

Tutkimuksen tietoaineistojen saatavuuden parantaminen: metatietomallit ja kokonaisarkkitehtuuri

Riitta Maijala

Opetus- ja kulttuuriministeriö
Undervisnings- och kulturministeriet

Mitä hyötyä yhteiskunnalle tuottaa tieteellisten julkaisujen ja aineistojen parempi saatavuus?

- Käytetään paremmin hyödyksi aiempia tutkimustuloksia
 - tutkimustulosten parempi laatu; tieteen kyky korjata itseään paranee
- Tarjotaan laajempi aineisto yhteiskunnallisen päätöksenteon tueksi
 - tieteeseen perustuva päätöksenteko helpottuu
- Tietoon pohjautuva osaaminen kehittyy
 - helpompi saatavuus tuo tieteellisen tutkimuksen tulokset laajemmin opetuksen ja osaamisen kehittämisen välineeksi
- Nopeutetaan innovointia
 - nopeammin markkinoille = > nopeampi kasvu
- Edistetään yhteistyötä ja vältetään päällekkäistä työtä
 - suurempi tehokkuus
- Osallistetaan kansalaisia ja yhteiskuntaa
 - tieteellisen prosessin suurempi avoimuus

EU-tason linjauksia

- EU:n tutkimuksen ja innovoinnin puiteohjelma
 - tieteellisten julkaisujen avoimen saatavuuden pilottihankkeen laajentaminen koskemaan koko Horisontti 2020 -ohjelmaa
 - tutkimusdatan avointa saatavuutta koskeva pilottihanke Horisontti 2020 -ohjelmassa
- Jäsenmaat
 - **määrittelevät selkeät toimintalinjat** julkisrahoitteisen tutkimustyön tuloksena syntyvien tieteellisten julkaisujen sekä tutkimusdatan levittämistä ja avointa julkaisemista varten
 - **varmistavat**, että julkisrahoitteisesta tutkimuksesta syntyvät tieteelliset julkaisut julkaistaan avoimesti mahdollisimman pian ja **että tutkimusaineistot ovat yleisesti saatavilla ja käytettävissä**
- Korkean tason asiantuntijaryhmän raportti: Riding the Wave, How Europe can gain from the rising tide of scientific data, 2010
- Komission tiedonanto: Kohti tieteellisen tiedon parempaa saatavuutta (COM(2012) 401, 17.7.2012)
- Komission suositus jäsenmaille tieteellisen tiedon saatavuudesta ja säilyttämisestä (C(2012) 4890, 17.7.2012)

the WHITE HOUSE PRESIDENT BARACK OBAMA

BLOG PHOTOS & VIDEO BRIEFING ROOM ISSUES the ADMINISTRATION

Home • The Administration • Office of Science and Technology Policy

Office of Science and Technology Policy

About OSTP | OSTP Blog | Pressroom | Divisions | R&D Budgets | Resource Library | NSTC

