



# Mistä puhumme, kun puhumme tutkimusdatan hallinnasta

Päivi Rauste, Data Management Office  
4.10.2022



# Luotettava tutkimusaineisto?

- Yksiselitteisesti ymmärrettävä!
- Läpinäkyvää & toistettavissa!



## → Riittävästi dokumentoitu

- Lähde ja provenanssi
- Tekniset ominaisuudet
- Konteksti & semantiikka
- Standardit ja formaatit
- Muuttujat

## → Eheä – ei muuttunut!

- Jos muuttunut, muutos dokumentoitu
- Jos ei saatavilla, metatiedot saatavilla
- Tarkistussumma & versiointi

# Mitä on tutkimuksen toistettavuus?

- Tieteellisen menetelmän peruseriaate eli "hyvää tieteellistä käytäntöä".
- Dokumentaatio mahdollistaa replikoinnin
- Tällöin hyvin dokumentoitu data on myös uudelleenkäytettävää

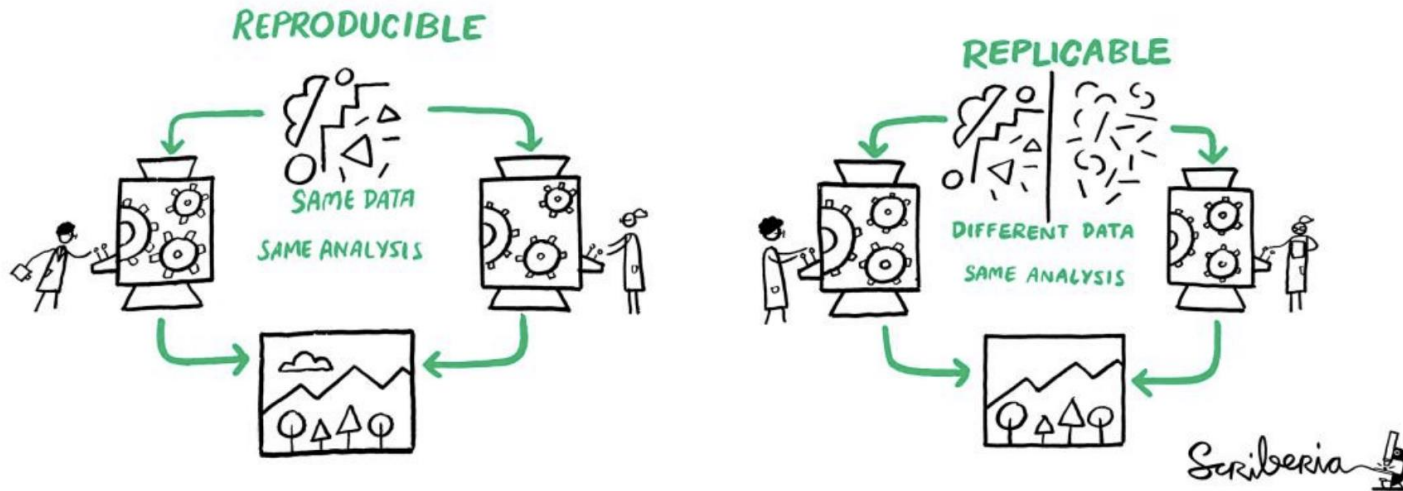
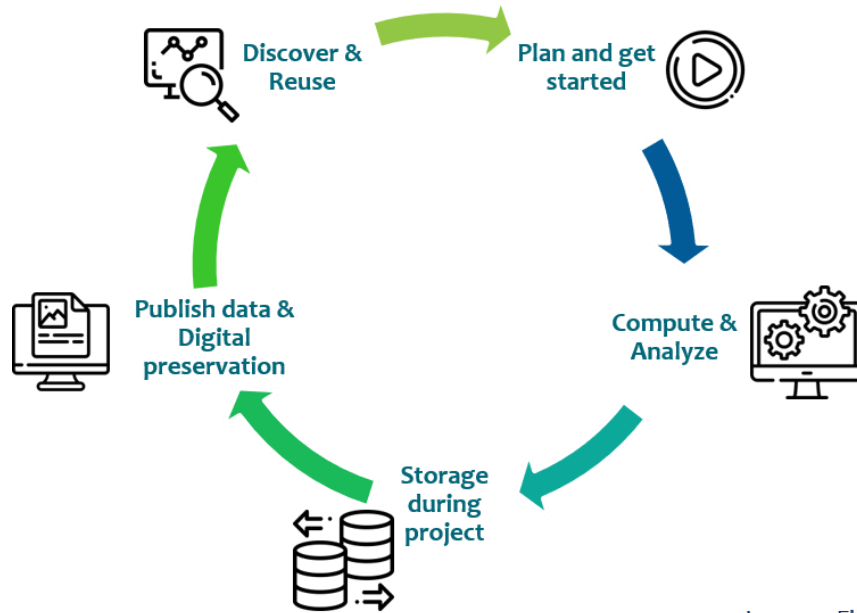


