

Tutkimuksen tietoaaineistot -seminaari
20.9.2012

Arto Mustajoki

Avoim tiede: mitä, miksi ja kuka?

Avoin tiede: kova kuhina

Raportteja ja kannanottoja syntyy kuin sieniä sateella:

- High Level Expert Group on Scientific Data (European Commission): Riding the Wave 2010.
- THE LERU Roadmap towards open access. 2011.
- Bruegel Policy Contribution: Michał Grajek, ICT for growth: A targeted Approach
- TNS BMRB: Open data dialogue. Final report
- The Royal Society: Science as an open enterprise
- ETLA: Julkinen tieto käyttöön

Avoim tiede: käsiteviidakko

scientific e-infrastructure, open science,
interoperativity

open data, open access, citizen science

Open Knowledge is ‘any kind of information – sonnets to statistics, genes to geodata – that can be freely used, re-used, and redistributed’ (Open Knowledge Foundation definition).

Open Scholarship refers to research that generates Open Knowledge.(LERU)

Avoin tiede: suuria haaveita

Vahvaa retoriikkaa: taikasana **sharing**

Mrs Neelie Kroes : “Data is the new gold”

Ongelmia ei ole, mutta on haasteita (to be overcome”)

Scientific e-infrastructure – some challenges to overcome (Riding the wave)

Collection How can we make sure that data are collected together with the information necessary to reuse them?

Trust How can we make informed judgements about whether certain data are authentic and can be trusted? How can we judge which repositories we can trust? How can appropriate access and use of resources be granted or controlled?

Usability How can we move to a situation where non-specialists can overcome the high barriers to their being able to start sensible work on unfamiliar data, perhaps using intelligent automated tools for an initial investigation?

Scientific e-infrastructure – some challenges to overcome (Riding the wave)

Interoperability How can we implement interoperability within disciplines and move to an overarching multidisciplinary way of understanding and using data? How can we find unfamiliar but relevant data resources beyond simple keyword searches, but involving a deeper probing into the data? How can automated tools find the information needed to tackle unfamiliar data?

Diversity How do we overcome the problems of diversity – heterogeneity of data, but also of backgrounds and data-sharing cultures in the scientific community?

How do we deal with the diversity of data repositories and access rules – within or between disciplines, and within or across national borders?

Security How can we guarantee data integrity? How can we avoid data poisoning by individuals or groups intending to bias them in their interest? How can we react in the case of security breaches to limit their impact?

Through Open Access, Europe's populations can be better informed, not only by their own efforts at seeking out specific research information on topics of interest, but through better-informed media bringing to their attention new developments and findings from basic research. Knowledge societies can be built around the world much more strongly and effectively if knowledge is easily accessed and spread. Open Access is a key to this transformation. (LERU)

Avoin tiede: paradokseja

”Auvoiseen tilaan päästään kun kaikkien tuottama data on kaikkien käytettävissä”,

tosin

nytkin on jo tarjolla niin paljon dataa, että sitä kaikkea ei osata / ehditä / voida hyödyntää

Avoim tiede: paradokseja 2

”Poliittinen päätöksenteko nousee uudelle tasolle kun päätökset perustuvat luotettavaan tietoon”,

tosin

tälläkin hetkellä suurin osa tarjolla olevasta tiedosta jää käyttämättä, koska politiikkaa ohjavat muut lainalaisuudet.

Avoim tiede: paradokseja 3

Viime vuodet valtionhallintoa on kaikkialla maailmassa liikelaitoistettu ja sinne on istutettu markkinamekanismeja

kun taas

tietoaineistojen vapaa jakaminen edustaa päinvastaista ideologiaa, tietynlaista sosialismia.

