

# Avoim tiede – tutkimustulosten saatavuus

Juha Haataja

Opetus- ja kulttuuriministeriö  
Undervisnings- och kulturministeriet

# Esityksen teemat

- Mitä hyötyä yhteiskunnalle tuottaa tieteellisten julkaisujen ja aineistojen parempi saatavuus
- Mitä OKM on aihepiirin suhteen tekemässä
- Tutkimuksen tiedonhallinnan prosessi – onko se hallussa?
- Tutkimusaineistojen pitkäaikaissäilytys

# 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.

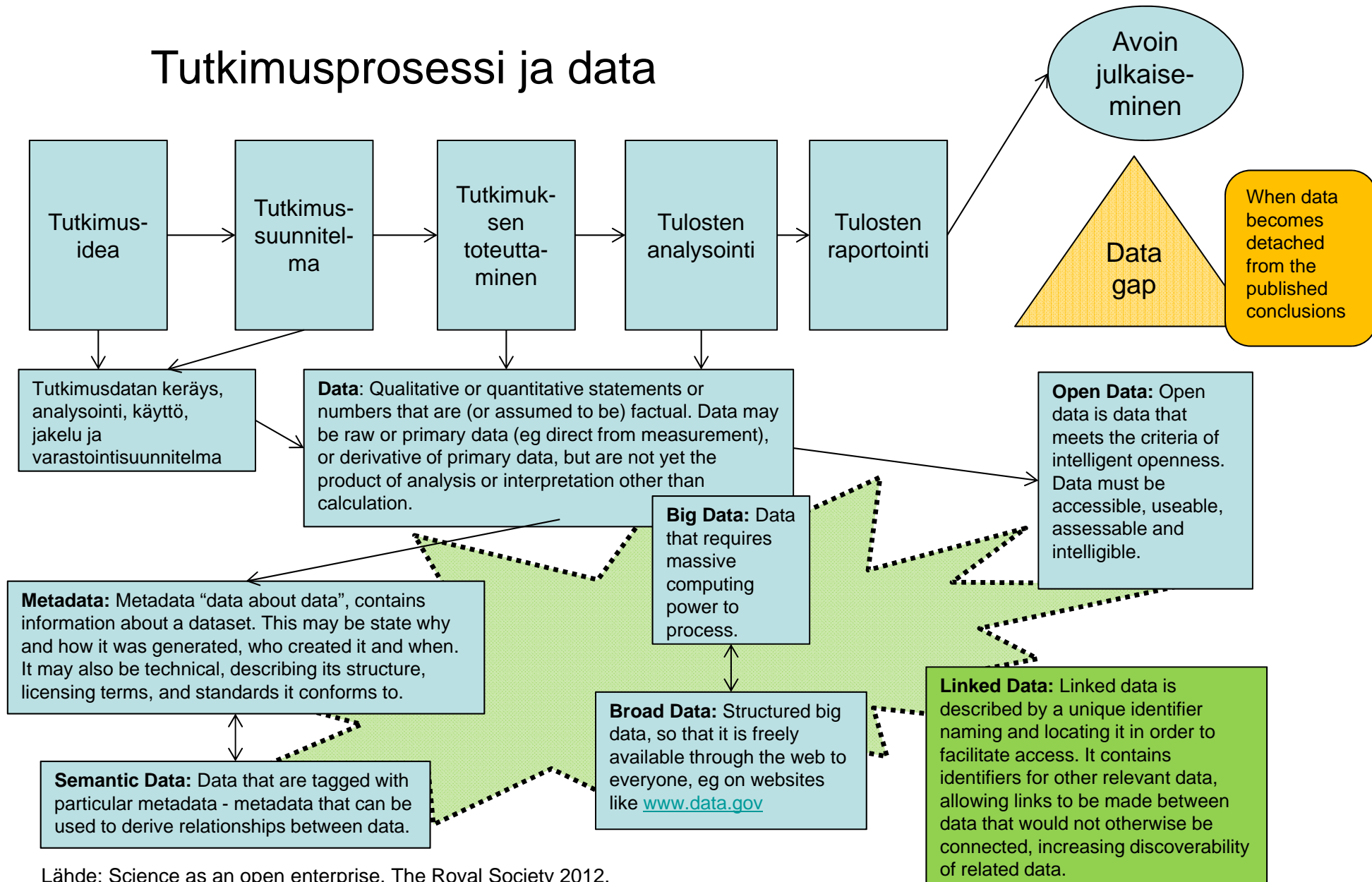
To see the new policy memorandum, please visit: [http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

“... 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...**”