Figure 1. Reproducible and replicable research results<sup>3</sup>

# Datan käytettävyyden A ja O on dokumentaatio!

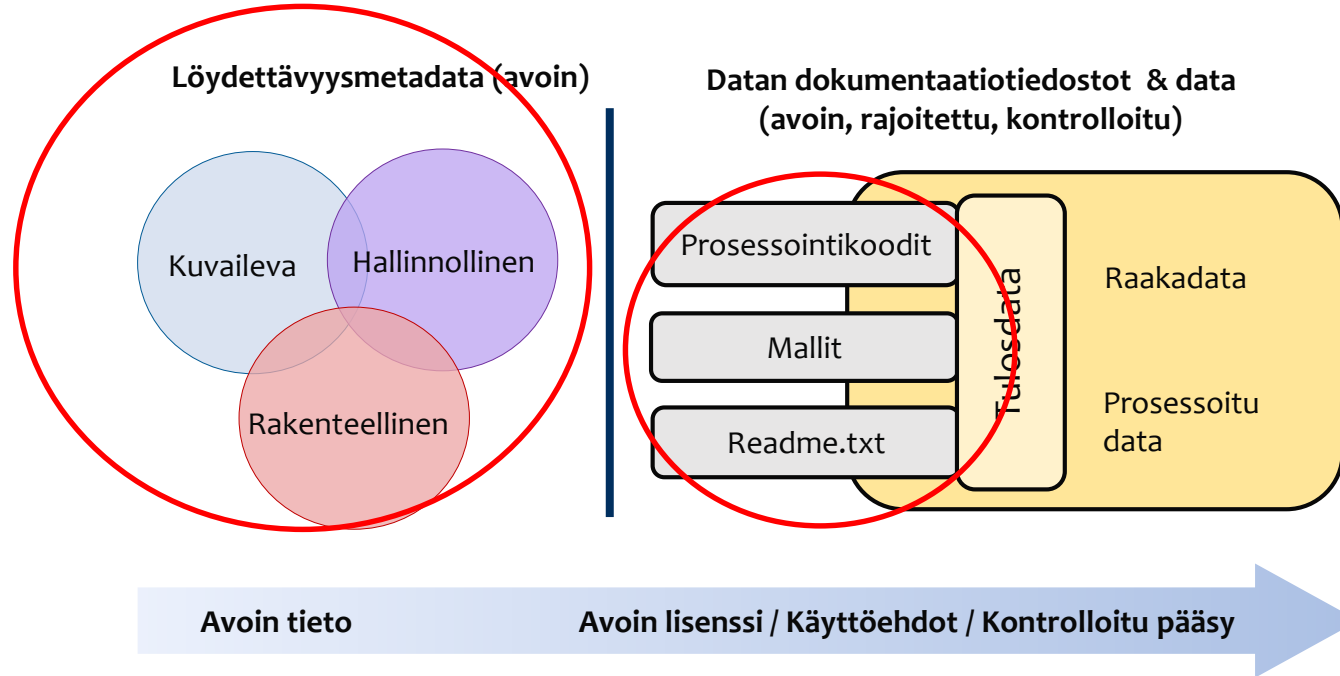


Jokaisesta tutkimuksen vaiheesta syntyy dokumentaatiota!

