

How to make sense of research data – the Finnish perspective

Dipoli Congress Center
3.12.2013

Riitta Maijala
Director, Science policy

Ministry of Education and Culture
Ministère de l'Éducation et de la culture



How to make sense of research data – the Finnish perspective

- Societal challenges and need to change approaches
- Supporting open data and open science
- Challenges and solutions

Societal challenges and change of science paradigm

- New approaches to research practises
 - Open Science, e-Science, data-driven science, citizen science
- Globalization of science and education
 - Open Access, visiting researchers
 - MOOCs (Massive Open Online Courses) and OERs (Open Educational Resources)
- Emphasis on open society
 - Open Government Data, transparency initiatives and decision making
- New demands on research impact
 - Benefits to innovation
 - Skills development at large
 - Tackling grand challenges such as aging populations and climate change

Challenging and changing environment

Changing drivers of R&D&I&HEI-policies

- Globalisation, emerging countries, emerging knowhow
- Grand societal and environmental challenges
- Financial crises
- Need of a broad based innovation concept and multidisciplinary approach
- National policies versus European policies? National versus regional policies? Local and organisational strategies?

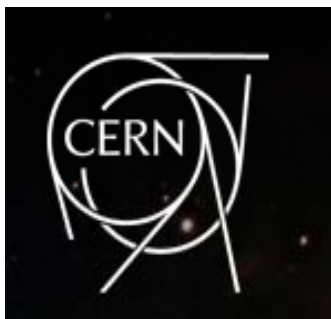
Changing mechanisms in STI&HEI

- Researcher careers/ Tenure track
- Research infrastructures
- Modern universities and RTOs
- Joint programmes and joint degrees
- Distribution of knowledge
- Joint pooling of funding (virtual common pots, real common pots, others)
- Joint evaluation and assessment activities
- Virtual learning and science

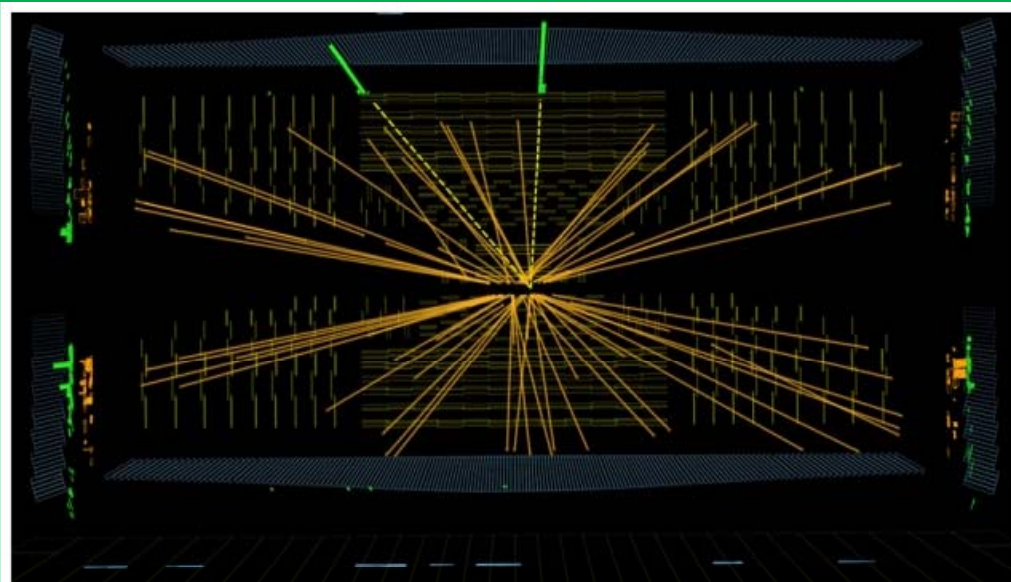
Open science and citizen science

- **Open science:** [...] to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional
- **Citizen science:** (also known as crowd science, crowd-sourced science, or networked science) is scientific research conducted, in whole or in part, by amateur or nonprofessional scientists, often by crowdsourcing and crowdfunding.

