

# TETO KÄYTTÖÖN: Saatavuus, säilyvyys ja hyödynnettävyys

FT Pirjo-Leena Forsström, CSC, Kehitysjohtaja



*CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus*

# Tieto käyttöön: mitä huomioida?

## • Saatavuus

Lait ja  
käyttötarkoitukset

Lisenssit ja  
pääsyoikeudet

Koneluettavuus, rajapinnat,  
tunnisteet, linkitykset, viittaukset...

## • Säilyvyys

Metatiedot:  
aineiston sisältö,  
historia, alkuperä,  
käyttö

Tiedostomuodot

Eheydestä  
huolehtiminen

## • Hyödynnettävyys

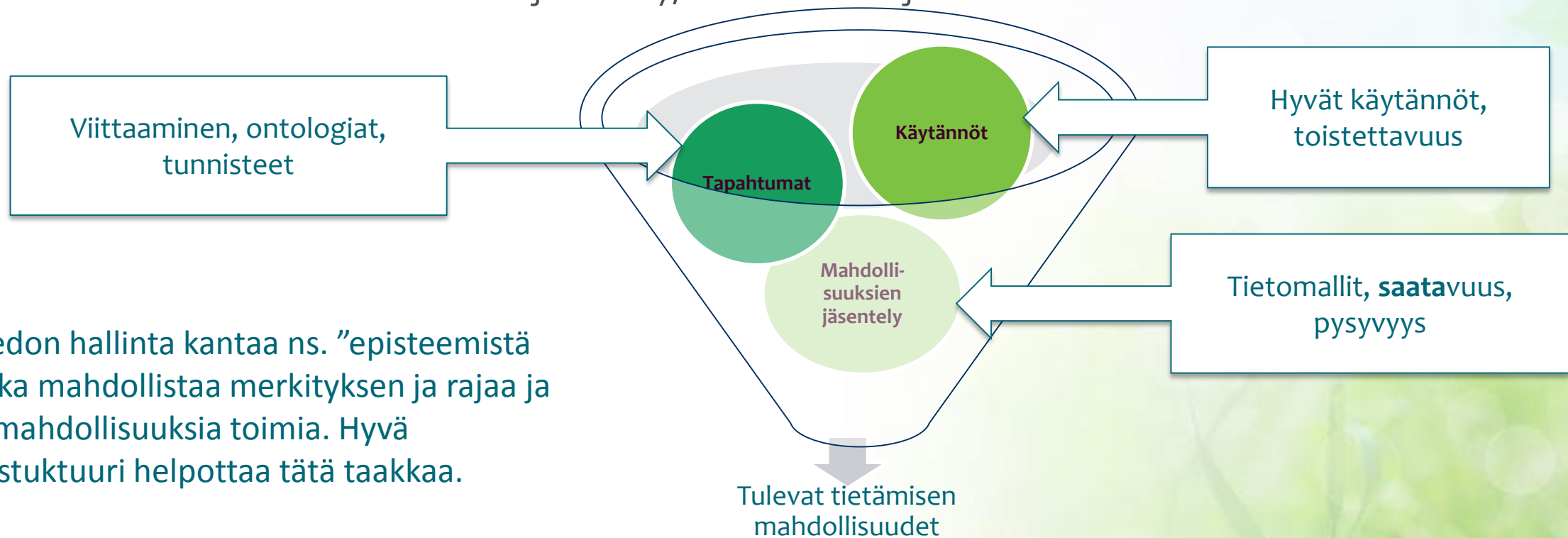
Semanttinen yhteentoimivuus:  
tietomallit, sanastot, koodistot etc.

Ymmärrettävyys, rajapinnat,  
tunnisteet, linkitykset, viittaukset...

Joustavuus ja  
laajennettavuus

# Tietäminen ja tieto

- Mielekkään merkityksen, ajattelun ja kommunikaation horisontti, jossa:
  - Tiedon käytännöt: menetelmät, dokumentaatio, huolellisuus, tarkkuus, kriittisyys, objektiivisuus, eettisyys, aineistot ja dokumentaatio
  - Tietämisen tapahtumat: aiemman käytön tuntemus, harhojen tunnistaminen ja hallinta, johtopäätöksien tekeminen
  - Tulevien tietämisen mahdollisuuksien jäsentely, toistettavuus ja todennettavuus



Laadukas tiedon hallinta kantaa ns. "episteemistä taakkaa" joka mahdollistaa merkityksen ja rajaa ja ehdollistaa mahdollisuuksia toimia. Hyvä palveluinfrastruktuuri helpottaa tätä taakkaa.

