



Tutkimuksen tietoineistot -hanke (TTA)

Tutkimuksen trendejä



- Yksilövetoisuudesta tutkimusryhmä/organisaatiotasolle
- Moniosaajien yhteistyö, joukkoistaminen
- Avoin tiede
- Yhteiskunnallinen hyöty
- Kustannustehokkuus
- Tietomassojen kasvu

Avoin tiede:



- Tutkimuksen tuottamat tietoaineistot käytettävissä
- Avoin pääsy tutkimuksen tuottamiin julkaisuihin
- Mahdollisuus käyttää tutkimusmenetelmiä joilla aineisto tuotettu
- Avoin infrastruktuuri jatkokäyttöä varten

Tutkimusaineistojen hallinta



- Systemaattinen hallinta oleellista jotta tieto ja sen evidenssiketju voidaan turvata
 - Tukee tutkimuksen ydintoimintoja
 - Vaikuttaa moniin tutkimusprosessin vaiheisiin
 - mahdollistaa luotettavuuden arvioinnin, ja siten tuo uskottavuutta tutkijalle, tutkimusryhmälle, organisaatiolle, rahoittajille



Ongelman yksi ulottuvuus



Tutkimuksen tietoineistot TTA



TTA-hanke edistää tietoineistojen **kuvausten yhtenäistämistä, säilytystä ja käyttöä.**

TTA-hankkeessa tuotetaan tätä tukemaan:

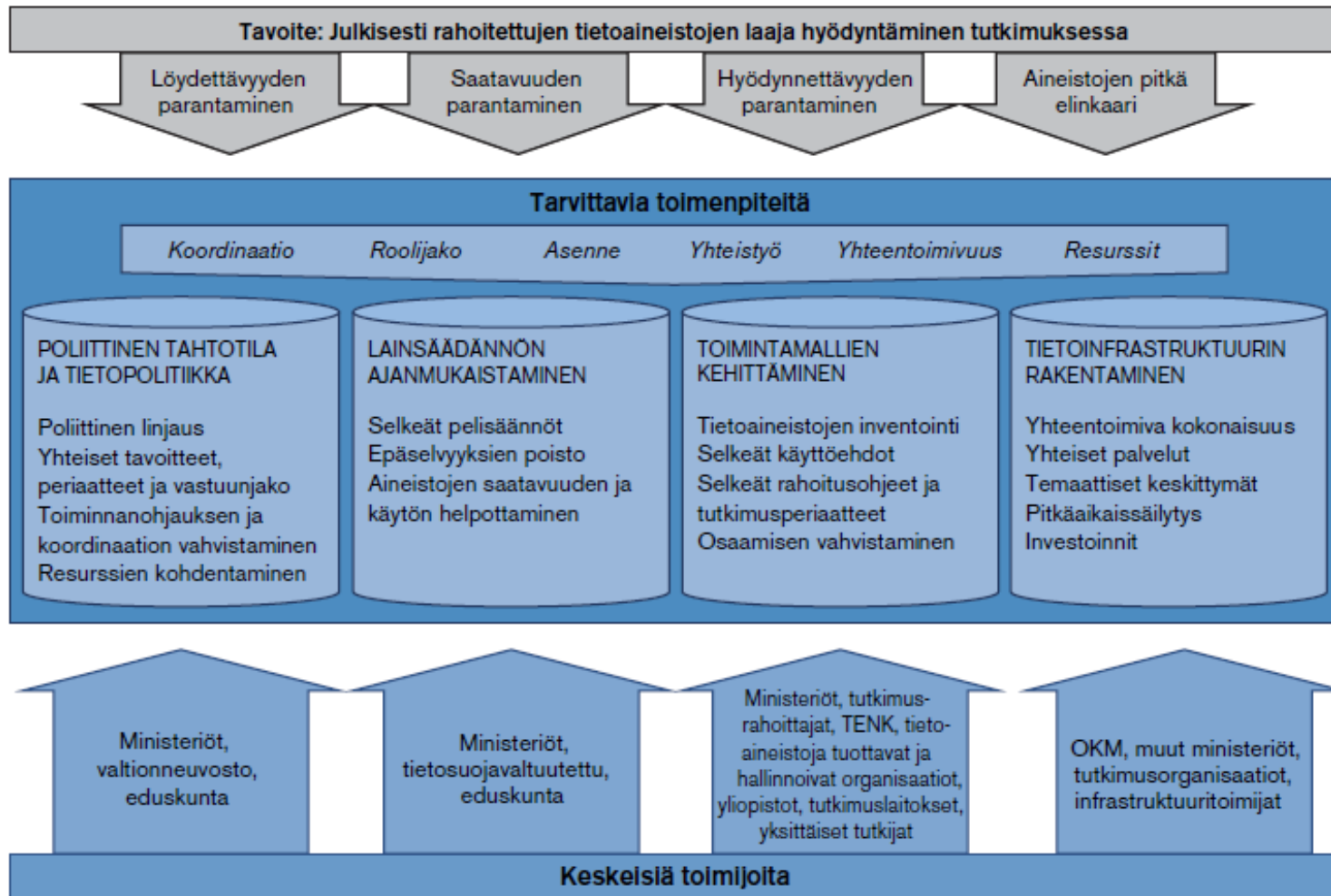
- *tutkimuksen tietoineistojen prosessikartta*
- *tietoinfrastruktuurin palvelukokonaisuuden suunnittelu*
- *metatietomalli tutkimustiedon hallinnalle*
- *yhteinen metatietokatalogi/hakupalvelu tutkimusaineistoille*
- *tallennuspalvelu tutkimuksen tietoineistoille*
- *yhteisen pitkäaikaissäilytysratkaisun valmistelu*

