

Helsingin yliopiston tutkimusdatapolitiikka 2021

Sisältö

1. Johdanto

- 1.1. Mitä tutkimusdatapolitiikka koskee
- 1.2. Tutkimusaineistojen hallintaan liittyvät muut linjaukset Helsingin yliopistossa
- 1.3. Tutkimussopimukset ja oikeuksista sopiminen
- 1.4. Tutkimusdatapolitiikan seuranta ja päivittäminen

2. Tutkimusaineistojen hallinnan periaatteet ja tavoitteet

- 2.1. Yleiset tavoitteet
- 2.2. Aineistonhallinnan suunnittelu ja valmistelu
- 2.3. Tutkimusaineiston dokumentointi ja käsittely
- 2.4. Tutkimusaineiston julkaiseminen ja saavutettavuus
- 2.5. Tutkimusaineistojen kaupallistaminen
- 2.6. Tutkimusaineiston tuhoaminen tai pitkäaikaissäilytys

3. Tutkimusaineistojen hallintaan liittyvät vastuut

- 3.1. Tutkijan vastuulla on
- 3.2. Tiedekunnan tai tutkimusta harjoittavan erillislaitoksen vastuulla on
- 3.3. Yliopiston vastuulla on

4. Sanasto

1. Johdanto

Tutkimusaineistoilla on keskeinen merkitys tieteelle. Aineistojen tuottaminen vaatii pitkäaikaista työtä sekä teknisiä ja taloudellisia resursseja. Tutkimusaineistojen uudelleenkäytettävyydestä onkin muodostunut entistä tärkeämpi kysymys sekä tieteen että tutkimuksen vaikuttavuuden kannalta.

Helsingin yliopistolle, sen yksiköille ja tutkijoille tutkimusaineistot ovat myös strateginen resurssi ja kansainvälinen kilpailutekijä. Yliopiston tavoitteena on edistää tutkimusaineistojen vastuullista hallintaa, millä on ratkaiseva merkitys tutkimukseen perustuvan tiedon löydettävyyden, saavutettavuuden ja jatkokäytön kannalta. (*Tutkimusaineiston ja tutkimusaineistojen hallinnan määritelmät, ks. sanasto.*)

Avoimuuden periaatteet ja avoimet tutkimusaineistot ovat tärkeässä roolissa Helsingin yliopiston strategiassa 2021–2030. Yliopiston tavoitteena on tutkimusinfrastruktuurien ja tutkimusaineistojen avoin saatavuus, FAIR-periaatteiden toteuttaminen tutkimusaineistojen hallinnassa, sekä suuriin ja avoimiin aineistoihin liittyvän osaamisen kehittäminen.

Tutkimusdatapolitiikan keskeisenä tavoitteena on ohjata kaikkia tutkimusaineistojen kanssa toimivia perehtymään ajantasaisiin aineistohallinnan vaatimuksiin ja ottamaan hyvät ja vastuulliset käytännöt osaksi tutkimuksen arkea. Tavoitteena on myös tehdä tutkimusaineistojen hallinnasta yksittäisen tutkijan kannalta entistä selkeämpää.

Organisaation näkökulmasta tutkimusdatapolitiikka kuvaa ne tavoitteet, joiden pohjalta Helsingin yliopiston datanhallinnan palveluita kehitetään. Tavoitteena on, että Helsingin yliopiston tutkijoilla on käytettävissään vastuullisen tutkimusaineistojen hallinnan mahdollistavat infrastruktuurit ja palvelut, joita kehitetään tutkijoiden tarpeet huomioiden taloudellisesti kestäväällä tavalla.

Helsingin yliopiston uusi tutkimusdatapolitiikka ei perustavoitteidensa osalta merkittävästi poikkea aikaisemmasta, vuonna 2015 julkaistusta datapolitiikasta. Tutkimusdatapolitiikkaa on täsmennetty vastaamaan nykyistä lainsäädäntöä sekä kansallisia ja kansainvälisiä ohjeita ja suosituksia. Uusi tutkimusdatapolitiikka koskee kaikkea tutkimusdataa ja sen hallinnan periaatteita, aikaisemman rajautuessa vain digitaalisiin aineistoihin. Toinen tärkeä muutos aikaisempaan on vastuiden täsmentäminen ja esille tuominen.