Expanding Public Access to the Results of Federally Funded Research Subscribe

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EST

E-Mail Tweet Share +

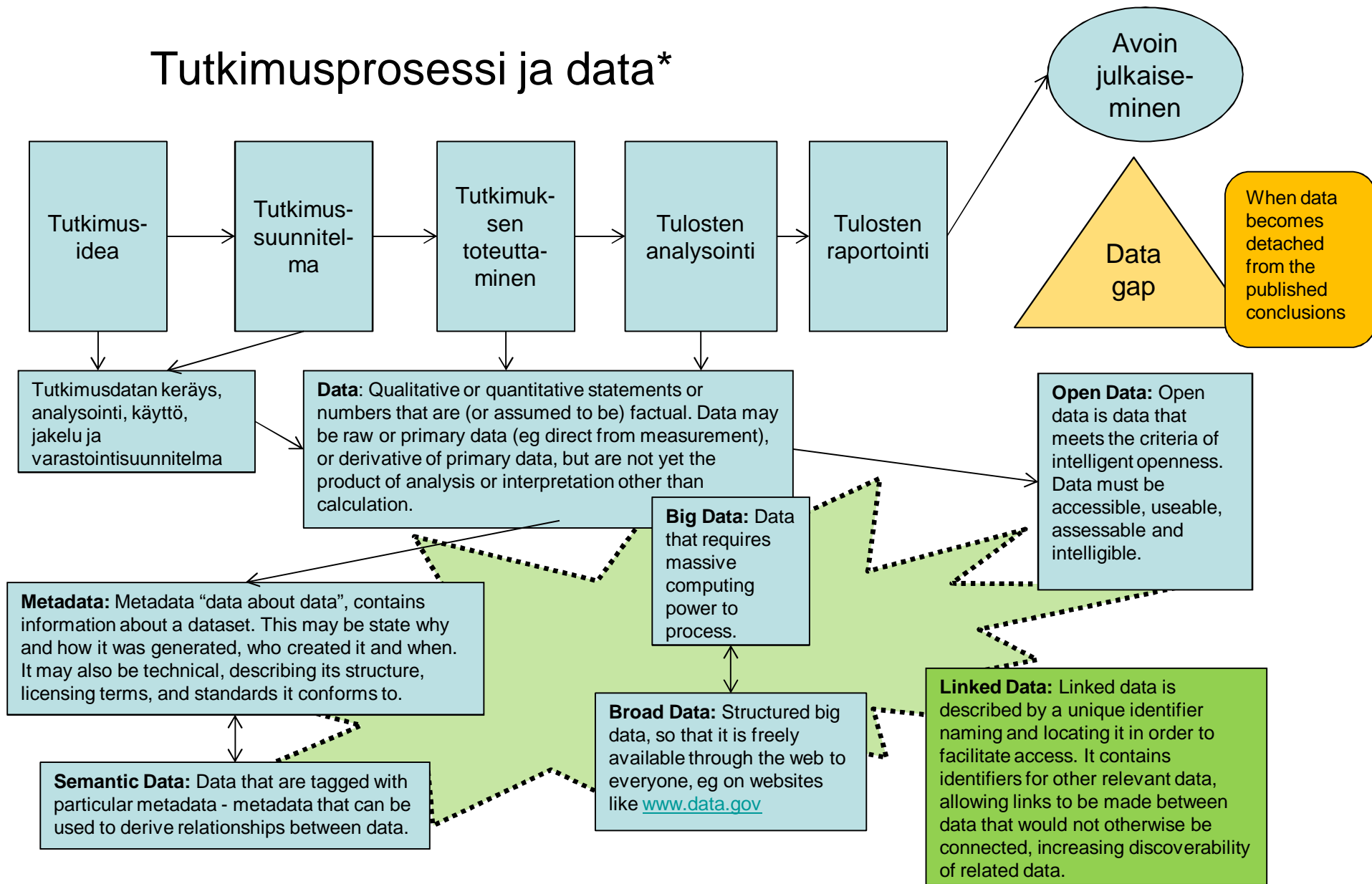
The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop a policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other members of the public—over 65 thousand of whom recently signed a *We the People* petition asking for expanded public access to the results of taxpayer-funded research.

To see the new policy memorandum, please visit: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

“... the results of unclassified research that are published in peer-reviewed publications directly arising from Federal funding should be **stored for long-term preservation and publicly accessible to search, retrieve, and analyze...**”

“... digitally formatted **scientific data** resulting from unclassified research supported wholly or in part by Federal funding should be **stored and publicly accessible** to search, retrieve, and analyze.”