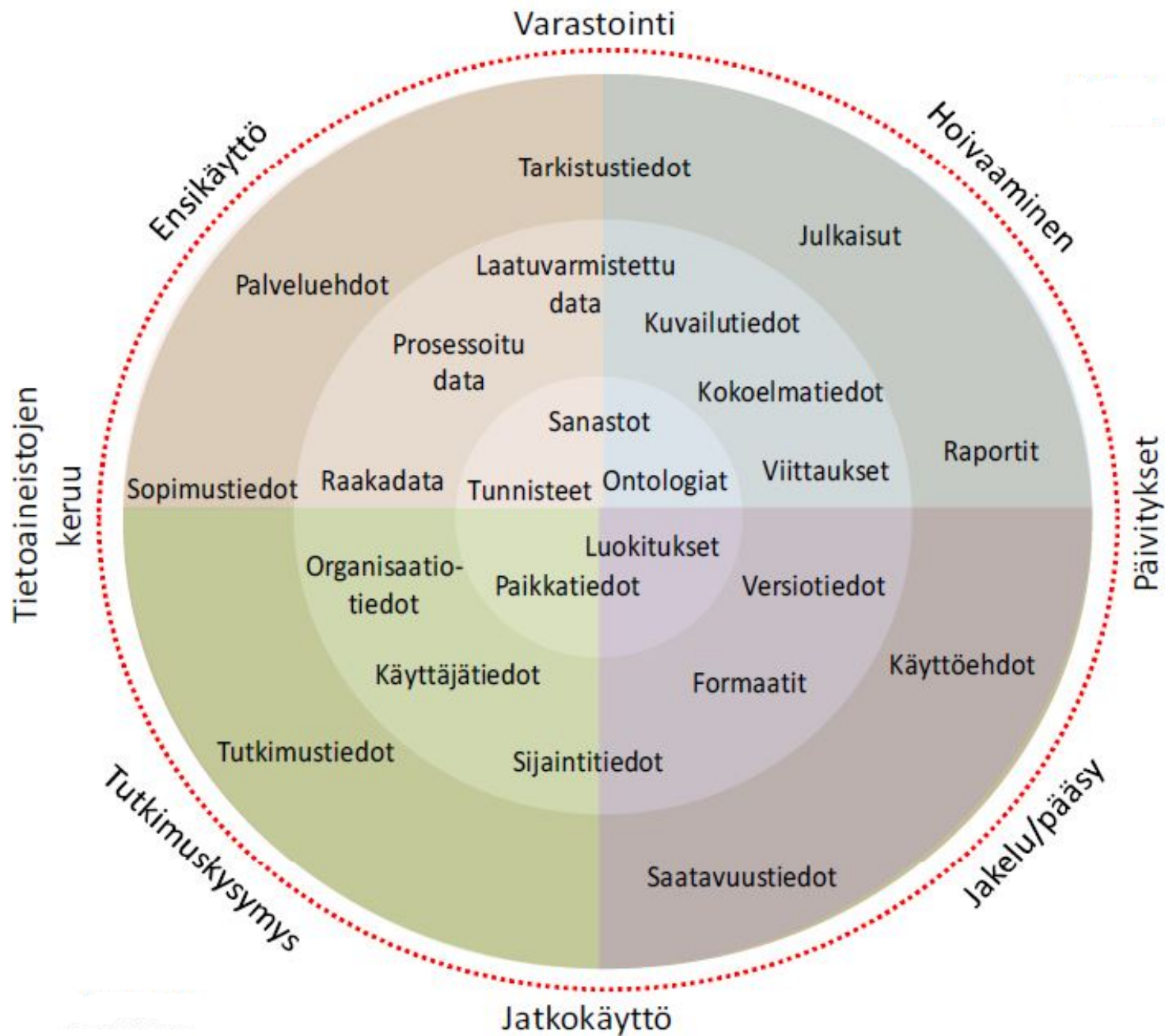
“... digitally formatted **scientific data** resulting from unclassified research supported wholly or in part by Federal funding should be **stored and publicly accessible to search, retrieve, and analyze.**”

# Tutkimusprosessi ja tiedon hallinta

# Tutkimusprosessi ja 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](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf)





# Mikä OKM on tekemässä asian hyväksi?

# 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)

Kokonaisarkkitehtuurityö OKM:n kohdealueilla  
Yhteentoimivuus, tietovarannot, kohdealueiden arkkitehtuurit

# Tutkimuksen tietoaaineistot -hanke (TTA)

## Koordinaatioryhmä

## Ohjausryhmä

Metatietoryhmä

Tietoinfrastruktuuri-  
ryhmä

PAS-selvitys  
(ja -työryhmä)