1.1. Mitä tutkimusdatapolitiikka koskee

Tutkimusdatapolitiikka koskee kaikkea Helsingin yliopiston piirissä tehtävää tutkimusta ja sen yhteydessä kerättyä ja tuotettua tutkimusaineistoa. Tämä tarkoittaa sekä digitaalisia aineistoja, että fyysisiä ja analogisia aineistoja, joihin viitataan tutkimusdatapolitiikassa sanoilla *tutkimusaineisto* ja *tutkimusdata*. (*Tutkimusaineiston määritelmä, ks. sanasto*)

Tutkimusdatapolitiikka koskee kaikkia tutkimusta tekeviä yliopistolaisia mukaan lukien yliopiston työntekijät ja opiskelijat, sekä niitä, jotka tekevät tutkimusta yliopiston nimissä.

1.2. Tutkimusaineistojen hallintaan liittyvät muut linjaukset Helsingin yliopistossa

Tutkimusdatapolitiikassa esitetyt periaatteet ja tavoitteet täydentävät Helsingin yliopiston muita tutkimusaineistojen hallintaan liittyviä linjauksia. Näitä ovat muun muassa [avoimen tieteen periaatteet](#), [tietosuojaperiaatteet](#), [tietoturvapoliittika](#), sekä Tutkimuseettisen neuvottelukunnan

ohjeet [hyvästä tieteellisestä käytännöstä](#). Tutkimusdatapolitiikka on linjassa myös yliopiston tiedonhallintakäytänteiden ja -periaatteiden kanssa.

Tutkimusaineistojen hallintaan liittyviä linjauksia Helsingin yliopistossa ohjaa ensisijaisesti EU-tason ja kansallisen tason lainsäädäntö. Linjauksilla ja ohjeilla yliopisto pyrkii turvaamaan tutkimustoiminnan lainmukaisuuden. Kansainvälisessä tutkimus- ja muussa yhteistyössä tutkimusaineistojen hallintaan voi vaikuttaa myös EU:n ulkopuolinen lainsäädäntö.

Myös yliopiston ulkopuoliset tutkimusrahoittajat ja yhteistyökumppanit voivat asettaa erityisehtoja tutkimusaineistojen hallinnalle.

Helsingin yliopiston omissa, kansallisissa tai kansainvälisissä tutkimusinfrastruktuureissa olevien aineistojen käsittelyä koskevat yleisen tason tavoitteet kuvataan Helsingin yliopiston tutkimusinfrastruktuuriohjelmassa (*Tutkimusinfrastruktuuri, ks. sanasto*). Tämän lisäksi kansallisilla ja kansainvälisillä tutkimusinfrastruktuureilla voi olla omia, yhteisesti sovittuja linjauksia infrastruktuureiden tuottaman tiedon käsittelystä. Kun tutkimusinfrastruktuurit laativat tai päivittävät omia aineistonhallinnan periaatteitaan, on huolehdittava, että ovat linjassa yliopiston tutkimusdatapolitiikan kanssa.

1.3. Tutkimussopimukset ja oikeuksista sopiminen

Tutkimussopimukset

Sopimusten ja sitoumusten laatiminen on tärkeä osa vastuullista tutkimusaineistojen hallintaa tutkimuksen rahoituslähteestä riippumatta. Sopimuksia tarvitaan aina silloin, kun tutkimusta tehdään yhteistyö- tai muussa suhteessa ulkopuolisen tahon kanssa, kuten esimerkiksi toinen yliopisto, yritys, tutkimusorganisaatio tai sairaanhoitopiiri. Sopimuksien avulla turvataan tutkijan ja Helsingin yliopiston edut, hallitaan tutkimukseen liittyviä riskejä ja varmistetaan toiminnan lainmukaisuus.

Tutkimuksen sopimuksista tulee huolehtia mahdollisimman varhaisessa vaiheessa ja mieluiten ennen tutkimusaineiston keruun tai käytön aloittamista. Sopimuksilla varmistetaan, että tutkimusryhmän yhdessä keräämä tutkimusaineisto on kaikkien tutkimukseen osallistuvien tutkijoiden käytettävissä. Sopimuksilla turvataan tutkimuksen jatkuvuus, ja tutkimusaineistojen jatkokäyttö esimerkiksi tutkimusprojektin päättyessä.

