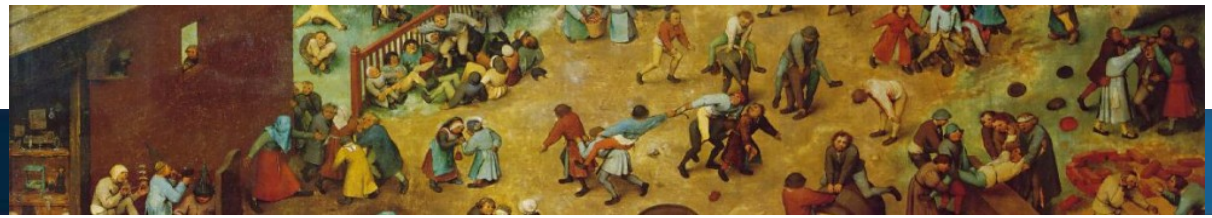
TTA
TUTKIMUKSEN TIETOAINEISTOT

TTA – National Research Data Initiative,

Pirjo-Leena Forsström, CSC

Digitalization of research processes

- Typical:
  - Growing volume of data and sources
  - Complexity of data processing
  - data is dynamic
  - High demand of data
- Most important challenges:
  - Managing,processing and combining exponentially growing datasets
  - Significant acceleration in analysis cycle
  - Persistence



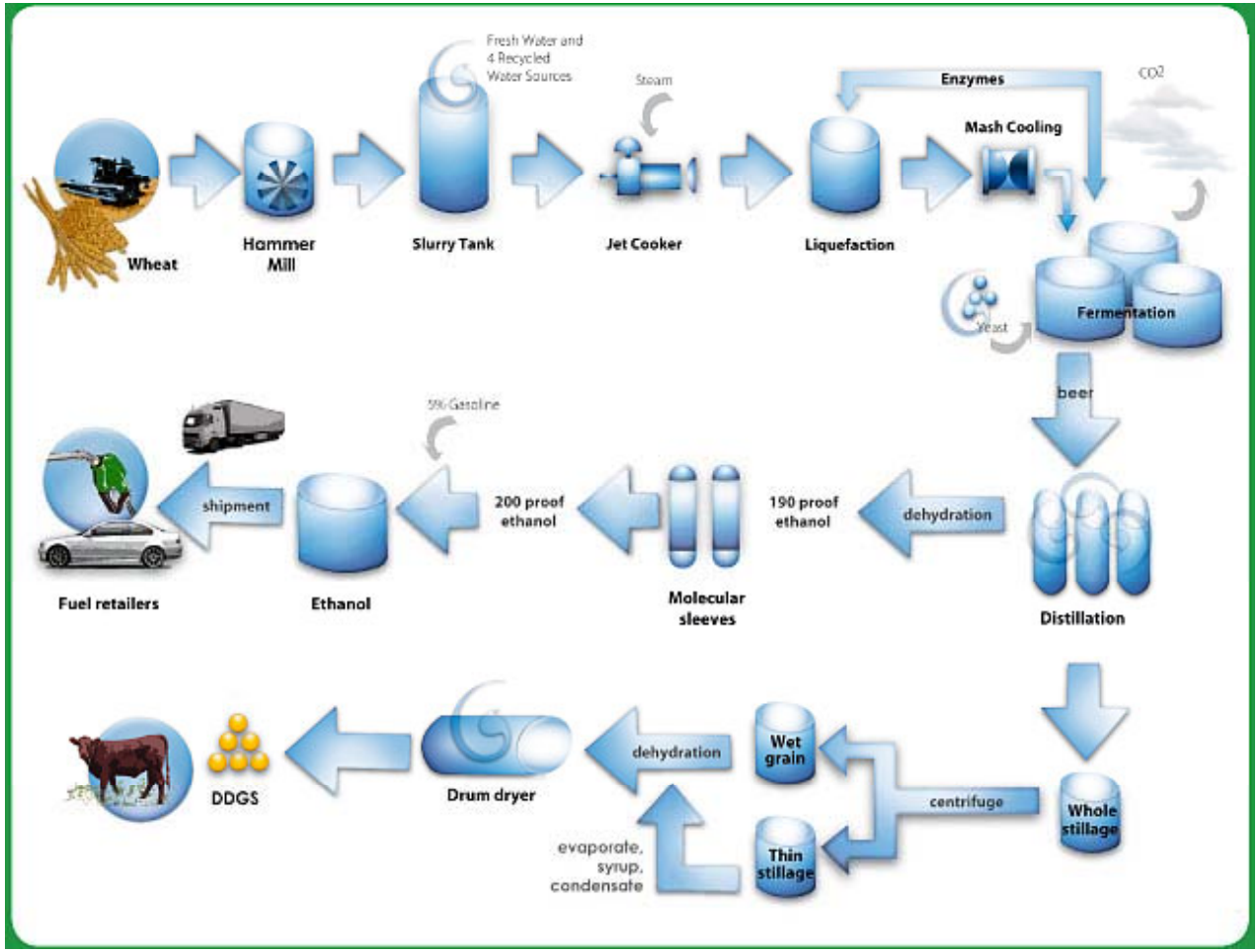
Source: wikipedia PD Image resources

Science as a process

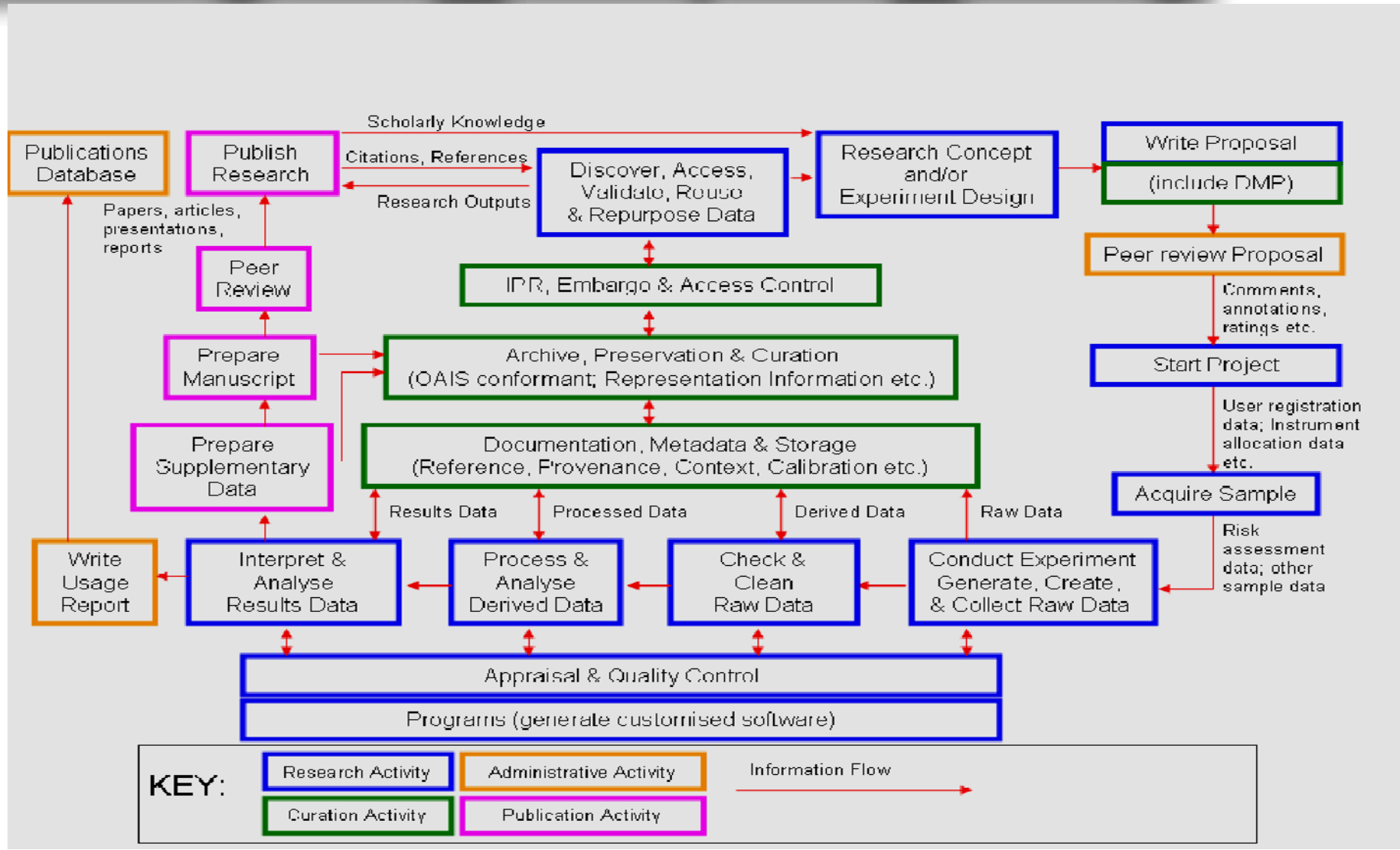
What is a Production System?

