International Activities of Data Centric Science

Kimmo Koski, CSC
February 24th, 2015
ROIS International Workshop, Tokyo

Agenda

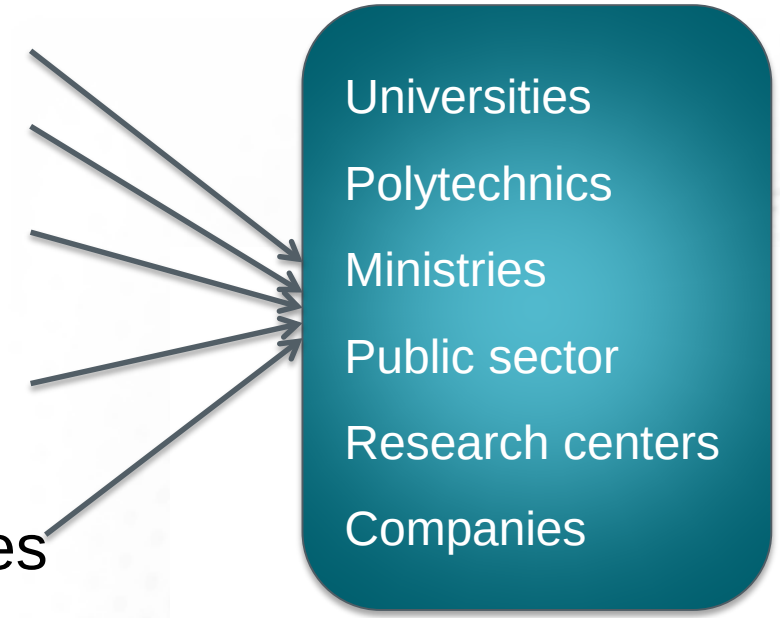
- Short presentation of CSC
- Motivation for collaboration in data centric science
- European landscape in Research Infrastructures
- From 7th Frame Program to Horizon2020: DCS related activities in EU
- European data infrastructures: EUDAT
- Global initiative: Research Data Alliance
- Building trust: need for global collaboration

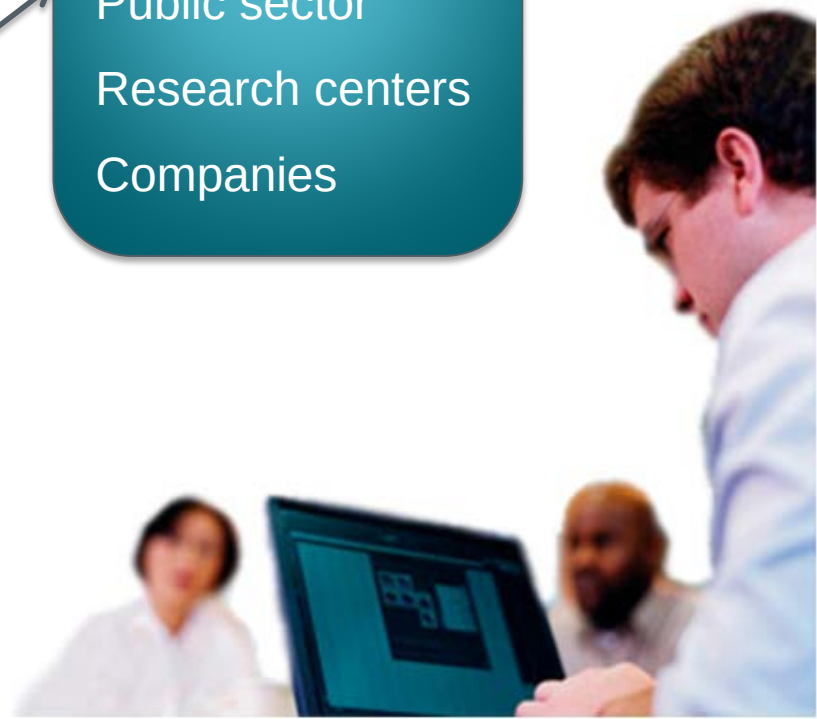
CSC at a Glance

- Founded in 1971 as a technical support unit for Univac 1108
- Connected Finland to the Internet in 1988
- Reorganized as a company, CSC – Scientific Computing Ltd. in 1993
- All shares to the Ministry of Education and Culture of Finland in 1997
- Operates on a non-profit principle
- Facilities in Espoo, close to Otaniemi campus (of 15,000 students and 16,000 technology professionals) and Kajaani
- Staff over 260
- Volume in 2014 about 35 MEUR (without investments)