Source of definitions: Wikipedia



CMS therefore decided to study a concrete use-case for its open data by launching a pilot project aimed at education. This project, partially funded by the Finnish Ministry of Education and Culture, will share CMS data with Finnish high schools and integrate them into their physics curriculum. These data will be part of a general platform for open data provided by [Finland's IT Center for Science \(the CSC\)](#).



Data from the Large Hadron Collider, such as this decay of a Higgs boson, could be made publicly available.

THOMAS MCCAULEY/LUCAS TAYLOR/CMS COLLECTION/CERN

A screenshot of the Nature journal website. The header is dark red with the "nature" logo in white. Below the logo, it says "International weekly journal of science". A navigation bar includes links for Home, News & Comment, Research, Careers & Jobs, and Current. Below this is a secondary navigation bar with links for Archive, Volume 503, Issue 7477, News, and Article. The main content area is white and features the headline "LHC plans for open data future" with the subtext "Researchers share results to keep them accessible." The author is listed as Elizabeth Gibney, and the date is 26 November 2013.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current

Archive > Volume 503 > Issue 7477 > News > Article

NATURE | NEWS

LHC plans for open data future

Researchers share results to keep them accessible.

Elizabeth Gibney

26 November 2013

Source: <http://www.nature.com/news/lhc-plans-for-open-data-future-1.14244> and <http://home.web.cern.ch/about/updates/2013/11/lhc-data-be-made-public-open-access-initiative>

Opening up LHC data at CERN

- CERN/LHC experiment is moving towards open data
 - 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
- Use of CMS data in Finnish high school curriculum is piloted in the AVAA project funded by the Ministry of Education and Culture
 - high school students can see what real particle physics data looks like
- Open research data allows multiple uses: research, citizen science, education, ...

Case: DigiTalkoot.fi at National Library of Finland (ended Nov 29, 2012)

- “...to index the library's enormous archives so that they are searchable ...”
- “... computers fail to recognize all the words. ... DigiTalkoot enabled volunteers to participate in this fixing work by playing games.”
- 110 000 participants, 8 million tasks

