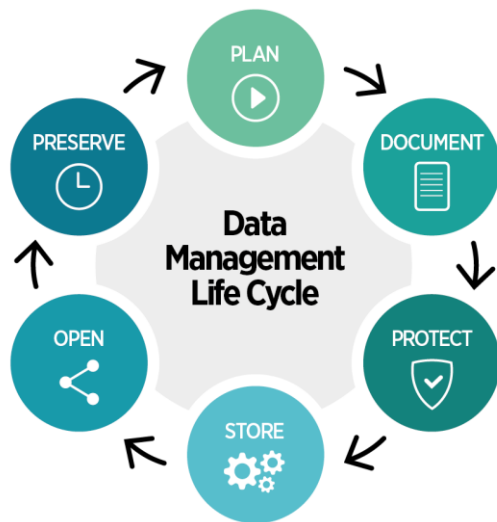


Research Data Management

Datanhallinnan perusteet

Viikki
2021

Tanja Lindholm, Liisa Siipilehto
University of Helsinki, Data Support
datasupport@helsinki.fi

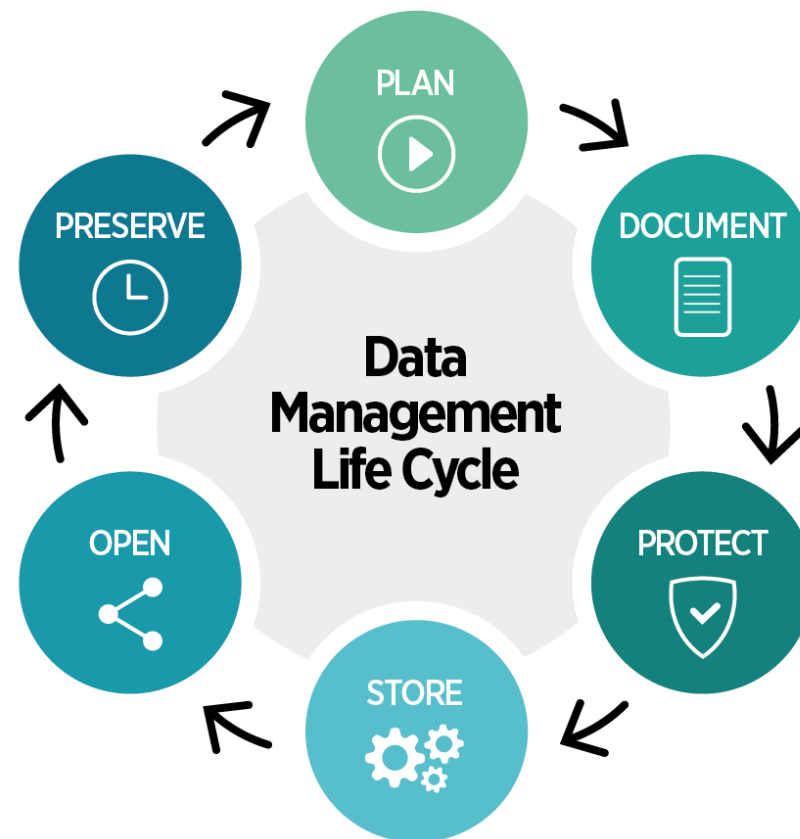


Sisältö

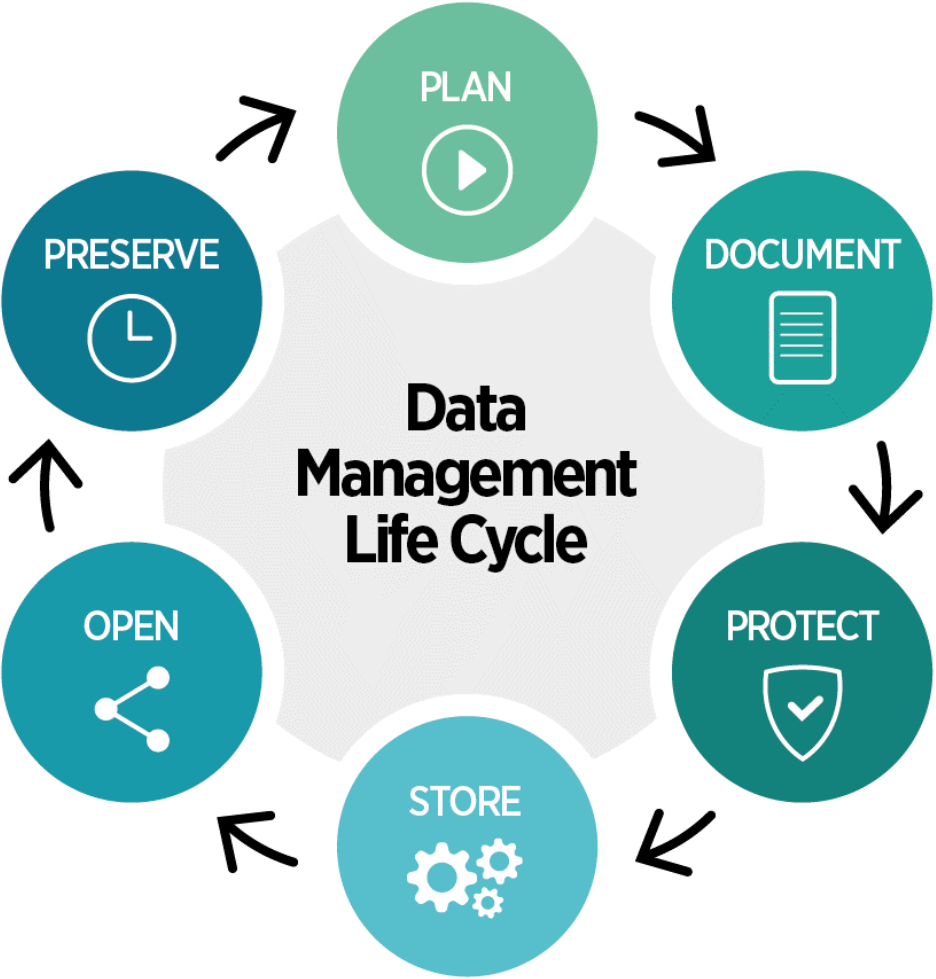
