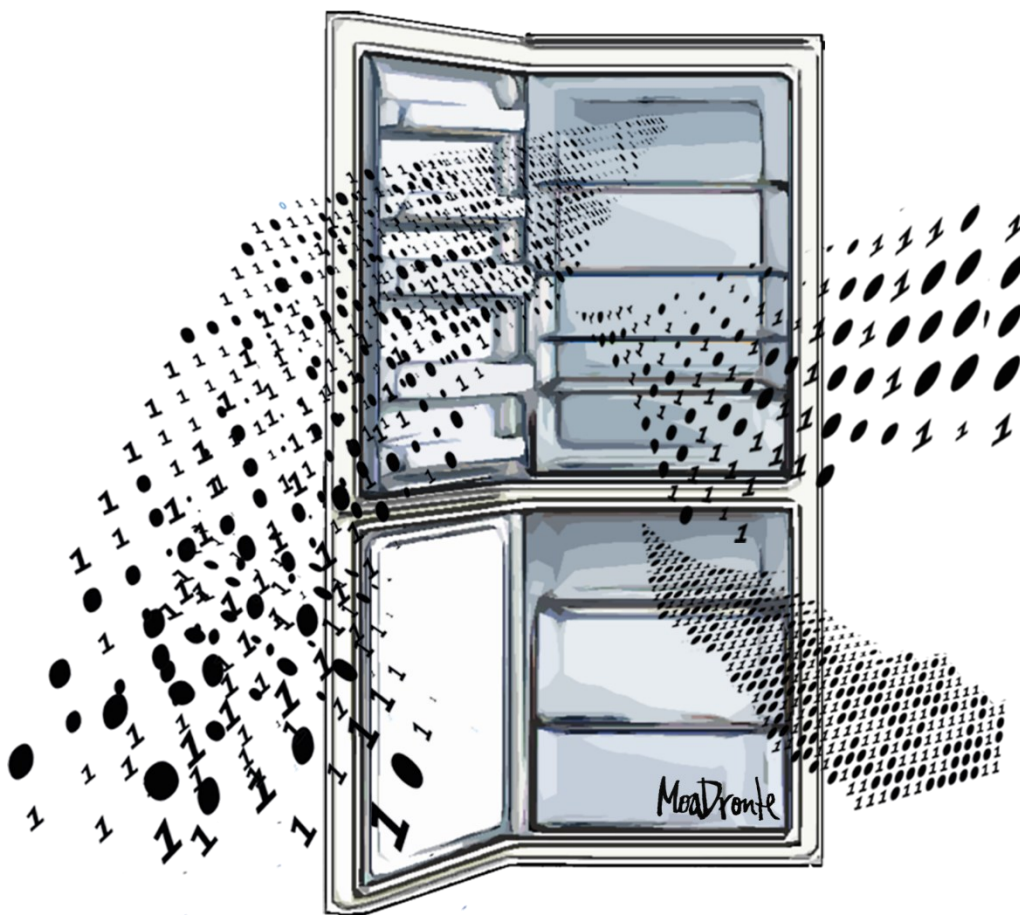


Sinustako säilyttäjä?

Helsingin yliopiston virallinen opas tutkimusaineistojen vastuulliseen säilyttämiseen

Timo Lahtinen, Matilda Mela, Mikko Mäkelä, Niina Nurmi, Mari Elisa Kuusniemi



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



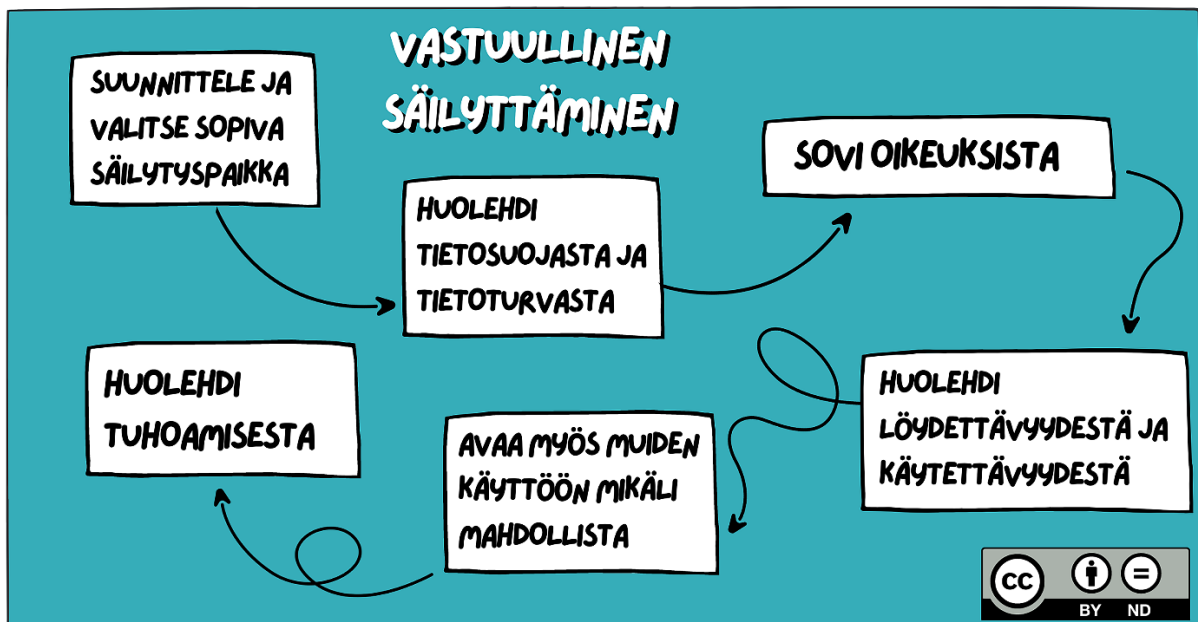
Vastuullisen säilyttämisen opas

Avoin tiede — kaikki hyötyvät

Tämä opas on koottu tukemaan tutkimusaineistojen säilyttämistä aineiston keräämisen ja aktiivisen analyysivaiheen jälkeen. Oppaassa kerromme mitä täytyy tehdä ennen tutkimusta ja sen aikana, jotta tuotetun aineiston säilyttäminen olisi ylipäätään mahdollista. Alussa on **muistilista**, johon olemme koonneet aineiston tallentamisen kannalta tärkeimmät asiat. Seuraavissa luvuissa käymme asioita läpi tarkemmin, ja oppaan loppuun olemme koonneet linkit lisätietoihin.

Tutkimusaineistojen vastuullinen säilyttäminen vaatii hieman vaivaa, mutta se myös palkitsee. Ilmastomuutoksen torjumiseksi tehtävä työ ja koronarokotteen nopea kehittäminen ovat esimerkkejä **yhteistyön voimasta**, mistä tässäkin on kyse: hyödynnäme tutkimusaineistot mahdollisimman tarkasti ja teemme yhteistyötä.

Lukuohje: Linkit on kerätty kunkin kappaleen loppuun lukemisen helpottamiseksi. Ohjeen alussa olevan muistilistan kaikkia vieraita käsitteitä ei ole avattu. Ne avataan oppaassa myöhemmin.



Kuva 1: Vastuullisen säilyttämisen vaiheet.

