

# Mihin tutkimusaineistojen hallintaa tarvitaan?

Antti Auer

Tietohallintopäällikkö, D.Sc.

Yliopistopalvelut

Strateginen kehittäminen

# Oxfordin yliopiston tutkimusdatahanke

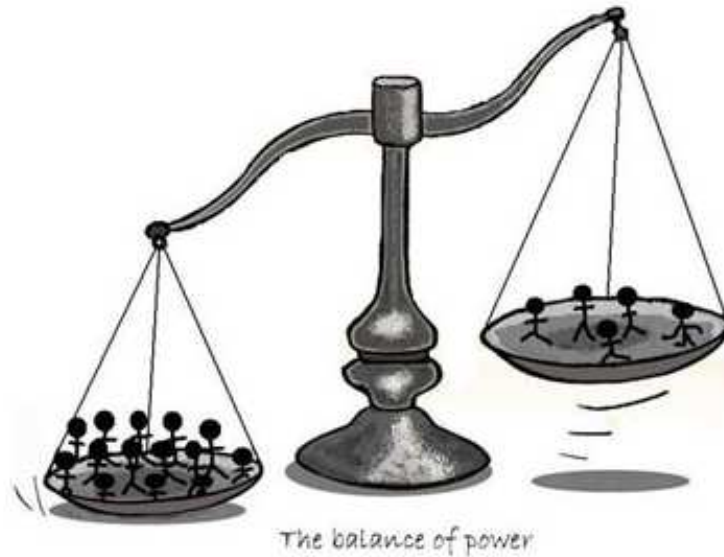
<http://www.ariadne.ac.uk/issue65/wilson-et-al>

Data management sharing plans	Legal & ethical	Best formats & best practice	Secure storage	Metadata	Access & discovery	Computation analysis & visualization	Restricted sharing	Data cleaning	Publication	Assess value	Preservation	Add value
Support												
Infrastructure and tools												
Policy												
Business model												

*"Data repositories tend to think of it primarily in terms of preservation and curation, but researchers are more likely to associate the term with structuring data in a database, or the organisation of files and folder. At its broadest, research data management involves all the processes that information from research inputs undergoes as it is manipulated and analysed en route to becoming a research output."*

# Aineiston koolla on väliä

- Aineistoja tulisi kerryttää
- Aineistojen kerryttäminen nostaa tutkimuksen laatua
  - Vain riittävän suurissa aineistoissa voidaan havaita variaatiot
  - Kehitys (pitkittäisaineistot)
  - Alueelliset/kansalliset ja sosioekonomiset erot
  - Vertailuaineistot
- Pienten aineistojen todistusvoima voi olla kyseenalainen
- Datamassojen syntyminen ja uudet analyysimenetelmät
  - Paljon dataa ei ole big data, kun aineistot ovat hajallaan eri puolilla erilaisissa tallennusvälineissä
- Aineistojen huoltaminen, kuvailu ja julkaiseminen keskitetysti ja yleisten standardien mukaan mahdollistaa uudelleenkäytön
- Hyvin hoidettu ja julkaistu aineisto kasvattaa impact factoria ja meritoi tutkijaa sekä tutkimusryhmää



Studies with larger sample sizes have more statistical power and can detect smaller, more subtle effects. Photograph: Kate Button

Small studies with low power undermine the reliability of science....

<http://www.theguardian.com/science/sifting-the-evidence/2013/apr/10/unreliable-neuroscience-power-matters>

[http://www.theregister.co.uk/2013/04/12/brain\\_science\\_low\\_power\\_junk/](http://www.theregister.co.uk/2013/04/12/brain_science_low_power_junk/)