Production System are those activities of an organization where

- resources flowing within a defined system
- are combined and transformed in a controlled manner
- to add value
- in accordance to the policies communicated by management



Research process



Source: JISC report on research data, 2009

Digital processes break easily

	Lifecycle
Short-period funding	1-3 years
Software lifecycle: code, formats, interfaces, …	3-10 years
Dependent on expert knowledge	10-40 years
Thin documentation and metadata	1-100 years

=> **Weak persistence**

Source: : wikipedia PD Image resources

Persistence

Holy grail of preservation & information management more generally

PID ISNI OID URN URL

Orchid Onthology Vocabulary

- What does persistence mean?
- How long it persists?
- What persists?
- What is “guaranteed” to be accessible?

Source: : wikipedia PD Image resources

Availability and findability

- Principles of availability: free, licensed, machine readable
- Findability: metadata, linked data, IDs

Future Aim:

Research and cultural data routinely deposited in well-documented form, regularly and easily consulted and analyzed, and openly accesible while suitably protected and reliably preserved.

Needs PERSISTENCE

- Coherent organizational framework?
  - Ownership
  - Curation
- Flexible technical architecture:
  - Standard open protocols and interfaces
  - Flexible user access, analysis and visualization of data
  - Address issues of autenthication, authorization, security
  - Supports workflows

Excellency and knowledge

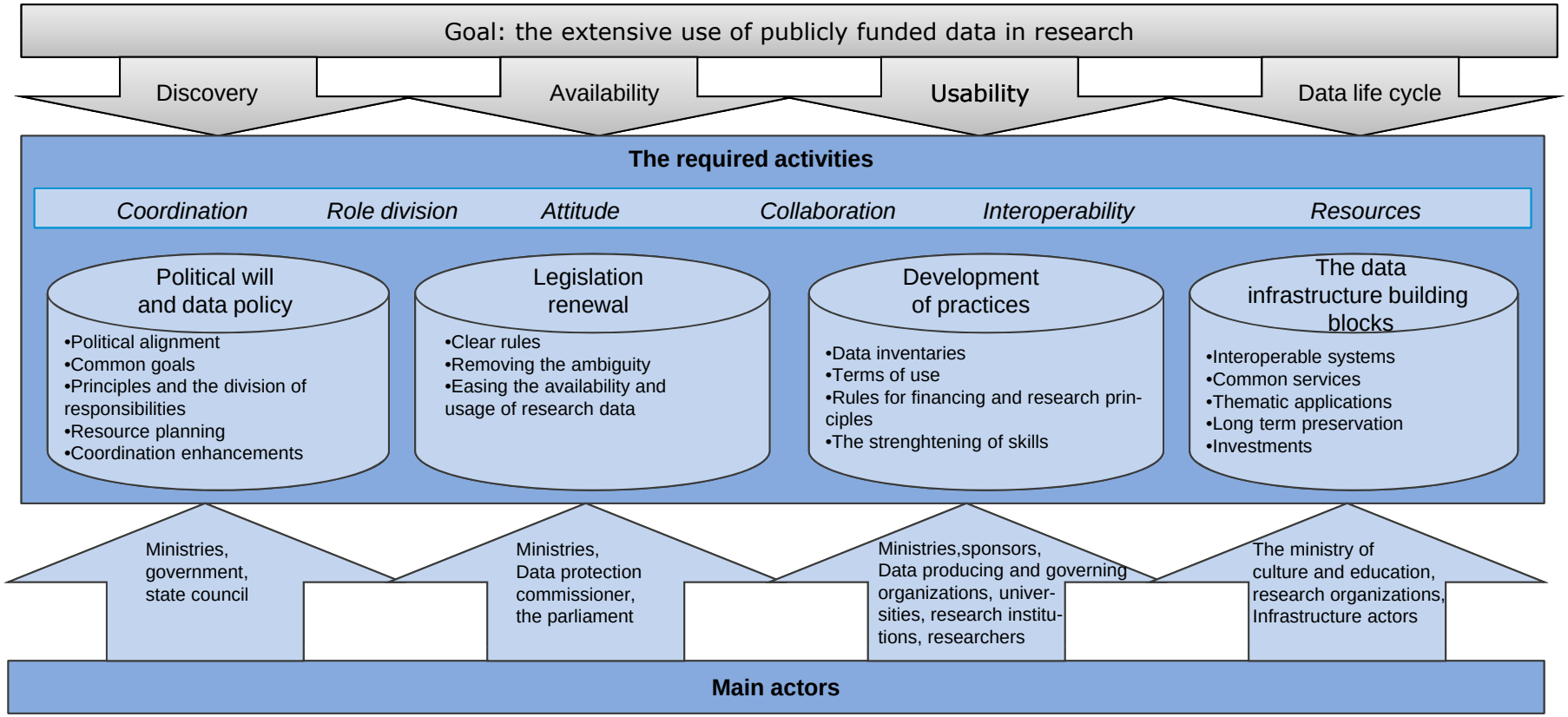
- Excellent people will in turn attract commercial investment.
- UK: “With over one million new science, engineering and technology professionals and technicians required by 2020, the supply of high quality STEM skills in the UK will be even more important than it is today.The UK’s ability to nurture domestic talent and attract the best international researchers will be an important component of the response to this shortfall.” (1)
- FiDiPro; postdocs

1) Fuelling prosperity:Research and innovation as drivers of UK growth and competitiveness
Monday 22 April 2013

TTA phases

- Phase I (2009-2011): TUTA, results in guide for policy-makers and generic roadmap
- Phase II (2012-2013): TTA, entreprice architecture for information infrastructure, metadata roadmap, basic services, gap analysis
- Phase III (2014-2017): building interoperability, supporting open science, establishing collaborative processes

Tasks for enhancing the usage of research datasets



The objectives of the TTA infrastructure

Science is becoming ever more open and data-driven. Large integrated datasets can provide a deeper understanding of nature and society.