Sisältö

Tutkimusdatan vastuullisen säilyttämisen muistilista	3
1. Miksi tutkimusaineistoja säilytetään?	4
1.1. Vastuullisen säilyttämisen periaatteet.....	5
1.2. Käyttöesimerkkejä.....	7
2. Miten valitset säilytettävän aineiston?	9
3. Säilytysaika ja tuhoaminen.....	12
4. Miten aineisto valmistellaan säilytykseen?	14
4.1. Valmistelussa huomioitavaa	14
4.2. Säilytyskelpoiset tiedostoformaatit	16
5. Missä aineistoja säilytetään?	18
5.1. Periaatteita säilytyspalvelun valintaan.....	19
5.2. Miten löydän sopivan säilytyspalvelun?	21
5.3. Data-arkiston valinnan vuokaavio.....	22
6. Ohjeen kootut linkit	23

Tutkimusdatan vastuullisen säilyttämisen muistilista

Mistä aloittaa?

- Etsi aineistollesi sopiva säilytyspalvelu. Katso tarvittaessa apua vuokakaaviosta, luvusta 5 “Missä aineistoja säilytetään?” tai [re3datasta](#).
- Tarkista, että sinulla on riittävät käyttöoikeudet luovuttaa data säilytettäväksi. Käyttöoikeudet tulee tarkistaa erityisesti silloin, jos se on peräisin muista lähteistä tai kerätty yhteistyössä muiden henkilöiden tai tahojen kanssa.
- Jos repositorio ei sovellu arkaluonteisen aineiston säilyttämiseen, varmista, että aineistosi ei sisällä sensitiivisiä henkilötietoja tai muita luottamuksellisia tietoja.

Datan valmistelu

- Suosi pitkään säilyviä, avoimia ja yleisiä tiedostomuotoja aina kun mahdollista.
- Tee data-aineistolle selkeä kansiorakenne.
- Nimeä tiedostot ja kansiot johdonmukaisesti ja kuvaavasti.

Datan kuvailu

- Varmista datan ymmärrettävyys ja jatkokäyttö kuvailemalla se kattavin metatiedoin.
- Anna data-aineistolle kuvaava otsikko.
- Tee aineistolle erillinen README.txt-tiedosto.
- Jos datasettisi sisältää taulukkomuotoisia tiedostoja, määrittele ja selitä sarakkeiden otsikot, muuttujien saamat arvot, mahdolliset lyhenteet ja käytetyt koodit.
- Liitä aineiston mukaan sen ymmärrettävyyttä ja jatkokäyttöä tukeva dokumentaatio, kuten kuvaus tutkimusmenetelmistä, kyselylomakkeet, suostumuslomakepohja tai sähköiset muistiinpanot.
- Lisää aineiston metatietoihin linkit muihin tutkimustuotoksiin. Näitä ovat esimerkiksi julkaistun artikkelin DOI, projektin nettisivujen osoite tai tutkimuksen aikana tuotettujen koodien tai ohjelmistojen linkit.

Lisenssit

- Valitse aineistollesi sopiva lisenssi. Katso lisätietoja HY kirjaston [lisenssioppaasta](#).

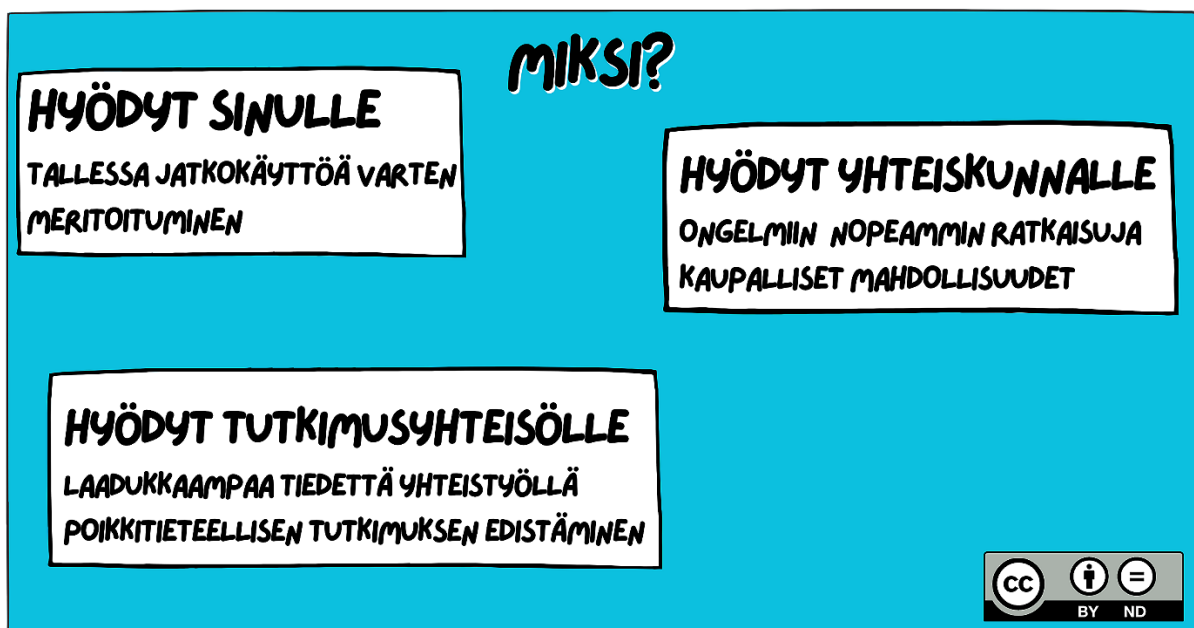
Mitä datan tallentamisen jälkeen?

- Jos valitsemasi repositorio antaa datasetillesi DOI:n tai muun pysyvän tunnisteeseen, sisällytä se julkaisemiesi artikkelien saatavuusselosteeseen (eng. *data availability statement*).
- Kerro muillekin jakamastasi datasta! Voit esimerkiksi lisätä siitä maininnan linkeineen Helsingin yliopiston TUHAT-profiiliisi.

Jos mikään säilyttämiseen liittyvä asia mietityttää, voit aina kysyä apua [datatuelta](#).

1. Miksi tutkimusaineistoja säilytetään?

Tutkimusaineistojen vastuullinen säilyttäminen liittyy muun muassa avoimen tieteen tavoitteisiin¹, joiden edistämiseen myös Helsingin yliopisto on sitoutunut. Maailmassa, jossa tieteellinen tieto lisääntyy ennennäkemätöntä tahtia, on erityisen tärkeää tukea sen järjestelmällistä säilyttämistä ja löydettävyyttä. Näin kertaalleen kerätyt aineistot saadaan hyödynnettyä mahdollisimman monipuolisesti tutkimuksen ja yhteiskunnan hyväksi.



Kuva 2: Vastuullisesta säilyttämisestä on paljon hyötyä.

Hyödyt sinulle

Tutkijana sinulle on merkittäviä hyötyjä tutkimusaineiston säilyttämisestä:

- Aineistosi tallennetaan turvalliseen säilytyspalveluun ja pääset näppärästi dataan käsiksi kauan tutkimushankeen päätyttyäkin.
- Tutkimusasetelmasi, tutkimusjulkaisu tai tutkimuksesi rahoittaja saattaa edellyttää aineiston säilyttämistä.
- Aineistoosi viittaaminen helpottuu, kun datasetillesi annetaan pysyvä tunniste (DOI).