Volunteer top 10

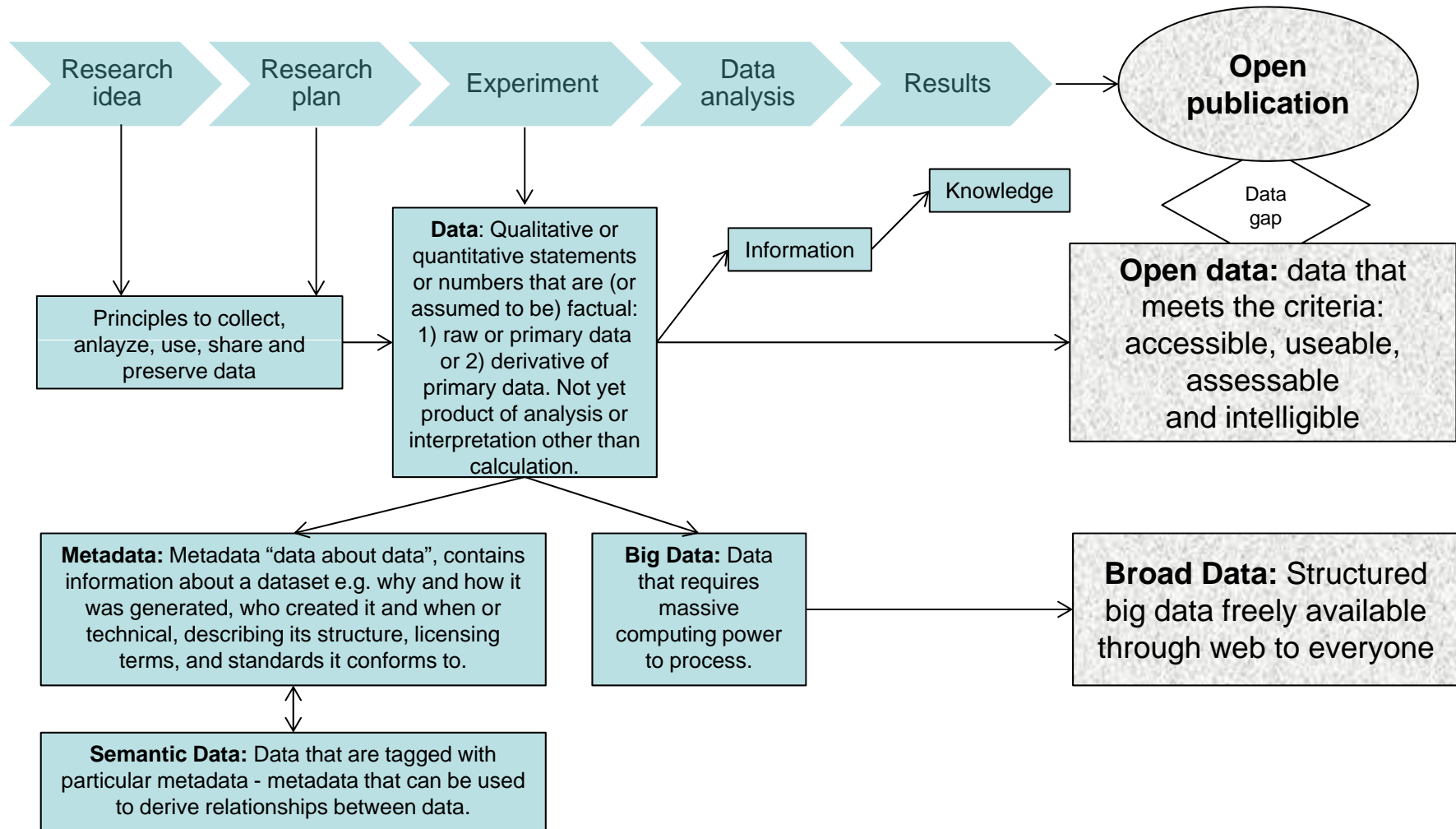
Name	Tasks	Hours	Name	Tasks	Hours
1. Petri M	348 422	395	6. Timo H	96 838	48
2. Kari K	188 128	169	7. Tea R	90 346	76
3. Vesa S	176 519	209	8. Jukka N	73 476	61
4. Anja S	133 600	106	9. Mikko V	61 129	94
5. Kirsti K	98 533	50	10. Anne-Maarit J	40 763	51



How to make sense of research data – the Finnish perspective

- Societal challenges and need to change approaches
- **Supporting open data and open science**
- Challenges and solutions

Research process and openness of data



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

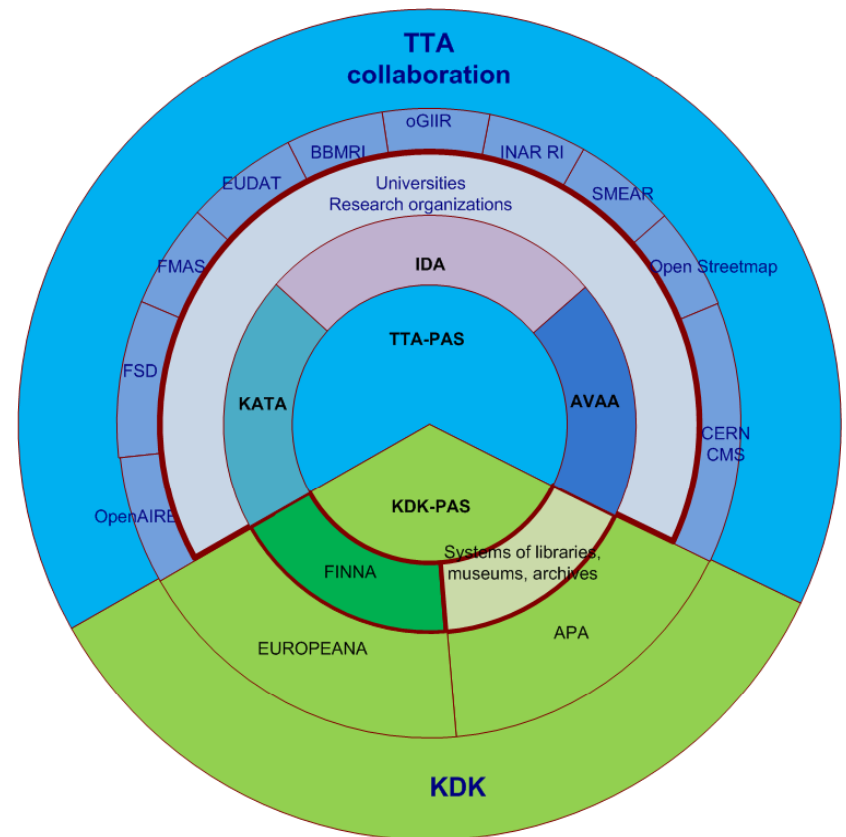
Opening the Data of the Public Sector – Open Knowledge Program 2013-2015