TTA-palvelukokonaisuus edistää tutkimuksen tietoineistojen kansallista, eurooppalaista ja kansainvälistä **yhteentoimivuutta.**

Tiedon pyramidi



Toimenpiteitä tietoaineistojen hyödyntämisen parantamiseksi



Hankkeen hyödyt



- Keskitettyä tietoa tutkimuksen tietoaaineistoista: helpompi löytää, helpompi käyttää
- Yhtenäisempiä käytäntöjä aineistojen hallintaan
- Yhteentoimivuuden lisääntyminen: metatietomalli, rajapinnat
- Monipuolisen palvelukokonaisuuden
- Tietoaaineistojen säilymisen turvaaminen



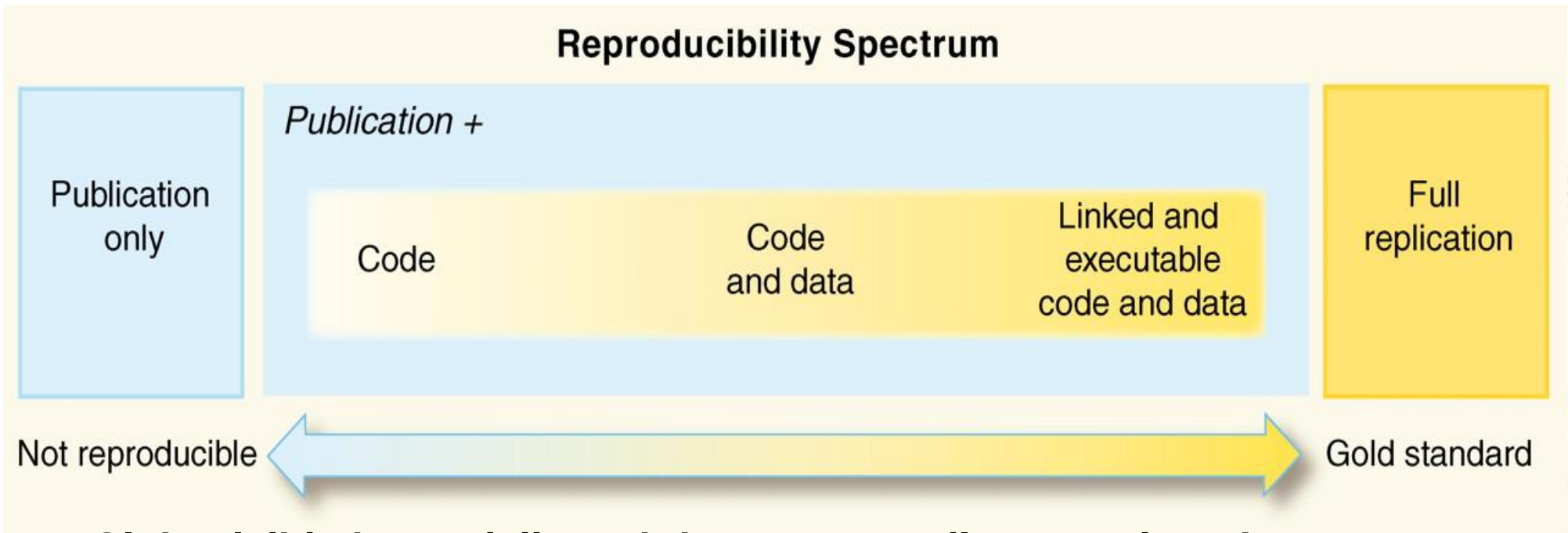
Tier 1 – International data services

Tier 2 – National data services

Tier 3 – Institutions (Universities & Institutes)

Tier 4 – “Small science” researchers & research groups

The spectrum of reproducibility.



- **Globaaleihin haasteisiin mahdoton vastata ilman avointa dataa**
- **Datavirtojen sulkeminen estää myös virtojen sulkijoiden etenemisen**