Tiedon  
saatavuus

TUTKIMUS-IDA

DATAKATALOGI

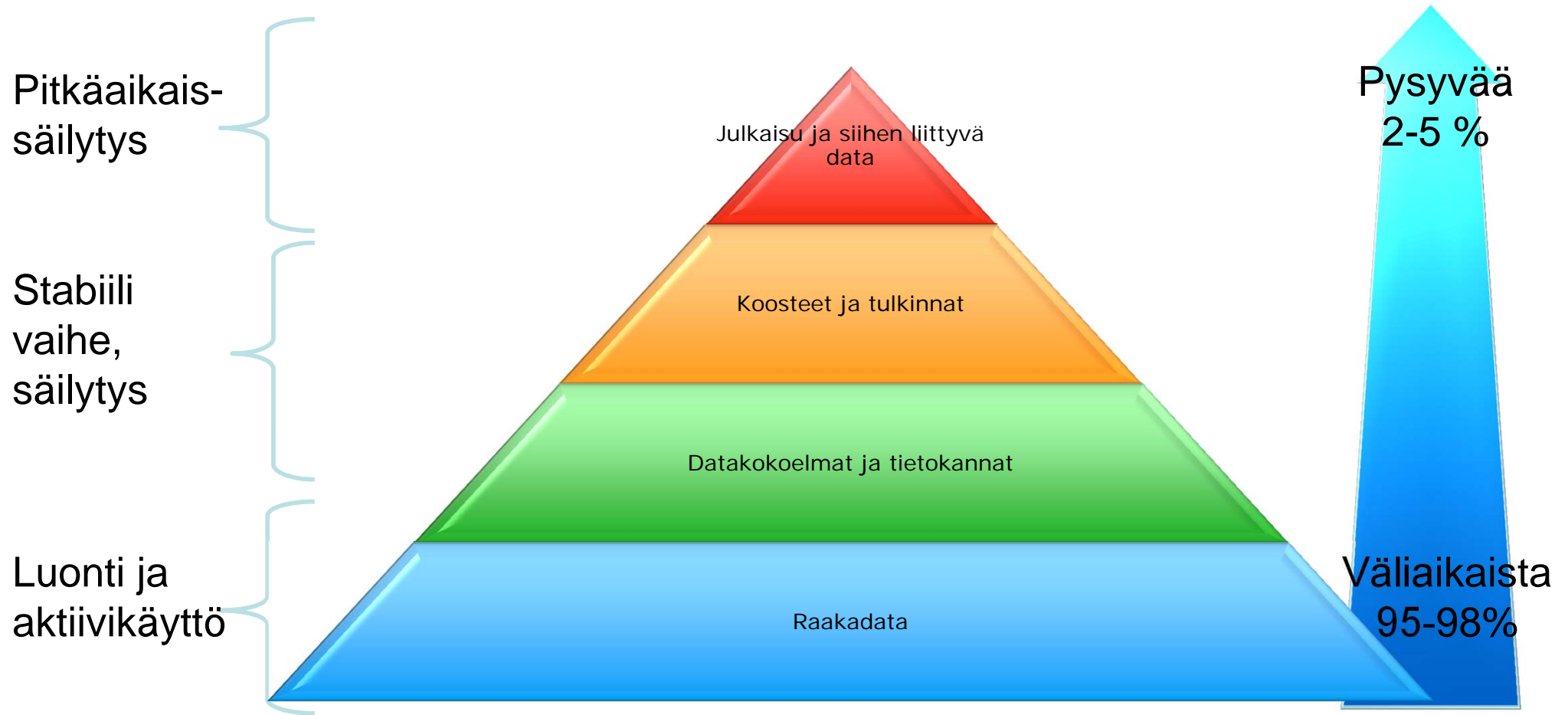
DATANHALLINNAN  
PILOTTI

AVOIN (Avoimen  
tiedon instrumentit)

# Tutkimusaineistojen pitkäaikaissäilytys

- Digitaalisten aineistojen pitkäaikaissäilytys tarkoittaa digitaalisen informaation **luotettavaa säilyttämistä kymmenien tai jopa satojen vuosien ajan**
  - Laitteet, ohjelmistot, mallit ja tiedostomuodot vanhenevat, mutta tästä huolimatta informaation täytyy säilyä ymmärrettävänä

# Tutkimuksen tietoaineistopyramidi



Communication from the Commission (17.7.2012)  
*Towards better access to scientific information:  
Boosting the benefits of public investments in research*

**4.3. Preservation of scientific information**

There are considerable economic and social benefits to be gained from the long-term preservation of information, knowledge and know-how for use by future generations. The UK research funding organisation JISC carried out a cost/benefit analysis of the preservation of research data. It found that preservation efforts lead to a four-fold return in terms of cost saving alone.

Member States are currently amending their legislation on the depositing of digital material.

Special attention must also be given to preserving scientific software and models in order to keep information re-usable and re-producible. Open standards, formats and open source software solutions can help ensure this.

## Commission recommendation of 17.7.2012 on *access to and preservation of scientific information*

(1) The Communication from the Commission Europe 2020 puts forward the development of an economy based on knowledge and innovation as a priority.

(11) Preservation of scientific research results is in the public interest. It has traditionally been under the responsibility of libraries, especially national legal deposit libraries. The volume of research results generated is growing tremendously. Mechanisms, infrastructures and software solutions should be in place to enable long-term preservation of research results in digital form. Sustainable funding for preservation is crucial as curation costs for digitised content are still relatively high. Given the importance of preservation for the future use of research results, the establishment or reinforcement of policies in this area should be recommended to Member States.

# ”Lööppiotsikoita”

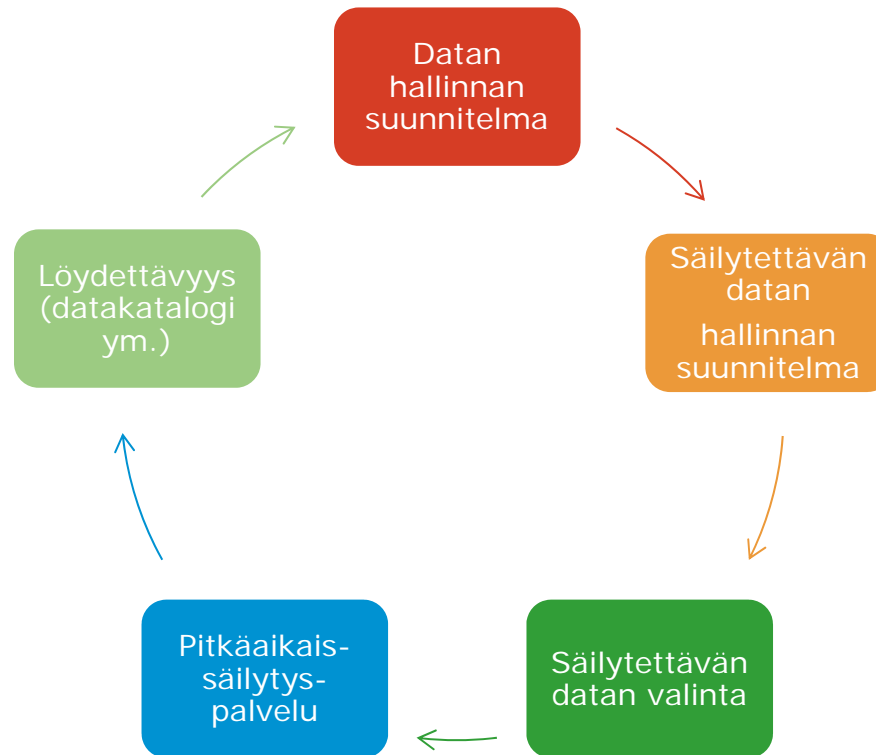
- ”Koululaisten oppimistulosten aikasarjat viime vuosikymmeniltä katosivat mystisesti”
- ”Kaikki vanhat Suomi-filmit menetettiin viime yönä”
- ”Kuka kävi manipuloimassa suomalaisten terveystietodataa?”
- ”Pitkäaikaissäilytyksen ICT-kustannukset räjähtivät moninkertaisiksi”
- ”Tutkimusaineistossa käytettyä tiedostoformaattia ei osata enää lukea”