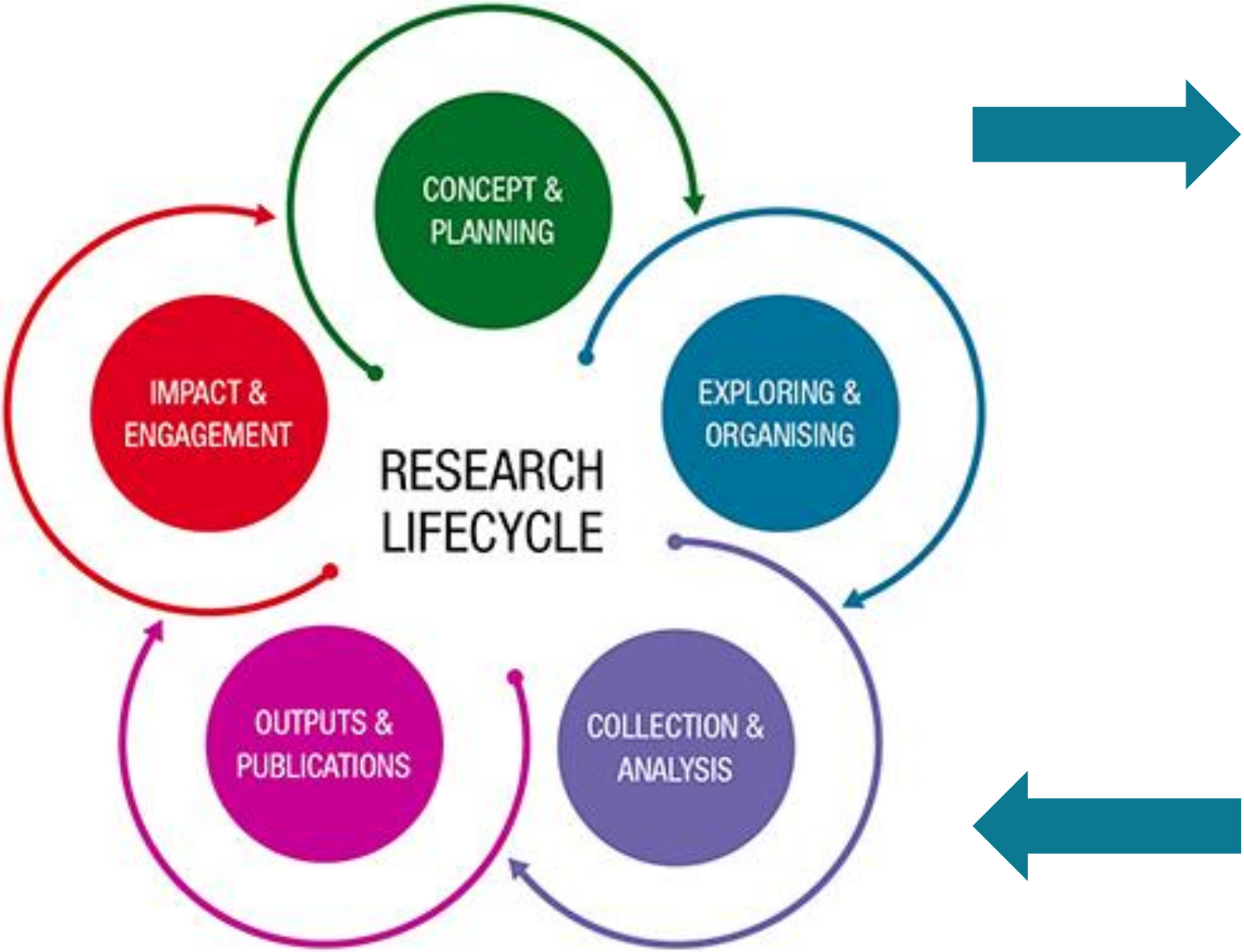
Miksi ja Mitä – Tutkimus Datan Hallintaa

1. Aineistot
2. Eettiset periaatteet ja lainsäädäntö
3. Dokumentaatio
4. HY:n tallennuspalvelut
5. Aineiston avaaminen, julkaiseminen ja arkistointi
6. Aineistohallinnan vastuut ja resurssit