Tutkimusaineistoihin liittyvistä oikeuksista sopiminen

Useat kansainväliset ja kansalliset tutkimusrahoittajat edellyttävät tutkimusaineistojen ja tutkimustulosten avointa saatavuutta rahoittamassaan tutkimuksessa. Rahoittajien asettamat velvoitteet tutkijoille ja yliopistolle edellyttävät, että tutkimusaineistojen oikeudet luovutetaan siinä laajuudessa kuin se on tarpeen rahoitusehtojen täyttämiseksi. Lisäksi oikeuksien jakamista tarvitaan, jotta esimerkiksi tutkimusaineistojen avaaminen, jatkokäyttö ja arkistointi ovat mahdollisia. Pääsääntöisesti tutkijalla säilyy rinnakkainen oikeus, jonka turvin tutkija voi jatkaa tutkimusaineiston käyttöä.

Tutkimusaineistoihin liittyvistä oikeuksista ja niiden jakamisesta on tarkoitus sopia jo tutkijan työsuhteen alkaessa työsopimuksen yhteydessä.

Tutkimusaineistoihin liittyvistä oikeuksista sopiminen on lainsäädännön ja hyvän tieteellisen käytännön mukaista. Se ei muuta tutkijan vastuuta tutkimusaineistojensa hallinnassa, vaan on osa sitä. Oikeuksien jakaminen ei muuta tutkijan oikeutta tulla mainituksi tutkimusaineistojen kerääjänä tai tuottajana.

1.4. Tutkimusdatapolitiikan seuranta ja päivittäminen

Tämä dokumentti korvaa Helsingin yliopiston vuonna 2015 hyväksytyn tutkimusdatapolitiikan. Päivitetylle tutkimusdatapolitiikalle laaditaan erillinen toteuttamissuunnitelma vuonna 2022.

Tutkimusdatapolitiikan toteutumista seurataan säännöllisesti tehtävillä arvioinneilla, joista ensimmäinen toteutetaan vuonna 2023.

Tutkimusdatapolitiikka päivitetään vuoteen 2025 mennessä.

2. Tutkimusaineistojen hallinnan periaatteet ja tavoitteet

2.1. Yleiset periaatteet

Seuraavassa tutkimusaineistojen vastuullisen hallinnan periaatteet ja tavoitteet on kuvattu suhteessa tutkimuksen elinkaareen, alkaen aineistonhallinnan suunnittelusta ja päättyen aineistojen pitkäaikaissäilytykseen tai tuhoamiseen. Aineistojen jatkokäytön mahdollisuudet on otettava huomioon jo tutkimusprojektin aikana.

Tutkimusaineistojen vastuullista hallintaa ohjaavat yleisellä tasolla FAIR-periaatteet, joiden mukaan tutkimusdatan tulisi olla löydettävää (*Findable*), saavutettavaa (*Accessible*), yhteentoimivaa (*Interoperable*) ja uudelleenkäytettävää (*Reusable*). (*FAIR-periaatteet, ks. sanasto.*) Toinen ohjaava periaate koskee tutkimusdatan avoimuutta: niin avointa kuin mahdollista, niin suljettua kuin välttämätöntä.

2.2. Aineistonhallinnan suunnittelu ja valmistelu

- Kaikille tutkimusprojekteille laaditaan [aineistonhallintasuunnitelma](#). Suunnitelmaa päivitetään tutkimusprojektin aikana, kun sisältöön tulee merkittäviä muutoksia. (*Aineistonhallintasuunnitelma, ks. sanasto*)
- Aineistonhallintasuunnitelmassa huomioidaan lainsäädännölliset näkökohdat ja [tutkimuseettiset kysymykset](#).