¹ Avoin tiede: [Mitä on avoin tiede?](#) (28.9.2023)

Hyödyt tutkimusyhteisölle

Datan säilyttäminen vahvistaa tutkimusyhteisöä. Jos myös jaat aineistosi:

- Tarjoat tutkijoille ja opiskelijoille pääsyn laadukkaaseen dataan.
- Parannat tutkimuksen todentamista ja toistettavuutta.
- Tuet poikkitieteellistä ja organisaatioiden välistä tutkimusta.

Hyödyt yhteiskunnalle

Tutkimusaineiston säilyttäminen on usein halvempaa kuin tutkimuksen toistaminen. Siksi tutkimuksen tulokset on tärkeää säilyttää ja jakaa tarkoituksenmukaisesti tuottamaan laajempaa yhteiskunnallista hyötyä, esimerkiksi tutkimuksen pohjalta syntyvien innovaatioiden kautta.

On vaikea arvioida mikä tieto on arvokasta tulevaisuudessa. Tutkimuksen tuottaman tiedon arvo saatetaan huomata vasta vuosien kuluttua. Jotta tutkimuksesta olisi mahdollisimman laajaa yhteiskunnallista vaikutusta, on tärkeää, että sen tuottamat aineistot säilyisivät ja tulevat sukupolvet voisivat hyödyntää kerättyä tietoa.

Säilytyspalveluiden monet nimet

Data-arkisto, datarepositorio ja datapankki liittyvät kaikki aineistojen tutkimuksen jälkeiseen tallentamiseen. Niillä on kuitenkin tiettyjä käyttötarkoitukseen liittyviä laadullisia eroja keskenään. Tässä oppaassa käytämme näistä kaikista yhteisesti termiä **säilytyspalvelu**.

1.1. Vastuullisen säilyttämisen periaatteet

Hyvä tieteellinen käytäntö

Miten hyvä tieteellinen käytäntö näkyy tutkimusaineistojen vastuullisessa säilyttämisessä? Todennäköisesti sinulla on paras tuntemus oman alasi osalta, sillä käytännöt vaihtelevat tieteenalojen kesken esimerkiksi suositeltujen säilytyspalveluiden ja -vaatimusten sekä

tutkimustulosten verifiointiin edellyttämän säilytysajan suhteen. Joillakin aloilla myös tutkimusaineistojen avaaminen on arkipäivää, kun toisilla sitä vasta aloitellaan.

Onneksi kaikkea ei tarvitse osata yksin. Alkuun pääset jo sillä, kun tunnet oman alasi ja tutkimusetiikan perusteet ja noudatat niitä.

Voiko arkaluonteisia tai luottamuksellisia aineistoja säilyttää?

Arkaluonteinen ja luottamuksellinen tieto on sellaista, joka voi paljastuessaan aiheuttaa vahinkoa. Siksi näiden aineistojen säilyttäminen on erityisen vaativaa.

Aineistot kerännyt tutkija vastaa arkaluonteisen ja luottamuksellisen tiedon turvallisesta säilyttämisestä. Toistaiseksi on olemassa vain vähän palveluita, jotka ottavat vastaan tällaista aineistoa tutkijoiden puolesta säilytettäväksi. Useimmissa palveluissa edellytetään kaiken arkaluonteisen tiedon poistamista ennen säilytykseen siirtämistä.

Tietosuojalainsäädäntö säätelee arkaluonteisia henkilötietoja sisältävien aineistojen säilyttämistä. Aineistoja voidaan säilyttää tunnisteellisina pitkäänkin, mikäli tutkittavia on informoitu asiasta ja aineistoa luovutetaan vain alkuperäiseen käyttötarkoitukseensa. Tällaisten aineistojen keruussa olisikin hyvä ennakoida omat ja mahdollisten muiden tutkijoiden tulevat tietotarpeet. Käyttötarkoitus onkin hyvä määritellä aika labeaksi. Lainsäädäntö on tältä osin uutta ja sen tulkinnat vakiintumattomia. Tämän vuoksi säilyttämisen palveluiden kehittäminen ei ole helppoa.

Muun kuin henkilötietoja sisältävien luottamuksellisten aineistojen säilyttäminen on helpompaa. Toki siinäkin täytyy huomioida tarkasti tietoturvanäkökulmat. Helsingin yliopisto tarjoaa luottamuksellisten aineistojen pitkäaikaiseen säilyttämiseen sopivaa palvelua.

Kysymyksiä pohdittavaksi arkaluonteisten aineistojen säilyttämisestä

Pohdi, onko aineisto enää käyttökelpoinen kaiken arkaluonteisen tiedon poistamisen jälkeen. Jos ei, selvitä löytyykö aineistolle turvallinen säilytyspaikka, jossa aineisto voidaan säilyttää tunnisteellisena.

Tarkista seuraavat asiat ennen arkaluonteisen aineiston tallentamista:

- Oletko informoinut tutkittavia aineiston säilyttämisestä?
- Mitä tutkittaville on luvattu aineiston käyttötarkoituksesta?
- Jos aineiston luvattu anonymisoida, onko aineistosta poistettu sensitiivinen tieto ja henkilöiden suorat ja epäsuorat tunnisteet? Onko myös aineiston kuvailutiedot tehty niin, että tutkittavia ei pysty tunnistamaan?

Lue lisää sensitiivisten aineistojen vastuullisesta säilyttämisestä:

- ❖ Avoin tiede: [Sensitiivisten aineistojen vastuullinen avaaminen](#)
- ❖ Vastuullinen tiede: [Turvaa tutkittavan anonymiteetti!](#)
- ❖ Tietoarkisto: [Tunnisteellisuus ja anonymisointi](#)

1.2. Käyttöesimerkkejä

Kun tutkija lähtee pois Helsingin yliopistosta

Miten tutkimusaineistosi pysyy laadukkaana ja käyttökelpoisena tutkimusryhmän nykyisille ja tuleville jäsenille sen jälkeen, kun tehtävät kutsuvat sinua muualla?

Tarkista ainakin seuraavat asiat:

- Sovi kirjallisesti tutkimusaineistosi oikeuksista. Kysy lisää Helsingin yliopiston lakimiehiltä osoitteesta: tutkimuksenjuristit@helsinki.fi.
- Huolehdi siitä, että tutkimusaineistosi on organisoitu hyvin, kuvailutiedot on tehty huolella ja tiedostoformaattit ovat kelvolliset — tulevan tallennuspaikan vaatimusten mukaan. Dokumentointi on silloin riittävä, kun aineiston ensikertaa avaava uusi tutkija ymmärtää sen sisällön ilman sinua. Tällöin aineistosi on itsensä selittävä.
- Tallenna aineistosi kuvailutietoineen sopivaan säilytyspalveluun tai tietokantaan.
- Varmista, että aineistollasi on sopiva käyttölisenssi.

Julkaisuviive

Voiko tutkimusaineistoon ja metadataan pääsyä rajoittaa julkaisuviiveen eli embargon ajaksi? Totta kai! Aikarajoitus määritellään etukäteen tutkijan toiveen mukaisesti. Voit tarkistaa lehden mahdollisesti asettaman julkaisuviiveen näppärästi [Sherpa Romeosta](#). Huomaathan, että osa tutkimusaineistojen säilytyspalveluista myöntää julkaisuviiveen vain perustellusta syystä.