1. Developing a Finnish *sustainable information infrastructure* for research and cultural data;
2. Providing selected services to this infrastructure cost-effectively and sustainably
3. Enabling and encourage *sharing and re-use* of scientific and cultural data;
4. Ensuring *preservation of digital data*,
5. Providing tools for data management, both on organizational, discipline and user level; and
6. Contributing to *unification of interfaces and metadata.*
7. Strengthen the *capasities* of research institutes and infrastructures, thus enhancing the *competiveness* of Finnish research.

Networked Readiness Index 2013: Top 10

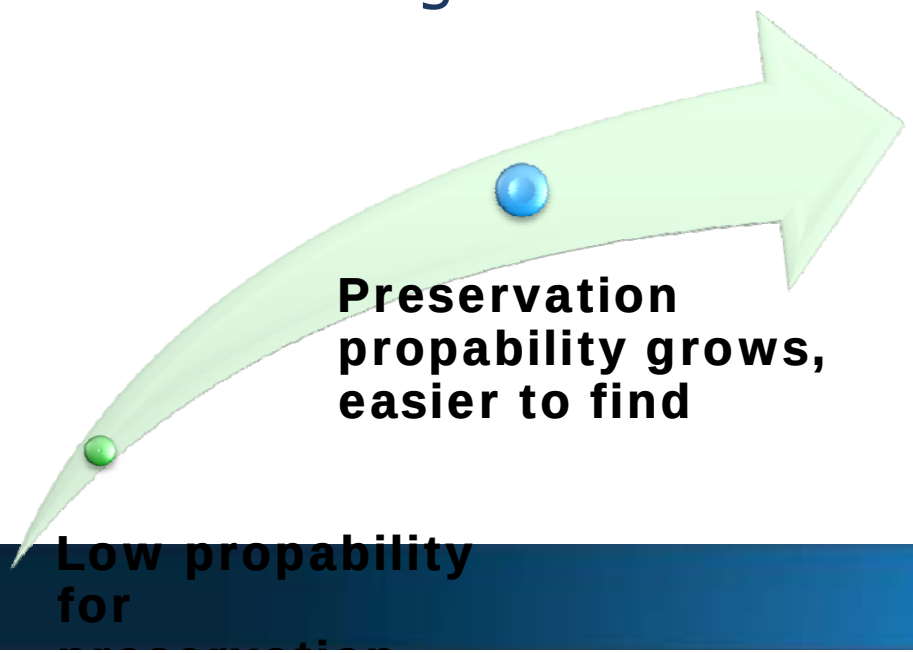
Economy	2013	2012	
Finland	1	3	↑
Singapore	2	2	→
Sweden	3	1	↓
Netherlands	4	6	↑
Norway	5	7	↑
Switzerland	6	5	↓
United Kingdom	7	10	↑
Denmark	8	4	↓
United States	9	8	↓
Taiwan (China)	10	11	↑

The added value of TTA to Finland

- Through enhanced
  - findability and availability of data
  - quality of data
  - data services including analysis and processing services
  - platform for interdisciplinary collaborative research
- Getting wider and more integrated research community
  - researchers in research institutes
  - scientists and students of different disciplines working in academia
  - IT and data service professionals
- To achieve
  - better use of Finland's unique and open data resources in science and education
  - scientific excellency and breakthroughs in many different research fronts
  - societal and environmental progress through evidence-based decision making
  - technological innovations and products

Benefits

- All national research data from one point: easy to find, easy to use
- Common practises for data management
- Interoperability: metadata, interfaces
- Versatile service collection
- Preservation of relevant information
- Tools to global connections



Tier 1 – International data services

Tier 2 – National data services

Tier 3 – Institutions (Universities & Institutes)

Tier 4 – "Small science" researchers & research groups

The international relevance of TTA

- makes Finland to be an attractive site and active actor in international research community
- draws international experts to work in Finland using Finnish data
- provides models of infrastructure functionalities that have not been utilised elsewhere
- contributes to the networking of data infrastructures in European and broader international levels (EUDAT etc.)

- With improvements in
  - scientific progress and science-based education
  - the culture of open science (access, reproducibility, cross-checking)
  - the quality of information used in decision-making
  - the diversity and quality of using scientific information in overall societal functioning and everyday life
  - the conditions that boost commercial innovations and business

Current blueprint for information infrastructure

TTA
TUTKIMUKSEN TIETOAINEISTOT

I Make it / I Find it

- Store, Study, Discuss
- Find Existing Data

I Make it Ready

- Describe, Package, Authorize

I Make it Available

- Manage, Market, Preserve

Guidelines & Support

IDA Storage Service

Analyzing & Visualization

Collaboration Tools

Finding Aids

AVAA Publishing Service

Metadata Support Services (Identifiers, Ontologies…)