# Miksi keskitetty tutkimusaineistoinfra?

- Tutkimusresurssit kohdistettava tutkimukseen, ei tukipalveluihin
  - Jyväskylän yliopiston Tutkimus-, innovaatio- ja infrastruktuuriohjelma 2012-2017 *”Jyväskylän yliopistossa luodun tutkimusaineiston omistaa Jyväskylän yliopisto.”*
  - Omistajuus aiheuttaa yliopistolle velvollisuuden hoitaa omaisuutta
  - Tutkimuksen peruspalvelut hankkeille automaattisesti
- Rahoittajat edellyttävät aineistojen hallintaa ja uudelleenkäyttöä
- Kustantajat edellyttävät aineiston läpinäkyvyyttä ja tulosten verifioitavuutta
- Tietoturva ja aineistojen säilyminen – aineistojen varmuuskopiointi automaattisesti
  - *”Tutkimus suunnitellaan, toteutetaan ja siitä raportoidaan ja siinä syntyneet tietoaineistot tallennetaan tieteelliselle tiedolle asetettujen vaatimusten edellyttämällä tavalla”* Julkaisueettiset periaatteet 15.5.2013
  - Tietosuoja – ihmistieteissä käsitellään usein yksilöityä henkilödataa
  - Pääsynhallinta roolien ja oikeuksien mukaan
- Tutkimusprosessin tuki ja tutkimuksen aikainen aineistojen käytettävyys
  - Aineiston hallinta tehostaa tutkimusta
  - Aineiston kuvailu ja versiointi
- Aineistojen tallennuksen ja jakelun hallittu ja läpinäkyvä ulkoistaminen
  - erityisesti CSC:n IDA-tallennuspalvelu ja Yhteiskuntatieteellinen tietoarkisto
  - Pitkäaikaissäilytyksen ulkoistaminen CSC:lle
- Virtuaaliset tutkimus-, oppimis- ja kehitysympäristöt
  - Aineistojen yhdistäminen tutkimushankkeen sisällä ja tutkimushankkeiden välillä
  - Aineistojen jakaminen yli organisaatorajojen – tutkimus- ja oppimisverkostot
  - Uudet tallennustavat ja tietokantapalvelut, uudet analyysimenetelmät
  - Pohja keskitetyille laskenta-, mallinnus- ja visualisointipalveluille (esim. CSC)
  - Datankäsittelyprosessien automatisointi
  - Aineistot innovaatioekosysteemin perustana

# Läpinäkyvyys (verifioitavuus, auditoitavuus)

- **“Research fraud exploded over the last decade” (2012)**

<http://arstechnica.com/science/2012/10/research-fraud-exploded-over-the-last-decade/>  
<http://www.pnas.org/content/109/42/17028>

