöiden yksityiselämästä, voi tällaisten tietojen arkistointi olla suhteeton puuttuminen yksityisyyden suojaan. Tällaisissa tilanteissa voidaan oikeasuhtaisuutta parantaa myös asettamalla aineistoon pääsulle tavallista ankarampia rajoituksia.

Kokonaisarvioinnista

Arvonmäärittämisestä tehtäessä kyse on aina lopulta kokonaisarvioinnista. On tietysti mahdollista saada jokaisesta arvonmäärittämisessä ulottuvuudesta niin selkeä tulos, että arvio on lähtökohtaisesti positiivinen, mutta usein eri ulottuvuuksien arviot voivat vetää eri suuntiin.

Lähtökohtaisesti tärkeintä on todeta, että aineistolla on tiedehistoriallista ja/tai tutkimuksellista merkittävyyttä jo tehdyn alkuperäisen tutkimuskäytön ulkopuolella. Tätä merkittävyyttä voidaan joutua kuitenkin puntaroimaan tarkemmin, mikäli aineisto sisältää tunnisteellisia erityisen arkaluonteisia henkilötietoja, sen käytettävyydessä on haasteita tai sen arkistointi on poikkeuksellisen kallista.

Osa ulottuvuuksista on sangen kategorisesti arvioitavia, joten mikäli aineiston arkistointi on oikeuskysymysten takia mahdotonta tai aineisto on kertakaikkisen sekavasti ja dokumentoimattomasti järjestetty, vaarantuu aineiston tutkimuksellinen arvo siinä määrin, ettei arkistointia voida puoltaa. Muissa tapauksissa eri ulottuvuuksia on lähdettävä arvioimaan hienosyisemmin.

Henkilötietojen kannalta tärkein kysymys on, voidaanko tieteelliset tai historialliset tutkimustarkoitukset täyttää siten, ettei ole enää mahdollista tunnistaa rekisteröityjä. Jos on, nämä tarkoitukset on täytettävä tällä tavoin. Toisaalta aina on arvioitava aineiston tulevaa käyttöä ja mahdollisuutta yhdistellä sitä muihin aineistoihin. Aineistojen täysi anonymisointi ei välttämättä ole mahdollista, jolloin tulee arvioida muita minimoinnin keinoja, erityisesti pseudonymisointia ja karkeistamista – sekä niiden vaikutusta aineiston tulevaan käyttöön. Jos henkilötiedot sisältävät erityisiä henkilötietoryhmiä, eikä tietojen anonymisointi ole mahdollista, on tämän harkinnan oltava huolellisempaa ja ylitettävä rima korkeampi.