KATA Data Catalogue

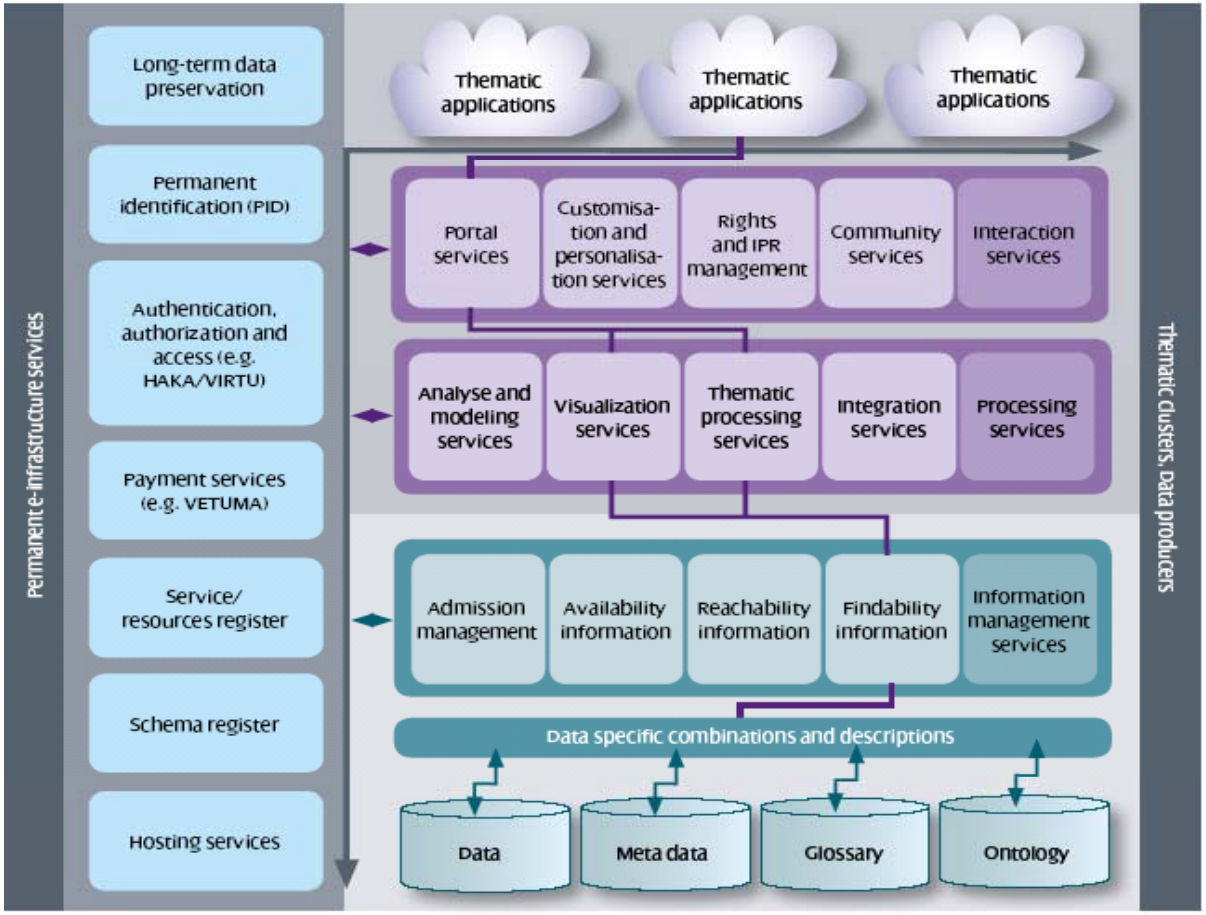
REMS Entitlements Mgmt

PAS Preservation

Startegy | Governance | Development Path

Dimensions of information infrastructure

- Embeddedness
- Transparency
- Reach of scope
- Links with conventions of practice.
- Embodiment of standards
- Build on an open platform: Infrastructure does not grow de novo; it wrestles with the "inertia of the installed base" and inherits strengths and limitations from

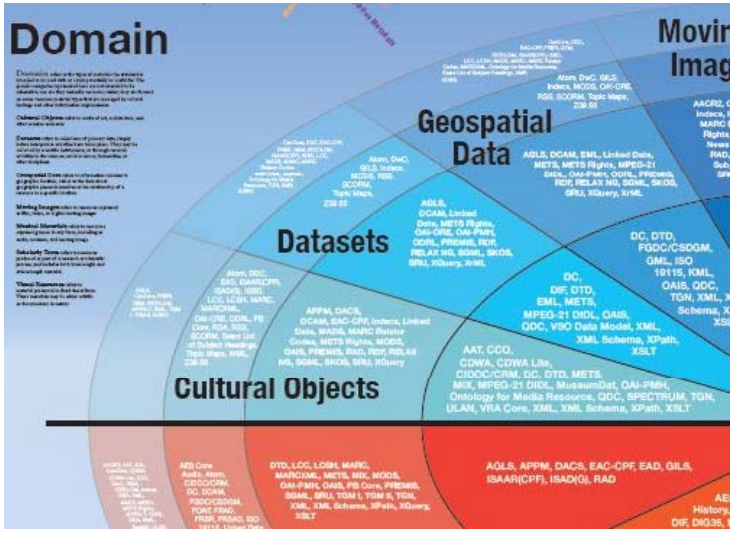


Open Architecture for Research Data

- A key conceptual approach
- Defined components
- Standard interfaces, protocols & objects
- Community responsibility (in the long-term) for development & maintenance of standards
- Bottom-up, grass-roots evolution as much as practicable
- Resulting in many capabilities that can all work together

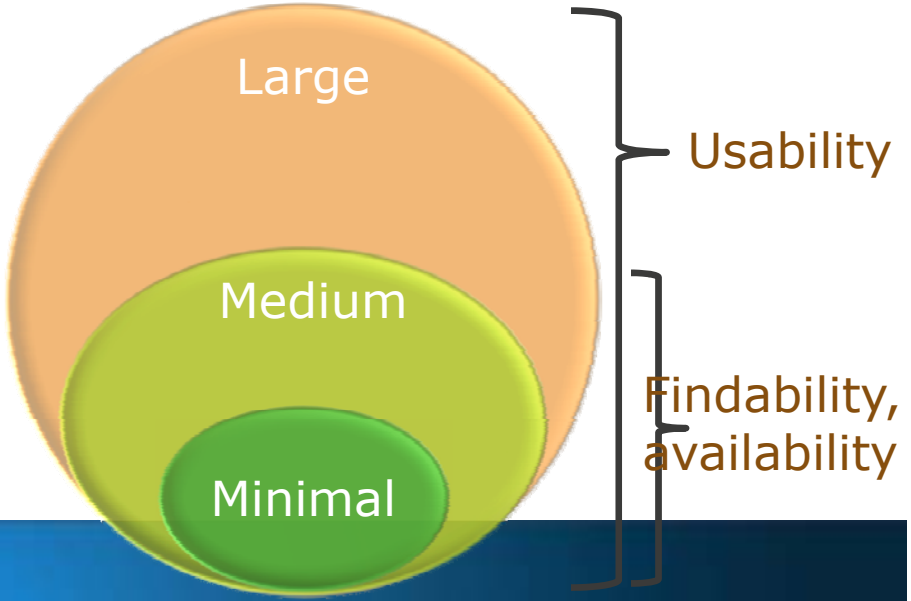
Metadata

J. Riley: Seeing Standards (2010)
http://www.dlib.indiana.edu/~jenlrile/metadatamap/



- No single metadata standard fits all!
- TTA metadata principle: reuse of metadata; utilize existing metadata formats and services
- International and national co-operation is important

- TTA metadata challenges: to improve data findability, availability and usability
- Three metadata tiers: minimal, medium, large
- TTA metadata model defines metadata that must be attached to each dataset and the structure of that metadata



TTA: permanent infrastructure and a network of networks

- heterogenous, connects many scientific disciplines
- All participating organisations are highly committed
- Integration into permanent national infrastructures
- exceptionally broad and diverse research infrastructure initiative
- TTA is already a functional infrastructure that involves
  - National Digital Library
  - open data and open science initiatives in public sector and science
  - Ministry of Education and Culture, other ministries
  - Universities
  - Research Infrastructures
  - Research institutes

Open scientific methodologies
Digital preservation
Metadata management and services
Incentives and policies
Culture and attitude
Planning and management
Common and distributed services

TTA-services 2013

- IDA – data storage (in use)
- KATA –data catalog (piloting)
- AVAA -open data platform (in use)
- PAS – Long Term Preservation (2015), pilots starting 2014