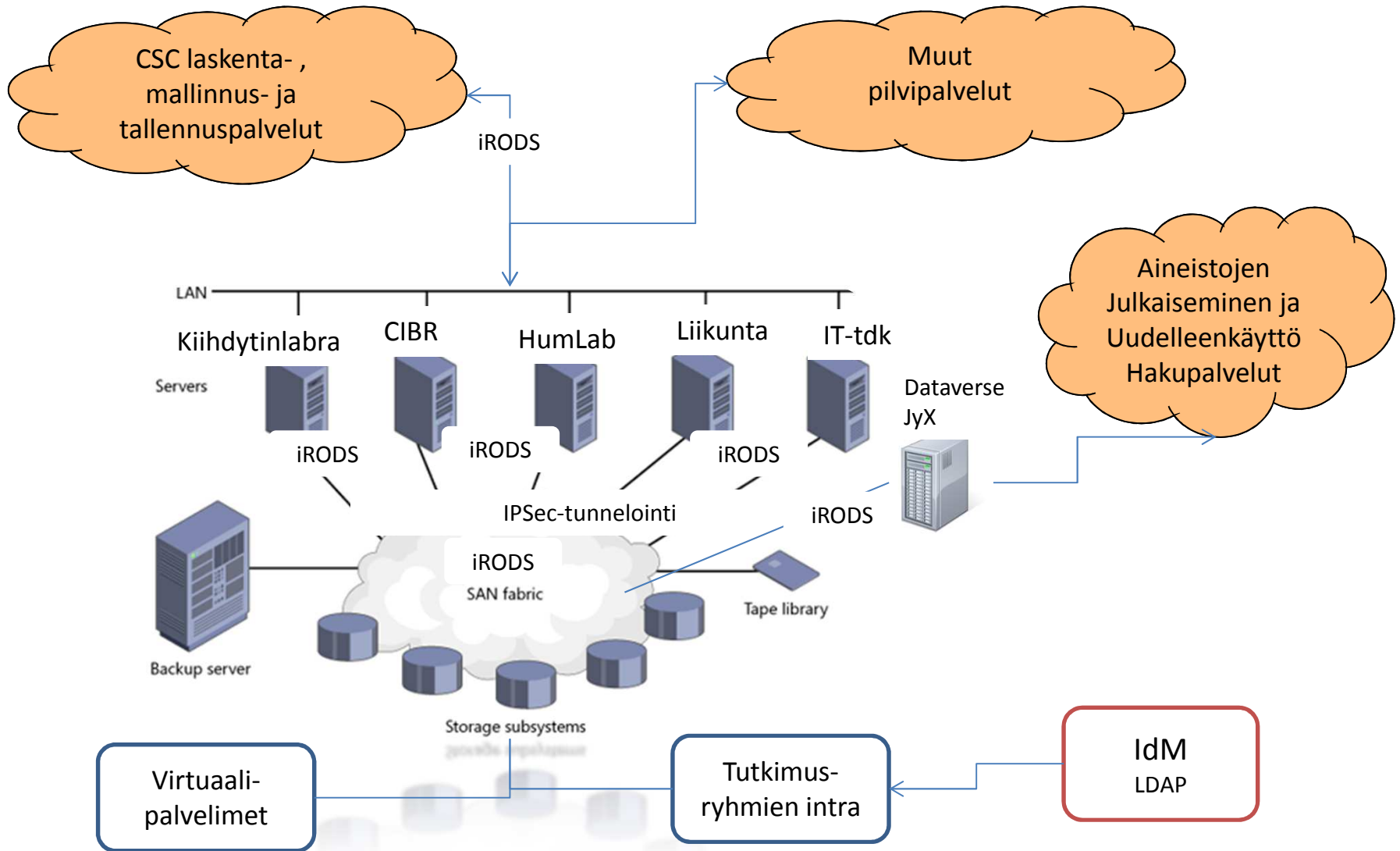
2,047 biomedical and life-science research articles indexed by PubMed as retracted . 67.4% of retractions were attributable to misconduct, including fraud or suspected fraud (43.4%), duplicate publication (14.2%), and plagiarism (9.8%).