Vähemmän koe-eläimiä prekliinisissä tutkimuksissa

Huolellisesti säilytetty data ja sen uudelleenkäyttö vähentävät tarvetta toistaa samoja testejä uusilla koe-eläimillä, erityisesti kontrolliryhmissä. Kehittyvät laskennalliset menetelmät ja havaintoihin perustuva mallintaminen ja simulaatiot voivat vähentää tarvetta entisestään. Tutkimusaineistojen hajanaisuus ja keskenään vaihtelevat tiedostoformaattit ovat kuitenkin merkittävä hidaste tälle.

Julkaise tutkimuksen nollatuloksella tallentamalla tutkimusaineisto

Lähtökohtaisesti kaikki tutkimusaineistot kannattaa pyrkiä säilyttämään, jotta niitä ei tarvitse kerätä uudelleen. Tämä antaa myös mahdollisuuden niin sanottujen negatiivisten tai nollatulosten julkaisuun tutkimusdatan avulla, mikä ei perinteisesti ole ollut mahdollista tutkimusartikkeleissa. Kun säilytät myös sellaisen aineiston, joka ei toiminutkaan odotusten mukaan, edistät tiedon kehittymistä ja tulevia tutkimuksia.

2. Miten valitset säilytettävän aineiston?

1. *Tunnista arvokas aineisto*, jota voidaan hyödyntää muissakin tutkimuksissa.
2. *Tunnista aineisto, joka sinun kuuluu säilyttää.*



Kuva 3: Säilytettävän aineiston tunnuspiirteitä.

Laita säilytykseen **eheää** aineistoa. On tavallista, että tutkimuksen aikana aineistosta luodaan kopioita tai eri versioita, jotka sisältävät raakadataa, analysoinnin välivaiheissa syntyneitä dataa ja tulosdataa. Harvoin säilytetään kaikkea, vaan on mietittävä, mitä pidetään ja mitä tuhotaan, jotta saadaan eheä kokonaisuus ja vältetään turhaa säilyttämistä. Aikaa ja rahaa säästyy, kun säilyttäminen on **suunniteltu** etukäteen aineistohallintasuunnitelmassa ja aineistoa siivotaan jo projektin aikana.

Tunnista arvokas aineisto

Arvokasta aineistoa on esimerkiksi:

- ainutkertainen ja **ainutlaatuinen** aineisto
- aineisto, joka kasvattaa tutkijan tai yliopiston **mainetta**
- aineisto, jota voi hyödyntää **kaupallisesti**
- kiinnostava aineisto, jonka voi **avata** kaikkien käyttöön
- aineisto, jonka avulla voi **parantaa maailmaa**

Aineisto saattaa olla myös hyödynnettävissä **opetuksessa** tai tulevissa **uusissa projekteissa**: ehkä keksitään uusia näkökulmia tai menetelmiä tai tehdään seurantaa, ehkä saadaan uusi rahoitus.

Tunnista aineisto, joka kuuluu säilyttää

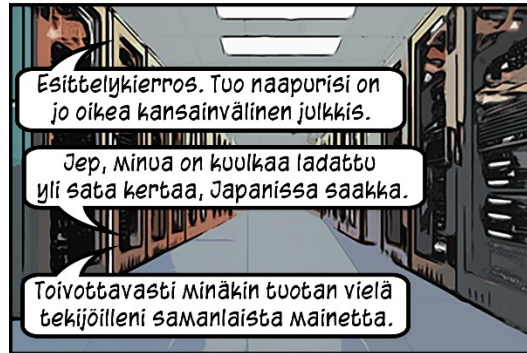
On aineistoa, joka on säilytettävä:

- **lainsäädännön** perusteella
- tutkimustulosten **verifiointia** varten
- tutkimuksen myöhempää **jatkamista** varten
- **oppiaineen suositusten** perusteella
- **informanteille** tehdyn lupauksen perusteella

Lisätietoa säilytettävän aineiston valitsemisesta:

- ❖ Digital Preservation (Fairdata-PAS): [Guidelines for UH Evaluators](#)
- ❖ DCC: [Five steps to decide what data to keep](#)
- ❖ UK Data Service: [Collections development selection and appraisal criteria](#)

Aineiston elämää: uusi aineisto saapuu säilytykseen



UNIVERSITY OF HELSINKI



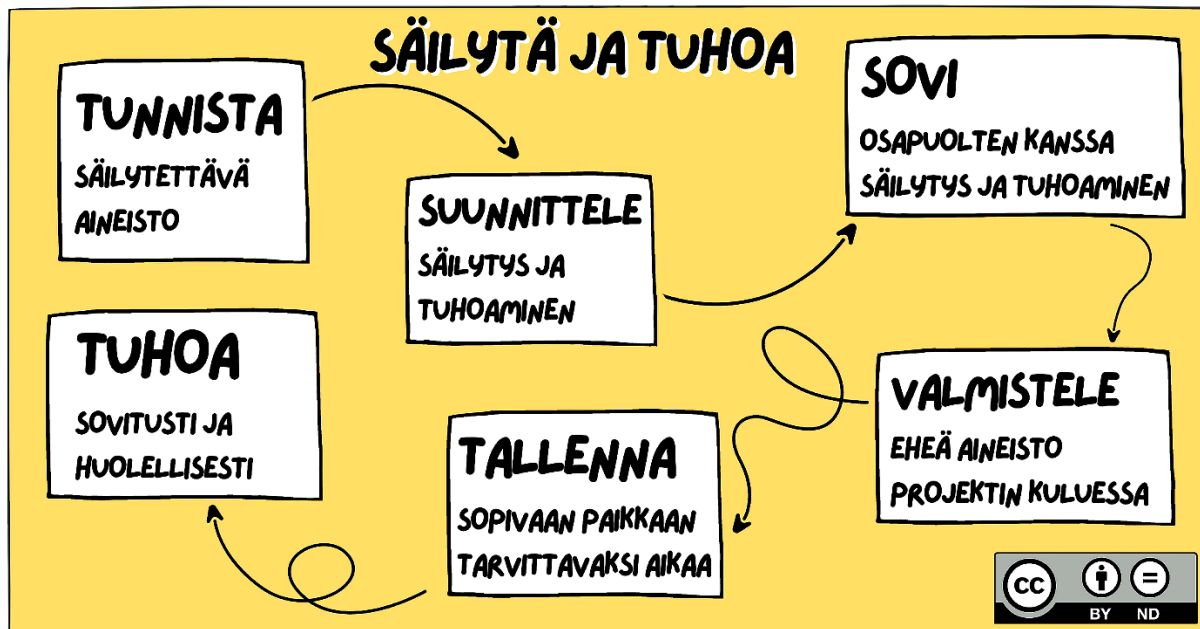
MoADRonte

datasupport@helsinki.fi

Kuva 4: Uusi aineisto saapuu säilytykseen.

3. Säilytysaika ja tuhoaminen

1. Kuinka **pitkään** aineistoa säilytetään?
2. Huolehdi myös aineiston **tuhoamisesta**.



Kuva 5: Säilyttämisen ja tuhoamisen vaiheet.

Kuinka pitkään aineistoa säilytetään?