- Henkilötietojen käsittelyssä noudatetaan yliopiston [tietosujoaohjetta](#) mm. tutkittavien informoinnista ja tarvittavien suojoitoimenpiteiden varmistamisesta ja tarpeen mukaan tehdään erillinen tietosujoa koskeva [vaikutustenanviointi](#). (*Henkilötiedot, ks. sanasto*)
- Tutkimukseen liittyvistä oikeuksista ja vastuista sovitaan tutkimusryhmän kesken ja yliopiston ulkopuolisten tahojen kanssa tehdään tarvittavat sopimukset ennen tutkimusaineistojen keräämisen tai käytön aloittamista.
- Arkaluonteisen aineiston suojaamiseen ja tarvittaessa anonymisointiin sekä salassapidettävien tietojen poistamiseen varaudutaan, ja tähän kohdennetaan tutkimusprojekteja suunniteltaessa riittävästi resursseja. (*Arkaluonteiset ja luottamukselliset aineistot, ks. sanasto*)
- Tutkimukseen liittyvät sopimukset, suostumukset, sitoumukset ja muut dokumentit arkistoidaan, niin että ne ovat aina tarvittaessa tutkijoiden ja tukipalveluiden saatavilla.
- Etenkin laajasti palveluita tuottavien tutkimusinfrastruktuureiden on syytä laatia aineistonhallintapolitiikka, jossa kuvataan infrastruktuurin tuottaman aineiston käsittelyn, tallentamisen, jakamisen, säilyttämisen sekä hävittämisen periaatteet, sekä aineistojen käsittelyyn liittyvät vastuu- ja omistajuuskysymykset.

2.3. Tutkimusaineiston dokumentointi ja käsittely

- Tutkimusaineistojen dokumentoinnin ja metatietojen tulee noudattaa tieteenalan standardeja, jotta aineistoa voidaan hyödyntää ja rikastaa myöhemmissä tutkimushankkeissa. (*Metatiedot, ks. sanasto*)
- Tutkimusaineistojen tallentamisessa ja käsittelyssä on huomioitava asianmukainen [tietoturva](#).
- Yliopisto tarjoaa tutkijoille peruspalvelut tutkimusdatan tallennukseen ja käsittelyyn. Mikäli dataa on hyvin paljon tai tarvitaan erityistä laskentatehoa, käsittelyn ja tallennuksen kustannukset tulee huomioida jo tutkimuksen suunnitteluvaiheessa.

2.4. Tutkimusaineiston julkaiseminen ja saavutettavuus

- Helsingin yliopiston piirissä tuotettu, julkaistuihin tutkimustuloksiin liitetty tutkimusdata on lähtökohtaisesti yhteiskäyttöistä ja avointa. Tutkimusaineistojen avaamisessa noudatetaan *niin avointa kuin mahdollista, niin suljettua kuin välttämätöntä* -periaatetta.
- Tutkimusdata julkaistaan data-arkistoissa, jotka turvaavat aineiston löydettävyyden ja mahdollistavat siihen viittaamisen. (*Data-arkisto, ks. sanasto*)
- Tutkimusaineistojen metadata tulee julkaista aina kun se on mahdollista, joko kansallisissa tai kansainvälisissä metadatapalveluissa.

- Avatusta tutkimusdatasta tulisi olla aina saatavilla yksilöivä tunniste sekä aineiston jatkokäytön mahdollistava lisenssi. (*Pysyvä tunniste, ks. sanasto*)
- Mikäli mahdollista, myös arkaluonteiset aineistot saatetaan muiden tutkijoiden käyttöön, kuitenkin huomioiden lainsäädännölliset, eettiset ja sopimukselliset rajoitteet. Tässä hyödynnetään palveluita, jotka mahdollistavat tutkimusdatan turvallisen säilytyksen ja rajoitetun pääsyn tutkimusluvalla.
- Yliopistolla on käytännöt ja palvelut kerättyä, tuotettua ja avattua tutkimusdataa koskevaa tiedonkeruuta ja seurantaa varten.

2.5. Tutkimusaineistojen kaupallistaminen

- Kaupallisesti arvokkaiden aineistojen osalta huolehditaan siitä, että aineistojen käyttö ei vaaranna kaupallistamista. Esimerkiksi tutkimustulosten julkaisemisessa on noudatettava rahoittajan asettamia ehtoja salassapidosta ja yliopiston ohjeistusta kaupallistamisesta.
- Yliopisto tarjoaa tutkijoille tukea ja selkeän palveluprosessin, kun tutkimuksessa on syntynyt kaupallisesti arvokasta tutkimusdataa.