- The public institutes and agencies are asked and guided to evaluate their information resources from the point of view of open data as a part of their information architecture and the frame budget formulation procedure
- Developing the practices, such as the general terms of use and licences (Public Administration Recommendation is drafted)
- Developing structures and information architecture, common metadata model for open data, common public sector data portal, supporting services and training
- Developing methods to monitor the development and the impacts of open data

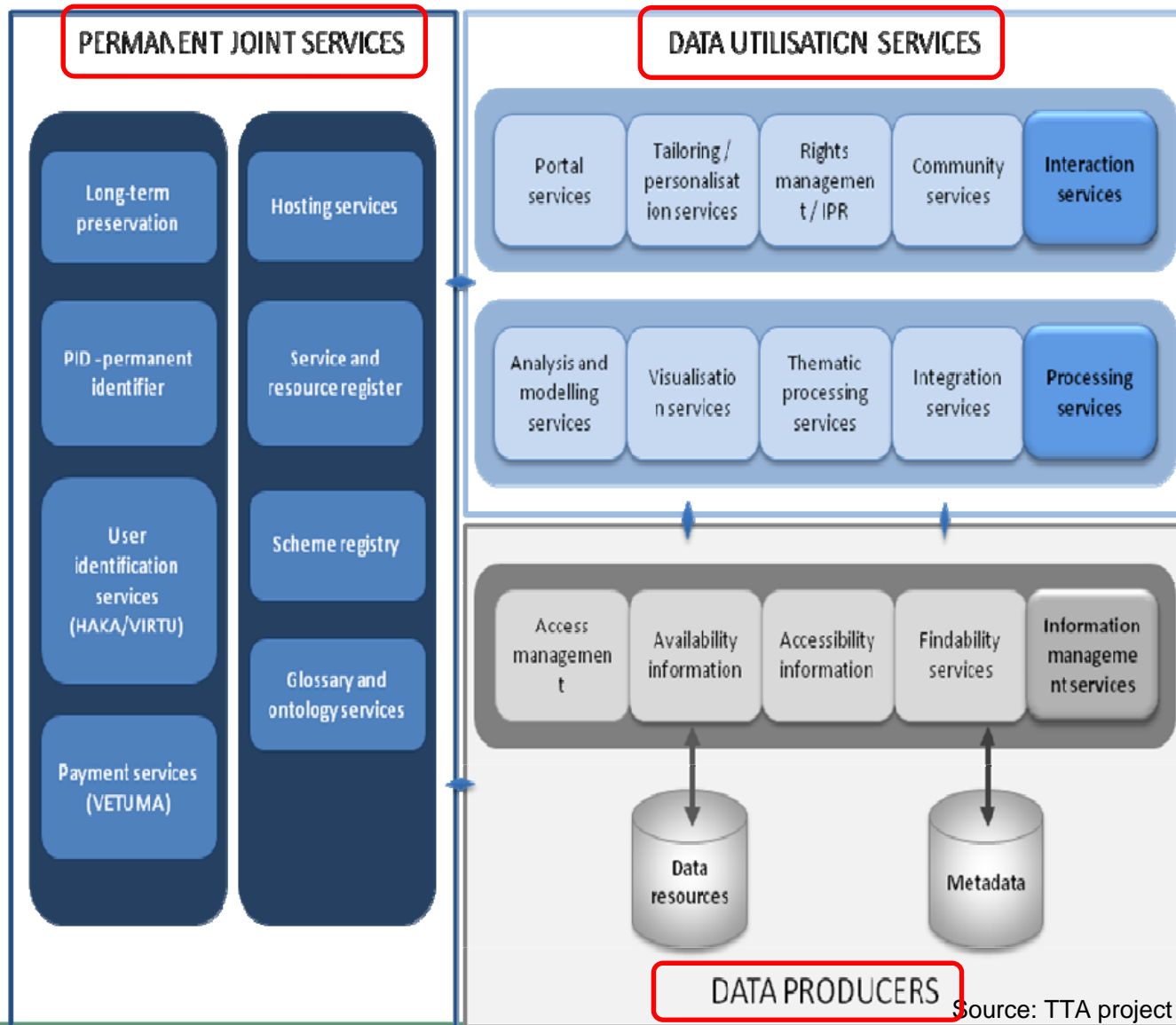
Source: Anne Kauhanen-Simanainen, Ministry of Finance