R D Peng Science 2011;334:1226-1227



Koordinaatioryhmä

Ohjausryhmä

Metatietoryhmä

Tietoinfrastruktuuri-
ryhmä

TURE-
pienryhmä

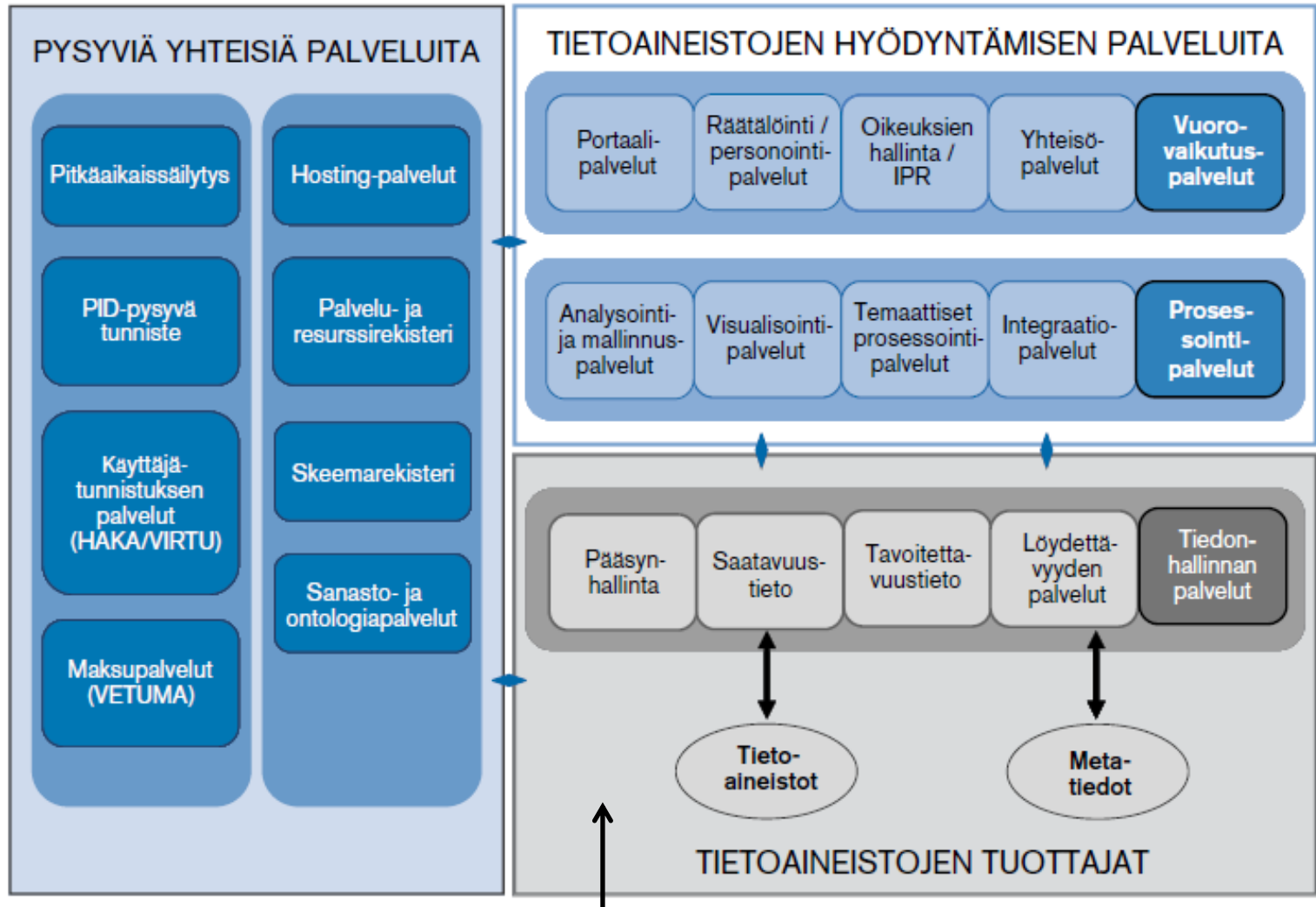
PAS-selvitys

TUTKIMUS-IDA

DATAKATALOGI

DATANHALLINNAN
PILOTTI

TUTKIMUKSEN TIETOINFRASTRUKTUURI



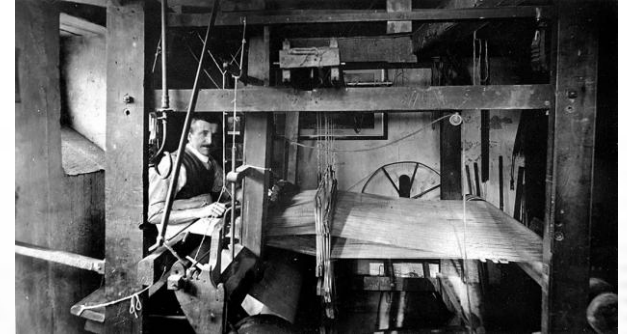
TTA-palvelut:
Tutkimus-IDA, Datakatalogi,
Datanhallinnan pilotti