Yleinen säilytysaika tulosten verifiointiseksi on viisi vuotta, mutta lääketieteessä se voi olla esimerkiksi 15 vuotta, mikäli tutkimuksella on pitkäaikaisvaikutuksia potilastietoihin. Aika siis vaihtelee aineistoittain ja tieteenaloittain. Helsingin yliopistolla ei toistaiseksi ole ohjetta tutkimusaineistojen säilytysajoista.

Esimerkiksi Helsingin ja Uudenmaan sairaanhoitopiiri HUS linjaa seuraavasti:

- Hyvän kliinisen tutkimustavan periaatteiden mukaisesti tutkimusaineistoja säilytetään pääsääntöisesti 15 vuotta tutkimuksen päättymisestä ja vuoden 2021 alusta 25 vuotta.
- Jos kyseessä oleva tutkimus liittyy lääkkeen myyntilupaan EU-alueella, materiaalia on säilytettävä vähintään 2 vuotta sen jälkeen, kun on saatu viimeinen myyntilupa EU-alueella tai 2 vuotta sen jälkeen, kun valmisteen tutkimus on lopetettu.

Ota huomioon mahdolliset rahoittaja- ja alakohtaiset suositukset säilytysajasta. Tarkista myös, miten pitkään kustantaja edellyttää säilyttämään julkaisuun liittyvää aineistoa. **Sovi** säilytysajasta osapuolten kanssa hyvissä ajoin. Mahdollisia osapuolia ovat esimerkiksi tutkimusryhmän jäsenet, tutkittavat, rahoittajat ja kustantajat.

Lisätietoa säilytysajasta:

- ❖ Tietosuojavaltuutetun toimisto: [Linjaus henkilötietojen säilyttämisestä](#)
- ❖ Deakin University: [What data do I need to keep and for how long?](#)

Huolehdi myös aineiston tuhoamisesta

Melkein kaikki aineistot tuhotaan lopulta. Mieti ja päätä, milloin ja miten. Ei ole hyvä toimintatapa, että mitään ei koskaan tuhota ja levytilat täyttyvät tiedostoilla, joita kukaan ei enää pysty hyödyntämään. Aineiston säilytysaikaa ja tuhoamista on syytä suunnitella etukäteen aineistohallintasuunnitelmassa. Tuhoamisesta on tärkeää keskustella koko tutkimusryhmän kesken, jotta vältetään ikäviltä väärinkäsityksiltä. Säilytyspalveluissa aineistolle saatetaan sopia tuhoamispäivämäärä ja myös syy tuhoamiselle.

Erityisesti arkaluonteinen aineisto on hävitettävä huolellisesti, kun sen säilyttäminen ei enää ole tarpeen käyttötarkoituksen kannalta. Pelkkä poistaminen (delete) ei tuhoa dataa.

Lisätietoa aineiston tuhoamisesta:

- ❖ Helsingin yliopiston Helpdesk: [Tiedostojen turvallinen poistaminen](#)

4. Miten aineisto valmistellaan säilytykseen?

Miksi valmistelutyö on tarpeen?

Tutkimusaineiston valmistelun tarkoituksena on varmistaa aineiston **tallennuksen lainmukaisuus, säilyminen ja jatkokäyttö** tulevaisuudessa.



Kuva 6: Valmisteltavia asioita.

4.1. Valmistelussa huomioitavaa

Käyttöoikeudet, sopimukset ja lisenssit

Käyttöoikeuksista on sovittava kaikkien aineiston kokoamiseen osallistuneiden tahojen kesken. Sopimukset on helpointa tehdä jo ennen aineiston keräämistä. Ohjeita tarvittavista sopimuksista löytyy Flammasta. Jos aineisto on saatu käyttöön muualta, täytyy aina varmistaa, millainen oikeus tutkijalla on aineiston säilyttämiseen. Erilaisilla säilytyspalveluilla on omat käyttöehtonsa, joihin kannattaa tutustua ajoissa. Käyttöehdoilla säilytyspalvelut varmistavat, että siellä säilytetään vain aineistoja, johon palvelu soveltuu. Säilytyspalvelut eivät useinkaan ota vastaan arkaluonteista henkilötietoa tai yritysten omistuksessa olevaa aineistoa. Käyttäjän vastuulla on noudattaa näitä ohjeita.

Säilytettävälle aineistolle annetaan lisenssi, jolla kerrotaan muille, minkälaisia käyttöoikeuksia säilytettävään aineistoon myönnetään. Lisenssillä voidaan joko estää tai mahdollistaa aineiston jatkokäyttö.

Uudelleen- vai jatkokäyttöä?

Kun aineistoa käytetään **samaan** käyttötarkoitukseen, puhutaan aineiston uudelleenkäytöstä. Tätä on esimerkiksi verifiointi. Jos käyttötarkoitus on **eri**, kyseessä on aineiston jatkokäyttö. Useimmiten aineistoja säilytetään nimenomaan jatkokäyttöä varten.

Aineiston valmisteluun liittyviä syventäviä ohjeita:

- ❖ Flamma: [Sopimukset aineiston avaamisesta yhteistyökumppanien kesken](#)
- ❖ Flamma: [Oikeuksiensiirtosopimus](#)
- ❖ Helsingin yliopiston kirjasto: [Lisenssit \(CC\), koodien ja ohjelmistojen lisenssit](#)

Tietosuojaja

Ennen aineiston tallentamista tai avaamista on varmistettava, ettei aineisto sisällä sellaisia henkilötietoja tai muuta salassa pidettävää materiaalia, jota ei ole mahdollista tallentaa valittuun palveluun. Tarkista siis ainakin seuraavat seikat:

- Mitä **henkilötietoja** aineisto sisältää? Kuuluvatko henkilötiedot GDPR:n määrittelemiін erityisiin henkilötietoryhmiin?
- Voidaanko aineisto **anonymisoida** tai **pseudonymisoida**?
- Miten tutkittavia on informoitu? Tutkittavien informoinnin yhteydessä tutkittaville kerrotaan aineiston projektinjälkeinen säilytyspalvelu, aineiston jatkokäyttötarkoitus sekä se, kenelle aineisto voidaan luovuttaa myöhemmin. Näiden linjausten tulisi olla muotoiltu siten, että niiden perusteella aineisto on mahdollista säilyttää tutkimuksen jälkeen.
- Sisältääkö aineisto muuta **salassa pidettävää** materiaalia, kuten liikesalaisuuksia tai tiettyjen uhanalaisten eliölajien esiintymistietoa?
- Jos aineistoa ei voida jakaa muiden käyttöön, on sen **metatiedot** kuitenkin hyvä avata.

Lisätietoja tietosuojasta:

- ❖ Helsingin yliopiston Datatuki: [Onko minulla henkilötietoja? -testi](#)
- ❖ Helsingin yliopiston kirjasto: [Tekijänoikeudet ja lisenssit](#)
- ❖ Tietosuojavaltuutetun toimisto: [Erietyiset henkilötietoryhmät](#)

4.2. Säilytyskelpoiset tiedostoformaattit

Avoimet tai **tieteenalalla vakiintuneet** tiedostoformaattit eli tiedostomuodot edesauttavat aineiston säilymistä käytettävänä myös useiden vuosien päästä sekä mahdollistavat laajemmat jatkokäyttömahdollisuudet. Aineistojen jatkokäytön kannalta on tärkeää, että käytettyjen tiedostoformaattien voidaan ennustaa säilyvän käyttökelpoisina myös useiden vuosien päästä. Tällaisia tiedostoformaatteja ovat todennäköisimmin **standardoidut** ja laajasti käytetyt formaattit. Jos aineisto on tallennettava jonkin tieteenalan erikoisformaattiin, tällöin on hyvä varmistua siitä, että kyseinen formaatti on tieteenalalla vakiintunut ja hyvin dokumentoitu.