Developing the Finnish information infrastructure

- TTA (Research data) and KDK (National Digital Library) collaborations
 - Foundation of information architecture
 - Wide collaboration between actors and policy-makers
- Establishment of key services for researchers
 - Availability of data: KATA, IDA and AVAA
 - Long-term preservation: TTA-PAS and KDK-PAS
- Wide collaboration nationally and internationally
 - Universities and research organizations
 - National research infrastructure roadmap (FIRI)
 - Projects such as OpenAIRE, EUDAT, BBMRI, RDA etc.



Information infrastructures for open data – a Finnish example



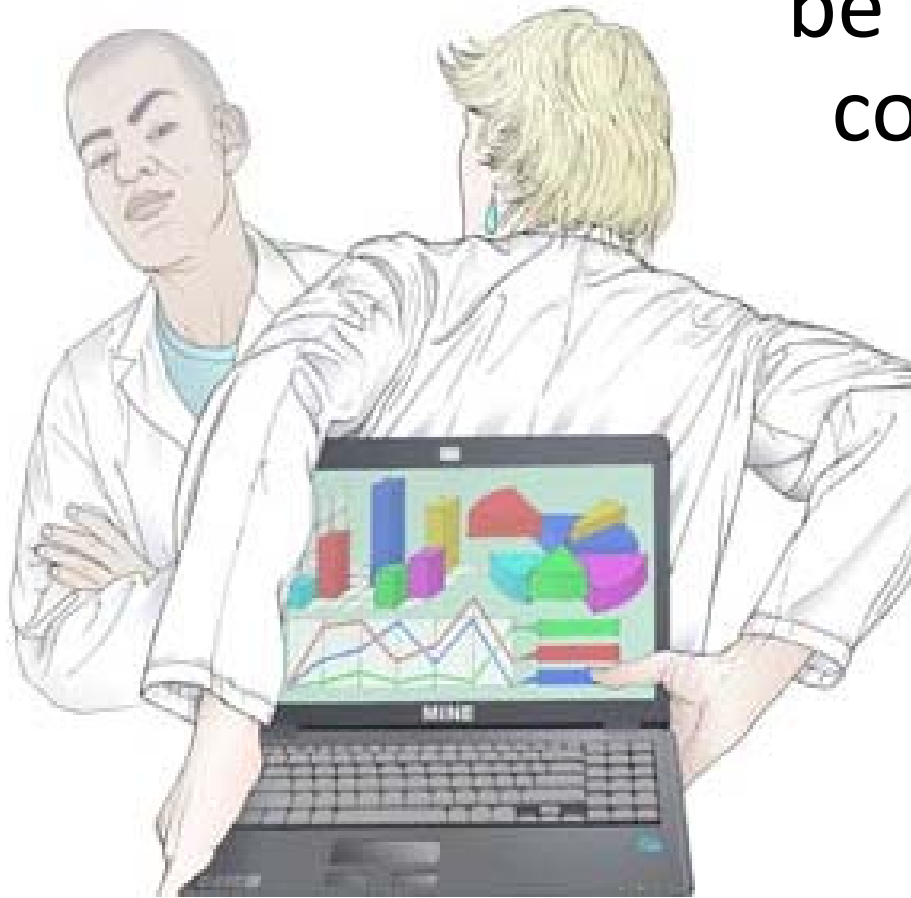
TTA WG Proposal for the policy guidance of improving access to results of scientific research