Tutkimustiedon infrastruktuurin

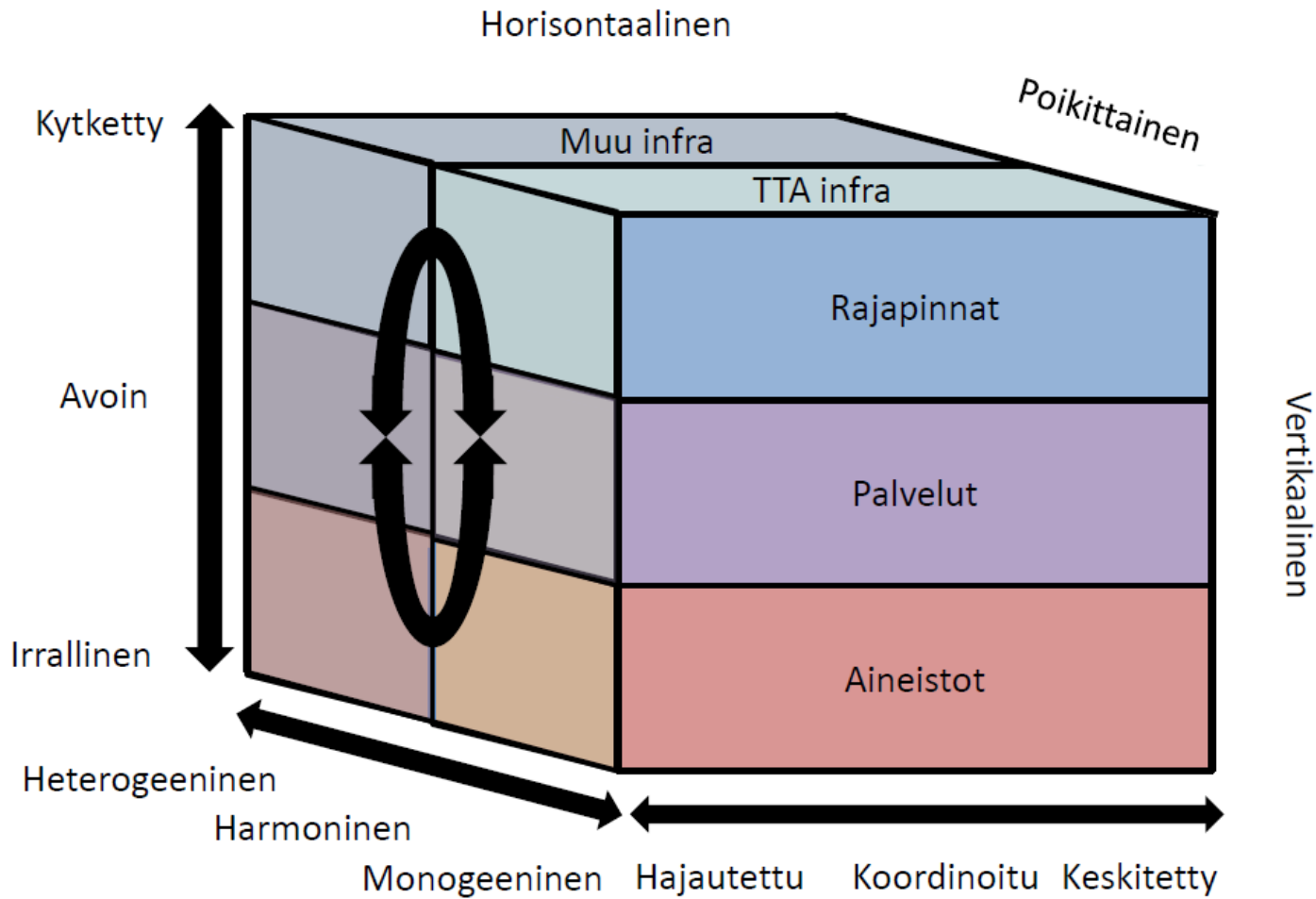


täytyy tukea:

- koko tutkimusprosessia
- tietoaaineistojen siirtoa tieteenalojen välillä
- avoimen datan tuottoa
- työvuopohjaista toimintaa
- julkaisujen ja tietoaaineistojen linkitystä
- tiedepolitiikan kehikkoa



Tutkimustiedon infrastruktuuri



Tutkimuksen tietoaineiston elinkaari



PITKÄAIKAISSÄILYTYS

“Leave a trail of evidence for scientists of the future”