- Code books
- Data dictionary
- Version control
- Readme.txt
- Data standards

Images: Flaticon.com




# Metadatan tulee olla ymmärrettävää



Data voi olla avointa, rajoitetusti saatavilla tai kontrolloitua. Käyttöehdot (tutkimukseen) on kuvattu hallinnollisissa metatiedoissa ja koneellisesti luettavassa lisenssissä.

# CSC:n tutkimusdatanhallinnan osaamiskeskus

**CSC-Datatukiverkosto**

- Sähköpostilista tiedotukseen 
- Datakahvit Zoomissa kerran kuukaudessa 
- Data-chat: chat.csc.fi, #cscdatatukiverkosto 
- Koulutusmateriaaleja vapaasti hyödynnettäviksi 



Yhteistyö ja tiedonvaihto kansainvälisissä kehityshankkeissa



Koulutusta datanhallinnan palveluista ja ajankohtaisista teemoista

Haluamme tukea asiakasorganisaatioitamme järkevän datanhallinnan toteuttamisessa ja datapalveluiden käytössä.

[csc.fi/asiakaskoulutus](https://csc.fi/asiakaskoulutus)

<https://wiki.eduuni.fi/display/csckorkeakoulut/CSC+-+Datatukiverkosto>

# Tutkimuksen datanhallinnan tukipalvelut ja konsultointi

Tukea loppukäyttäjille

Tieteenalakohtaiset  
tiimit tukemassa  
tutkimusryhmiä

Perustukea ja  
neuvontaa  
datanhallinnan  
suunnittelussa

Mukana RI-hankkeissa  
kehittämässä ja  
tukemassa vastuullista  
datanhallintaa

Räätälöityä  
konsultointia erillisellä  
sopimuksella

[servicedesk@csc.fi](mailto:servicedesk@csc.fi)

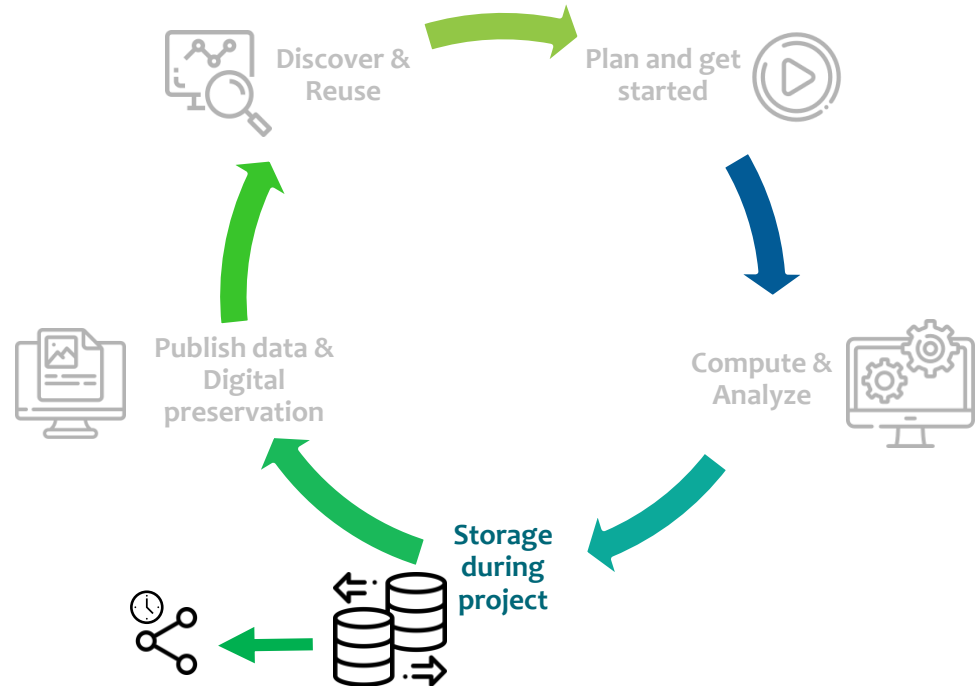
# Yleisesti palveluiden käyttöpolitiikasta, rooleista ja vastuista

- CSC ei vaadi omistusoikeutta tai mitään immateriaalioikeuksia palveluissamme oleviin aineistoihin.
- On suositeltavaa sopia oikeuksista tutkimusdataan liittyen tutkimusryhmässä ja tutkimusorganisaatiossa jo varhaisessa vaiheessa.
- On datan omistajan vastuulla päättää, mikä palvelu soveltuu kyseiselle datalle ja että salaus - jos se on tarpeen - tehdään asianmukaisesti.
- Datan omistaja päättää avoimuus- ja käyttöperiaatteista.