- Basic approach
 - Research data and publications are available openly via open interfaces
- Some proposals
 - All members of the Finnish research community share the scientific publications and research data produced openly in information network
 - The re-use of research data and publications is not unnecessarily restricted and the terms of use should be clearly available
 - The openness will be carried out according to the ethical guidelines for the responsible conduct of research and with respect to the valid legislation.
 - The research data should always be provided openly, unless otherwise specified by the law or contracts
 - The contracts and funding decisions concerning research should support the open availability of publications and research data

How to make sense of research data – the Finnish perspective

- Societal challenges and need to change approaches
- Supporting open data and open science
- **Challenges and solutions**

“Data management should be woven into every course in science”



Attitudes to sharing!

Nature **461**, 145 (10 September 2009) | doi:10.1038/461145a

Finnish universities: Small survey on the state of data management in 2013

– indicative results in TTA project

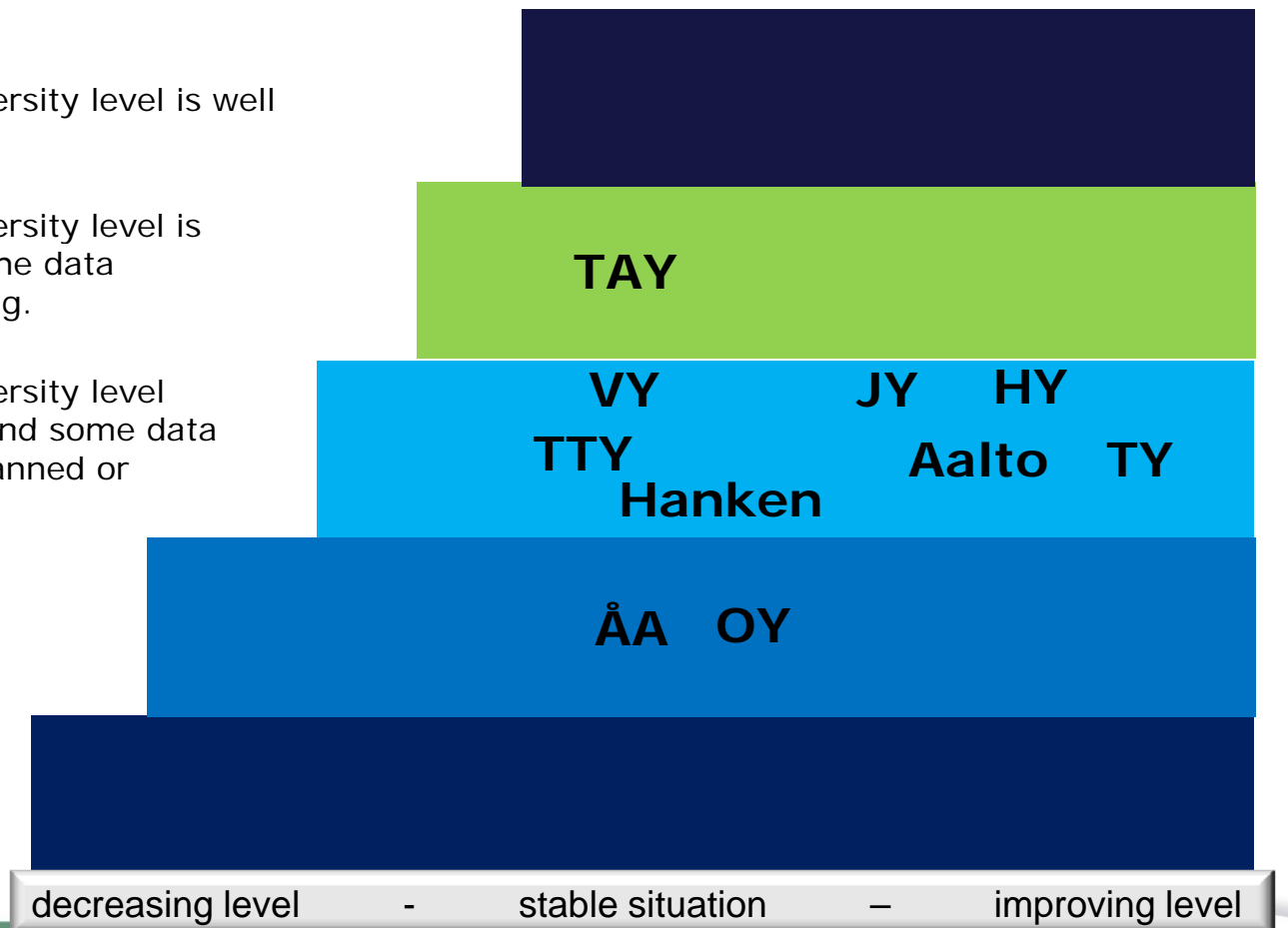
Level 4: Data management at university level is well planned and implemented.