- Käytä **avoimia** ja yleisti käytettyjä tiedostoformaatteja.
- Tarkista alallasi käytetystä aineiston säilytyspalvelusta, mitä tiedostoformaatteja ne ottavat vastaan.
- Voit käyttää tieteenalallasi vakiintuneita erikoisformaatteja, jos ne ovat hyvin dokumentoituja. Riittävän dokumentaation löytää usein Wikipediasta.

Lisätietoja säilytyskelpoisista tiedostomuodoista:

- ❖ Wikipedia: [Lista avoimista tiedostoformaateista](#) (englanniksi)
- ❖ Digital Preservation: [Pitkäaikaissäilytykseen soveltuvat tiedostomuodot](#) (PDF, suositus yli 25 vuotta säilytettäville aineistoille)

Säilytettävän aineiston kuvailu ja metatiedot

Kattavat kuvailutiedot auttavat muita löytämään, ymmärtämään ja käyttämään aineistoa. Pyri järjestämään aineisto niin, että siitä muodostuu **itsensä selittävä, eheä kokonaisuus**.

- Järjestä tiedostot selkeän kansiorakenteen mukaan.
- Pyri tiedostojen ja kansioden nimeämisessä selkeyteen ja yksiselitteisyyteen.

- Huolehdi, että aineistossa käytetyt muuttujat, lyhenteet sekä koodit on selitetty ymmärrettävästi.
- Täydennä aineiston kuvailua laatimalla **README**-tiedosto.
- Valitse aineistolle säilytyspalvelu, joka mahdollistaa antamaan aineistolle on riittävät löydettävyyttä tukevat metatiedot ja pysyvät tunnisteet, kuten DOI:n.

Lisätietoja aineiston kuvailusta:

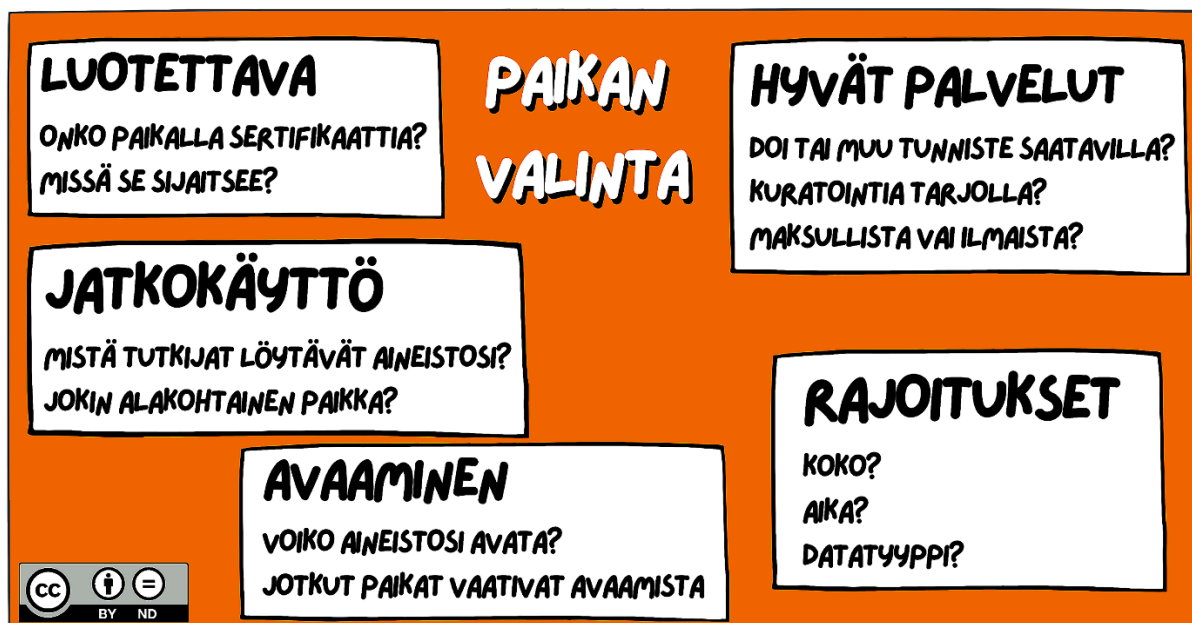
- ❖ Fuchs & Kuusniemi (2018): [Making a research project understandable](#)
- ❖ Esimerkki löydettävyyttä tukevista minimimetatiedoista: [DataCite Schema – Mandatory fields](#)

5. Missä aineistoja säilytetään?

Tutkimusdatalle löytyy laaja valikoima säilytyspalveluita, jotka voidaan karkeasti jakaa yleisiin ja tieteenalakohtaisiin. Lisäksi organisaatioilla voi olla oma säilytyspalvelu tutkijoiden tai tutkimusinfrastruktuurien tuottamille aineistoille.

Tieteenalakohtaiset data-arkistot, repositoriot ja tietokannat ovat kehittyneet tietyn tutkimusyhteisön tai datatyypin tarpeista. Näin ollen ne tarjoavat hyvän lähtökohdan säilytyspalvelua valittaessa. Toinen tärkeä lähtökohta on palvelun luotettavuus. Joissain tapauksissa organisaation tarjoama säilytyspalvelu voi olla sopivin ratkaisu.

Helsingin yliopisto tarjoaa vuoden 2024 alusta HY Datapankki säilytyspalvelua, jonne tutkimusaineistoja voi tallentaa 5–15 vuodeksi tarpeen ja sopimuksen mukaan. Aikaa voidaan hakemuksesta myös jatkaa. Yleisesti käytetty tallennuspalvelu Zenodo lupaa tallennusta 25 vuodeksi. Tutkimusaineistoille, joiden tulisi säilyä tuleville sukupolville, on olemassa Fairdata PAS-palvelu. Siellä aineistot säilyvät käyttökelpoisina jopa satoja vuosia aktiivisen aineiston hoivaamisen ansiosta.



Kuva 7: Ota nämä asiat huomioon, kun valitset säilytyspaikkaa.

5.1. Periaatteita säilytyspalvelun valintaan

Tiedon säilyttämisen ammattilaiset jakavat palvelut karkeasti kahteen kategoriaan, tutkimusaineistoja tallentaviin (*eng. storing*) ja säilyttäviin (*eng. preserving*) palveluihin. Tallentavat palvelut säilyttävät tiedostot juuri sellaisina kuin ne tallennettiin palveluun. Säilyttävät palvelut taas pystyvät pitämään tietoa ymmärrettävänä, vaikka tiedostoformaattit ja ohjelmistot ajan kuluessa muuttuvat. Alanmukaiset, vain tietyssä formaatissa aineistoa hyväksyvät palvelut, ovat usein tietoa säilyttäviä palveluita, kun taas yleiset, kaikenlaisessa muodossa dataa hyväksyvät palvelut, pystyvät vain tallentamaan datan sellaisenaan. Mitä pidempään aikaa kuluu, sitä todennäköisempää on, että käytettävien ohjelmistojen kehittyessä tallennetut tiedostot muuttuvat käyttökelvottomiksi. Jos siis halutaan tiedon säilyvän hyvin pitkään, on syytä suosia säilyttäviä palveluita.