# Pitkäaikaissäilytyksen osat

Riskienhallinta

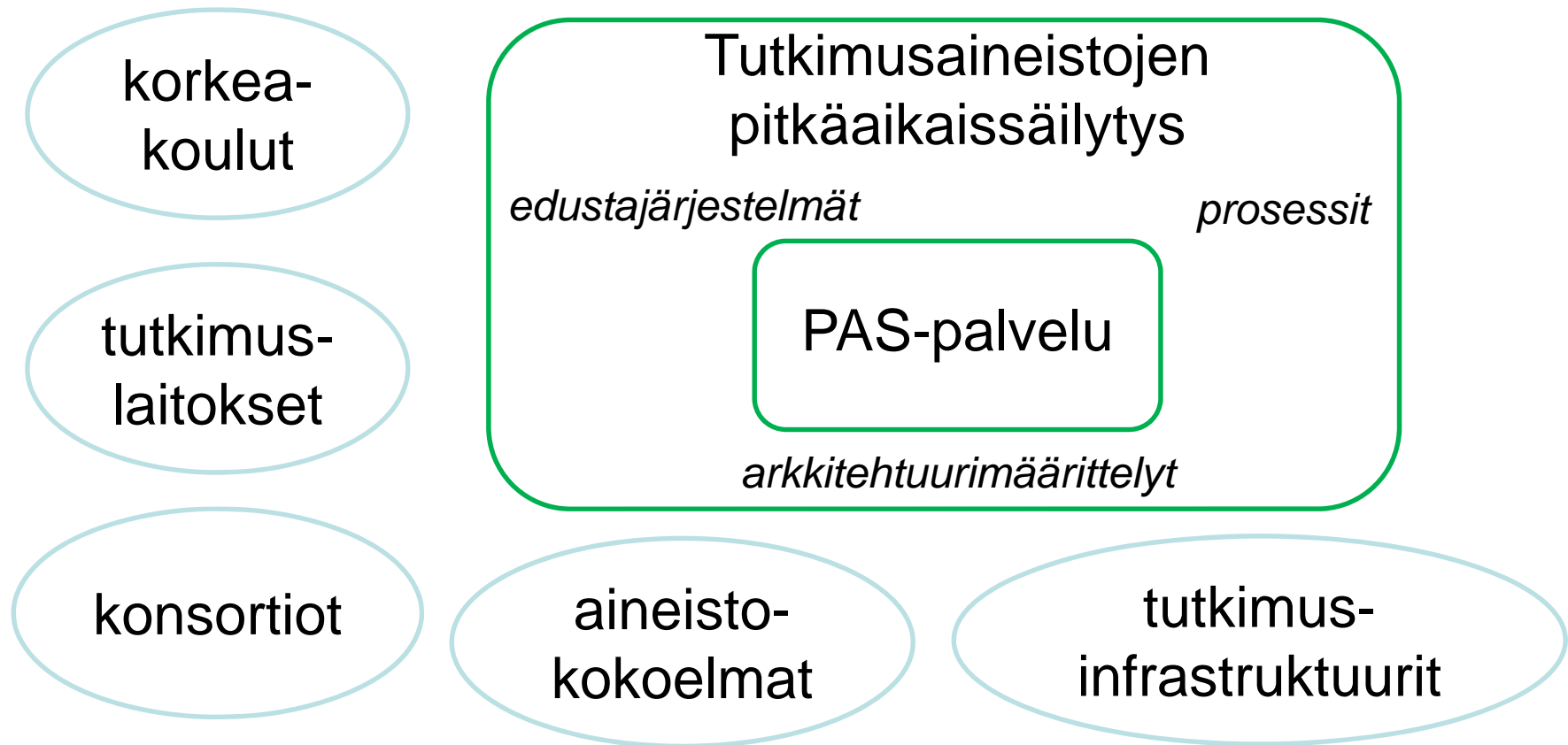
Datanhallinnan  
politiikat ja  
strategiat