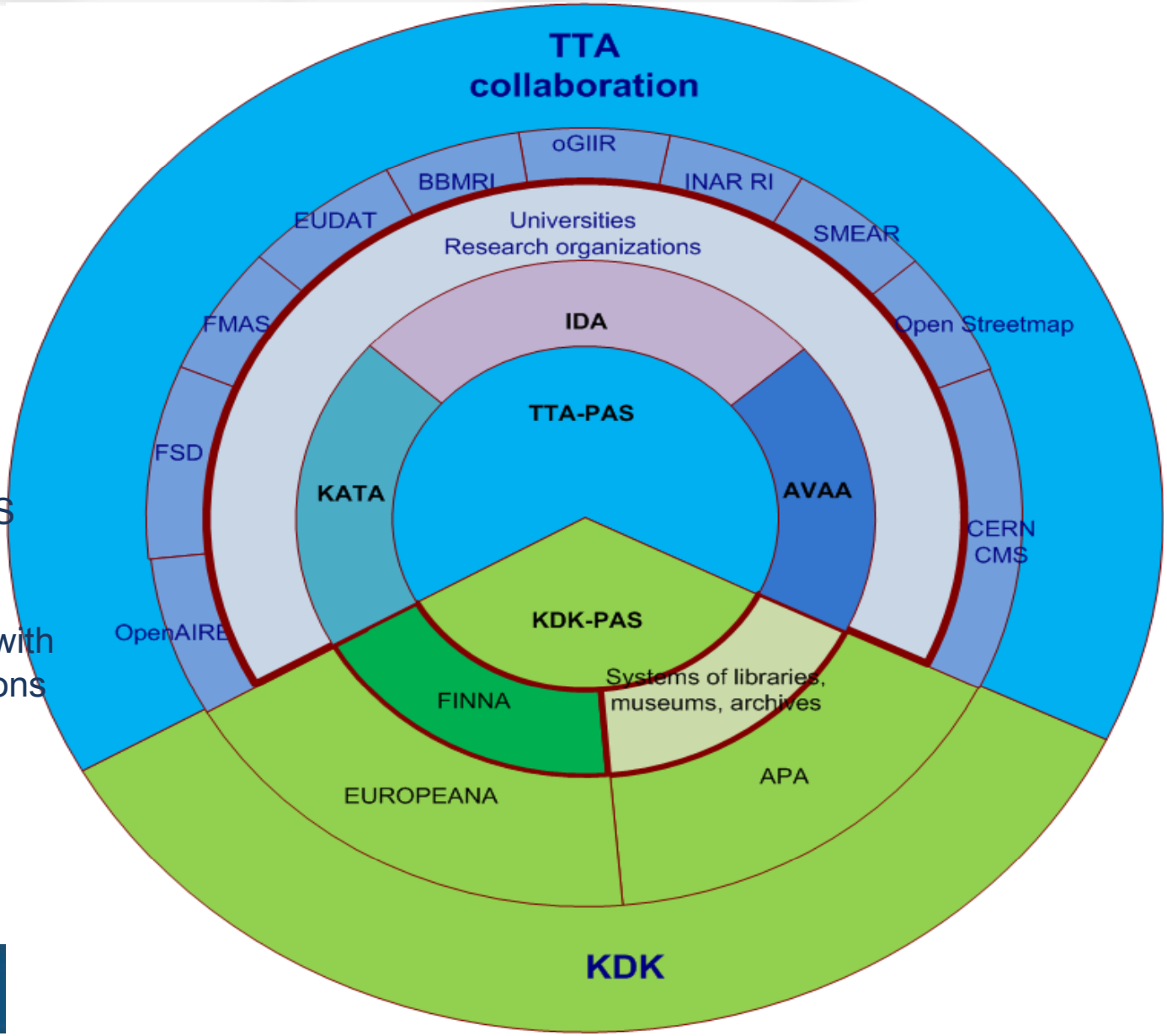
FIRI: TTA/KDK collaboration

- TTA and KDK services form backbone of the Finnish information infrastructure

- The data publication platform AVAA is connected to national public sector initiative Open Knowledge, coordinated by the Ministry of Finance; CERN CMS collaboration

-EU-project level collaboration with several projects and organizations

-Especially: EUDAT



Components of TTA

- *Collaboration*

- *Services*
  - efficient data and metadata management
  - research data repository
  - user oriented development and interface of all services

- *Human resources*
  - networked researchers and infrastructure constructors
  - professionals needed in infrastructure development, training, etc.

- *Hardware and software*
  - shared and networked facilities
  - Open source software

Integrated Data Archive (IDA) storage service

- Joint storage service for preserving digital research data and increasing its re-use
- Safe preservation of data and metadata
- Data preservation in intact and unchanged format by means of managing copies and their integrity

The KATA metadata catalogue

- makes it easier to find data that is available for research
- produces information about the existence of data for funders
- enables the creation of a joint terms of use and rights of use culture (ownership and utilisation rights information in the data catalogue)
- enables generation of merit for the researcher
- helps to identify and find data for long-term preservation
- requires the projects to commit to the agreed principles (description, openness)

Tervetuloa - Kata metadata catalogue - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Tervetuloa - Kata metadata cat...

https://kata.csc.fi/fi/

Most Visited Red Hat Customer Portal Documentation Red Hat Network

KATA

Viestit 0 | Kirjaudu ulos | Maija Meikäläinen | Omat tietoaineistot

Etsi

Ohjeet FAQ Stats

TIETOAINEISTOT KOKOELMAT TIETOA

suomi

Mikä Kata on?

Kata-metadatakatalogissa on metadataa tutkimusdatasta. Tässä palvelussa voit etsiä ja löytää tutkimusdataa käyttöösi, ja myös esitellä oman datasi muiden jatkohyödynnettäväksi.

Mitä on metadata?

Lyhyesti määriteltynä metadata on dataa datasta. Metadata kuvailee datan ominaisuuksia ja kuvailut saattavat koskea esimerkiksi tekijöitä, aihetta, formaattia ja lisenssiä. Metadatan sisalto riippuu suuresti datasta, mutta tama palvelu tarjoaa yhtenäisen rakenteen tutkimusaineistojen metadatalle: TTA:n metatietotyöryhmän julkaiseman minimimetatietomallin mukaiset kentät.

Onko tämä palvelu minulle?

Kata-palvelu pyrkii keräämään dataa useista paikoista ja useilta tieteenaloilta, jotta yhdestä paikasta pääsisi kiinni eri tutkimusalojen ja tutkimuslaitosten datoihin. Etsiäksesi dataa sinun ei tarvitse olla ammattitiltasi tiedemies tai tutkija - hyödyllistä dataa voi tuottaa ja käyttää kuka tahansa ja missä tahansa.

Onko kaikki data avointa?

Lähes kaikki metadata on avointa, mutta datan omistaja päättää miten ja kuka dataan pääsee käsiksi. Datan julkaisemista avoimena kuitenkin suositellaan.

Työkalut ja lisätietoa

Minimimetatietomalli määrittelee aineistokuvailuille yhteisiä metadatakenttiä. Tutustu TTA:n metatietotyöryhmän tuottamaan minimimetatietomalliin:

Minimimetatietomalli (In Finnish)

TTA:n suosittelema lisenssi on Creative Commons 3.0 -lisenssiperhe. Lisätietoa lisensseistä:

Hae Dataa

esim. ilmasto

Uusimmat tietoaineistot

Nuorten politiikka 2009: esseevastaukset
Aineisto sisältää Nuorten politiikka 2009...

Koulun hyvinvointiprofiili 2012-2013: henkilökunta
Tässä aineistossa tarkastellaan henkilökunnan...