Level 3: Data management at university level is well planned and at least a part of the data management services are functioning.

Level 2: Data management at university level has been planned to some extent, and some data management services have been planned or implemented.

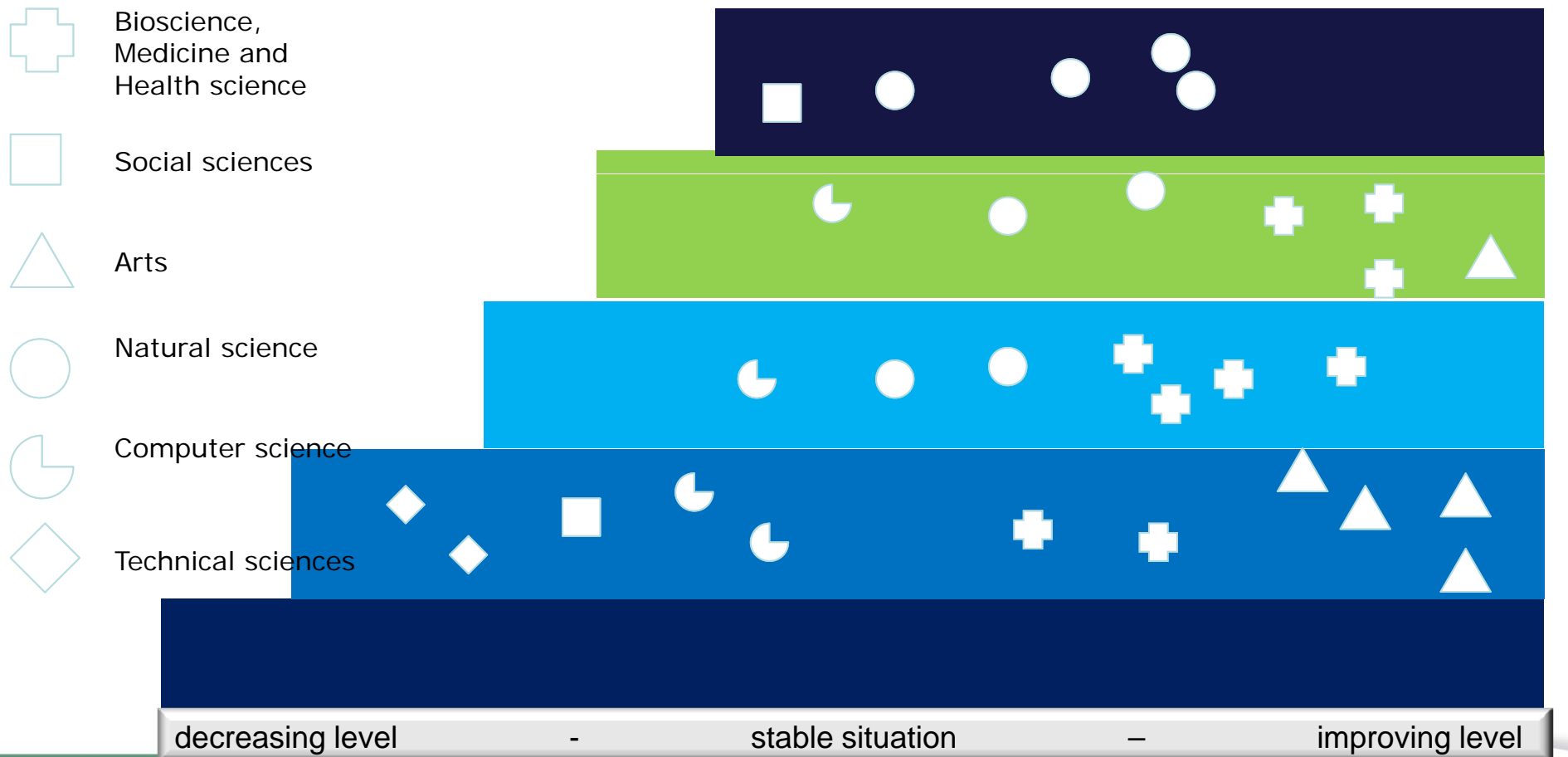
Level 1: The need to plan university level data management has been recognized and actions have been taken to develop a plan on data management.