Tutkimusaineistojen pitkäaikaissäilytyks



- Tutkimustiedon **kumuloituminen** on tutkimuksen ja tutkimusorganisaatioiden toiminnan ydin
 - Nykyisin ei ole käytössä hallittua ja toimivaa tapaa huolehtia digitaalisesta tiedosta
 - Pitkäaikaissäilytyks on tapa huolehtia tiedon säilymisestä pitkällä aikavälillä
- Digitaalisten aineistojen pitkäaikaissäilytyks tarkoittaa digitaalisen informaation **luotettavaa säilyttämistä kymmenien tai jopa satojen vuosien ajan**
 - Laitteet, ohjelmistot ja tiedostomuodot vanhenevat, mutta tästä huolimatta informaation täytyy säilyä ymmärrettävänä.
- Tutkimusaineistojen pitkäaikaissäilytyksessä tehdään tiivistä **yhteistyötä** Kansallisen Digitaalisen Kirjaston (KDK) kanssa

Kolme aineistoluokkaa

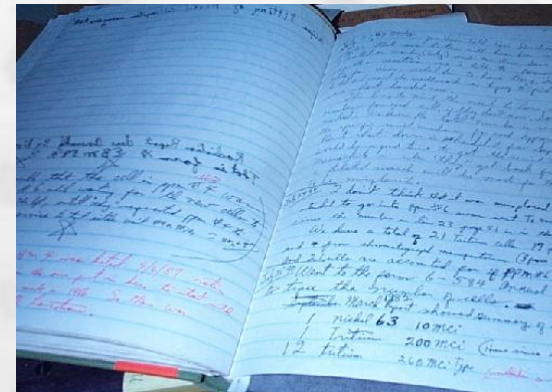
1. Tärkeät säilytettävät aineistot

- Välttämättömiä organisaation toiminnan jatkuvuudelle
- Liittyy korkea uudelleen käytön potentiaali
- Ei liity merkittäviä käyttörajoituksia

2. Mahdollisesti säilytettävät aineistot

- organisaation tehtävään keskeisesti liittyvät aineistot
- Liittyy uudelleen käytön potentiaalia
- Sallitaan jonkin verran käyttörajoituksia

3. Muut organisaation aineistot



Ehdotus keskeisiksi aineistotuottajien sidosryhmiksi

- Korkeakoulujen vararehtorit tai tietohallintojohtajat
- FIRIn rahoittamat hankkeet
- Suomen Akatemian huippuyksiköt
- Biotieteiden ja biolääketieteen teema
- Ilmakehätieteiden teema
- Yhteiskuntatieteiden teema
- Fysiikan, materiaalitieteen ja energian teema

Kyvykkyudet datahallintaan

14 kategoriaa

Dataset and workflow discovery	Data access services	Tools to Probe, Validate, Verify, and Visualize Data	Data security and trust
Workflow execution management	Data management within workflows	Metadata for workflow and data sets	
Numerical methods and software engineering	Modeling standards and frameworks	Modeling capabilities within Cloud, Grid, HPC, and Science Portals	
Policy enforcement processes	Best practices & Governance Models for definitions & standards	Broad Participation: NGO, industry, domain partners, international	Continuity, Sustainability, & Evolution

“But I am a scientist, NOT a records manager!”



Rooleja tietoaineistokentässä

- ‘Tutkija’: Datan ja informaation tuottaminen ja käyttö
- ‘Palvelukeskus’: säilyttäminen ja pääsynhallinta
- ‘Käyttäjä’: uudelleenkäyttö, informaation lisääminen
- ‘Rahoittaja’: Tutkimuksellinen/Yhteiskunnallinen hyöty
- ‘Omistaja’: sisällön ymmärrettävyyden vastuu