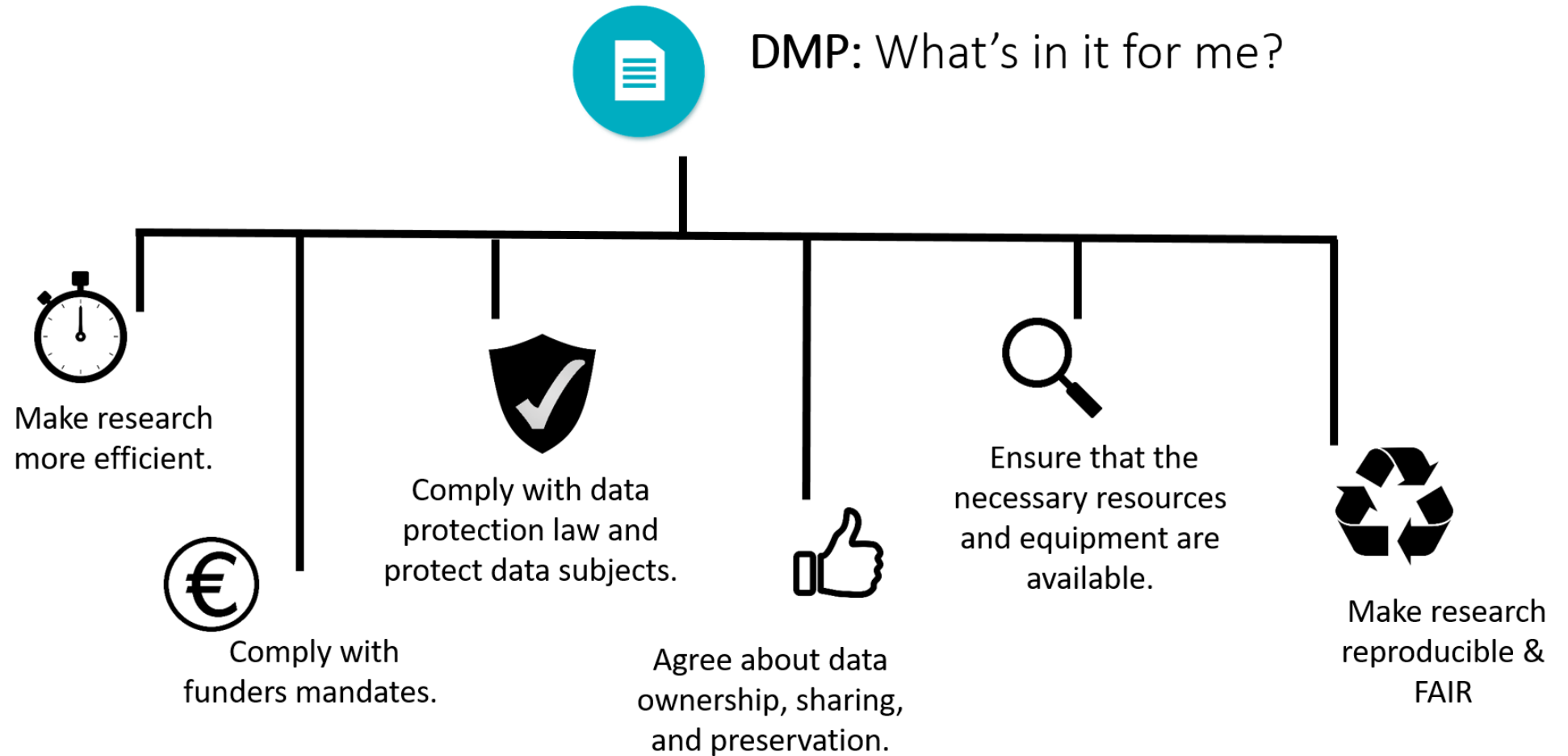
Hyvä tietää! (Linkejä ja muuta)



Mitä datanhallinta on?



Miksi aineistohallintaa?





1. Aineistot

1. Aineistot

- Tutkimus materiaali = / ≠ data, yleinen termi vai ainoastaan digitaalinen materiaali
 - Eri metodein kerättävä aineisto
 - (Biologiset) näytteet, mittaukset, kartoitukset, haastattelut, kuvat, kokoelmat jne.
 - Tutkimuksen tuloksena kerättävä aineisto
 - Analyysien tulokset, sekvenssit, päiväkirjat, koodit jne.
 - Aiemmin kerätty aineisto/Avoimen aineistot
 - Biopankin näytteet, arkistomateriaalit, repositorioiden materiaalit, kartta-aineistot, koodit jne.

1. Aineistot – Minun Aineistot

Data type	Source of the data	File format	Size estimate
Majavien pesät	Re-used	csv	
MS-NFI	Re-used	tif	
Corine Land Cover 2012	Re-used	tif, shp	
Digital map	Re-used	shp	
Habitat variables	Produced	csv, shp, tiff	

2. Eettiset periaatteet ja lainsäädäntö

Mitä eettisiä seikkoja aineistosi hallintaan liittyy?

Miten aineiston omistajuuteen, tekijänoikeuksiin ja immateriaalioikeuksiin liittyviä asioita hallitaan?



2. Eettiset periaatteet ja lainsäädäntö

Miten eettisiä asioita hallitaan?

- Sisältääkö aineistosi henkilötietoja?
- Teetkö **eläinten** kanssa töitä, jotka vaativat [eettisen](#) luvan?
- Käsitteletkö muuta **salassa pidettävää** tai **luottamuksellista aineistoa**
 - Arkaluontoinen lajitieto, esim. uhanalaiset eläimet
 - Muu luottamuksellinen tieto, esim. maanpuolustukseen liittyvä tieto, liikesalaisuudet, bioturvallisuus
 - *Aineiston arkaluontoisuutta on tarkasteltava tapauskohtaisesti*

2. Eettiset periaatteet ja lainsäädäntö – Henkilötiedot

- EU:n tietosuoja-asetus ([GDPR](#)) asettaa minimivaatimukset sille, miten henkilötietoja pitää käsitellä EU-maissa.
- [6 kohtaa jotka antavat oikeuden käsitellä henkilötietoja:](#)
 - A. Rekisteröidyn suostumus, jolla rekisteröity hyväksyy henkilötietojensa käsittelyn
 - B. Yleistä etua koskeva tieteellinen tai historiallinen tutkimus tai tilastointi
- Tietosuoja-asetus ei koske kuolleita henkilöitä tai anonymisoitua aineistoa

Tunnisteelliset tiedot = kaikki ne tiedot, joista henkilö on yksilöitävissä joko suoraan tai epäsuorasti.