CSC's Services

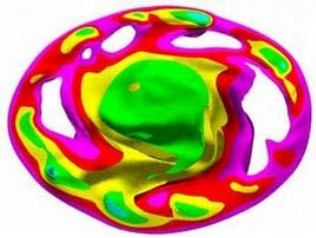
- Funet Services
- Computing Services
- Application Services
- Data Services for Science and Culture
- Information Management Services

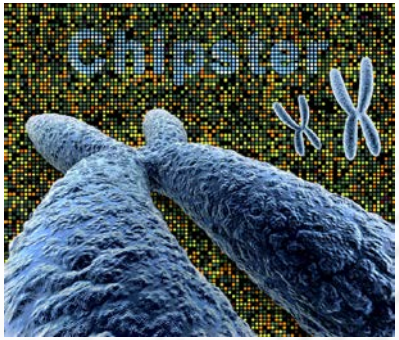




International (EC) and national projects

CSC: One-stop shopping for e-infrastructure





EUDAT

PRACE

FINNISH GRID INFRASTRUCTURE

FUNET

CSC's Services

- Funet Services
- Computing Services
- Application Services
- Data Services for Science and Culture
- Information Management Services

Elmer

- Universities
- Polytechnics
- Ministries
- Public sector
- Research centers
- Companies

More information

- www.csc.fi

- @CSCfi
- www.facebook.com/CSCfi
- www.youtube.com/channel/UCFv-76jNZlBFp6O9umdnyDA
- www.linkedin.com/company/csc---it-center-for-science
- www.slideshare.net/CSCfi

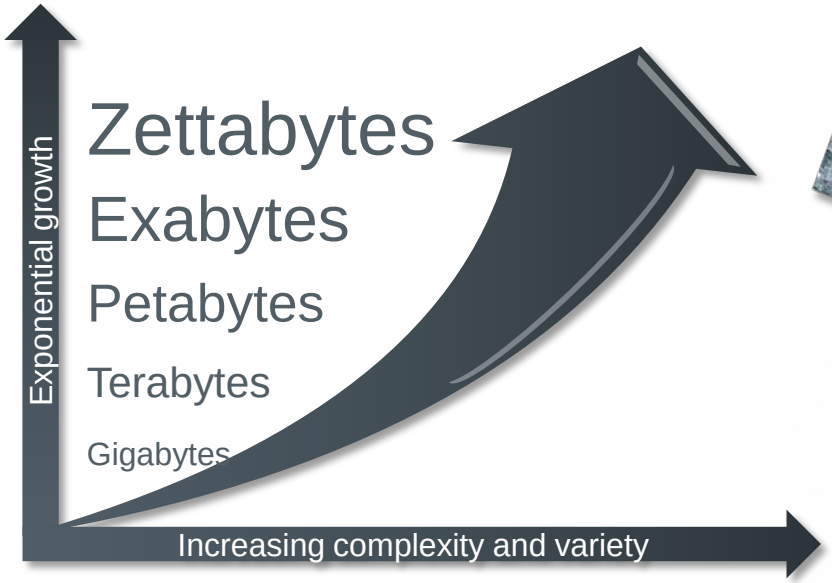
Research Infrastructure

the term Research Infrastructure refers to facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields.

e-Infrastructure

The term e-Infrastructure refers to a new research environment in which all researchers - whether working in the context of their home institutions or in national or multinational scientific initiatives - have shared access to unique or distributed scientific facilities (including data, instruments, computing and communications), regardless of their type and location in the world.

Things get more complex…

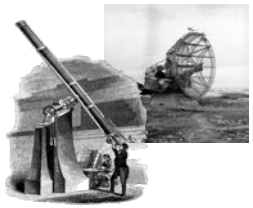


- Where to store it?
- How to find it?
- How to make the most of it?

Research Infrastructures

Research Infrastructure trends:

- Internationalization
- Diversification



ESFRI

European Ris:

- Around 300
- € 100 billion investment

middle age | 19th century | 20th century | 21st century

Looking for synergy

- The worst case scenario: Every Research Infrastructure builds an incompatible self-made ICT system of their own
- What can we do to promote collaboration and re-use of e-infrastructure?