Level 0: Data management is not on the agenda of the university.



Finnish universities: Small survey on data management of research projects in 2013

– indicative results in TTA project (n=29)

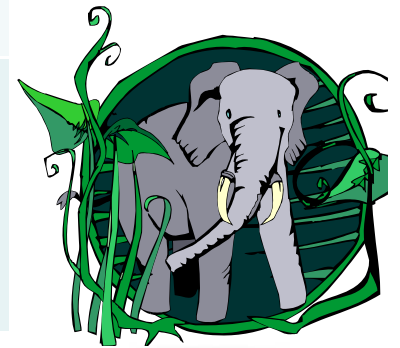


There are many challenges in open data

- Services: ontologies, metadata, long-term preservation
- Cost estimation and management
- Data skills and infrastructure skills
- Attitudes to openness
- Datasets containing personal information
- Ownership issues of open data differentiate between disciplines and countries
- Balance between individuals and the macroeconomic benefits
- Data could be misused to threaten security, public safety or health
- Priority setting for decision making

Gradual and practical approach in Finland

Challenges	Solutions
Availability of storage	Services such as TTA-IDA
Harmonization of metadata	Developing standards
Open access, licence policy	Proposals submitted to the Ministry
Cultural change towards openness	Seminars, training, guidance
International collaboration	Standardization, making use of researchers' networks, active role in key initiatives



Ministry of Education and Culture

– Key targets in open data

- Ensuring competitiveness of Finnish scientific environment
- Developing a Finnish sustainable information infrastructure for research and cultural data
- Ensuring preservation of digital data
- Providing tools for data management, both on organizational, discipline and user level
- Providing services to this infrastructure cost-effectively and sustainably
- Enabling and encourage sharing and re-use of data
- Contributing to unification of interfaces and metadata across sectors
- Actively engaging with international collaboration

Conclusions for making sense of research data

- Responding to grand challenges and changes in society requires multidisciplinary and crosssectional data
- Digitalization of science multiplies the amount of data
- Open data may boost economic and societal growth
- Solutions include both hard and soft elements
 - Identification of needs and benefits for individuals, groups and society
 - Identification of problems and finding solutions
 - Financial and other support and incentives
 - Building up infrastructures, harmonization of metadata etc.
 - Changing cultures, trust building
 - Collaboration and open dialogue essential