Ensisijaisesti kannattaa suosia yleisesti tunnistettuja, tieteenalakohtaisia data-arkistoja. Myös sertifikaatin saaneet data-arkistot tarjoavat hyvän vaihtoehdon. Kaikissa tapauksissa sopivan säilytyspalvelun löytäminen aineistolle ei ole itsestään selvää. Tällöin apuna voidaan käyttää erilaisia periaatteita, joita olemme listanneet alempana hyödyntäen viittä eri kategoriaa: luotettavuus, sopivuus, jatkokäyttö, säilytysaika ja avoimuuden tasot. Lisäksi voit käyttää apuna säilytyspalvelun valinnan vuokaaviota, joka löytyy kohdasta 5.3.

Luotettavuus ja palvelut (sertifikaatti, kuratointi, pysyvä tunniste):

- Tarkista, antaako valitsemasi data-arkisto pysyvän tunnisteiden, kuten DOI:n (Digital Object Identifier) tai arkistotunnuksen (Accession number). Pysyvä tunniste on koneellisesti tulkittava, ainukertainen objektin yksilöivä merkkijono, joka mahdollistaa tutkimusaineiston löydettävyyden ja siihen viittaamiseen.
- Valitsemalla kuratoidun data-arkiston voit edesauttaa datasi jatkokäyttöä. Osaan data-arkistoista tutkijat tallettavat datan itse, ilman laadunvarmistusta. Kuratoiduissa data-arkistoissa datan laatu varmistetaan muun muassa tarkistamalla tiedostomuotojen toimivuus ja rikastamalla metatietoja.
- Muista tarkistaa data-arkiston maksullisuus, kuten kuratointimaksu tai maksimikoon ylittävien datasettien lisämaksut.

- Tarkista onko harkitsemallasi data-arkistolla sertifiointia, kuten CoreTrustSeal. CoreTrustSeal-sertifiointi myönnetään arkistoille, joiden säilytyskäytännöt on todettu riittäväksi ulkopuolisten arvioijien toimesta. Arvioinnissa tarkistetaan muun muassa kuratoinnin taso ja palvelun jatkuvuuden takeet.
- Data-arkiston taustalla toimivan palvelimen maantieteellisellä sijainnilla voi joissain tapauksissa olla merkitystä datan säilyttämisen kannalta. Henkilötietoja ei saa siirtää EU:n ulkopuolelle, mikäli tästä ei ole informoitu tutkittavia.

Uudelleenkäyttö ja jatkokäyttö (tieteenalakohtaisuus, metatiedot):

- Tarkista onko tutkimusalallasi jokin yleisesti käytetty data-arkisto. Valitsemalla tieteenalakohtaisen data-arkiston edistät tehokkaampaa datan uudelleenkäyttöä, jatkokäyttöä ja löydettävyyttä.
- Yleiset data-arkistot ottavat vastaan eri tieteenalojen dataa ja erilaisia datatyyppejä. Ne tarjoavat hyvän vaihtoehdon eritoten silloin, kun tutkimusalalla ei ole vakiintunutta, tieteenalakohtaista data-arkistoa.
- Jos data ei sovi yleiseen data-arkistoon esimerkiksi liian suuren kokonsa puolesta, organisaatiokohtainen säilytyspalvelu voi olla sopivin ratkaisu.
- Tarkista ajoissa, mitä vaatimuksia valitsemallasi data-arkistolla on aineiston kuvailun suhteen (muun muassa asiasanat, käytetyt menetelmät, muuttujien kuvaus).

Sopivuus (kokorajoitukset, datatyypit, tiedostomuodot):

- Tutkimusaineiston koko vaikuttaa arkiston valintaan — muista tarkistaa mahdolliset kokorajoitukset ja lisämaksut.
- Tarkista, sopiiko arkisto tarvitsemallesi aineistotyyppille tai tiedostomuodolle. Kaikki data-arkistot eivät välttämättä vastaanota esimerkiksi koodia tai ohjelmistoja.

Säilytysaika (minimisäilytysajat, pitkäaikaisuus):

- Data-arkistojen antamat tiedot säilytysajoista vaihtelevat. Tarkista, onko arkisto ilmoittanut sivuillaan säilytyksen pitkäaikaisuudesta tai minimisäilytysajoista.

Avoimuuden tasot (käyttöoikeudet, julkaisuviiveet):

- Useat data-arkistot tukevat avoimia lisenssejä. Osa arkistoista edellyttää talletuksen ehtona CC0-lisenssiä. CC0-lisenssi on Creative Commons (CC) -lisensseistä avoin, eli se ei sisällä käyttörajoituksia ja näin ollen tekijä luovuttaa kaikki oikeudet aineistoon.
- Jos et voi avata dataa ja tarvitset sille kuitenkin pitkäaikaisen säilytyspaikan, tarkista, voiko valitsemassasi data-arkistossa säilyttää data-aineistoa niin, että sitä avataan muille vain rajatuilla käyttöoikeuksilla, kuten tutkimusluvalla.
- Jos tarvitset datalle julkaisuviiveen ennen sen avaamista, tarkista, että valitsemasi data-arkisto tarjoaa mahdollisuuden julkaisuviiveeseen.

Lisätietoja säilytyspalvelun valintaan:

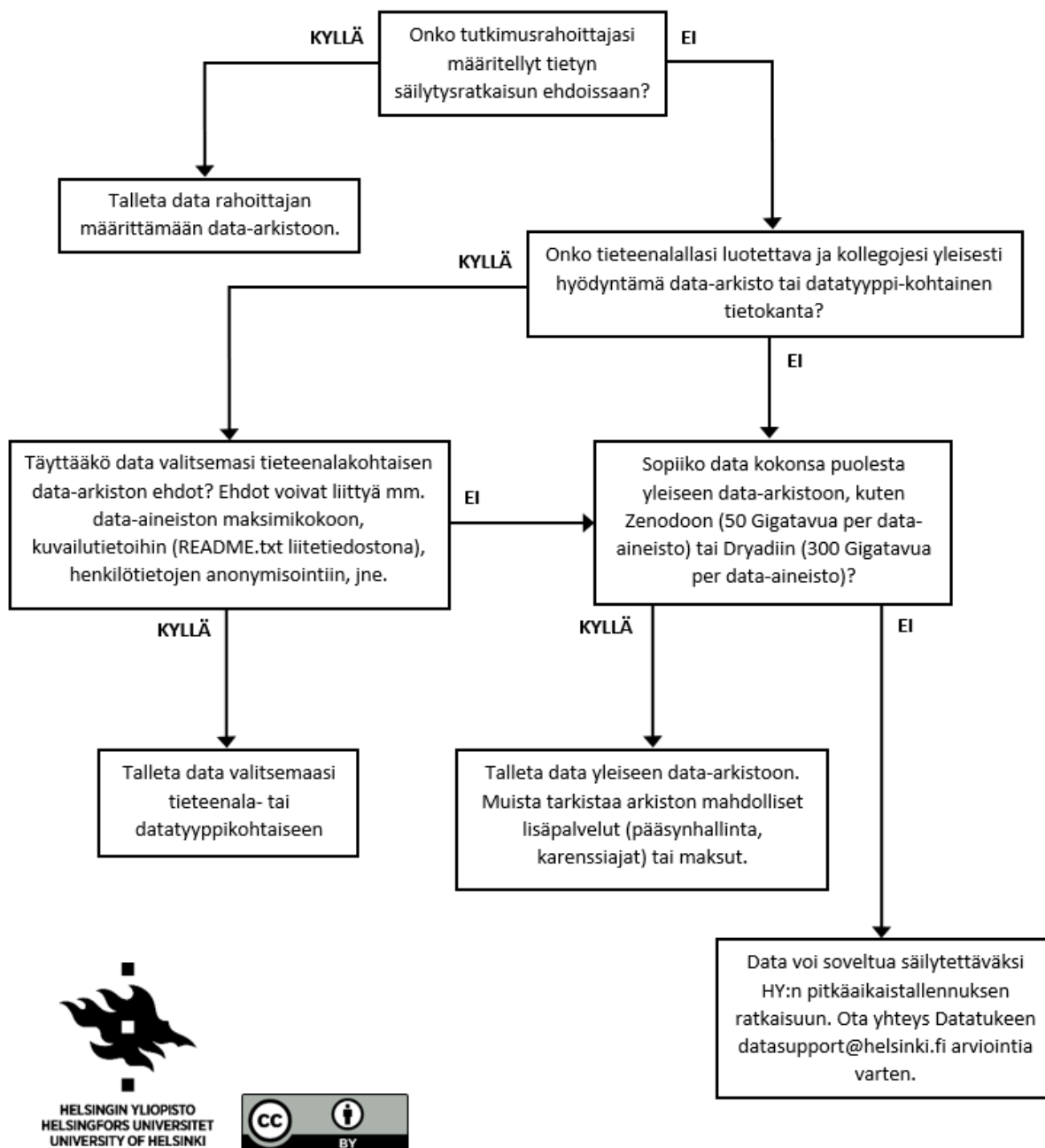
- ❖ Zenodo data-arkisto: [Usein kysytyt kysymykset](#) (englanniksi)
- ❖ Dryad data-arkisto: [Usein kysytyt kysymykset](#) (englanniksi)
- ❖ The Software Sustainability Institute: [Software Deposit: Guidance for Researchers](#)
- ❖ CoreTrustSeal: [CoreTrustSeal-kriteerit](#) (englanniksi)

5.2. Miten löydän sopivan säilytyspalvelun?

Tieteenalakohtaisen data-arkiston etsinnässä kannattaa hyödyntää tutkimusyhteisöä. Kysy kollegoiltasi, mitä arkistoja he yleensä käyttävät tai tarkista alasi keskeisistä artikkeleista, missä dataa on avattu. Kannattaa huomioida, että yleisenä käytänteenä pidetty datan avaaminen pelkkänä artikkelin lisäliitteenä (*eng. supplementary material*) ei ole omiaan edistämään datan löydettävyyttä tai jatkokäyttöä.

Tieteenalakohtaisen data-arkiston löytämistä helpottaa hakualusta Registry of Research Data Repositories ([re3data](#)).

5.3. Data-arkiston valinnan vuokaavio



Kuva 8: Data-arkiston valintaa ohjaava vuokaavio.

6. Ohjeen kootut linkit

1. Miksi tutkimusaineistoja säilytetään

- ❖ Avoin tiede: [Mitä on avoin tiede?](#)
- ❖ Unesco (2021): [Recommendation on Open Science](#)

Vastuullisen säilyttämisen periaatteet:

- ❖ Helsingin yliopisto: [Luottamuksellisten aineistojen pitkäaikaissäilytys](#)
- ❖ Avoin tiede: [Sensitiivisten aineistojen vastuullinen avaaminen](#)
- ❖ Vastuullinen tiede: [Turvaa tutkittavan anonymiteetti!](#)
- ❖ Tietoarkisto: [Tunnisteellisuus ja anonymisointi](#)

Käyttöesimerkkejä:

- ❖ Helsingin yliopiston tutkimuksen lakipalvelut: tutkimuksenjuristit@helsinki.fi
- ❖ Helsingin yliopiston tuki tietosuojaan (GDPR) liittyen: tietosuoja@helsinki.fi
- ❖ Julkaisuviive: [Sherpa Romeo](#)

2. Miten valitset säilytettävän aineiston?

Säilytettävän aineiston valitseminen:

- ❖ Digital Preservation (Fairdata-PAS): [Guidelines for UH Evaluators](#)
- ❖ DCC: [Five steps to decide what data to keep](#)
- ❖ UK Data Service: [Collections development selection and appraisal criteria](#)

3. Säilytysaika ja tuhoaminen

Säilytysaika:

- ❖ Tietosuojavaltuutettu: [Säilytyksen rajoittaminen](#) (henkilötietojen säilyttämisestä)
- ❖ Deakin University: [What data do I need to keep and for how long?](#)

Aineistojen tuhoaminen:

- ❖ Helsingin yliopiston Helpdesk: [Tiedostojen turvallinen poistaminen](#)

4. Miten aineisto valmistellaan säilytykseen?

Aineiston valmistelun syventävät ohjeet:

- ❖ Flamma: [Sopimukset aineiston avaamisesta yhteistyökumppanien kesken](#)
- ❖ Flamma: [Oikeuksiensiirtosopimus](#)
- ❖ Helsingin yliopiston kirjasto: [Lisenssit \(CC\), koodien ja ohjelmistojen lisenssit](#)

Tietosuoja:

- ❖ Helsingin yliopiston Datatuki: [Onko minulla henkilötietoja? -testi](#)
- ❖ Helsingin yliopiston kirjasto: [Tekijänoikeudet ja lisenssit](#)
- ❖ Tietosuojavaltuutetun toimisto: [Erietyiset henkilötietoryhmät](#)

Säilytyskelpoiset tiedostomuodot:

- ❖ Wikipedia: [Lista avoimista tiedostoformaateista](#) (englanniksi)
- ❖ Digital Preservation: [Pitkäaikaissäilytykseen soveltuvat tiedostomuodot](#) (PDF, suositus yli 25 vuotta säilytettäville aineistoille)

Aineiston kuvailu:

- ❖ Fuchs & Kuusniemi (2018): [Making a research project understandable](#)
- ❖ Esimerkki löydettävyyttä tukevista minimimetatiedoista: [DataCite Schema – Mandatory fields](#)

5. Missä aineistoja säilytetään?

Periaatteita säilytyspaikan valintaan:

- ❖ Zenodo data-arkisto: [Usein kysytyt kysymykset](#) (englanniksi)
- ❖ Dryad data-arkisto: [Usein kysytyt kysymykset](#) (englanniksi)
- ❖ The Software Sustainability Institute: [Software Deposit: Guidance for Researchers](#)
- ❖ CoreTrustSeal: [CoreTrustSeal-kriteerit](#) (englanniksi)

Miten löytää sopiva säilytyspaikka?

- ❖ Data-arkistojen hakualusta: [re3data](#)