# FAIR-periaatteet (EU Council Conclusions on Open Science)

Fairdata.fi



FAIR eli REILU-periaatteet ovat tutkimuksen hyvää tiedonhallintaa

## F (Findable) LÖYDETTÄVYYS

- Miten hyvin aineisto löytyy hakukoneilla ja muin tavoin

## A (Accessible) SAAVUTETTAVUUS

- Pääsy aineistoihin, periaatteellisesti ja teknisesti

## I (Interoperable) YHTEENTOIMIVUUS

- Saumaton toiminta mm. palveluiden ja prosessien välillä sekä tietoja yhdisteltäessä

## R (Re-usable) UUDELLEENKÄYTETTÄVYYS

- Aineiston käyttö ja käsittely muiden toimesta ja muuhun kuin alkuperäiseen käyttötarkoitukseen

## 1. Resurssille annetaan ainutkertainen ja ikuisesti pysyvä tunnus (persistent identifier, PID).

Jotta kaksi resurssia ei saisi samaa tunnusta, täytyy tunnuksien antamisen olla standardoitua ja hyvin hallittua. Tutkimusdatan ja tieteellisten julkaisujen alalla käytetyimpiä ovat [DOI](#)-, [URN](#)- ja [Handle](#)-tunnusjärjestelmät.

Pysyvyys edellyttää, että sekä tunnusten että resurssien hallinta on kestävästi järjestetty. Resurssilla täytyy olla ylläpitäjä, joka huolehtii siitä, ettei se katoa verkosta ja että tunnus osoittaa aina oikeaan resurssiin.

## 2. Resurssille annetaan kattavat metatiedot.

Kuvaileva metatieto kuvaa resurssin sisältöä ja luonnetta.

Hallinnollinen metatieto sisältää hallinnolliset ja tekniset tiedot, esimerkiksi tutkimusaineiston käyttöehdot ja käytössä tarvittavat tekniset tiedot.

Rakenteellinen metatieto kuvaa resurssin rakenteen, esimerkiksi tiedostojen suhteen toisiinsa.

## 3. Resurssi kuvaillaan ja rekisteröidään hakupalveluun.

Hakupalvelut voivat olla mm. organisaatiokohtaisia, tieteenalakohtaisia tai kansallisia.

Usealla tieteenalalla käytetään sen erityispiirteet huomioon ottavia metatietomalleja. Nämä metatietomallit ohjaavat sitä, mitä metatietoja ja missä täsmällisessä muodossa esim. tutkimusaineiston kuvailussa vaaditaan. Metatietomalli voi olla myös yleispätevä.

## 4. Pysyvä tunniste käy ilmi metatiedoista.

Viittauksen tulisi aina sisältää viitattavan resurssin pysyvä tunniste.

**1. Resurssi tai sen metatiedot ovat noudettavissa tunnisteiden avulla standardia yhteyskäytäntöä noudattaen.**

Kaikki aineistot eivät voi olla avoimia esimerkiksi tutkimuseettisistä syistä. Pyrkimyksenä on, että tutkimusaineistot ovat niin avoimia kuin mahdollista ja niin suljettuja kuin on tarpeellista.

**2. Yhteyskäytäntö on avoin, maksuton ja yleisesti käyttöön otettavissa.**

Saavutettavuuden ei tulisi edellyttää esim. maksullisten ohjelmistojen käyttöä.

**3. Yhteyskäytäntö tukee tunnistautumista ja valtuuttamista, kun se on tarpeen.**

Kaikki aineistot eivät voi olla avoimesti saatavilla. Aineistojen saavutettavuutta voi silti parantaa esimerkiksi sillä, että käyttöluvan anominen ja myöntäminen on helppoa.

**4. Metatiedot ovat saavutettavia, vaikka aineisto ei enää ole.**

Metatietojen tulee olla sellaisia, että ne voidaan pitää avoimesti saatavilla, myös sen jälkeen kun itse aineisto poistetaan.

**1. Resurssin ja sen metatiedon sisältö esitetään määrämuotoista, monikäyttöistä, saatavilla olevaa ja jaettava kieltä käyttäen.**

Aineiston tulisi olla sekä ihmis- että koneluettavaa.

Tietosisältöjä täytyy voida siirtää järjestelmien välillä.

**2. Resurssissa ja sen metatiedoissa käytetään sanastoja, jotka noudattavat FAIR-periaatteita.**

Suosi olemassa olevia, avoimia ja koneluettavia sanastoja, ontologioita ja koodistoja.

**3. Resurssissa ja sen metatiedoissa on tyypiteltyjä viittauksia muihin resursseihin.**

Tyypittely tarkoittaa, että viittauksen luonne on määritelty; esim. A on B:n aiempi versio tai B on käyttänyt A:ta lähteenä.