Koulun hyvinvointiprofiili 2012-2013: toisen asteen oppilaitokset
Tässä aineistossa tarkastellaan toisen asteen...

Koulun hyvinvointiprofiili 2012-2013: yläluokat 7-9
Tässä aineistossa tarkastellaan...

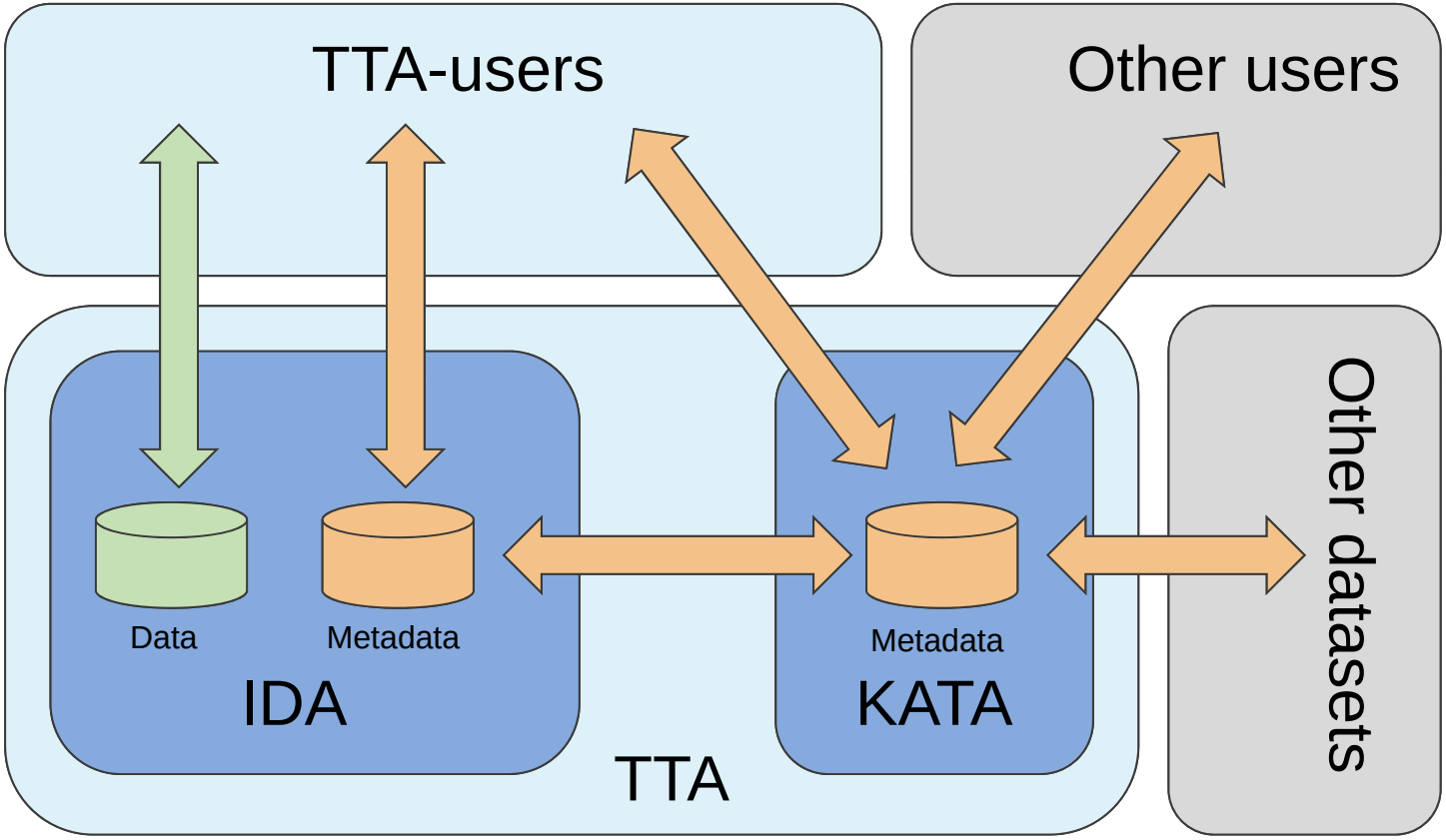
Koulun hyvinvointiprofiili 2012-2013: alaluokat 4-6
Tässä aineistossa tarkastellaan...

Kata on osa Tutkimuksen tietoaineistot (TTA) -palveluita

Palvelun tuottaa CSC - Tieteen Tietotekniikan Keskus Oy

TTA TUTKIMUKSEN TIETOAINEISTOT CSC

IDA-KATA interfaces



Luo tietoaineisto - Kata metadata catalogue - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Luo tietoaineisto - Kata metada...

https://kata.csc.fi/fi/dataset/new

Google

Most Visited Red Hat Customer Portal Documentation Red Hat Network

Viestit 0 | Kirjaudu ulos | Maija Meikäläinen | Omat tietoaineistot

KATA

Etsi

Ohjeet FAQ Stats

TIETOAINEISTOT KOKOELMAT TIETOA

suomi

/ Tietoaineistot / **Luo tietoaineisto**

Pakolliset tiedot

Perustiedot

Otsikot

Otsikko: Tutkimus

Kieli: Finnish

Laatijat ja organisaatiot

Laatija: Meikäläinen Maija

Organisaatio: Tutkimuslaitos

Laatija: Meikäläinen Matti

Organisaatio: Tutkimuslaitos

Avainsanat

Etsi avainsanoja (tageja) YSO-ontologiasta. Valitse valittu avainsana näppäimistösi enter-painikkeella tai hiiren klikkauksella.

tutkimus × kyselytutkimus ×

Etsi: f fi

fajanssi
fajanssiteollisuus
fakkikieli → ammattikieli
faksi → telekopiointi
faksi → telekopiolaitteet
faksimile → näktöispainokset
faktatietokannat
faktorianalyysi

Jakelijan / julkaisijan yhteystiedot

Nimi: Maija Meikäläinen

Puhelinnumero: +35844122123

Sähköposti: maija@yliopisto.fi

Kotisivu: http://esimerkki.fi/contact

Mitä ovat tietoaineistot?

Tietoaineistot kerätävät yhteen asiaan liittyvää dataa ja tämän metadataa yhteisen osoitteen alle.

Kenttä on pakollinen, jos valintaruutua ei ole valittu.

Anna pilkulla erotettu lista datassasi käytetyistä kielistä ISO 639-2 T muodossa, esimerkiksi "eng, fin, swe".

Etsi tietoaineistoa - Kata metadata catalogue - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Etsi tietoaineistoa - Kata metad...

https://kata.csc.fi/fi/dataset?ext_date-metadata_modified-start=&tags=poliittis

Google

Most Visited Red Hat Customer Portal Documentation Red Hat Network

Viestit 0 Kirjaudu ulos Maija Meikäläinen Omat tietoaineistot

KATA

Etsi

Ohjeet FAQ Stats

TIETOAINEISTOT KOKOELMAT TIETOA

suomi

/ Etsi tietoaineistoja

Lisää tietoaineisto

Etsi Tarkennettu haku

Kaikki kentät | Kysely | ADD

OR | Laatija | M* | DEL

AND | Organisaatio | Yhteiskuntatieteellinen* | DEL

Vuosi: Aloitusvuos - Lopetusvuo

Etsi

30 tietoaineistoa

Järjestä: Viimeksi muokattu

Avainsanat: poliittiset instituutiot

YLE Uutisten kysely presidentin valtaoikeuksista, tammikuu 2009

Kyselyssä tiedusteltiin mielipiteitä presidentin toimivallasta. Kysymykset käsittelivät mm. ulkopolitiikan johtamista, puolustusvoimien ylipäällikkönä toimimista, pääministerin...