Vastaansanomattomat periaatteet

- **(Perus)tutkimuksen avoimuusprinsiippi:** kaiken tiedon on oltava vapaata, kaikki tuloksiin johtava perusdata tulee olla muiden tutkijoiden tarkistettavissa
- **Demokratia:** kaikilla on oikeus päästä käsiksi tietoon ja dataan
- **Veronmaksajan näkökulma:** kaikki verovaroin tuotettu tieto tulee olla kansalaisten käytettävissä

Avoimen tieteen ”lajit”

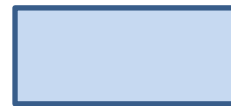
- Julkaisujen avoin saatavuus (**open access**)
- Avoin data (**open data**)
 - Julkisten data-aineistojen avoin saatavuus
 - Tutkijoiden tuottaman datan avoin saatavuus
- Kansalaistiede (**citizen science**)
(talkoistaminen)

HYÖTYJÄT

	Kansalaiset	Tiede	Elinkeinoelämä	Päätäjät
avoin julkaiseminen		hyvin merkittävä hyöty	jonkin verran hyötyä	
avoin julkinen data	jonkin verran hyötyä	hyvin merkittävä hyöty	merkittävä hyöty	jonkin verran hyötyä
avoin tieteellinen data	jonkin verran hyötyä	hyvin merkittävä hyöty	jonkin verran hyötyä	
kansalaistiede	jonkin verran hyötyä	merkittävä hyöty	jonkin verran hyötyä	jonkin verran hyötyä