Pysyvien tunnisteiden käytöllä voidaan turvata viittausten kestävyys pitkällä aikavälillä.

## 1. Resurssille on annettu runsaasti tarkkoja ja hyödyllisiä attribuutteja.

- Kattavat metatiedot edistävät löydettävyyttä. Niiden tulee myös auttaa arvioimaan, onko resurssi varmasti käyttökelpoinen tietyssä asiayhteydessä. Metatiedon tulisi siksi sisältää tietoja mm. tietojen keruutavasta, tunnetuista virhelähteistä yms.

## 2. Resurssi ja sen metatiedot on julkaistu käyttäen selkeää käyttölisenssiä, jonka ehdot ovat saatavilla.

- Lisenssi/käyttöehdot ovat tärkeä metatieto. Epäselvät käytön ehdot saattavat estää uudelleenkäytön. Koneluettava formaatti kuten [CC-lisenssi](#) on suositeltava.

## 3. Resurssi on aina yhdistettävissä alkuperäänsä ja elinkaareensa.

- Uudelleen käyttäjän pitäisi tietää, mistä tietoaaineisto on tullut ja keneen sitä käyttäessä tulee viitata.
- Myös syntyprosessin kuvailu on tärkeää: kuka ja miten keräsi aineistoa; miten sitä käsiteltiin; mitä julkaisuja siitä on tuotettu; onko siinä hyödynnetty muita datan lähteitä.

## 4. Aineisto täyttää oman alansa tiedeyhteisön vaatimukset.

- Noudata tieteenalasi hyviä käytäntöjä ja kansallisia sekä kansainvälisiä ohjeistuksia



- Pääpainona informaation jatkuva käytettävyys

- Tasoja:

## 1) Bittien säilyttäminen

- Alkuperäinen bittijono voidaan toistaa nykyaikaisilla laitteilla, mutta ei takeita siitä voidaanko sisältöä tulkita
- Säilyttämisen perusta ja optio tulevaisuuteen: antaa mahdollisuuden nostaa vaatimustasoa myöhemmin jos tarvetta ja rahaa on

## 2) Sisällön ymmärrettävyyden säilyttäminen

- Teksti, kuvat ja muu olennainen sisältö voidaan tulkita
- Käytön ja uuden tiedon tuottamisen kannalta yleensä tärkein taso

## 3) Alkuperäisen käyttökokemuksen säilyttäminen

- Kalvoesityksen tehosteet toistuvat alkuperäisessä muodossaan
- Vanha tietokonepeli näyttää ja kuulostaa samalta kuin 30 vuotta sitten
- Vaativin, mutta historiantutkimuksen kannalta mielenkiintoinen taso

# SÄILYVYYDEN RISKIT



- **Tekniset riskit**

- Mm. tallennusmedian tai laitteiston rikkoutuminen, ohjelmistovirheet

- **Inhimilliset riskit**

- Ylläpitäjän virhe

- Hyökkäys järjestelmää vastaan (organisaation ulkoa tai sisältä)

- **Eryteisesti pitkäaikaissäilytykseen liittyvät riskit**

- Tiedon ymmärrettävyyden häviäminen

- Piilevät virheet ("bit rot")

- **Katastrofit**

- Luonnonilmiöt: tulva, tulipalo ym.

- Yhteiskunnalliset: Infrastruktuurien sortuminen, rahoituksen loppuminen ym.

# Tietointensiivisen infrastruktuurin "blueprint"

## Tuota ja käytä

- Kerää, tallenna, käytä, keskustele
- Löydä muuta relevanttia dataa

## Datan valmistelu ja hallinta

- Kuvaile, paketoi, ylläpidä ja huolehdi elinkaaresta

## Datan jakelu

- Luo jakelumalli ja rajapinnat, jaa metatietoa

Koulutus, ohjeet ja tukipalvelut

## Tiedon tallennus

Optimoidaan käyttönopeutta

Optimoidaan säilyvyyttä

Optimoidaan saavutettavuutta

Analyysityökalut

Yhteistyöalustat

Hakutyökalut

Metatiedot tuki: tietomallien hallinta, tunnisteet, ontologiat

Pitkäaikaissaatavuus: paketointi, sopimus, seuranta

Avoimet rajapinnat

Käyttövaltuutus

Jakeluportaalit

Lait/Politiikat /Aineistonhallintasuunnitelmat /Säilytysuunnitelmat

# LOPUKSI

*“Ei ole yhdentekevää, millaisista  
tulevaisuuksista me unelmoimme”*

*M. Dufva*