- *Suorat tunnisteet:* nimi, puhelinnumero, sosiaaliturvatunnus, kuva, ääni, sormenjälki, hammaskartta
- *Epäsuorat tunnisteet:* sukupuoli, ikä, koulutus, ammatti, kansalaisuus, työhistoria, järjestelmän logitiedot siviilisääty, asuinpaikka, rekisterinumero

2. Eettiset periaatteet ja lainsäädäntö – Henkilötiedot vs Arkaluontoiset henkilötiedot

- Osa **henkilötiedoista** on **arkaluontoista** ja sen käsittelyä määrää GDPR:n artiklat 9 ja 10 ([Eriyisten henkilöryhmien käsittely](#))

1. Rotu tai etninen alkuperä
2. Poliittinen mielipide
3. Uskonnollinen tai filosofinen vakaumus
4. Ammattiliiton jäsenyys
5. Geneettinen tai biometrinen henkilön tunnistamista varten
6. Terveyttä koskeva tieto
7. Seksuaalinen suuntautuminen tai käyttäytyminen
8. Rikosrekisteri

Tutkimuksen tekijöiden vastuulla on tunnistaa tiedot, joiden paljastumisesta saattaisi olla haittaa tutkittaville.



DNA strands that are unique/ whole genome, MRI images, bones etc.
→ Any character that links the data to only one person.

2. Eettiset periaatteet ja lainsäädäntö – Arkaluontoiset henkilötiedot

Anonyymi tieto:

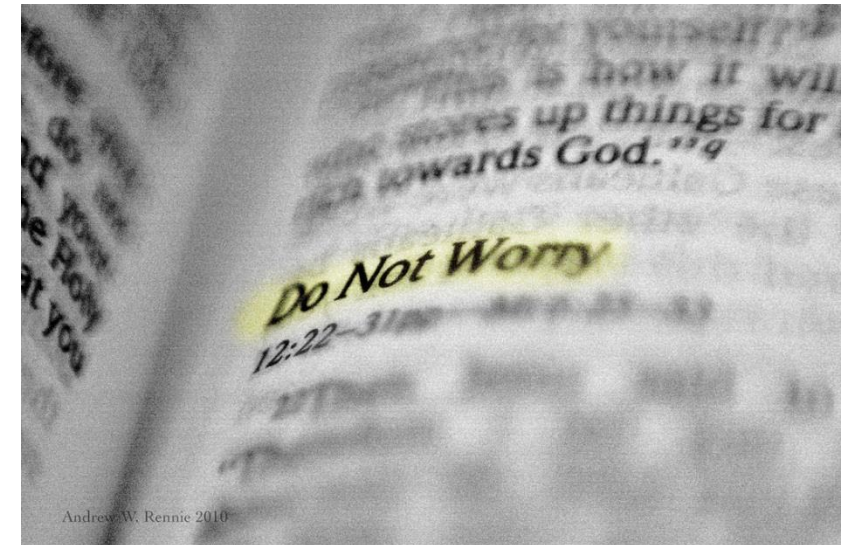
Yksittäinen havainto ei ole kohtuullisin keinoin **tunnistettavissa** annettujen tietojen perusteella tai niitä muihin yhdistämällä. Anonyymi tieto ei ole henkilötietoa, eikä tietosuoja-asetus koske sitä.

Pseudonyymi tieto:

Yksittäinen havainto ei ole **tunnistettavissa pseudonymisoitujen tietojen perusteella ilman ulkopuolisia lisätietoja**. *Pseudonymisointi* on aineiston tunnistellisten tietojen poistamista tai korvaamista peitetiedolla tai koodeilla, jotka prosessin jälkeen säilytetään erillään aineistosta ja suojataan organisatorisesti ja teknisesti. Aineisto on pseudonyymi niin pitkään kuin erillään oleva tieto tuhotaan.

2. Eettiset periaatteet ja lainsäädäntö – Henkilötiedot vs Arkaluontoiset henkilötiedot

- Opastus ja tuki
 - [Tutkijan tietosuoja-asiat HY](#) (Flamma)
 - [Tietosuoja Yammer](#) (Avoin tietosuojan tukiryhmä)
 - [Tietoarkisto henkilötiedot](#)



2. Eettiset periaatteet ja lainsäädäntö – Arkaluontoinen aineisto

Muista huomioida, kun käsittelet arkaluontoista aineistoa:

Tallennus

- Tarpeeksi turvallinen säilytyspaikka
- Salaus
- Pseudonymisaatio, anonymisaatio

Julkaisu

- Ainoastaan anonymisoitu aineisto voidaan julkaista!
- Metadatan voi aina julkaista!

2. Eettiset periaatteet ja lainsäädäntö – Lainsäädäntö ja sopimukset

Mitä on sovittu aineiston omistajuudesta, tekijänoikeuksista ja immateriaalioikeuksista?

Kuka on tuottanut eri aineistot?

- Kerääjä / tuottaja omistaa aineiston jos muuta ei ole sovittu.
- Useat rahoittajat vaativat aineiston omistajuuden siirtämistä Helsingin Yliopistolle
 - HY **ei omista** aineistoa ellei siitä ole sovittu erikseen
 - PI:t ovat vastuussa sopimuksista.
- [Instruction on concluding an agreement](#)
- tutkimuksenjuristit@helsinki.fi

2. Eettiset periaatteet ja lainsäädäntö – Aineiston omistajuus, IPR ja sopimukset

Mitä on sovittu aineiston omistajuudesta, tekijänoikeuksista ja immateriaalioikeuksista?

Kuka on **tuottanut** aineistot?

→ Mitä on sovittu *aineistoista / koodeista / ohjelmistoista*?