# Allas

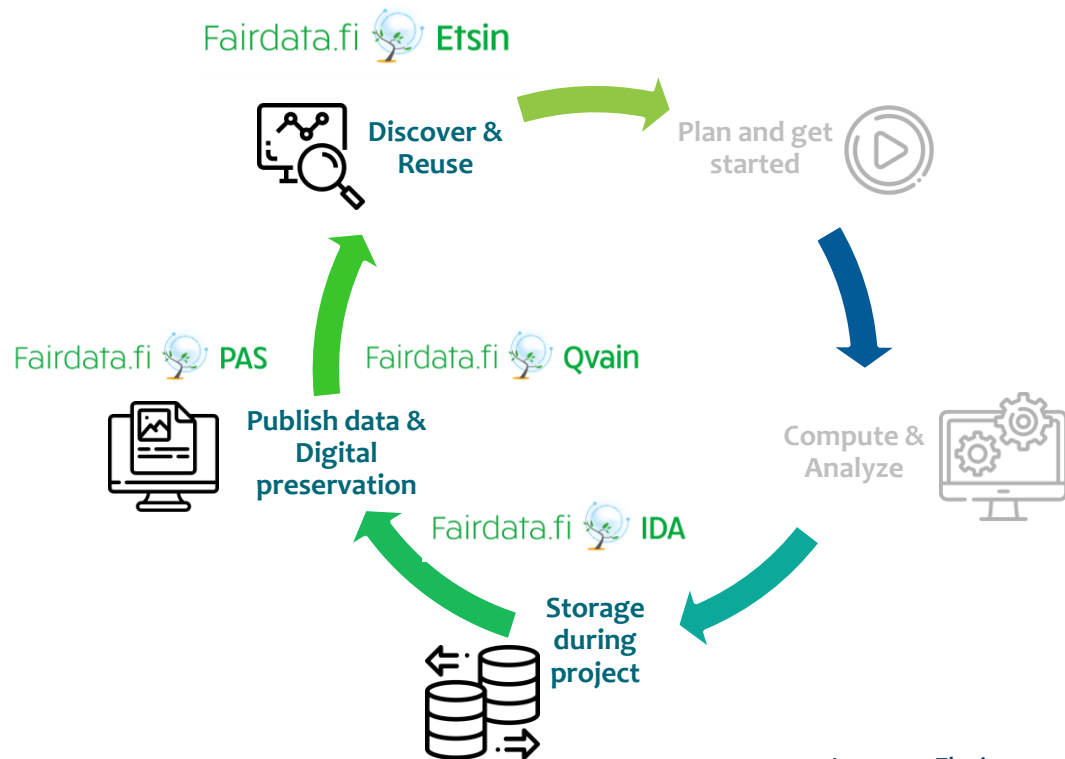
- **Palvelun tarkoitus:**  
Säilytys ja jakaminen – etenkin suurille datamäärille – tutkimusprojektin aktiivisessa vaiheessa
- **Hyödyt:**  
Datan prosessointi voidaan tehdä käyttäen avoimia rajapintoja mistä tahansa. Data voidaan helposti siirtää CSC:n pilvi- ja laskentaympäristöihin ja niistä pois.
- **Lisätietoa:**  
<https://research.csc.fi/en/-/allas>