Tutkimusprosessi ja data*



Osaamistarpeiden muutokset muuttuvassa tutkimusprosessissa

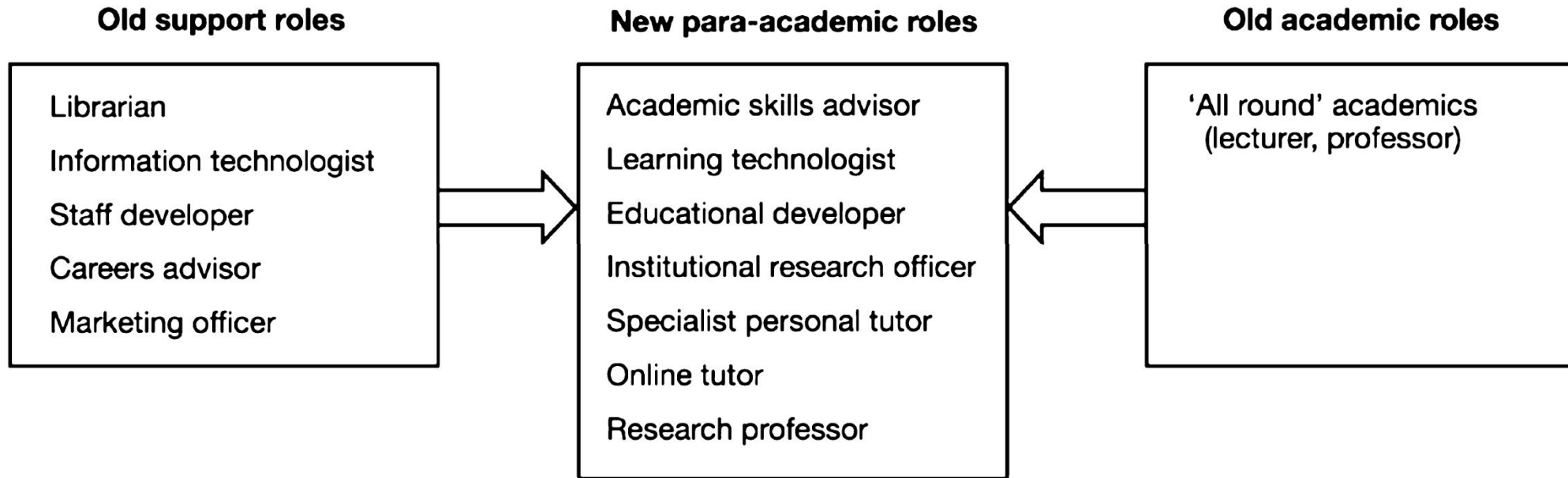


Figure 2 The origins of the para-academic

The Morphing of Academic Practice: Unbundling and the Rise of the Para-academic. Bruce Macfarlane, University of Hong Kong, bmac@hku.hk; Higher Education Quarterly, 0951–5224
DOI: 10.1111/j.1468-2273.2010.00467.x
Volume 65, No. 1, January 2011, pp 59–73

Keskeisiä käsitteitä

- **Data:** Qualitative or quantitative statements or numbers that are (or assumed to be) factual. Data may be raw or primary data (eg direct from measurement), or derivative of primary data, but are not yet the product of analysis or interpretation other than calculation.
- **Big Data:** Data that requires massive computing power to process.
- **Broad Data:** Structured big data, so that it is freely available through the web to everyone, eg on websites like www.data.gov
- **Data-gap:** When data becomes detached from the published conclusions
- **Data-intensive science:** Science that involves large or even massive datasets

Lähde: Science as an open enterprise, The Royal Society 2012,
http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Keskeisiä käsitteitä...

- **Open Data:** Open data is data that meets the criteria of intelligent openness. Data must be accessible, useable, assessable and intelligible.
- **Linked Data:** Linked data is described by a unique identifier naming and locating it in order to facilitate access. It contains identifiers for other relevant data, allowing links to be made between data that would not otherwise be connected, increasing discoverability of related data.
- **Metadata:** Metadata “data about data”, contains information about a dataset. This may be state why and how it was generated, who created it and when. It may also be technical, describing its structure, licensing terms, and standards it conforms to.
- **Semantic Data:** Data that are tagged with particular metadata - metadata that can be used to derive relationships between data.

Lähde: Science as an open enterprise, The Royal Society 2012,
http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

VM:n Avoimen tiedon ohjelma 2013-2015



Tehtävä: koordinoidaan ja vauhditetaan julkisen hallinnon tiedon avaamiseen ja käytön lisäämiseen tärkeitä toimenpiteitä.