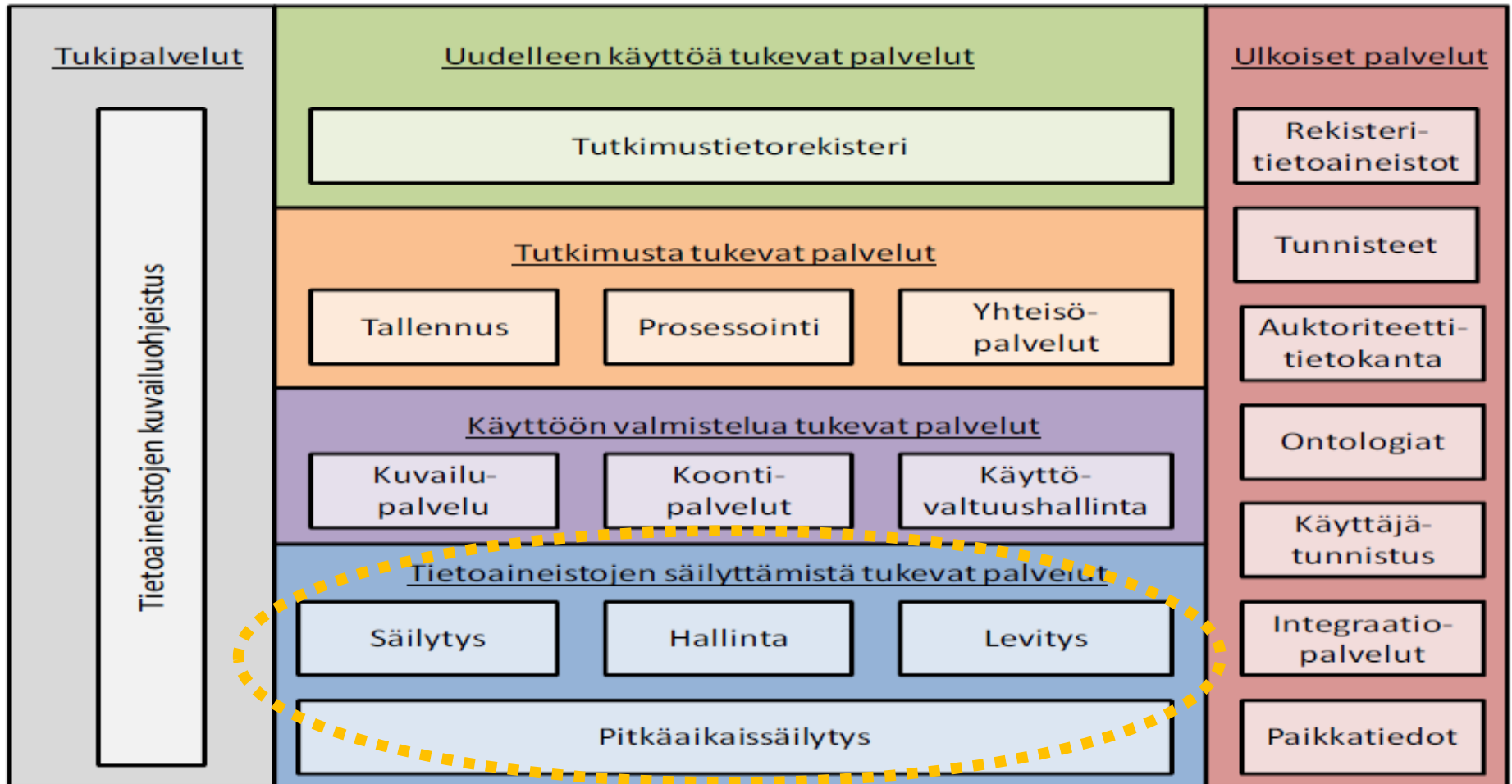
Rahoitusmallit

Opastus, koulutus, tuki

# Pitkäaikaissäilytyksen toimijoita ja rooleja



# TTA-infrastruktuuri



# TTA-PAS työryhmän perustaminen

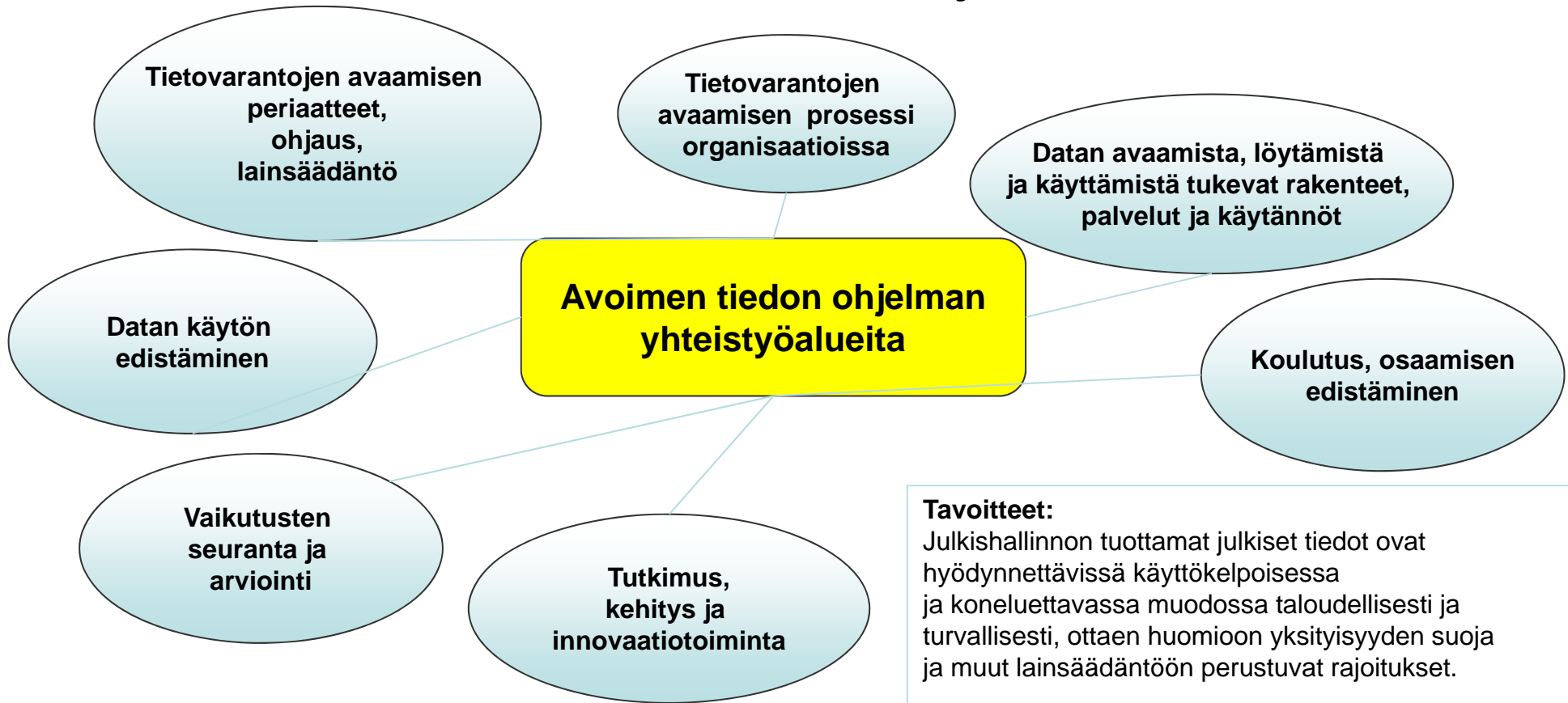
- Tehtävinä
  - Pitkäaikaissäilytyksen hyödyntäjien ja hyödyntämismallin kypsyystason tunnistaminen
  - Pitkäaikaissäilytykseen tarvittavien metatietojen määrittely yhdessä metatietotyöryhmän kanssa
  - Pitkäaikaissäilytysprosessin täsmennys yhdessä tietoinfrastruktuurityöryhmän kanssa
  - Pilotoinnin ja vastaanottopalvelun alustava suunnittelu
  - Ehdotus etenemisestä
- Jäseniä mahdollisesti seuraavien toimijoiden joukosta: UNIFI, FUCIO, AAPA, TSV, Tiedeakatemia, tutkimuslaitokset, ...