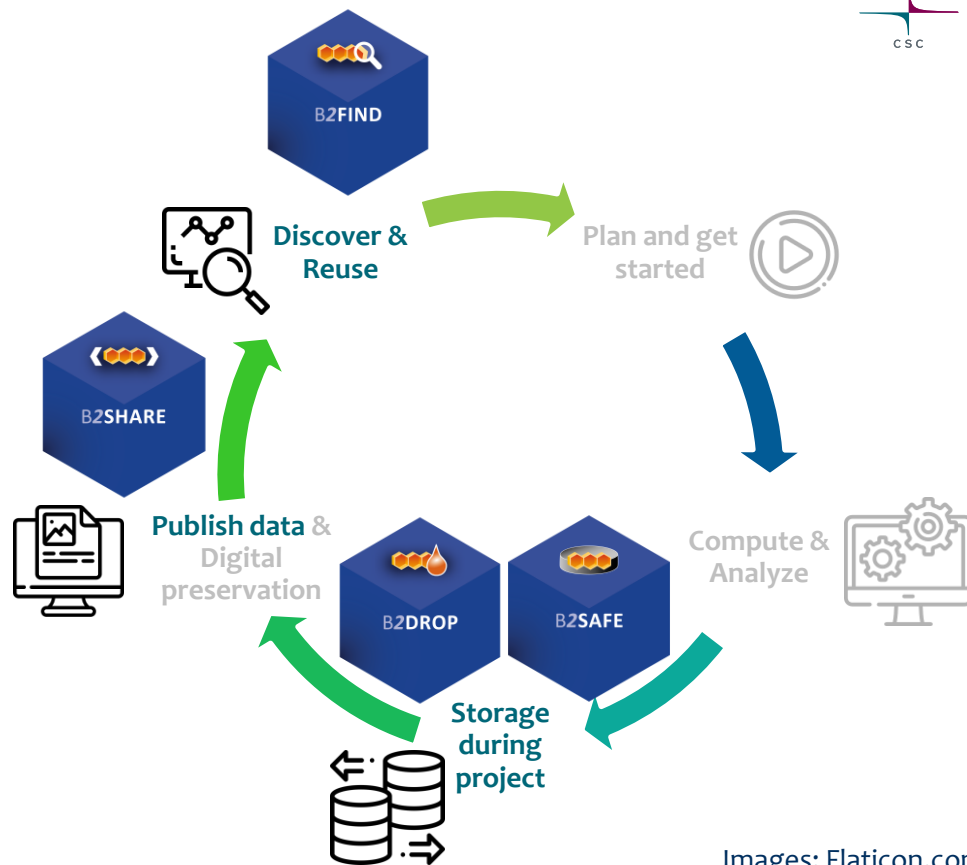
- **Palvelun tarkoitus:** Säilytys, kuvailu, julkaisu, löytäminen ja tutkimusaineistojen pitkäaikaissäilytys

## Hyödyt:

- Jopa suurten tai kasvavien tutkimustietoaineistojen julkaiseminen FAIR-periaatteiden mukaisesti maksutta helppokäyttöisten verkkotyökalujen avulla.
  - Tutkimusaineisto voidaan siirtää tutkimusaineistojen pitkäaikaispalveluun.
  - Research.fi-portaalin kautta lisänäkyvyyttä kansallisesti.
- **Lisätietoa:** <https://www.fairdata.fi>



- **Palvelun tarkoitus:** Sama kuin Fairdata, sillä erolla että EUDAT
  - tarjoaa helpon pääsyn kansainvälisiin infrastruktuureihin, palveluihin ja resursseihin.
  - Ei tarjoa tutkimusaineistojen pitkäaikaispalvelua
- **Hyödyt:**
  - Voidaan räätälöidä vastaamaan yhteisön erityistarpeita metatietojen, tallennustilan, käyttäjähallinnan ja brändin osalta.
  - Yhteinen ydinskeema ja aineistokokonaisuuksien löydettävyys B2FINDin avulla. Tuki useiden kansainvälisten organisaatioiden väliselle käytölle.
  - Tukee linjausten mukaista datanhallintaa automaation ja avoimien rajapintojen avulla.