Tavoitteet:

Julkishallinnon tuottamat julkiset tiedot ovat hyödynnettävissä käyttökelpoisessa ja koneluettavassa muodossa taloudellisesti ja turvallisesti, ottaen huomioon yksityisyyden suoja ja muut lainsäädäntöön perustuvat rajoitukset.

Keskeisimmät julkisen hallinnon tietovarannot ovat avoimesti saatavilla verkossa ja niitä käytetään laajasti ja monipuolisesti.

Puretaan tiedon käytön esteitä ja luodaan kannustimia tiedon avaamiselle sekä palvelujen kehittäjäverkostojen ja ekosysteemien työlle.

Työtä tiedon saatavuuden hyväksi OKM:ssä

Hankkeita
tietovarantojen
avaamiseksi

Tutkimuksen
tietoaineistot (TTA)

Kansallinen digitaalinen
kirjasto (KDK)

Avoimen tiedon
instrumentit (AVOIN)

Pitkäaikaissäilytys
(TTA-PAS ja KDK-PAS)

Kokonaisarkkitehtuuryö OKM:n kohdealueilla

Yhteentoimivuus, tietovarannot, kohdealueiden arkkitehtuurit

Tutkimuksen tietoaaineistot -hanke (TTA)

Koordinaatioryhmä

Ohjausryhmä

Metatietoryhmä

Tietoinfrastruktuuri-ryhmä

PAS-selvitys

Tiedon
saatavuus

TUTKIMUS-IDA

DATAKATALOGI

DATANHALLINNAN
PILOTTI

AVOIN (Avoimen
tiedon instrumentit)

Tilaisuuden tavoitteet ja jatkotoimet

- Ymmärryksen lisääminen metatiedon käsittelystä ja kokonaisarkkitehtuurityöstä
 - näkemyksiä eri näkökulmista, myös kriittisiä puheenvuoroja
- Tehdyn työn tuominen esille
 - mitä on saatu aikaan, mitä on vielä edessä
- Kehityskohteiden identifiointi ja palaute TTA-hankkeen työryhmille maaliskuun loppuun mennessä
 - Metatietoryhmä, pj. Mari Kleemola, Yhteiskuntatieteellinen tietoarkisto
 - Tietoinfrastruktuuriryhmä, pj. Esa-Pekka Keskitalo, Helsingin yliopisto

TTA-hankkeen tilaisuuksia keväällä 2013

- 4.3.
 - Avoin tiede -keskustelutilaisuus
- 10.4. klo 9-16:30 (Tieteiden talo)
 - Pitkäaikaissäilytyksen keskustelutilaisuus ja Kerro meille datastasi -työpaja
- 21.5.
 - AVOIN-hankkeen seminaari

OHJELMA: Metatietomallit ja kokonaisarkkitehtuuri

- 9:30 - 10:15 *Kahvi ja keskustelumahdollisuus*
- 10:15 - 10:30 *Seminaarin avaus* (Juha Haataja, OKM)
- 10:30 - 10:45 *TTA-hankkeen esittely* (Pirjo-Leena Forsström, CSC)
- 10:45 - 11:30 *TTA-kokonaisarkkitehtuuri* (Esa-Pekka Keskitalo, Kansalliskirjasto)
- 11:30 - 12:15 *TTA-minimimetatietomalli* (Mari Kleemola, Yhteiskuntatieteellinen tietoarkisto)
- 12:15 - 13:00 *Buffet-lounas*
- 13:00 - 13:30 *Julkishallinnon standardisointi ja PAS-metatiedot* (Juha Hakala, Kansalliskirjasto)
- 13:30 - 14:00 *Semanttinen tieto, ontologiat* (Matias Frosterus, Kansalliskirjasto)
- 14:00 - 14:30 *Kentän näkökulma* (Antti Auer, Jyväskylän yliopisto)
- 14:30 - 15:00 *Yhteenveto*