hyvin merkittävä hyöty



jonkin verran hyötyä



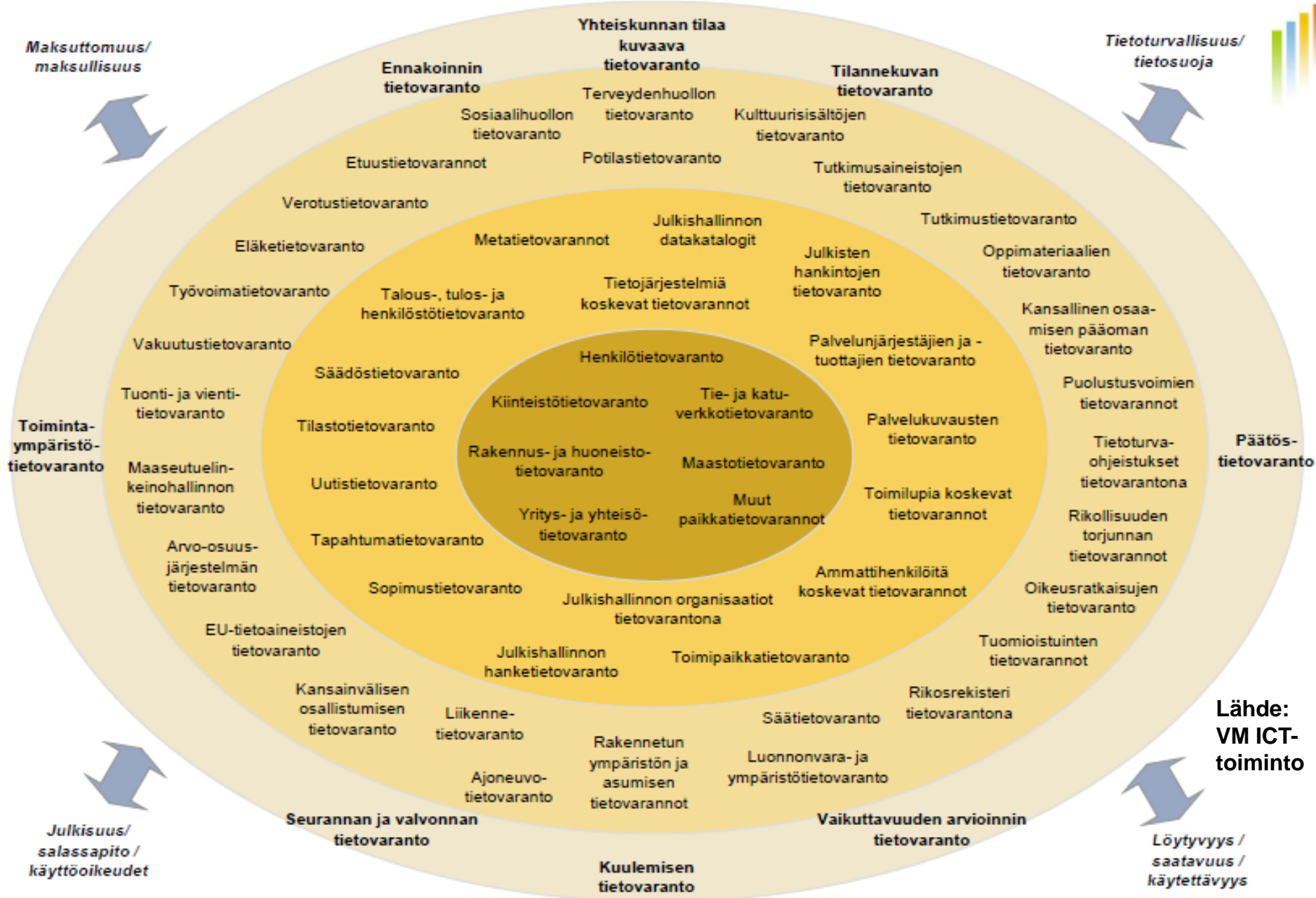
merkittävä hyöty



ei juurikaan hyötyä

”How **Europe** can gain from the rising tide of scientific data”

Valtionhallinnon tietoaisteistoja



Esteitä ja hidasteita

Pelot:

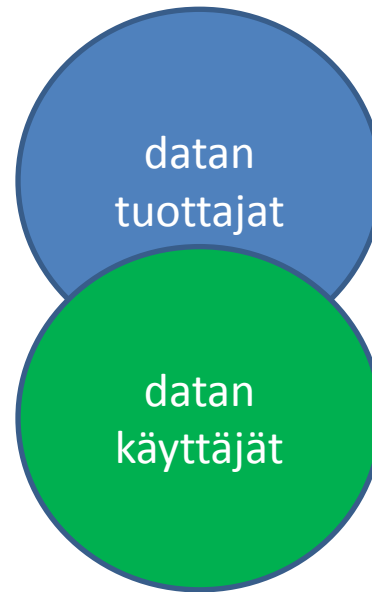
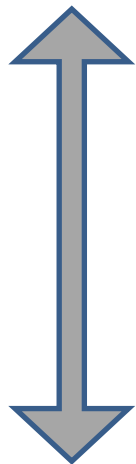
- väärinkäytön pelko
- väärin tulkintojen pelko
- vallan menetys

Hidasteet:

- vie ylimääräistä aikaa (tekijä itse ei voita mitään)
- ei ole välineitä

Metadata

- Dataa koskeva data
 - Kaksi funktiota:
 - Voidaan arvioida aineiston luotettavuus
 - Väline datan löytämiseen
 - Kuinka piilossa metatieto on? Pakotetaanko käyttäjä lukemaan se?
 - Voidaanko ennakoida sudenkuoppia datan käytössä? Estää sen väärinkäytön



CLARIN

Puitteiden luominen

- yhteensopivuus, formaatti (puhekielen litteroinnin tarkkuus)
- IPR- / copyright-ongelmat
- helppo pääsy, käyttäjäkategoriat

Kenelle tehdään? (Vrt. kirjaston rooli ?)

Mielekkyyys ja tuloksellisuus on kiinni kahdesta asiasta:

- haluavatko tutkijat aineistojaan sinne
- käyttävätkö tutkijat niitä (ja opettavatko oppilaansa käyttämään) – käyttävät vain jos he **TIETÄVÄT** niiden olemassa olosta ja **KÄYTTÄMINEN ON RIITTÄVÄN YKSINKERTAISTA**

INTERNET, GOOGLE

- **Internet** on erilaisen datan ja informaation valtava varasto
- **Tieteen datanet** (sciencenet) on tieteellisen data valtava varasto
- Internetin datavarasto saa merkityksen vain jos datan löytää eli tarvitaan **Google** (tai vastaava hakukone) jolloin 10 käyttäjän sijasta voi olla miljoona käyttäjää