- **“Research fraud needn't happen at all” (2002)**

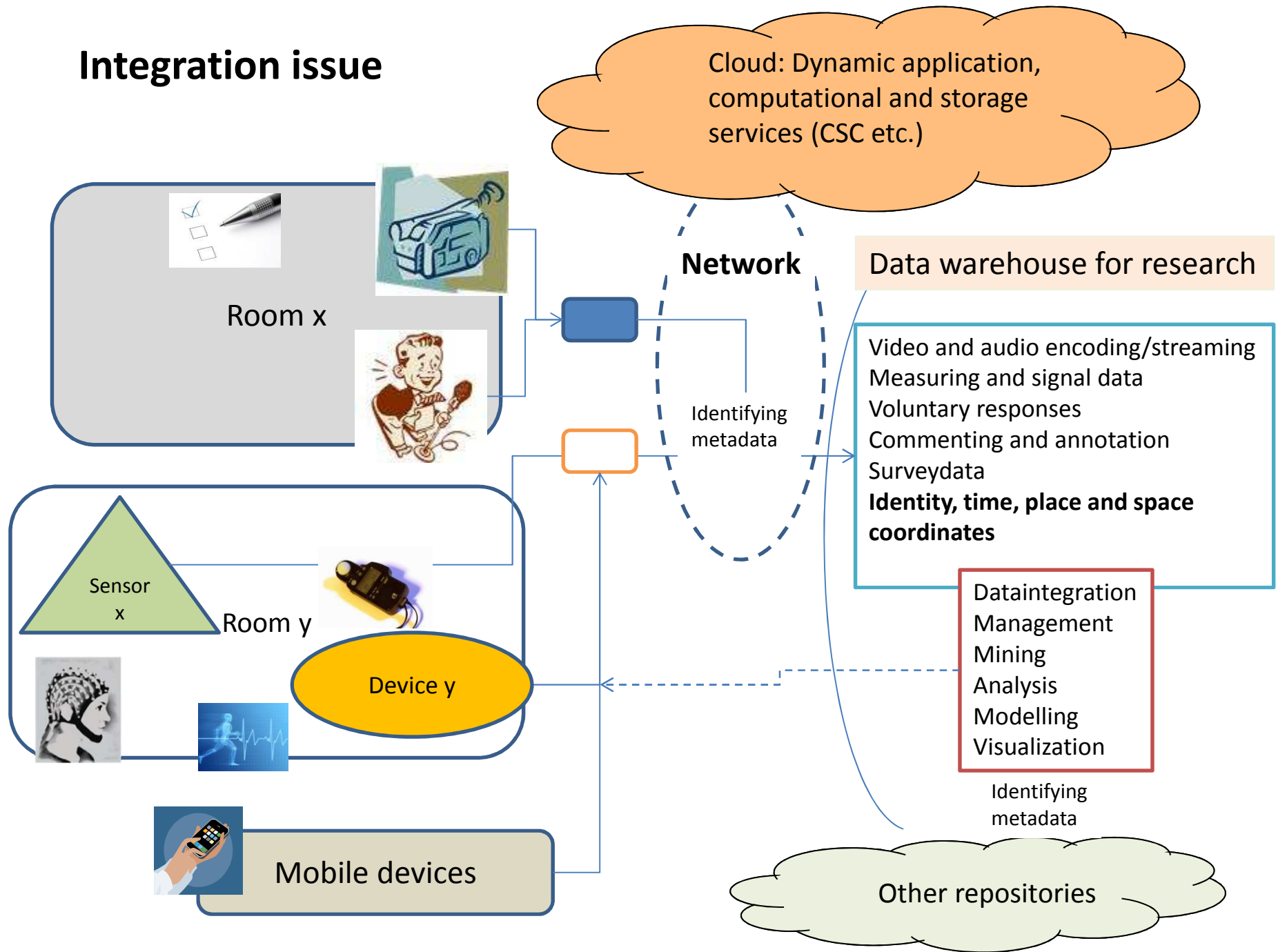
<http://www.apa.org/monitor/feb02/fraud.aspx>

- Experts say it is the shared responsibility of institutions, funding agencies, journal editors and principal investigators to prevent research fraud.
- **Setting clear rules for data collection and analysis.** Principal investigators ought to lay out guidelines for the type of data that will and won't be considered in a study.
- **Reviewing data.** Another pre-emptive strategy is checking the raw data collected by one's co-investigators before writing it up and submitting it for publication. Principal investigators should do this and journal editors could check raw data as well. But that would likely be a tough sell, given the amount of time it would require.
- **Explaining the data analysis process.** In any submitted journal article, researchers ought to delineate how they categorized and analyzed their data.

# Jyu.fi data-arkkitehtuuria



# Integration issue





# Tutkimusaineisto-oppaita

- <https://www.jyu.fi/tutkimus/tutkimusaineistot>
- Tutkimuksen tietoaineistot (TTA) –hankkeessa julkaistu Datanhallinnan opas  
<http://www.tdata.fi/tutkimusdatan-hallinta>
- Yhteiskuntatieteellisen tietoarkiston tutkimusaineistojen tiedonhallinnan käsikirja.  
<http://www.fsd.uta.fi/tiedonhallinta>
- University of Oxford Research Data Management  
<http://www.admin.ox.ac.uk/rdm/>  
(incl. [Data management planning checklist](#) and University of Oxford Policy on the Management of Research Data and Records)

Henkilöt vaihtuvat

**Infrastruktuurit pysyvät!**

Ja luovat tutkimukselle jatkuvuutta....

**Hyvin hoidettu aineisto on tärkeä  
osa tutkimuksen infrastruktuuria!**