If we do not collaborate, we will…

- … make a lot of overlapping work by re-inventing solutions in multiple places
- … do things inefficiently
- … loose money by unnecessary investments in hardware, software and services
- … run out of competent and experienced data specialists

e-infrastructure priorities

Back in 2013, when planning for Horizon2020

Data-centric science and engineering

Infrastructure for open access, management of extremely large research datasets, persistence and trust, as well as community-driven data infrastructures, and global coordination for research data

Computational infrastructure

Support to setting up of HPC Centres of Excellence, deployment of HPC Tier-0 services, support to open computing platforms and services

GÉANT

Continued development and operation of the GÉANT infrastructure, support to international links and opening and strengthening innovation activities

VRCs and virtual research environments

Supporting VRE's as an open call (bottom-up)

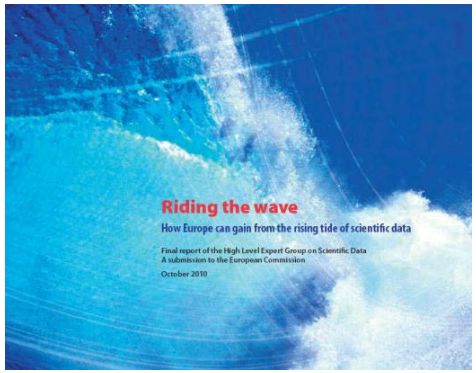
Policy development and international cooperation

Global reach and connectivity; governance; sustainability; coordination with MS


CSC
elixir Data for Life
EU SOLARIS
COPAL
European Community Airborne Laboratory
DARIAH-EU
EISCAT
EuroFEL
EUROPEAN FREE ELECTRON LASERS
EATRIS
ICOS
cta
cherenkov telescope array
SIOS
European Infrastructure for Translational Medicine
instruct
Integrating
Biology
EMBRC
EUROPEAN
MARINE
BIOLOGICAL
RESOURCE
CENTRE
EURO-BIOIMAGING
BBMRI
erinha
european
multidisciplinary
seafloor
HiPER
emso
EUROPEAN
SPALLATION
SOURCE
EU-OPENSCREEN
Chemical keys for life's locks
IFMIF
EPOS
EUROPEANPLATEOBSERVINGSYSTEM
EMFL
European Magnetic Field Laboratory
IAGOS
LifeWatch
eli
anaee
CLARIN
MYRRHA
ECRIN
ECCSEL
DASISH
BioMedBridges
ENVRI
CRISP
PRACE
EGI
EUDAT
HELIX NEBULA
THESCIENCECLOUD
GÉANT
OpenAIRE
Open Access Infrastructure for Research in Europe
HORIZON 2020

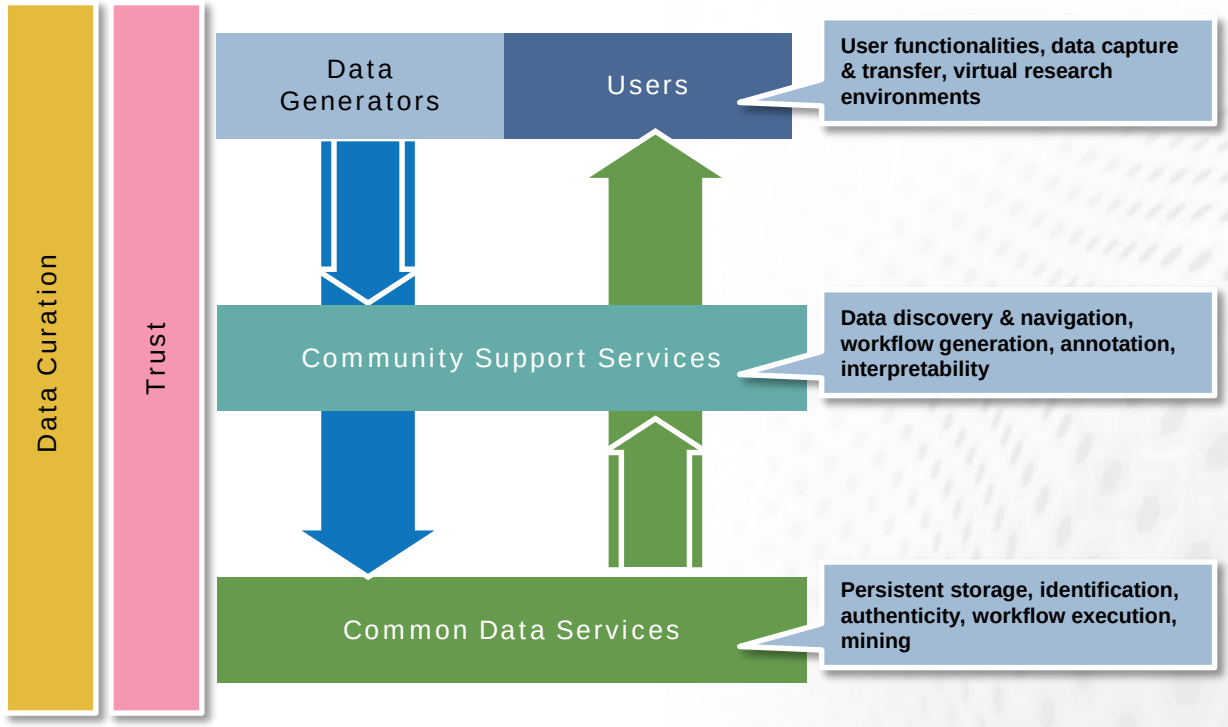
Data as infrastructure: Europe is Riding the Wave