# TTA-PAS työryhmän perustaminen...

- Jäseniltä toivotaan osaamista seuraavista asioista:
  - Tutkimusaineistojen metatieto
  - Tutkimusaineistojen säilytysratkaisut
  - Kokonaisarkkitehtuuri
  - Tutkimusinfrastruktuurien sähköiset palvelut
  - Tutkimushallinnon ja tutkimuksen prosessien tuntemus
  - Tutkimusaineistokokoelmat
- Määräaika vuoden 2013 loppuun (ja tästä eteenpäin uusi tehtävän määrittely)
- Lisätietoa
  - Pirjo-Leena Forsström, TTA-hankkeen pääsihteeri, CSC
  - Juha Haataja, OKM

# Avoim tieto julkishallinnossa

# VM:n Avoimen tiedon ohjelma 2013-2015



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

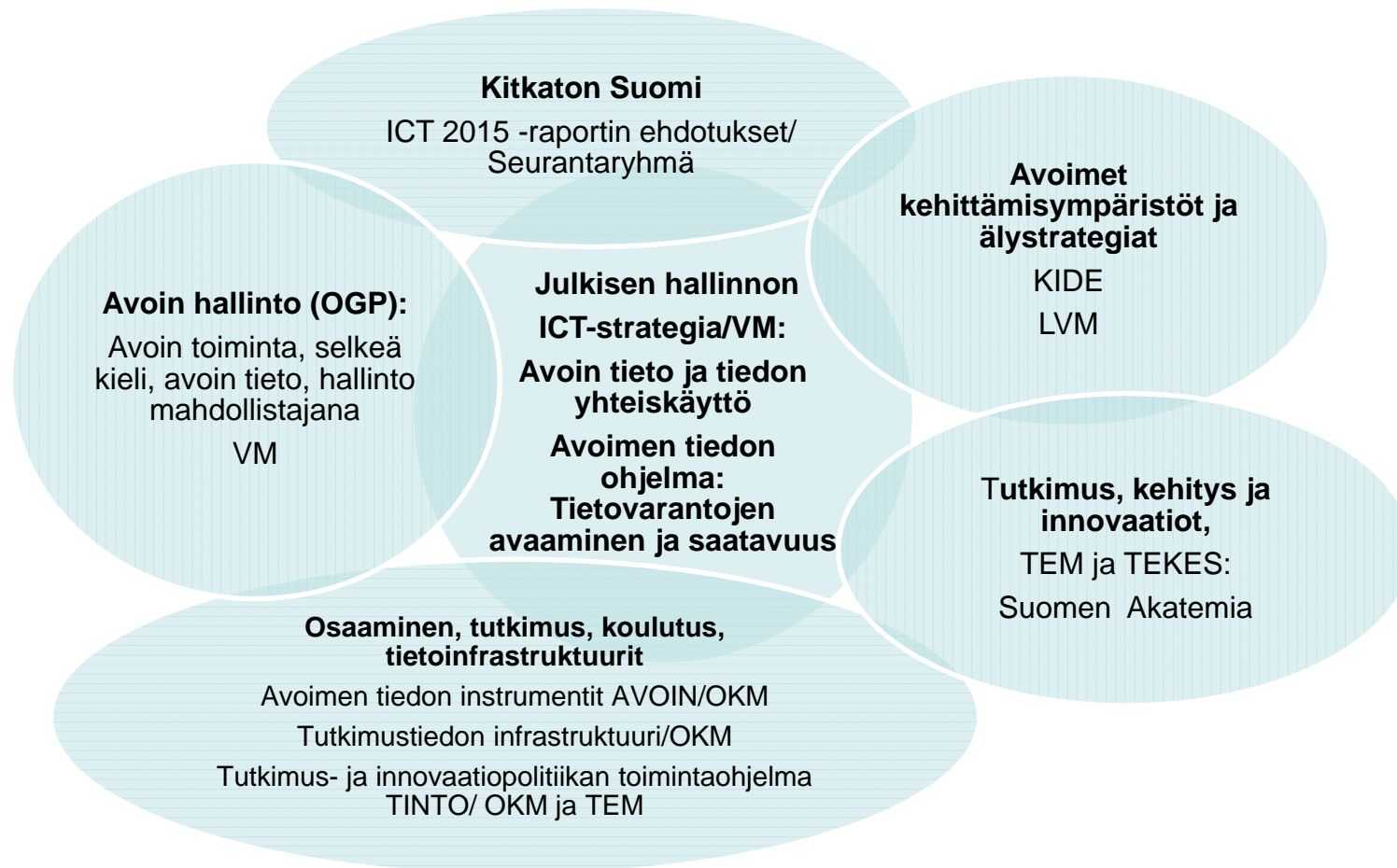
## **Tavoitteet:**

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.