# Miten Fairdata- ja EUDAT-palvelut voivat auttaa noudattamaan FAIR-periaatteita (Allas ei tue FAIR-periaatteita)



- Palvelut käyttävät monia yhteisiä sanastoja (F, I)
- Tallennettu data voidaan sisällyttää aineistokokonaisuuksiin, kuvailla ja julkaista (A, R)
- Julkaistulla datalla on pysyvä tunniste ja laskeutumissivu, jonka avulla data on saatavissa (F, A, I, R)
- Tutkimusdatan dokumentaatio on saatavissa yhteentoimivassa, koneluettavassa muodossa. (F, A, I, R)



## FINDABLE

- Described in relevant catalog with enough detail
- Landing page with globally unique persistent identifier



## ACCESSIBLE

- Can be retrieved over the internet
- Versioning and lifecycle are documented
- Tombstone page if data is deleted



## INTEROPERABLE

- Common, documented, and open formats
- Shared and curated semantic artefacts



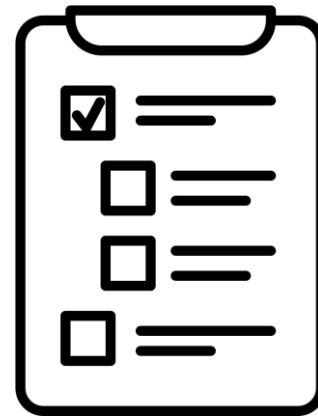
## RE-USABLE

- Well documented and intelligible
- Rights clearly stated

3

## Aineistonhallintasuunnitelma (Data Management Plan, DMP)


- on tutkijan työkalu, joka auttaa tutkimusprosessin valmistelussa ja potentiaalisten **riskien tunnistamisessa**, sillä
  - suunnitelmasta on suuri **apu eettisten ja lakiasioihin liittyvien yksityiskohtien konkretisoinnissa** tarkistuslistan avulla.
  - suunnitelma on ensimmäinen askel, joka **ohjaa kohti tietosuojan vaikutusten arviointia (DPIA)** ja tarvittaessa edelleen eettiseen arviointiin.
- on dokumentti, jonka avulla tutkimusorganisaatiot
  - saavat enemmän tietoja projekteista
  - tunnistaa **mitkä projektit käsittelevät aineistoja**, jotka edellyttävät tarkkavaisuutta
  - voivat **valmistautua auttamaan** tutkijaa projektin eri vaiheiden toteutuksessa
  - voivat **kehittää** tutkimuksessa tarvittavia palveluita tutkijoiden tueksi



**Suunnittelemalla datanhallintaa tutkijat voivat tunnistaa, ennakoida ja hallita aineistonhallintaan liittyviä riskejä.**

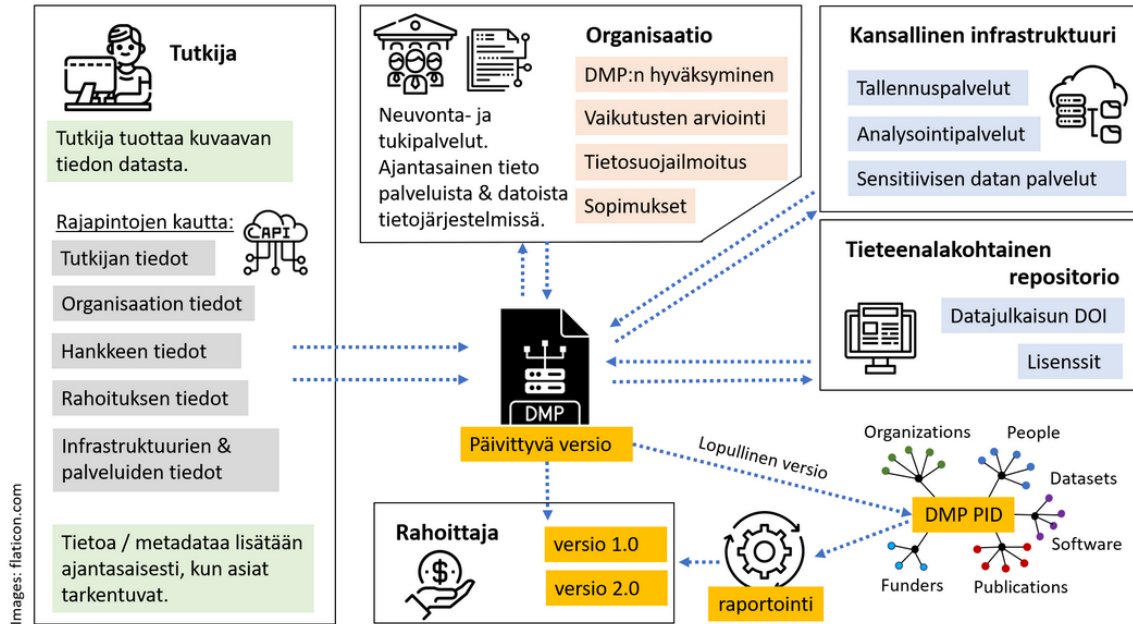
## DMP-työkalujen integrointi organisaation järjestelmiin, prosesseihin ja työvoihin hyödyntäisi kaikkia