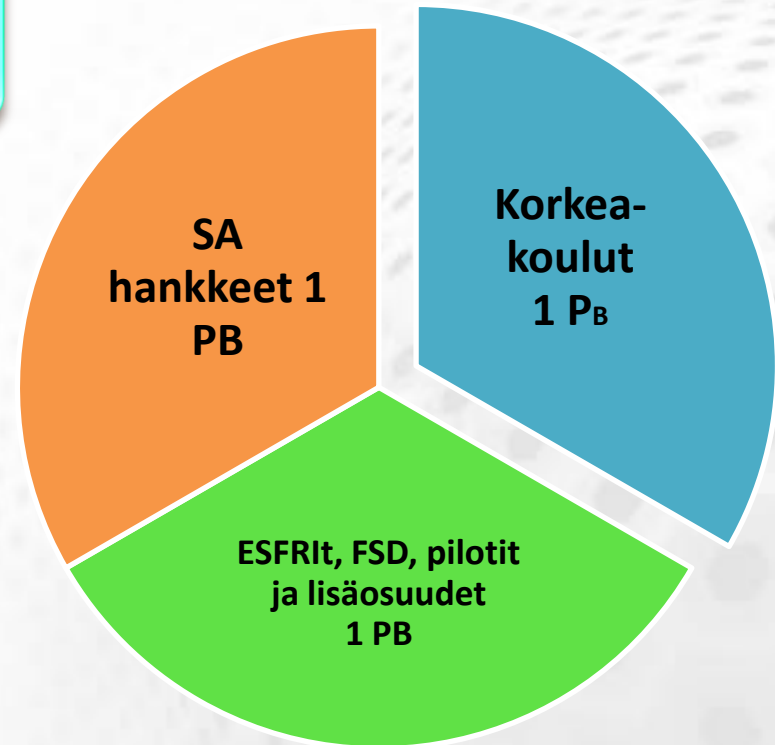
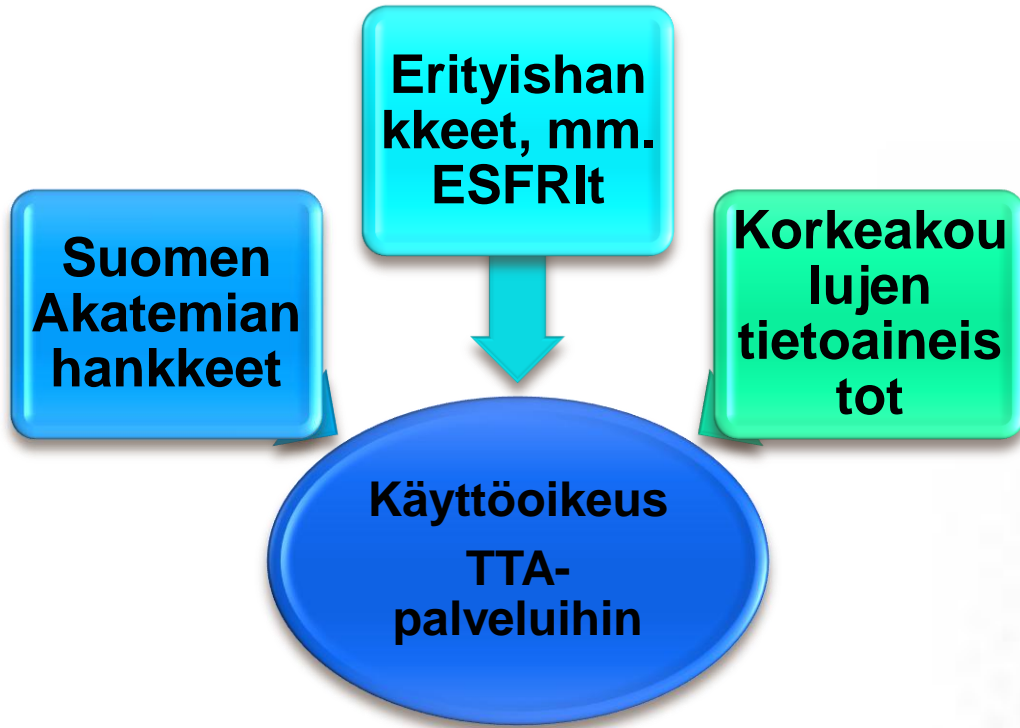
TTA –palvelujen tilanne