2.6. Tutkimusaineiston tuhoaminen tai pitkäaikaissäilytys

- Tutkimusprojektin aikana täsmennetään, mikä osa tutkimusaineistosta on arvokasta pitkällä tähtäimellä ja säilytetään, sekä mikä osa tuhotaan projektin päättyessä.
- Tarpeettomiksi tulleet tutkimusaineistot tuhotaan tarvittavan säilytysajan jälkeen. Tuhoamisessa noudatetaan erityistä huolellisuutta, jos tuhottava aineisto sisältää arkaluonteista tietoa.
- Yliopistolla on prosessi arvokkaan tutkimusdatan kuratoimisesta ja siirtämisestä digitaaliseen pitkäaikaissäilytykseen. (*Pitkäaikaissäilytys, ks. sanasto*)

3. Tutkimusaineistojen hallintaan liittyvät vastuut

3.1. Tutkijan vastuulla on

- Perehtyä vastuullista tutkimusaineistojen hallintaa koskevaan ohjeistukseen ja noudattaa sitä.
- Suunnitella ja toteuttaa aineistonhallinta oman tutkimuksensa osalta.
- Huolehtia, että tutkimuksessa tarvittavat sopimukset, sitoumukset ja suostumukset laaditaan.
- Perehdyttää tutkimusryhmän jäsenet vastuulliseen aineistonhallintaan toimiessaan tutkimushankkeen vastuullisena tutkijana.
- Huolehtia siitä, että ryhmän tai yhteistutkimuksen jaettaviksi tai yhteisiksi sovitut aineistot

ovat muiden saatavilla.

- Välittää vastuullisen aineistonhallinnan periaatteita ja hyviä käytäntöjä edelleen toimiessaan opinnäytetöiden ohjaajana.
- Päivittää tutkimusdatan hallinnan osaamistaan säännöllisesti.

3.2. Tiedekunnan tai tutkimusta harjoittavan erillislaitoksen vastuulla on

- Ylläpitää kokonaiskuvaa tutkimusaineistoista ja niiden hallinnasta yksikössä, mukaan lukien yksikön sopimussuhteet ja velvoitteet.
- Huomioida tutkimusdatapolitiikka toiminnan ja talouden suunnittelussa ja osoittaa siihen tarvittavat resurssit.
- Viedä yliopistotasoiset ohjeistukset ja linjaukset yksikön toimintaan siten, että tutkimusalaakohtaiset erot huomioidaan.
- Varmistaa, että akateeminen henkilöstö ja opiskelijat ovat perehtyneet tutkimusdatan hallintaan osana hyvää tieteellistä käytäntöä.
- Tarjota ja resursoida tutkijoille lähitukea tutkimusdatan hallintaan, yhdessä tutkimuksen tukipalveluiden kanssa.
- Toteuttaa ennakoivaa riskienhallintaa ja varautua mahdollisiin poikkeamatilanteisiin. (*Riskienhallinta ja poikkeamatilanne, ks. sanasto*)
- Varmistaa vastuullisen tutkimusaineistojen hallinnan toteutuminen tutkimusinfrastruktuureissa.

3.3. Yliopiston vastuulla on

- Luoda edellytykset vastuullisen tutkimusaineistojen hallinnan toteutumiselle yliopistossa.
- Varmistaa yliopistotasoisien tutkimusdatainfrastruktuurin toimivuus, ajantasaisuus ja riittävä resursointi siten, että palveluita on tarjolla tutkimuksen elinkaaren kaikissa vaiheissa. (*Tutkimusdatainfrastruktuuri, ks. sanasto*)
- Luoda yhteistyössä akateemisten yksiköiden kanssa arviointikäytäntöjä ja kannustimia, joilla tutkijoiden työ tutkimusaineistojen jakamisen ja jatkokäytön edistämiseksi sekä tutkimusaineistojen hallinnan osaaminen otetaan huomioon ammatillisena meritoitumisena.
- Luoda kannustimia yliopiston yksiköille tutkimusaineistojen vastuulliseen hallintaan.
- Edistää tutkijoiden ja tukipalveluiden asiantuntijoiden mahdollisuuksia erikoistua tutkimusdatan hallintaan vakiinnuttamalla tähän liittyviä asiantuntijarooleja ja kehittämällä urapolkuja.
- Koordinoida tutkimusaineistojen hallintaan liittyvää kehittämistyötä.
- Laatia yliopistotasoiset tutkimusaineistojen hallintaa koskevat linjaukset ja niiden toteuttamista tukevat prosessit.
- Laatia tarvittavat yliopistotasoiset tutkimusaineistojen hallintaa koskevat ohjeistukset.
- Toteuttaa ja tukea ennakoivaa riskienhallintaa ja varautumista mahdollisiin poikkeamatilanteisiin.

