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

- Universities
- Polytechnics
- Ministries
- Public sector
- Research centers
- Companies
International (EC) and national projects
CSC: One-stop shopping for e-infrastructure

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

Universities
Polytechnics
Ministries
Public sector
Research centers
Companies
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.
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

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
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? -

- Trust
- Data Curation
- Data Generators
- Users
- Community Support Services
- Common Data Services

User functionalities, data capture & transfer, virtual research environments
Data discovery & navigation, workflow generation, annotation, interpretability
Persistent storage, identification, authenticity, workflow execution, mining
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](http://www.eudat.eu))
  - European data infrastructure
- **RDA-Europe** ([europe.rd-alliance.org](http://europe.rd-alliance.org))
  - Research Data Alliance
- **PRACE** ([www.prace-ri.eu](http://www.prace-ri.eu))
  - Supercomputing
- **GEANT** ([www.geant.net](http://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


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
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 their 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)
Data Centers and Communities

26 European partners
User Forums + 30 communities
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.
User-Driven

Factor 6

Start EUDAT (5)

EUDAT H2020 (32)

Integrating

Pilots/testing

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
## RDA Member Growth

The table below shows the growth of RDA members from July 2013 to July 2014 for different regions:

<table>
<thead>
<tr>
<th>Country/Total</th>
<th>JULY 2013</th>
<th>NOV 2013</th>
<th>MARCH 2014</th>
<th>JULY 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU</td>
<td>178</td>
<td>416</td>
<td>774</td>
<td>949</td>
</tr>
<tr>
<td>AU</td>
<td>11</td>
<td>32</td>
<td>51</td>
<td>69</td>
</tr>
<tr>
<td>US</td>
<td>165</td>
<td>473</td>
<td>592</td>
<td>681</td>
</tr>
<tr>
<td>Others</td>
<td>33</td>
<td>98</td>
<td>168</td>
<td>244</td>
</tr>
<tr>
<td>TOTAL</td>
<td>387</td>
<td>1019</td>
<td>1585</td>
<td>1943</td>
</tr>
</tbody>
</table>
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
The Data Harvest: How sharing research data can yield knowledge, jobs and growth

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