# Avointa tietoa edistäviä ohjelmia ja hankkeita



Lähde: Anne Kauhanen-Simanainen, VM



# Avoimen tiedon instrumentit (AVOIN-hanke)

Palvelu-  
kehitys  
(esimerkkejä)

Avointen aineistojen  
julkaisun alusta

Tutkimustiedon  
avaamisen alusta

Avointen rajapintojen  
palvelu, auktoriteetti-  
tiedon palvelu ym.

Paikkatiedon liittäminen  
tutkimusaineistoihin

AVOIN-hanketta hallinnoidaan TTA-hankkeen yhteydessä

Työryhmiä: metatieto, tietoinfrastrukturi, tiedon saatavuus, ...

# Kerro minulle datastasi -työpajan 10.4. aineistokyselyn vastauksia ja tuloksia

- Auki 18.3. – 8.4., N=27
- Edustetut tutkimusalat
  - Geenitutkimus
  - Geotieteet
  - Tietojenkäsittelytiede
  - Humanistiset tieteet
  - Fysiikka
  - Tieteellinen laskenta
  - Biotieteet
  - Insinööritieteet
  - Historia
  - Taloustiede
  - Eläketutkimus
  - Lääketiede
  - Avaruusfysiikka

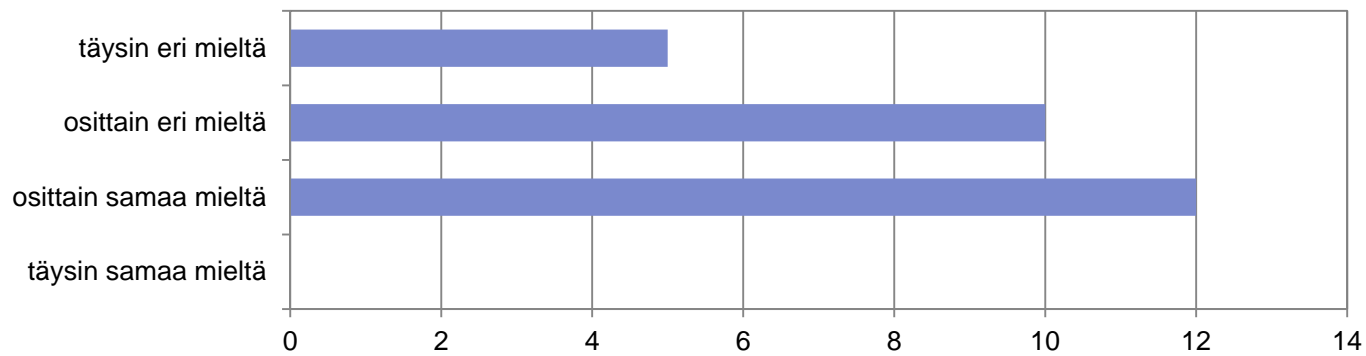
# Poimintoja aineistokyselyn tekstivastauksista: aineistojen löydettävyys ja hyödynnettävyys



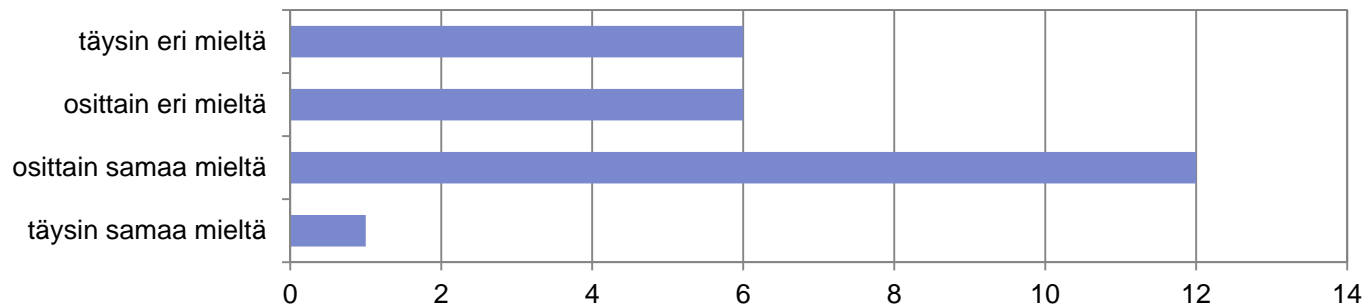
- Varsinaista dataa ei ole kovinkaan paljon saatavilla.
- Vaikeus on tietää mitä kaikkea on saatavissa.
  
- Dokumentoin datat omalla tavallani niitä luodessani, mutta dokumentaation ylläpito jää helposti leväperäiseksi
- Data ei elä päivästä toiseen. On standardiformaatissa, oman tieteenalan tutkijoiden ymmärtämässä muodossa.

# Epäsymmetrisyys!

## 1. Löydän helposti tietoa tutkimusalani jaetuista jo olemassa olevista tutkimusdatoista.



## 3. Saan paljon viittauksia ja yhteistyötarjouksia jakamieni tutkimusdatojen ansiosta.



# Onko ongelmia säilytyksessä?



- Meillä on heti ongelmia luottamuksellisten tietojen säilytyksessä ja käyttöoikeuksien hallinnassa.
- Olemme sopineet käyttöoikeuksista, mutta varsinainen protokolla ja todelliset omistajuudet ovat kyllä hieman heitteillä.
- Tallennuskapasiteetti on aina ongelma...
- Levytila riittää kyllä.
- Tämä ei ole vielä millään tavalla selvillä.