- Tallennuspalvelu IDA avataan käyttöpolitiikan mukaiseen käyttöön 20.9.
- Vaiheistettu käyttöönotto:
 - Korkeakoulujen kiintiöiden käyttöönotto
 - Suomen ESFRi-hankkeet: 5 TB/hanke
 - Suomen Akatemian hankkeet ja huippuyksiköt: alustava 2 TB/hanke:
 - Ehdotetaan että edetään kyvykkyyks- ja tarvelähtöisesti, mutta tasapuolisesti
 - Joulukuun OR/koordinaatioryhmien kokouksissa tarkistetaan alustavat kiintiöt



TTA käyttöpolitiikka



EDELLYTYS:

- ➊ Hankkeet sitoutuvat kuvailemaan tietoaaineistonsa sovittavien periaatteiden mukaan.
 - ➋ Hankkeet antavat tietoaaineistojen metatiedot avoimeen metatietokatalogiin
-
- ➌ **SA:n hankkeet** : 2 TB tallennustila/hanke.
 - ➍ **ESFRI-hankkeet**: 5 TB tallennustila/hanke
 - ➎ **Korkeakoulut**: Jaetaan osasta kapasiteettia yliopistokohtaiset kiintiöt, joista yliopistot voivat jakaa tallennustilaa.
 - ➏ **Tutkimuslaitokset**: Edetään aineistopilottien kautta. Tutkimuslaitosten käytölle pyritään aktiivisesti löytämään rahoitus ja politiikka, viimeistään vuoden 2013 aikana.
 - ➐ **SHOK-yhteistyö ja säätiöiden rahoittamat tutkimusaineistot**: Pyritään löytämään vuoden 2013 aikana rahoitus ja politiikka.

Käyttöoikeudet: Suomen Akatemian kautta



- IDA-palvelussa tarjotaan Suomen Akatemian hankkeiden tarpeisiin vuoden 2013 loppuun asti vähintään noin 1 petatavun (PT) tallennuskapasiteetti.
- Käyttöoikeus on maksuton Suomen Akatemian rahoituksen hankkeille (akatemiahankeet, huippuyksiköt, tutkimusohjelmat, tutkimusinfrastruktuurit).
- K.o. hankkeille tarjotaan 2 teratavun (TT) tallennusoikeutta hanketta kohden .

Käyttöoikeudet: korkeakoulut

Yliopisto	Osuus	Teratavuina
Aalto-yliopisto	0,12	120
Helsingin yliopisto	0,27	270
Itä-Suomen yliopisto	0,08	80
Jyväskylän yliopisto	0,07	70
Kuvataideakatemia	0	
Lapin yliopisto	0,01	10
Lappeenrannan teknillinen yliopisto	0,03	30
Oulun yliopisto	0,09	90
Sibelius-Akatemia	0	
Svenska handelshögskolan	0,01	10
Tampereen teknillinen yliopisto	0,06	60
Tampereen yliopisto	0,06	60
Teatterikorkeakoulu	0	
Turun yliopisto	0,14	140
Vaasan yliopisto	0,01	10
Åbo Akademi	0,04	40

Ammattikorkeakouluille varataan 10 TT kiintiö ammattikorkeakoulujen yhteisiin tarpeisiin.

Käyttöoikeudet: Lisäosuus

- ESFRI-hankkeiden sekä Suomen Akatemian rahoituksen hankkeiden keskuudesta ehdotetaan valikoitavaksi tarpeen ja kyvykkyyden mukaan hankkeet, joille voidaan myöntää tasakiintiötä suurempi kiintiö.
- Valikointiin nimitetään vastuutaho.
- Tähän varataan n. 1 PT kapasiteetista

Kiitos!



”Päämäärä ja keinot ovat
synonyymeja”

Mahatma Gandhi