Seurauksia

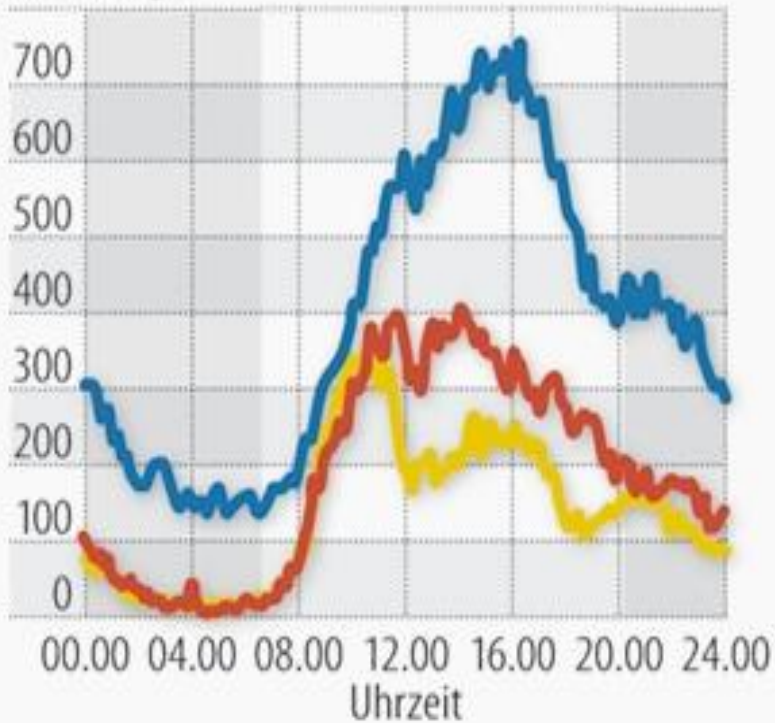
Avoimella tieteellä voi olla yllättäviä seurauksia koko tieteen tekemisen struktuurille:

- tieteen edistyminen
- kansalaisten / poliitikkojen käsitys siitä, mitä tiede on
- tieteen mittarit (jakaminen vs. itsekkyyys)
- tutkijakoulutus
- massiivisen datan avulla tehtävät yleistykset (vrt. Culturomics)
- ”numerovalta” kasvaa? (**data** - information - knowledge - wisdom)

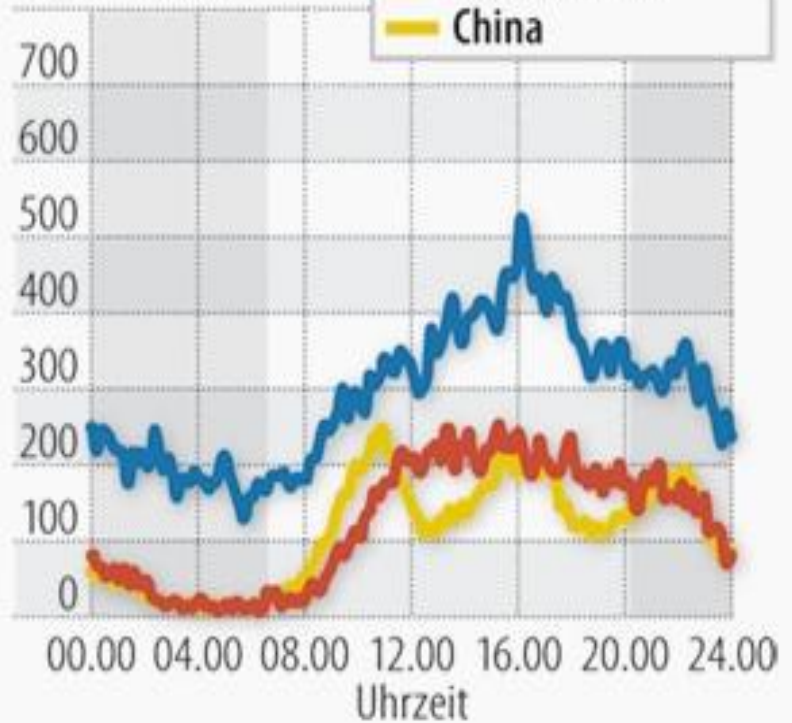
Forscher sind auch nachts und an Wochenenden aktiv

Zahl der Downloads beim Springer-Wissenschaftsverlag im Tagesverlauf

Werktage



Wochenende



Quelle und Grafikvorlage: Xianwen Wang et al., Exploring Scientists' Working Timetable: Do Scientists Often Work Overtime?, in Journal of Infometrics, 6 (4), S. 655–660

F.A.Z.-Grafik Heumann

Число публикаций по запросу относительно общего числа публикаций, %