- Kenellä on aineiston käyttöoikeudet ja kuka vastaa aineiston uudelleen käytöstä?
- Kuka päättää aineiston avaamisesta ja säilytyksestä?
- Onko kirjottajuuksista sovittu?

→ **Tee aina kirjallinen sopimus kaikista edellämmainituista kohdista**

2. Eettiset periaatteet ja lainsäädäntö – sinun aineistosi?

Data type	Sensitivity+ controller	Owner
Majavien pesät	No	LUKE/KK
MS-NFI	No	LUKE/Avoin aineisto
Corine Land Cover 2012	No	SYKE/Avoin aineisto
Digital map	No	Maanmittauslaitos/Avoin aineisto
Habitat variables	No	Minä itse

3. Data dokumentaatio

= tarkoittaa datan kuvailua

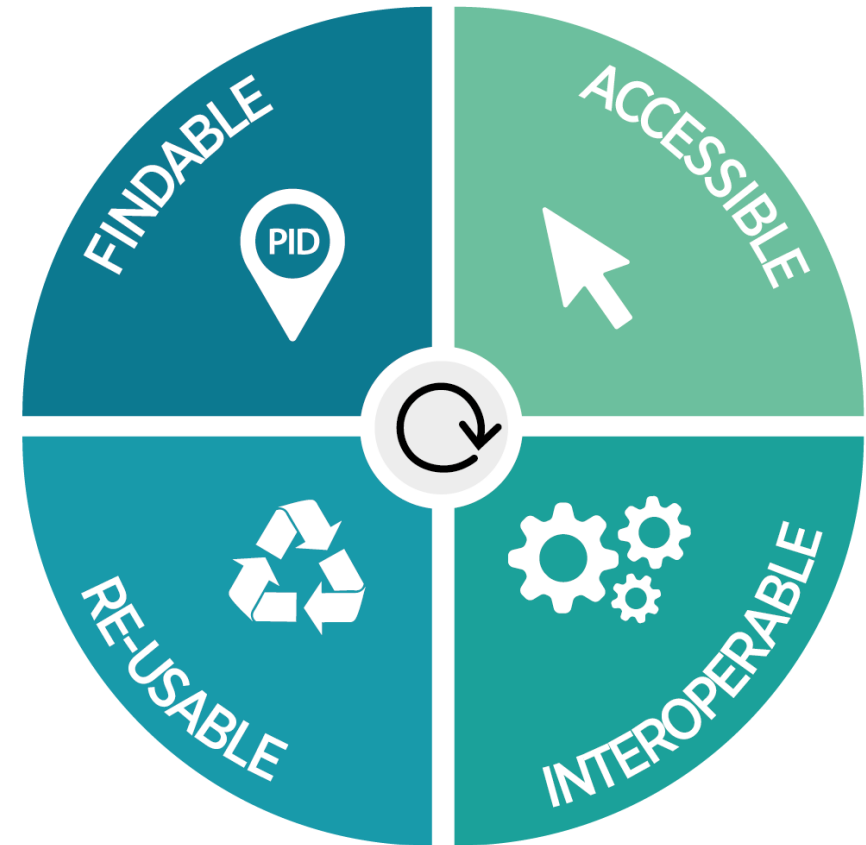


Miten varmistat että aineistosi on “FAIR” (löydettävä, saavutettava, yhteensopiva ja uudelleenkäytettävä) niin itsellesi kuin muille?

FAIR -periaatteet

Periaatteet tarjoavat ohjausta tieteelliseen datanhallintaan ja samalla edistävät tutkimusdatan käyttöä.

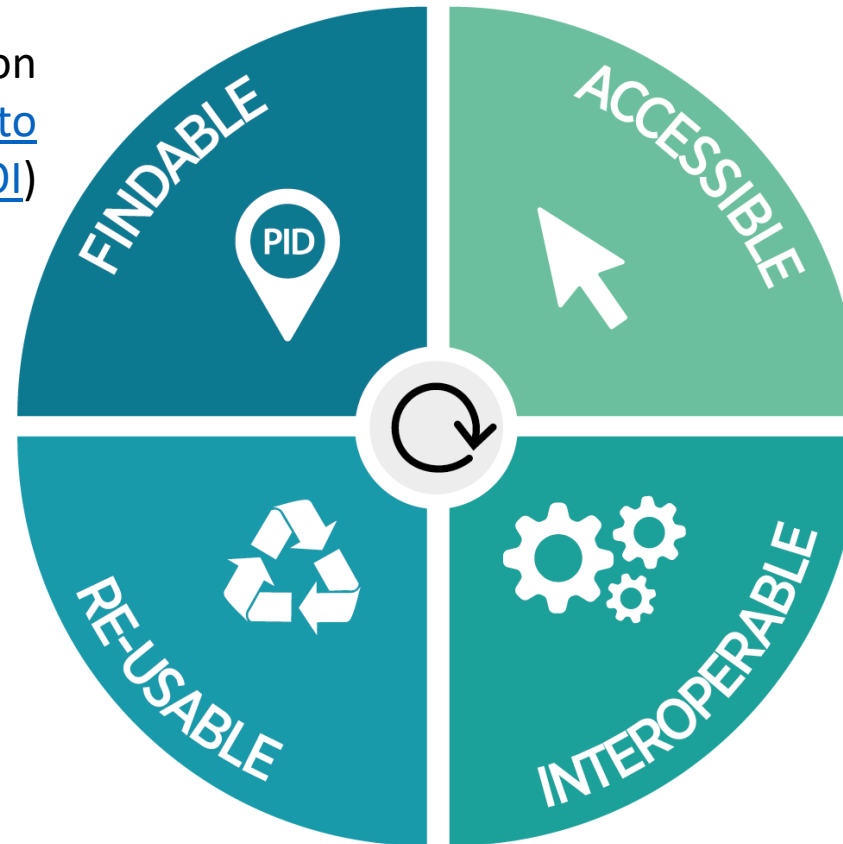
- EU:n neuvoston linjaus: Aineiston optimaalinen uudelleenkäyttö on mahdollista ainoastaan, kun aineisto on yhdenmukainen FAIR-periaatteiden kanssa
- Opetus- ja kulttuuriministeriö on sitoutunut FAIR-periaatteisiin



FAIR -periaatteet

Datalla ja sitä täydentävällä materiaalilla on riittävän runsas [Metatieto](#) sekä pysyvä tunniste (mm. [DOI](#))

Datalla on selkeä käyttölisenssi sekä tietoa sen alkuperästä.