- tulisi kehittää rakenteellisempaan suuntaan, mikä mahdollistaa tietojen integroinnin ja hyödyntämisen kotiorganisaatiossa koneluettavassa muodossa, esim.
  - arkaluonteisen datan tunnistaminen
  - tietosuojan vaikutusten arvioinnin
  - datan säilytyspalveluiden turvallisuusvaatimuksien tunnistaminen
  - datan siirtämistarpeet ja siihen liittyvät riskit
  - integroidut palvelukatalogit

- DMP-konsortio The logo for DMP Tuuli consists of the text 'DMP' in a small blue font, followed by 'Tuuli' in a large blue font with a stylized wind icon to its right.
- Etsii integraatiopilotteja, missä integroidaan DMP Tuuli organisaation omiin järjestelmiin ja tutkimustietovarantoon
  - Valmistelee toimintasuunnitelmaansa ja pohtii, miten voidaan siirtyä dynaamisempaan ratkaisuun

<https://wiki.eduuni.fi/display/csckorkeakoulut/DMP-konsortio>

# maDMP, CSC:n visio koneluettavasta DMP:stä



Pohjana on käytetty kesäkuussa julkaistua dynaamisen DMP:n työryhmän raporttia. Kuvaan tullaan viittaamaan, kun keskustellaan dynaamisen DMP:n kehittämisestä kansallisella tasolla.

# Palvelukatalogitietomalli-työryhmä

## Tehtävä:

Tutkijan käyttöön on tarjolla sekä tutkimusorganisaatioiden hankkimia, kansallisten infrastruktuurien tuottamia että kansainvälisten verkostojen ylläpitämiä palveluita. **Palvelukatalogi on yksi tapa tuoda nämä palvelut tutkijan tietoon. Palvelukatalogin ylläpitämisen helpottamiseksi tiedot tulisi olla mahdollista lukea avoimen rajapinnan kautta.**

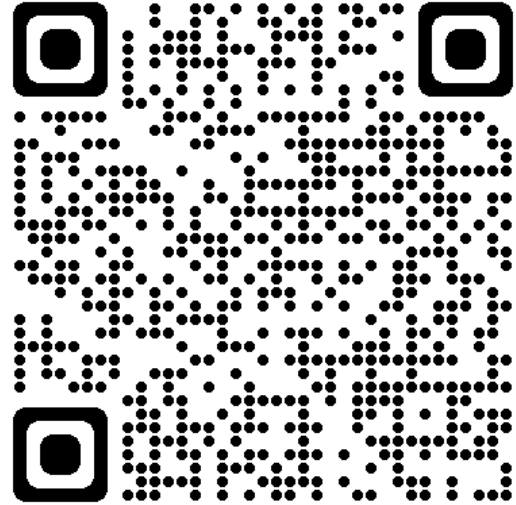
- > Palvelukatalogimetatietotyöryhmässä tullaan pohtimaan **yhteistä metatietomallia palveluille sekä yhteistä tapaa kategorisoida palvelut asiakaslähtöisesti.**

Lisätietoja: [tua.hindersson-soderholm@csc.fi](mailto:tua.hindersson-soderholm@csc.fi)



## Lisätietoja

- Service Catalog for Research  
<https://research.csc.fi/service-catalog>
- CSC Services for Research (2022) *Services for Sensitive Data*.  
<https://research.csc.fi/en/sensitive-data>
- Docs CSC (2022) *Working with Data*.  
<https://docs.csc.fi/data/datasets/dataset-sources/>



Tutkimusdatahallinnan  
osaamiskeskus