- Tarjota tutkimusdatan hallintaan liittyvää koulutusta ja perehdytystä tutkijoille ja muulle henkilöstölle.
- Tarjota tukea tutkimusaineistojen hallintaan tutkijoille ja akateemisille yksiköille.

4. Sanasto

Aineistohallintasuunnitelma

Dokumentti, jossa kuvataan tutkimusaineisto, jota hankitaan tai tuotetaan tutkimusprojektin aikana. Suunnitelmasta käytetään yleisesti myös kansainvälistä lyhennettä DMP (*data management plan*). Aineistohallintasuunnitelman yhteydessä sana *data* ymmärretään laajasti, eli se kattaa kaikki tiedot ja resurssit, joihin tutkimustulokset perustuvat. Suunnitelma kattaa myös menetelmäkuvaukset kuten koodit ja ohjelmistot.

Suunnitelmassa kuvataan myös se, miten tutkimusaineistoihin liittyviä oikeuksia hallitaan, mitä sopimuksia tarvitaan, miten tietosuojaa varmistetaan, miten tutkimusdata tallennetaan, miten tutkimusdata avataan tai muutoin mahdollistetaan aineistojen löytyvyys ja käyttö tutkimustulosten verifiointiksi tai jatkotutkimuksessa. Aineistohallintasuunnitelma laaditaan tutkimuksen suunnitteluvaiheessa, mutta se on luonteeltaan elävä dokumentti, jota tulee päivittää tutkimushankkeen edistyessä.

Arkaluonteiset ja luottamukselliset aineistot

Tutkimusaineistot, joiden tallentamista, käyttöä ja jakamista rajoittavat eettiset, juridiset, sopimukselliset tai taloudelliset syyt. Tällaista aineistoa on käsiteltävä ja suojattava erityisen huolellisesti.

Arkaluonteiset aineistot liittyvät esimerkiksi tietynlaisiin henkilötietoihin (*ks. erityiset henkilötietoryhmät henkilötiedon määritelmässä*), uhanalaisiin lajeihin, bioturvallisuuteen tai maanpuolustukseen. Luottamuksellisesti annettu tieto patenteista ja liikesalaisuuksista on niin ikään suojattavaa, ja sen paljastuminen voi johtaa korvausvaateisiin.

Data-arkisto (datarepositorio)

Virtuaalinen, yleensä tieteenalakohtainen arkisto tai tietokanta, johon tutkijat voivat siirtää tutkimusdatansa jakamista, raportointia ja jatkokäyttöä varten. Data-arkisto säilyttää tutkimusdataa, asettaa sen käytettäväksi ja järjestää sen loogisella tavalla. Data-arkistot helpottavat myös tutkimusdataan viittaamista, kun käytetään pysyviä tunnisteita.

FAIR-periaatteet

Eurooppalaiset tutkimusdatan ja sen metatietojen laatua koskevat yleisen tason tavoitteet. Lyhenne FAIR tulee sanoista *findable* (löydettävä), *accessible* (saavutettava), *interoperable* (yhteentoimiva) ja *reusable* (uudelleenkäytettävä). FAIR-periaatteet ohjaavat erityisesti metatietojen laatimista. Löydettävyys tarkoittaa, että tutkimusdatalla on pysyvä tunniste, joka

toimii linkkinä aineistoon, ja jonka avulla aineisto voidaan aina löytää, vaikka säilytyspaikka muuttuisi. Löydettävyyden on mahdollista toteuttaa myös ei-digitaalisten tutkimusaineistojen kohdalla, kun aineistojen metatiedot ovat avoimesti saatavilla. Saavutettavuus tarkoittaa, että aineistoon ja sen metatietoihin pääsee käsiksi verkkoselaimen kautta. Yhteentoimivuus tarkoittaa, että aineiston tallennuksessa on käytetty avoimia tiedostomuotoja ja yhteisiä standardeja. Uudelleenkäytettävyys tarkoittaa, että tutkimusaineistosta on olemassa laadukkaat metatiedot sekä aineiston jatkokäytön ehdoista kertova lisenssi.