Resurssi tai **metatieto** helposti noudettavissa.
Data on luotettavassa repositoriassa.
Ei vaadi maksullisia ohjelmistoja

Avoimet ja käytettävät tiedostomuodot.
Aineiston tulisi olla sekä ihmis- että koneluettavaa

3. Data dokumentaatio

= aineiston kuvailu

Tieto tiedosta = **metadata**

Tieto tiedosta. Aineiston sisältö ja rakenne. Lista muuttujista ja niiden kuvailu. Mitä aineistokansiot ovat, mistä ne löytyä, miten ne ovat nimetty, mitä muuttujat tarkoittavat

Projektin metadata= Aineiston “Merkki” , joka tulisi aina julkaista. Mitä on kerätty, milloin on kerätty ja kuka on kerännyt

Löydettävä metadata = Datasetin “merkki” + Pysyvä tunniste + Tutkimuksen liitteet (mm. hallinnolliset dokumentit, ja muut tutkimukseen liittyvät kuvailut)



3. Data dokumentaatio

- *Suunnittele jo aineistonhallinnansuunnitelman yhteydessä ennen aineistojen keräämistä*
- Ennakoi tulevaa, ei ole suuri tehtävä kun sen aloittaa ajoissa!
- Miksi?
 - Sinä ja muut ymmärtävät aineistosi=aineiston jakaminen on helpompaa.
 - Huolelliseen dokumentaatioon käytetty aika, säästää aikaa sekä projektin aikana, että sitä julkaistaessa.
 - Competitive advantage!

Standardoidut käytännöt lisäävät aineiston arvoa ja jatkokäytön mahdollisuuksia

3. Data dokumentaatio

1. Jos mahdollista, käytä [metadata standards](#) ja tunnettuja sanastoja.

- Kuvaile data käyttäen tunnettuja sanastoja
- Käytä alan sanastoa ja standardeja jos ne ovat olemassa
- Katso repositorion vaatimukset metadatan suhteen
- Mistä löydän:
 - [Digital curation centre](#) (DCC) has gathered discipline specific metadata standards
 - [EMBL-EBI Ontology](#) lookup service is a repository for biomedical ontologies.

3. Data dokumentaatio

2. If available, use [data management software](#), to make documenting easier.

- Software take various data in & convert it into a database. Metadata is generated automatically when new inputs are made.
 - Easy error spotting: inputs out of range can be automatically detected .
- Electronic laboratory notebooks ([Splice bio article 2019](#))
 - Easy to share & control access, usually have safe storage & search tools
 - E.g. [Scinote](#), [Benchling](#), [Rspace](#)



3. Data dokumentaatio

3. Tutustu seuraaviin menetelmiin ja valitse itsellesi sopiva:

Data dictionaries and Code books	Dictionaries explain variables used in a dataset. Codebooks are collections of codes, algorithms and calculations used
Directory structure	Create a folder structure to suit your project needs
Tagging files	Tags are keywords assigned to files, which enable organizing and searching files easier
File naming conventions	Create a meaningful but brief system with unique names
Version control	Automatic version control system preferred
Readme-files	Readme-files are text documents (e.g. format.txt) providing information about data files to ensure they are interpreted correctly

3. Data dokumentaatio

”Pitäisi toimia kuin jokainen aloitettu mittaus jatkuisi ikuisesti mutta mittauksen aloittajat ja datavirran ylläpitäjät siirtyisivät muihin töihin seuraavalla viikolla. Kaikki dokumentointi pitäisi tehdä muille, ei itselle.”

Pasi Kolari [Think Open Blog](#)

PDF-opas:

<https://doi.org/10.5281/zenodo.1683181>



4. HY:n tallennuspalvelut

3. Data dokumentaatio – säilytys?

Data type	Sensitivity + controller	Metadata/ Documentation	Storage during project
Majavien pesät	No	None?	UH Personal folder
MS-NFI	No		Rajapinta
Corine Land Cover 2012	No		Rajapinta
Digital map	No		Rajapinta
Habitat variables	No		UH Personal folder

5. Aineiston avaaminen, julkaiseminen ja arkistointi

Miten, missä, milloin ja kenelle aineisto avataan ja julkaistaan?

Mihin pitkällä aikavälillä arvokkaat aineistot arkistoidaan ja kuinka pitkäksi ajaksi?



5. Aineiston avaaminen, julkaiseminen ja arkistointi



Miten, milloin, missä ja kenelle aineisto avataan?