The High Level Expert Group on Scientific Data presented **Riding the Wave** in October 2010



Vision: "data e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the **data themselves become the infrastructure** a valuable asset on which science, technology, the economy and society can advance".

Collaborative Data Infrastructure -A framework for the future? -



Three priorities towards H2020

- We need to build trust between researchers and e-infrastructure providers
  - Synergy, efficiency, proper workload division
- What we do in European level need to link tightly to national and global activities
  - Most of the funding for e-infrastructure is national
  - Researchers don’t want to have too many layers
- The services need to be user driven
  - What to do by user communities, how to implement by technology providers
  - Co-design: user communities as partners in ICT projects, not only as customers

Science is global, as also the need for e-infrastructure and related services

- Research collaborations are global
- E-Infrastructures must react to the requirements
  - Examples from the networks
- Sustaining Global Forums
  - RDA, CODATA

What will happen in Horizon2020?

- The work with Research Infrastructures and e-infrastructures will continue
  - Continuation of successful projects from FP7
  - New initiatives
  - Update of ESFRI Research Infrastructure list
    - http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
    - ESFRI = European Strategy Forum for Research Infrastructures
- Targets include:
  - Excellent science
  - Industrial leadership
  - Addressing societal challenges
  - http://ec.europa.eu/programmes/horizon2020/h2020-sections

Some flagship projects in H2020

- EUDAT2020 (www.eudat.eu)
  - European data infrastructure
- RDA-Europe (europe.rd-alliance.org)
  - Research Data Alliance
- PRACE (www.prace-ri.eu)
  - Supercomputing
- GEANT (www.geant.net)
  - Research network
- Other, including Research Infrastructures, clusters of Ris and e-infrastructure projects

E-Infrastructures with data

- EGI, www.egi.eu
  - European grid initiative
- Helix-Nebula, http://www.helix-nebula.eu/
  - Cloud services
- OpenAIRE, https://www.openaire.eu/
  - Publications, open access
- Number of new initiatives starting with first H2020 calls

Previous EU projects

- 2 data infrastructure calls
  - About 15 disciplinary projects
- ESFRI cluster projects
  - Bringing Research Infrastructures from same area closed together
  - BioMedBridges for biomedical
  - DASISH for humanities
  - ENVRI for environmental sciences
  - CRISP for (physics) experiments
  - Continuation for part of these expected in H2020

ESFRI Research Infrastructures

- Newest roadmap update from 2010, projects on-going: http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf#view=fit&pagemode=none
- Six areas: social sciences and humanities, biological and medical sciences, environmental sciences, materials and analytical facilities, energy, physical sciences & engineering
- 38 projects, major investment in each
  - Not all funded, planning and prioritization on-going
- All 38 projects require data management: how to enable synergy and collaboration?