Henkilötiedot

Eläviin luonnollisiin henkilöihin liittyvät tiedot, eli kaikki tiedot, jotka liittyvät tai voidaan yhdistää tunnistettuun tai tunnistettavissa olevaan henkilöön. Henkilötietoja ovat myös sellaiset tiedot, joiden perusteella henkilö voidaan epäsuorasti tunnistaa, esimerkiksi yhdistämällä yksittäinen tieto johonkin toiseen tietoon, joka mahdollistaa tunnistamisen (pseudonymisoidut henkilötiedot). Henkilötietoja voi olla talletettuna esimerkiksi sähköisissä tiedostoissa, tietokannoissa, paperilla, kortistossa, mapeissa, kyselylomakkeissa, tai ääni- tai kuvatallenteella.

Tietosuojasääntely on kategorisoinut tietyt tiedot kuuluvaksi **ns. erityisiin henkilötietoryhmiin**. Näitä kutsutaan myös arkaluonteisiksi henkilötiedoiksi (*ks. myös arkaluonteiset ja luottamukselliset aineistot*),

- tiedot, joista ilmenee rotu tai etninen alkuperä, poliittisia mielipiteitä, uskonnollinen tai filosofinen vakaumus tai ammattiliiton jäsenyys
- geneettiset tiedot
- biometriset tiedot, jos niitä käsitellään henkilön yksiselitteistä tunnistamista varten
- terveyttä koskevat tiedot
- luonnollisen henkilön seksuaalista käyttäytymistä ja suuntautumista koskevat tiedot
- rikostuomioihin ja rikkomuksiin liittyvät tiedot

Edellä mainittujen tietojen lisäksi henkilötieto tai -tietojen käsittely voi olla myös muutoin arkaluonteista, jos tietojen käsittelystä saattaa aiheutua riskejä tutkittaville (*ks. riskienhallinta*)

Metatiedot (metadata)

Metadata on tietoa tiedosta. Tutkimusaineistoilla on useita erityyppisiä metatietoja, kuten kuvailevat, rakenteelliset, hallinnolliset, tilastolliset ja oikeuksiin liittyvät metatiedot, sekä viite- ja viittaustiedot. Metatietojen avulla huolehditaan siitä, että tutkimusaineisto on löydettävää ja jatkokäytettävää. Kun tutkimusaineistot kuvaillaan ja dokumentoidaan asianmukaisesti, muut käyttäjät voivat jäljittää ja ymmärtää tutkimuksen yksityiskohdat. Metadata helpottaa tutkimusdatan hakemista ja löytämistä, kun data on tallennettu datarepositorioon.

Pitkäaikaissäilytys

Digitaalisen informaation säilyttäminen ymmärrettävänä ja käytettävänä useiden kymmenien ja jopa satojen vuosien ajan. Viitataan myös lyhenteellä PAS. Pitkäaikaissäilytettäväksi valikoidaan arvokkaita aineistoja. Pitkäaikaissäilytyksen tavoitteena on varmistaa digitaalisten objektien saatavuus, aitous, ymmärrettävyys ja eheys pitkäksi aikaa, vaikka esimerkiksi laitteet, ohjelmistot

ja tiedostomuodot tänä aikana vanhenevat ja muuttuvat. Pitkäaikaissäilytys varmistaa tutkimusdatan pitkäaikaissaatavuuden.

Poikkeamatilanne

Poikkeamatilanteella tarkoitetaan normaalitoiminnasta poikkeavia tapahtumia tai olosuhteita, jotka esimerkiksi hidastavat, estävät tai vahingoittavat tutkimustoiminnan toteuttamista. Poikkeamatilanteiden luonne vaihtelee tutkimusaloittain.

Tietoturvapoikkeamalla tarkoitetaan tapahtumaa, jonka seurauksena tutkimusaineisto tuhoutuu, häviää, muuttuu tai aineistoa luovutetaan luvottomasti tai aineistoon pääsee käsiksi taho, jolla ei ole siihen oikeutta.