Suomen Keskustan sisäinen päätöksenteko 2001

Tutkimus selvittää Suomen Keskustan jäsenten näkemyksiä puolueen sisäisestä päätöksenteosta ja eri toimijoiden rooleista siinä. Kysely on osoitettu niin puolueen rivijäsenille...

KuntaSuomi 2004: kuntalaiskysely 2004

Kyselyn tarkoitus oli selvittää kuntalaisten näkemyksiä asuinkunnastaan, sen palveluista sekä päättäjistä. Kyselyn avulla selvitettiin myös kuntalaisten osallistumis- ja...

KuntaSuomi 2004: perusterveydenhuolto 2003: kuntayhtymiin kuuluvat kunnat

Tämä tutkimus on osa KuntaSuomi 2004 tutkimusohjelman Sosiaali- ja terveyspalvelujen tutkimusta. Tutkimuskohteina ovat kuntien perusterveydenhuollon ohjaus ja johtaminen,...

KuntaSuomi 2004: perusterveydenhuolto 2000: kuntayhtymiin kuuluvat kunnat

Tämä tutkimus on osa KuntaSuomi 2004 tutkimusohjelman Sosiaali- ja terveyspalvelujen tutkimusta. Tutkimuskohteina ovat kuntien perusterveydenhuollon ohjaus ja johtaminen,...

Avainsanat Tyhjennä

- poliittiset institu... (30)
- järjestelmät ja org... (30)
- hallitus (30)
- hallinto (30)
- kunnallistiede (28)
- kunnat (25)
- kehittäminen (22)
- palvelutuotanto (18)
- sosiaalipolitiikka (16)
- sosiaalipalveluiden... (15)
- Näytä lisää

Tiedostomuodot Tyhjennä

Haulle ei ole suodattimia

Tieteenala Tyhjennä

- tilastotied (30)
- Näytä lisää

Laatija Tyhjennä

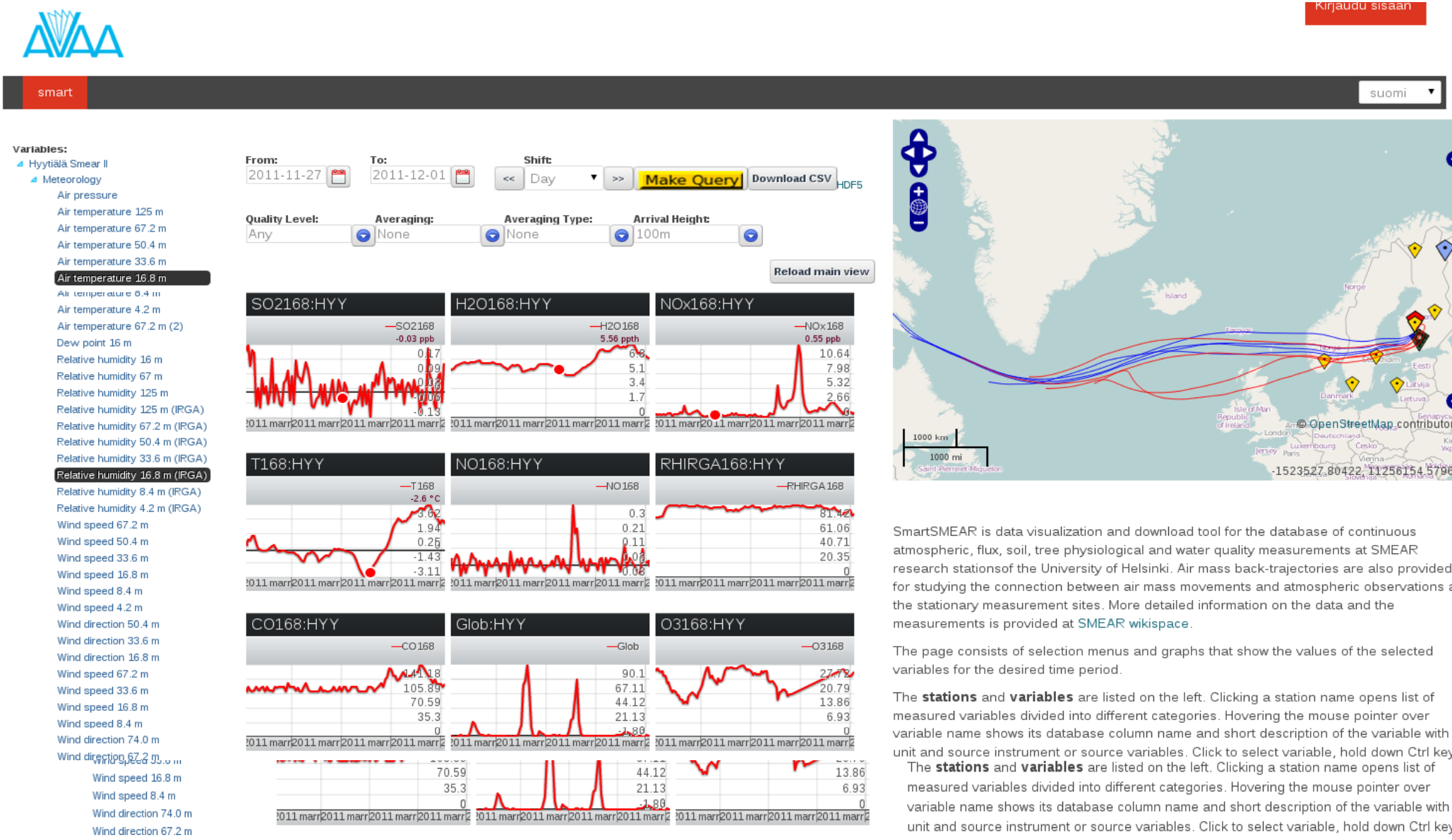
- Mäkelä, Pertti (6)
- Ikola-Norrbacka, Rinna (6)
- Vaasan yliopisto. H... (3)
- Mattila, Juha (3)
- Kauppi, Ulla (3)
- Heikkinen, Erja (3)
- Ikola, Rinna-Marika (2)
- Sjöblom, Stefan (1)
- Pekola-Sjöblom, Mar... (1)
- Latva, Anne (1)
- Näytä lisää

AVAA-project

- A platform of web-based tools for opening and publishing scientific data sets
- Building generic tools is challenging!
- Pilots => common tools and case examples
- Started with 3 pilots
- Target: to offer experience on formats and publication of different data sets

AVAA-project

- Educational application of Cern CMS **particle physics** data for schools (ready 2014 , see Cern link) ◦Includes publication of data in original format and simpler, open, machine-readable format (JSON)
- **Climate research** SMEAR datan visualization, JSON-interface and open download both as CSV and HDF5
- **GIS data interface**, as atest case OpenStreetMap WMS –service in ETRS-TM35FIN-coordinates
- Early 2014 WMS- and WFS-interfaces fo distribute **Kotimaisten kielten keskuksen murrekartta-aineisto**



AVAA-pilotti SmartSMEAR on Helsingin yliopiston SMEAR-tutkimusasemien (Station for Measuring Ecosystem-Atmosphere Relations) tietokannassa olevan mittaustiedon visualisointi- ja lataustyökalu.