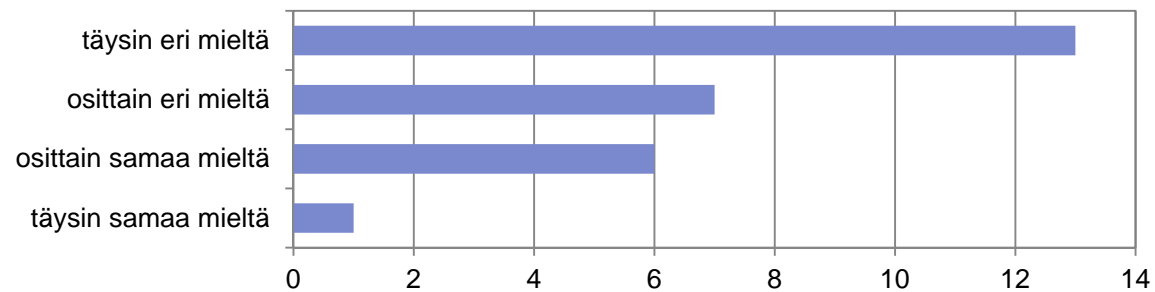
# Tutkimusaineistojen hallinnointi ja säilytys



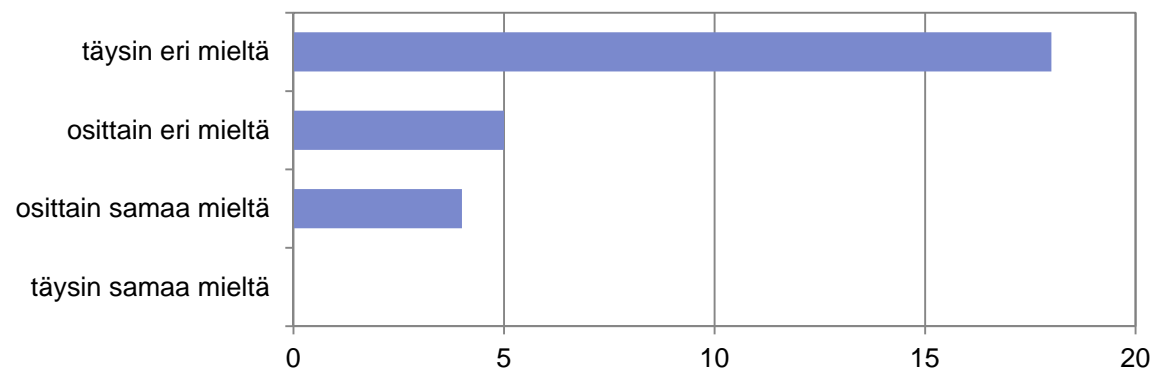
- Ei ole olemassa selkeitä ohjeita tähän puoleen.
- Osittain olemassa hyviä käytäntöjä ja standardeja, vielä tekemistä tällä saralla.
- Erilaisia standardeja on useita, mikä niistä valitaan?
- Oikeaa raakadataa ei missään nimessä saa edes luovuttaa Suomen rajojen ulkopuolelle. Sen sijaan tutkimuksessa tuotettua analyysitietoa voidaan jakaa ulkopuolelle. Tällöin ei ole kuitenkaan kyse datasta vaan raakadatatista johdetuista tuloksista.
- Minun tutkimusaineistoa ei voi todellakaan tuottaa uudestaan, koska se on uniikkiaineisto.

# Tarve säilytykseen

## 9. Tutkimusdatani siirtyvät kansainvälisiin arkistoihin pitkäaikaissäilytettäväksi.



## 10. Tutkimusdatojani ei tarvitse pitkäaikaissäilyttää, koska ne voi helposti tuottaa uudelleen.



# Kysyttävää ja keskusteltavaa



# TTA-hankkeen tilaisuuksia keväällä 2013

- Avoin tiede -keskustelutilaisuus (4.3.)
- Metatietomallit ja kokonaisarkkitehtuuri (19.3.)
- Pitkäaikaissäilytyksen keskustelutilaisuus ja Kerro meille datastasi -työpaja (10.4.)
- AVOIN-hankkeen seminaari (21.5. klo 9-12, Balderin sali, Helsinki)
- Lisätietoja ja materiaalia: <http://www.csc.fi/sivut/e-infra/tietoaineistot/>

# TTA-työpaja Kerro meille datastasi Jyväskylässä

- (Alustavaa hahmottelua...)
- Fokuksena aineiston hallinta
  - case kielitieteet ja fysiikka
- Järjestetään Jyväskylän yliopistolla viikolla 22
  - päivämäärä varmistuu vkolla 16
- Lisätietoa: Saara Värttö <Saara.Vartto@csc.fi>

# Miten tästä eteenpäin

- Mukaan tekemään ja vaikuttamaan
  - Työpajat ja keskustelutilaisuudet
  - TTA-hankkeen työryhmien työ
- Ota yhteyttä
  - Tiedon saatavuus -työryhmä, pj. Annikki Roos, HY
  - TTA:n kokonaisarkkitehtuurityö etenee tavoitetilan hahmottamisesta nykytilan kartoitukseen (lisätietoja: Esa-Pekka Keskitalo, Kansalliskirjasto)
  - TTA-minimimetatietomalli on suositus vähimmäismetatiedoista (lisätietoja: Mari Kleemola, FSD)
  - Tutkimusaineistojen PAS-ratkaisua kehitetään yhteistyössä KDK-hankkeen kanssa (lisätietoja: Pirjo-Leena Forsström, CSC)