“As open as possible, as closed as necessary”

- Jos mahdollista projektin lopuksi kaikki aineisto tulisi avata
- Vastuullisen tieteen periaatteet eivät kuitenkaan velvoita kaiken aineiston avaamiseen, eikä se ole aina mahdollista
 - **Sensitiivistä tai ainoastaan pseudonymisoitua aineistoa ei voi avata,** mutta aineiston **metadata** voidaan avata

Aineiston avaamisen hyödyt

Henk.koht hyödyt:

- Lisää näkyvyyttä
- Kontaktit & Yhteiset julkaisut
- Tieteellinen meritoituminen

Hyödyt yhteisölle:

- Organisaatio hyötyy sinun menestyksestä
- Aineiston läpinäkyvyys
- Rahoituksen tehokas hyödyntäminen

5. Aineiston avaaminen, julkaiseminen ja arkistointi

Miten, milloin, missä ja kenelle aineisto avataan?

1. Aineiston julkaisu ja jakaminen data repositoriassa/arkistossa
2. Aineiston julkaisu data julkaisussa



Why use online repositories

- Preservation of data beyond work contract length
- Access to data from anywhere
- Discoverability of data by search engines
- Citation system and PIDs
- Book keeping of data downloads
- Getting visibility to your work
- Funders & publisher require data to be made available online.

5. Aineiston avaaminen, julkaiseminen ja arkistointi

1. Aineiston julkaisu ja jakaminen data repositoriassa

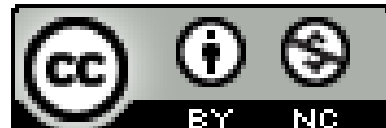
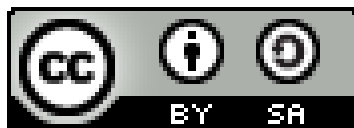
- Tee aineistostasi, löydettävä, saavutettava ja viitattava tai/ja täytä rahoittajan vaatimukset valitsemalla repositorio, missä:
 - **Pysyvä tunniste** (e.g. DOI, URN, DOI, Handle, ARk, PURL) ja pysyvä linkki, joka viittaa aineistoosi eli tekee aineistostasi löydettävän ja viitattavan;
 - **Lisenssi** joko annettu tai voi valita, edesauttaa selkeyttä ja helpottaa aineistosi käytettävyyttä.

5. Aineiston avaaminen, julkaiseminen ja arkistointi



1. Aineiston julkaisu ja jakaminen data repositoriassa

- Kun avaat aineiston, käytä lisenssiä
 - Lisensointi mahdollistaa tutkimukselle laajemman käytön sillä lisenssi kertoo käyttäjälle miten sitä voidaan jakaa
 - Creative commons: Choose [here](#).
 - [UH guide](#) & guide in Finnish [here](#)



5. Aineiston avaaminen, julkaiseminen ja arkistointi

1. Aineiston julkaisu ja jakaminen data repositoriassa

Valitse repositorio:

- [Re3data](#) Look for repositories!
- [ELIXIR Deposition Databases for Biomolecular Data](#)
- [B2Share](#)
- [IDA](#) (general) & [Etsin](#) (metadata) by CSC
- [Zenodo](#)
- [Dryad](#)
- [Figshare](#)
- [FSD \(The Finnish Social Science Data Archive\)](#),
questionnaires, interviews etc.

How to choose?

- Favour archives with a *certificate* for long-term preservation (e.g. Core Trust Seal-certified).
- What are the costs per dataset or gigabyte?
- What is the physical storage location of data? EU or US?
- What is the default license?
- Is long-term preservation guaranteed or not?

5. Aineiston avaaminen, julkaiseminen ja arkistointi

2. Julkaisu datajulkaisussa

- Julkaise datasetti/setit vertaisarvioidussa julkaisussa.
- Julkaisun rakenne on usein: abstrakti, johdanto, datan kuvaus sisältäen menetelmät, lyhyt yhteenveto ja uudelleenkäyttö mahdollisuudet.
- Esimerkkejä datajulkaisuista: [Scientific Data](#); [Data in Brief](#); [Data Science Journal](#).

5. Aineiston avaaminen, julkaiseminen ja arkistointi

Long-term preservation

- For tens and even hundreds of years = *over generations*
- Technical challenge
 - Hardware, software, and file formats age, the information must still be kept usable
 - Repositories specialized for long-term preservation will keep the bits safe.
- Data should be self explanatory = documentation well done
- [Fairdata-PAS](#) is coming. UH instructions [here](#).



“All information that is needed to replicate a study should be preserved, and everything that is potentially useful for others.”
– Sarah Jones /DCC

5. Aineiston avaaminen, julkaiseminen ja arkistointi – oma aineisto?

Data type	Opening	Long term archiving
Majavien pesätiedot	Metadata? LUKE?	?
MS-VMI	Avoimena	
Digitaaliset kartat	Avoimena	
Corine maankäyttö kartat	Avoimena	
Habitaatti muuttajat		



6. Aineistonhallinnan vastuut ja resurssit

6. Aineistohallinnan vastuut ja resurssit

- *Kuka* vastaa aineistohallintaan liittyvistä tehtävistä?
 - Kuka vastaa aineiston tietosuojasta ja -turvasta?
 - Kuuluuko edellinen yhdelle ihmiselle vai koko ryhmälle?
 - Edellyttääkö aineiston hallinnointi, säilyttäminen ja jakaminen asiantuntija-apua/palkattua työntekijää?
- *Kuinka paljon resursseja* (time & workload) aineistohallinta vaatii?
- Mitä kustannuksia voi olla?