Example: EISCAT3D

- The next generation European incoherent scatter radar system
- Construction 2014-16, operation 2016-46
- Construction 60-250 MEUR, Operation 4-10 MEUR/year
- Sweden, Norway, Finland, UK, Germany, Japan, China
- https://www.eiscat3d.se/

EUDAT2020

Data Services, Tools & Knowledge

EUDAT

A pan-European e-Infrastructure solution for pan-European RI data Challenges

- All Research Infrastructures are facing data challenges
  - Where to store the growing amount of data?
  - How to find it?
  - How to make the most of it?

- Many communities are developing own solutions
  - This is good…
  - … but we also need to make sure that the solutions remain interoperable

- EUDAT mission is to fill this gap
  - Providing a set of services to help RIs managing their growing amount of data
  - Providing these services across communities to ensure minimum level of interoperability
  - EUDAT also help to bring data and computing together (HPC centers core partners)

EUDAT

Data Centers and Communities

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

KIT Karlsruhe Institute of Technology

INGV

CERN

epcc

BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

DKRZ DEUTSCHES KLIMARECHENZENTRUM

CINES Centre Informatique National de l'Enseignement Supérieur

CINECA

RedIRIS

CSC

Max-Planck-Institut für Meteorologie

rzg RECHEN-ZENTRUM GARCHING

JÜLICH FORSCHUNGSZENTRUM

Science & Technology Facilities Council

maatG

CERFACS

umweltbundesamt ENVIRONMENT AGENCY AUSTRIA

POZNAN SUPERCOMPUTING NETWORKING

SNIC

Trust-IT Services Ltd Communicating ICT to markets

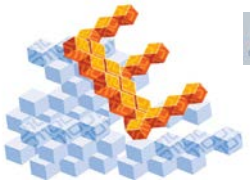
UNINETT sigma

SURF SARA

UCL

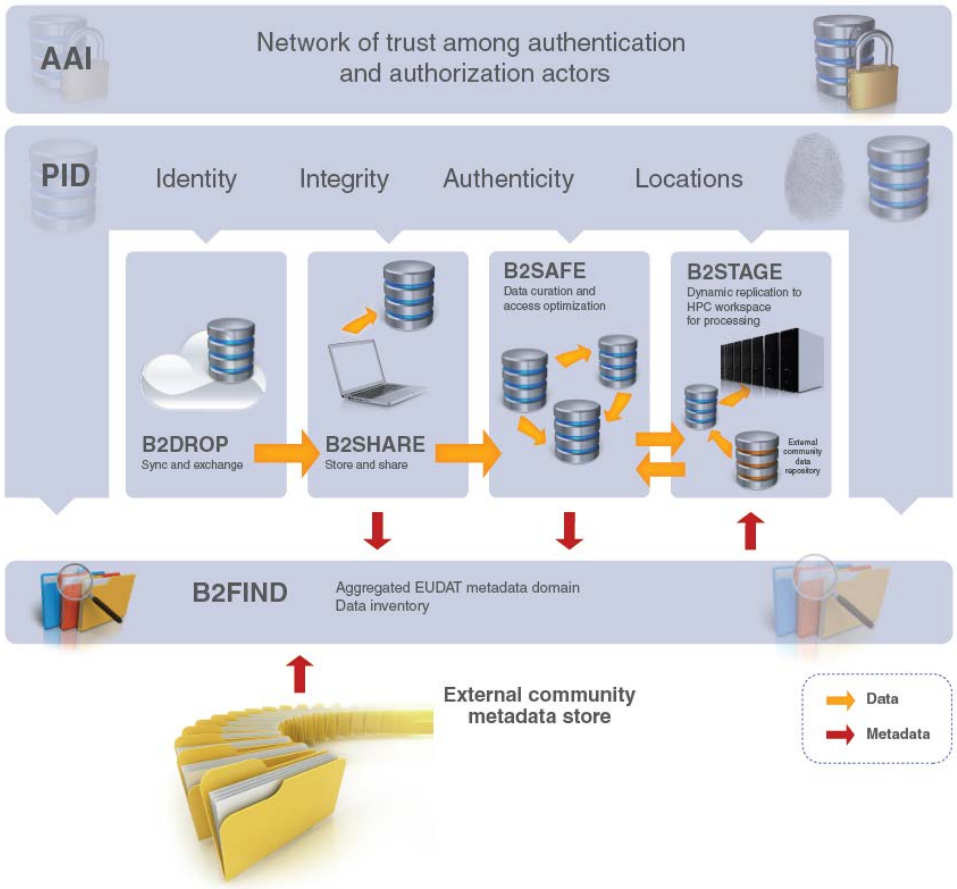
EUDAT

User Forums + 30 communities



1st User Forum
7-8 March 2012, Barcelona

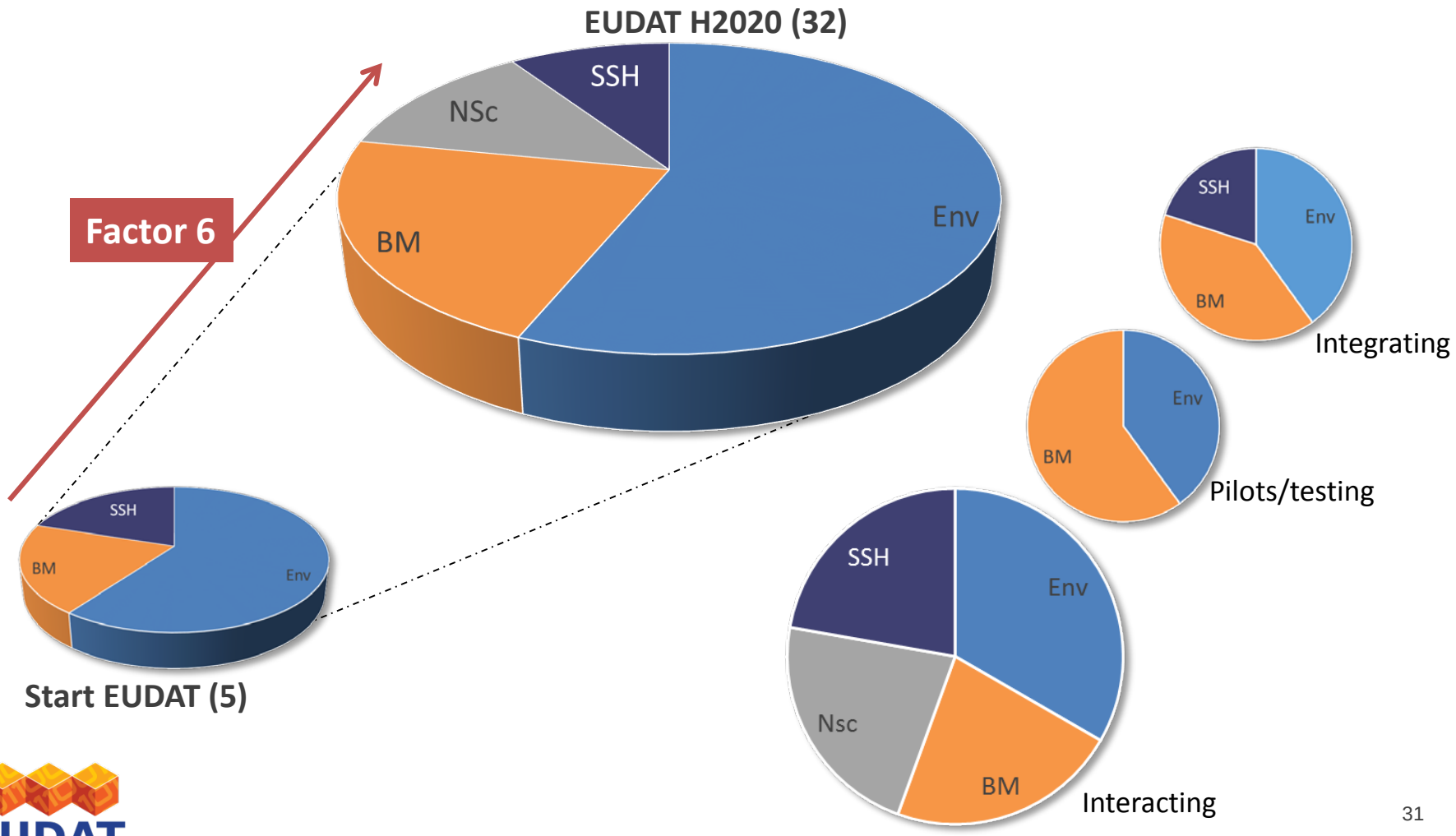
Services & Resources



Covering both **access** and **deposit**, from **informal data sharing to long-term archiving**, and addressing **identification**, **discoverability** and **computability** of both **long-tail and big data**, EUDAT's services will **address the full lifecycle of research data**