Pysyvä tunniste

Uniikki ja yksiselitteinen koneluettava nimi tutkimustuotokselle. Pysyvistä tunnisteista käytetään yleisesti kansainvälistä lyhennettä PID (*persistent identifier*). Tunniste muodostaa pysyvän linkin, joka vie aina esimerkiksi julkaisuun tai tutkimusaineiston kuvailusivulle. Pysyvä tunniste mahdollistaa digitaalisen tutkimustiedon pitkäaikaisen löydettävyyden.

Riskienhallinta

Ennakoiva varautuminen seurauksiltaan kielteisiin tapahtumiin. Riskienhallinta on koordinoitua toimintaa, jolla yliopiston, sen yksikön tai tutkijan toimintaa ohjataan, johdetaan ja valvotaan suhteessa riskeihin. Riskienhallinnan tavoitteena on auttaa yliopiston, sen yksiköiden ja tutkijan tavoitteiden saavuttamista ja päätöksentekoa

Henkilötietojen käsittelyssä (*ks. henkilötiedon määritelmä*) on arvioitava käsittelyyn liittyvät riskit ja varmistettava tietosuojaperiaatteiden toteutuminen. Henkilötietojen osalta riskiarvio on tehtävä tutkittavien näkökulmasta.

Tutkimusaineisto (tutkimusdata)

Tutkimusaineistot, jotka on kerätty, havaittu, mitattu tai luotu tutkimustulosten todentamiseksi tai joita pidetään tutkimusyhteisössä yleisesti tarpeellisina tutkimustulosten tarkistamiseksi. Konteksti tekee datasta tutkimusdataa. Mikä tahansa tietoaaineisto voi olla tutkimusdataa, jos sitä analysoidaan tutkimustarkoituksessa.

Tutkimusdataa ovat esimerkiksi mittaustulokset, koetulokset, haastatteluaineistot, ääni- ja kuvatallenteet sekä näytteet. Tutkimusdata on usein digitaalisessa muodossa, mutta se voi olla myös fyysistä tai analogista aineistoa. Tutkimusdata voi olla luonteeltaan raakadataa, prosessoitua dataa, kolmannen osapuolen hallussa olevaa dataa, jaettua dataa tai julkaistua dataa. Tutkimusdatan avoimuuden aste vaihtelee salatusta ja arkaluontoisesta datasta avoimesti saatavilla olevaan.

Sanoilla *tutkimusaineisto* ja *tutkimusdata* voidaan tarkoittaa eri asioita erilaisissa yhteyksissä tai ne voidaan ymmärtää synonyymeinä. Tässä dokumentissa niitä käytetään synonyymeinä.

Tutkimusaineistojen hallinta

Käytetään myös kansainvälistä lyhennettä RDM (*research data management*). Tutkimuksen elinkaaren aikainen prosessi, joka sisältää tutkimusaineiston keräämisen tai hankkimisen, organisoinnin, kuratoinnin, tallennuksen, (pitkäaikais-) säilyttämisen, suojaamisen, laadunvarmistuksen, pysyvien tunnisteiden ja muiden metatietojen käytön, lisensoinnin sekä aineiston jakamisen noudattaen tieteenalan sääntöjä ja menettelytapoja. (Euroopan komissio/ Horizon Europe).

Tutkimusdatainfrastrukturi

Prosessit, teknologiat ja palvelut, joiden kautta tutkimusdatan hallinta käytännössä tapahtuu. Tutkimusdatainfrastrukturi edellyttää organisaatiota, toimintakulttuuria ja pitkäjänteisesti toimivia sosiaalisia verkostoja, joiden varassa tekniset ja hallinnolliset ratkaisut sekä palvelut voidaan toteuttaa.

Tutkimusinfrastrukturi

Tutkimusinfrastruktuureilla tarkoitetaan välineitä, laitteistoja, tietoverkkoja, tietokantoja ja aineistoja sekä palveluita, jotka mahdollistavat tutkimustyön, edistävät tutkimusyhteistyötä sekä vahvistavat tutkimus- ja innovaatiokapasiteettia ja osaamista. (Suomen Akatemia)