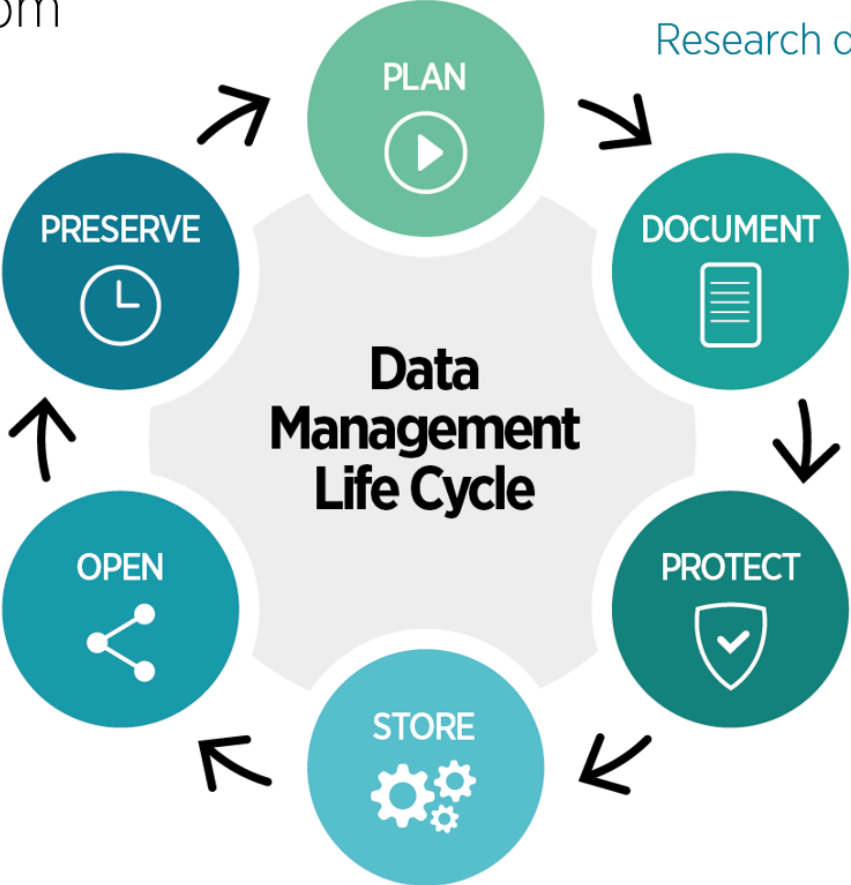
Hyvä tietää...



Data management services available from Data Support

datasupport@helsinki.fi

Long term preservation (PAS)



DMP tool: DMPTuuli & commenting service
Research data management guide
Courses & Workshops

Documentation guide

Legal affairs
Data Protection

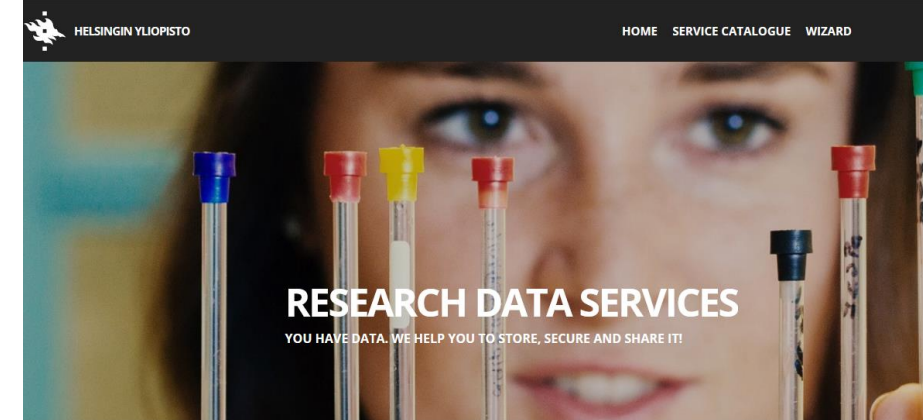
IT Services & CSC Services



RDM basic lecture → DMP workshop → datasupport@helsinki.fi → DMP review service →

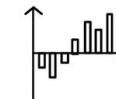
Oppaita & apua

- [Research data management guide](#) (UH)
- [Datanhallinnan perusopas](#)
- [Open access guide](#)
- [Research data services at UH](#)
- [DMPTuuli](#) for making a Data Management Plan
- <https://blogs.helsinki.fi/thinkopen/blogin-vuosikatsaus-2018/>
- Help in any matter regarding research data datasupport@helsinki.fi



These are the research data services provided by the University of Helsinki. You can browse the service catalogue to find the services that best meet your needs. The UH research data services aim to support you throughout the data life cycle from project planning to data archiving. Confidentiality and data protection issues should be taken into consideration early on. Read more about our services.

Awesome, let's go!



RESEARCH DATA SERVICES

Research data services support you through the entire life cycle of your data.

[Browse services](#) →



WIZARD TO IDENTIFY SERVICES

Answer these questions regarding your data needs that suits your specific needs.

[Browse Wizard](#) →

HELSINKI

NEWS RESEARCH ADMISSIONS COOPERATION UNIVERSITY

RESEARCH / RESEARCH ENVIRONMENT /

RESEARCH DATA MANAGEMENT

Research data management and its planning are an integral part of good research practices. It is important to control the risks involved, pay attention to data protection and information security, agree on data preservation, as well as ensure that the necessary resources and equipment are available. Our guide covers the mentioned topics and how to answer the general questions of a Data Management Plan (DMP).

- ↓ Data policy and support services
- ↓ Data management planning
- ↓ 1. Research data
- ↓ 2. Ethical and Legal Compliance

- ↓ 3. Documentation and metadata
- ↓ 4. Storing data and archiving
- ↓ 5. Opening data and access
- ↓ 6. Data management and security

TURN TO UH DATA SUPPORT!

IMPROVE YOUR ACADEMY OF FINLAND DMP

BY SENDING IT TO
DATASUPPORT@HELSINKI.FI

UH Data Support's Data Management Plan
commenting service is open until 15.9.*

*It is not guaranteed
that all DMPs get
comments, plans are
handled in arrival
order.