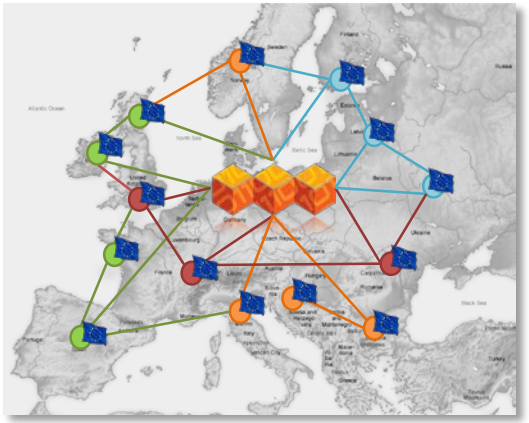
EUDAT

User-Driven


EUDAT H2020 (32)
SSH
NSc
BM
Env
Factor 6
SSH
BM
Env
Start EUDAT (5)
SSH
Env
BM
Integrating
Env
BM
Pilots/testing
SSH
Env
Nsc
BM
Interacting

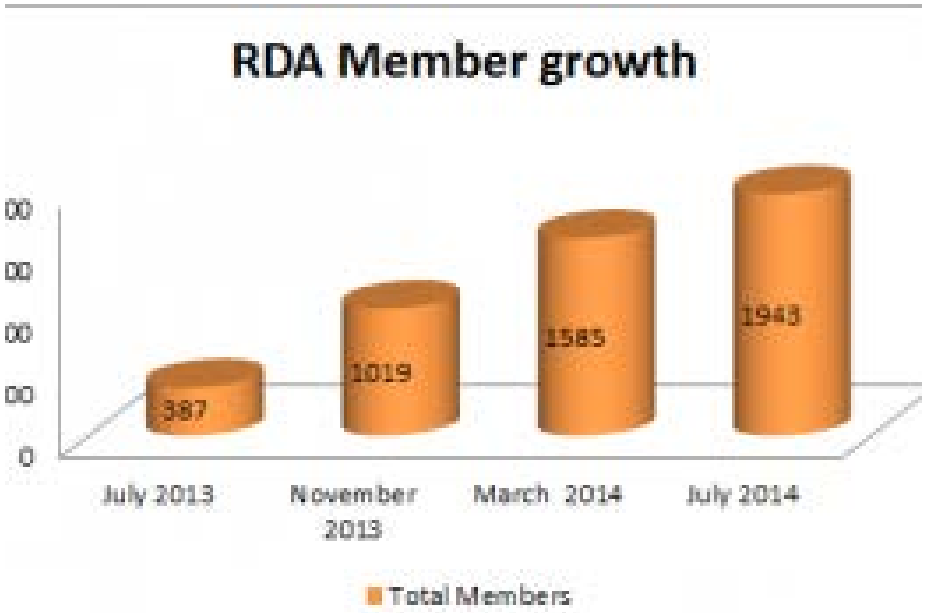
Bridging National and European solutions

- Making national resources more available and visible
  - Making visible valuable national collections through EUDAT
  - Access to European resources through national catalogues
- Enhancing cross-national collaborations
  - EUDAT provides a European extension to national solutions
- Research and infrastructures are still largely funded at national levels
  - Better coordination between funding schemes and roadmaps is needed



EUDAT

Research Data Alliance (RDA)
rd-alliance.org



Country/ Total	JULY 2013	NOV 2013	MARCH 2014	JULY 2014
EU	178	416	774	949
AU	11	32	51	69
US	165	473	592	681
Others	33	98	168	244
TOTAL	387	1019	1585	1943

RDA

The goal of RDA is to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange. This is achieved through the development, adoption, and deployment of infrastructure, policy, practice, standards, and other deliverables.

- **RDA Vision**

  Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

- **RDA Mission**

  RDA builds the social and technical bridges that enable open sharing of data.