AVAA project

CMS-pilot has been published:

- International Science Grid This Week (27.11.2013): LHC data to be made public via open-access initiative
- Nature - International Weekly Journal of Science (26.11.2013): LHC plans for open data future
- Cern web home site (15.11.2013): LHC data to be made public via open access initiative

Cern CSM –pilot 1

- A eaching application for high schools Cern CMS-experiment in particle physics
  - 2013 were the technical solutions for transferring data, performance, demand of capacity and suitable formats figured out: for example how much the applications are capable of reading data efficiently. Autumn 2013 a thesis-questionnaire for high school teachers at Helsinki University department of physics concerning the application. Planning and implementation of the application based on the questionnaire results.
    - → simple, fast executable exercises motivating studies of physics
    - → target: easily re-usable and generalised components are carried out for transferring and visualisation of data
  - In addition to original format published, data will be transformed into more simple, open and machine readable formats in order to be program usable (at the moment JSON)
  - The documentation of the complex understandable also for the not physicists (for example for the students)
    - --> Possibility for piloting Linked Data (?):
    - Linking the”general concepts” and additional information to metadatadescription of physicians for example existing ontology services

AVAA beta

- AVAA beta –pilotportal

  www.avaa.tdata.fi

- Questions? avaa@csc.fi

Data management guide
http://www.tdata.fi

- Data management and planning
- Data storage
- Data sharing and publication
- Finding and re-using data
- Data management and processing at CSC
- Research process and data

Lots of guidelines, examples, information and best practises

Open user forum for TTA users

- Support for all TTA-users
  - Connect the users, peer-to-peer
  - Get feedback
  - Share information
  - Collect development needs and ideas
  - Introducing services to new users
- Platform in connection with other TTA-services, opens early 2014
- User forum face-to-face -meetings 2014. First: IDA users

PAS: Long-term preservation of research data

- Accumulation of research information is foundation of research and research organisation activity.
- At this time, there is no controlled and functional way of handling digital information in the long term.
- Long-term preservation of digital data means the reliable preservation of digital information for several decades or even hundreds of years.
- Equipment, software, and file formats will become outdated, but despite this the information must be preserved in understandable form.
- Long-term preservation of research data involves close co-operation with the National Digital Library (NDL).

How to get access to the services?

- IDA – How to become an IDA user:
  - via university´s IDA contact person or
  - contact@csc.fi
    - projects funded by the Academy of Finland
    - polytechnics
- KATA – open for everybody
- AVAA - open for everybody
- PAS – will be defined later

METS-standardi päivittyy

- ineistojen paketoinnissa ja siirrossa hyödynnettävästä METS-standardista julkaistaan pian versio 1.10
- Uudessa versiossa huomioitu kaikki KDK:n esittämät muutospyynnöt
- Muutokset mahdollistavat pitkälti myös tulevien kansallisten tarpeiden lisäämisen
- Uudesta LoC:n METS-versiosta on jo olemassa skeemaluonnos
  - Julkaistaneen lopullisesti vielä tänä vuonna
-

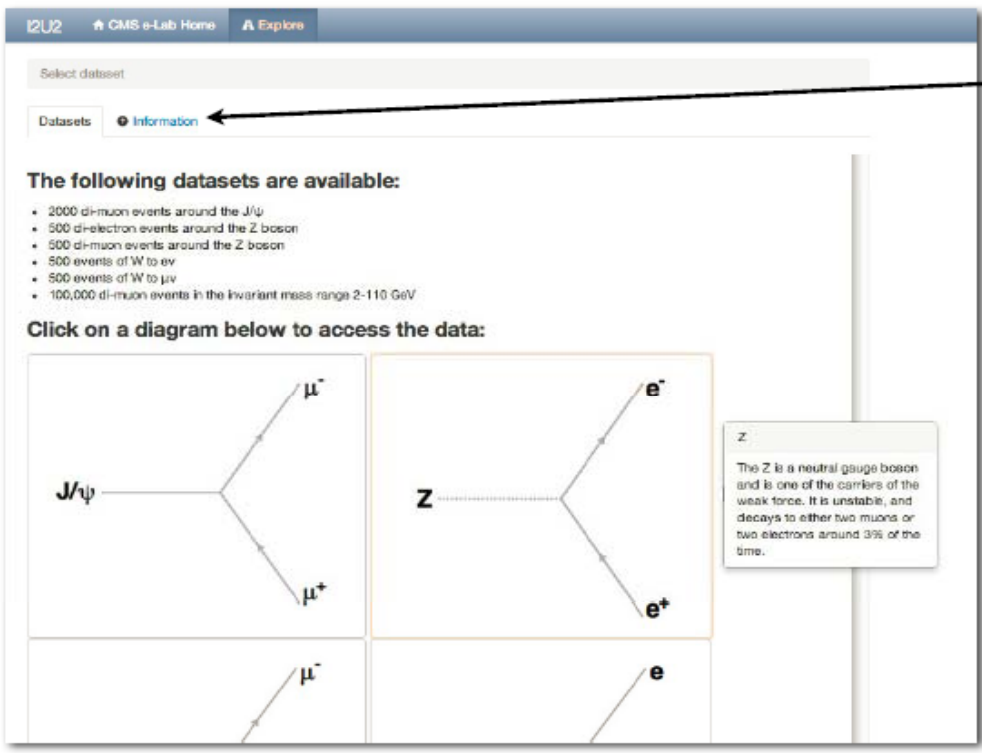
KDK:n METS-muutostoiveet

- Viime keväänä laadittiin 5 muutospyyntöä aineistojen paketoinnin pilottien tuloksina
- Nyt muutospyyntöjen perusteella METS:in seuraavasta versiosta (1.10) on tehty luonnos:
  - LIDO lisätään kuvailevan metadatan muotoihin
  - Paikallisia (esim. kansallisia) tarpeita varten tiettyihin elementteihin voidaan lisätä omia attribuutteja, ja nyt KDK:ssa siis seuraaville asioille:
    - Skeemaversiolle, esim. kdk:METSVERSION
    - Eri osien globaalisti uniikeille tunnisteille
    - Metadatan kieli
    - Epämääräisemmän luontiajan ilmaisumahdollisuus
    - Muutos mahdollistaa myös muiden attribuuttien lisäämisen, jos nähdään tarpeelliseksi

Cern CSM –pilot 2

- Example of the prototype of the teaching application:

More information (glossary of terms, images, etc.) available in tab



"Information" tabs: to contain more background information on particles, parameters, etc.

- An article about the pilot: http://home.web.cern.ch/about/updates/2013/11/lhc-data-be-made-public-open-access-initiative

Further information:

- www.tdata.fi
- www.csc.fi/tta
- www.csc.fi/ida