Ways to participate in RDA

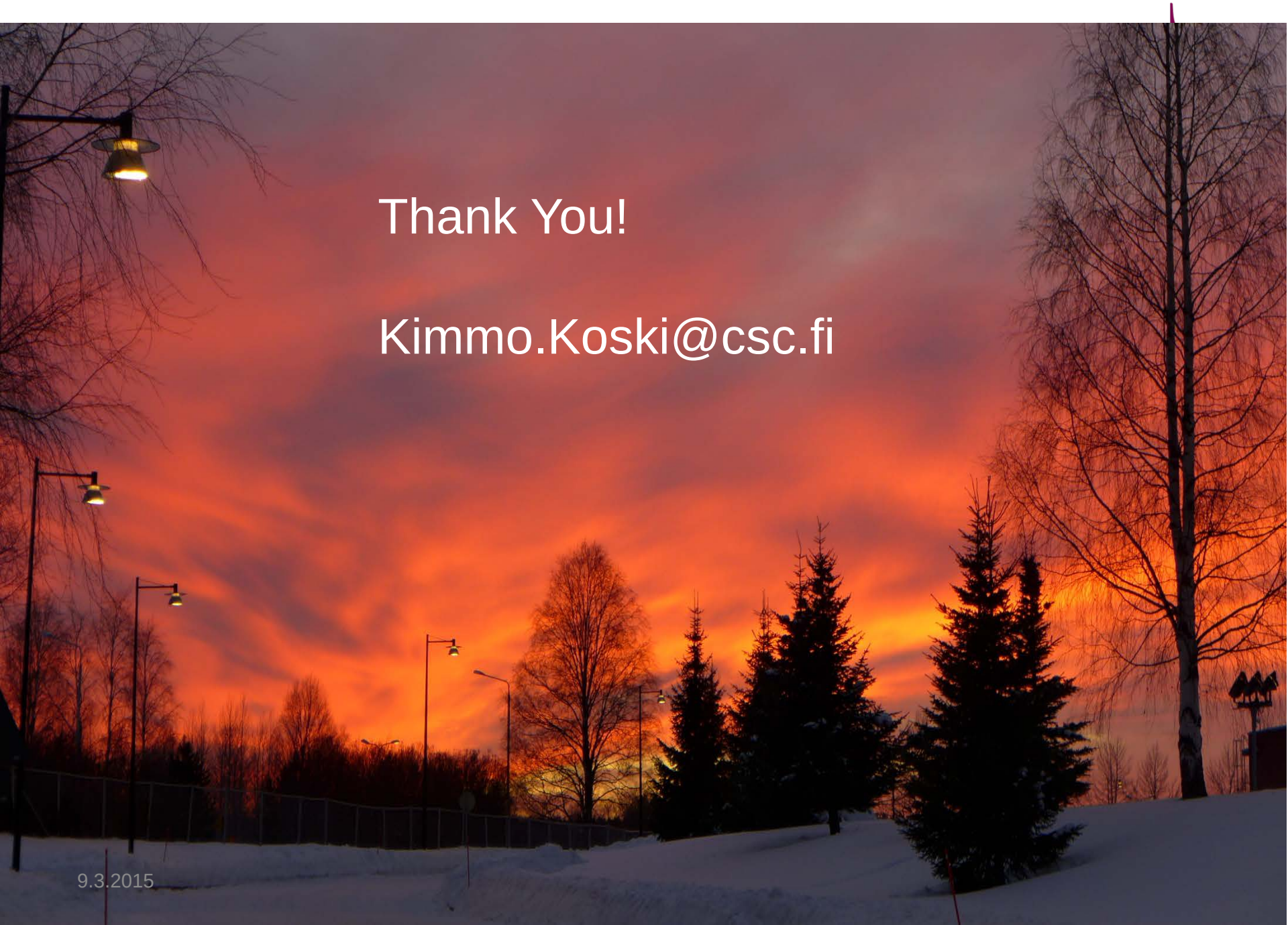
- Joining as a member
- Participation in working groups, interest groups or BoF groups
- RDA conferences twice a year
  - March 9-11th 2015 San Diego, US
  - September 2015 Paris, France
- Number of other events
  - https://rd-alliance.org/news-and-events.html

The Data Harvest: How sharing research data can yield knowledge, jobs and growth

https://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

Building trust and collaboration

- Bringing research and ICT closer together
  - Tools, methodology, data analysis
  - Optimizing workload, benefits from synergy
- New requirements for education system
  - Producing data scientists with multidisciplinary approach
- Planning National – Regional – Global
  - Collaboration in all levels
- Sharing best practices and exchanging knowledge


Thank You!
Kimmo.Koski@csc.fi