CSC computing resources

Pekka Lehtovuori, Tomasz Malkiewicz and Atte Sillanpää
CSC – IT Center for Science Ltd.

Program

- 14-15 How to utilize new CSC computing resources in your research
  - Interactive! Q&A welcome
- 15-15:30 Round robin / free discussion
- 15:30-> F2F meetings

Outline

- CSC at a glance, services for researchers
- CSC supercomputers Phase 2
  - *Compute in Sisu* (Cray XC30)
  - *Compute in Taito* (HP cluster)
- Services for sharing your data
- Virtualised compute Cloud and FGI
- Training

CSC at glance

- Founded in 1971
- Operates on a *non-profit* principle
- Staff ~255 people
- Facilities in Espoo and Kajaani
- Free of charge services for higher education institutions in Finland

Datacenter CSC Kajaani

CSC's Services

- FUNET Services
- Services for Research
  - Computing Services
  - Application Services
- Data Services for Science and Culture
- Information Management Services

Universities

Polytechnics

Ministries

Public sector

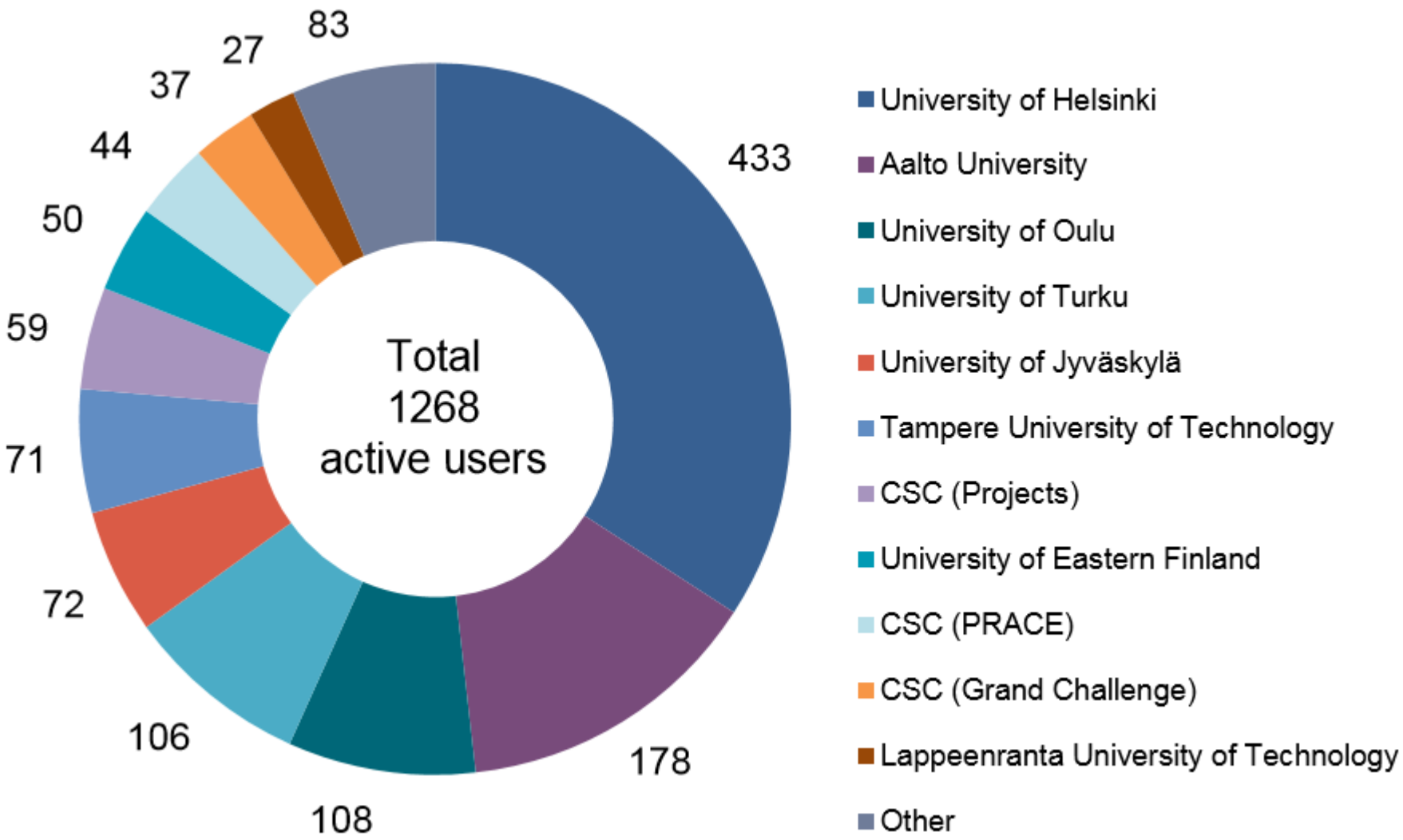
Research centers

Companies

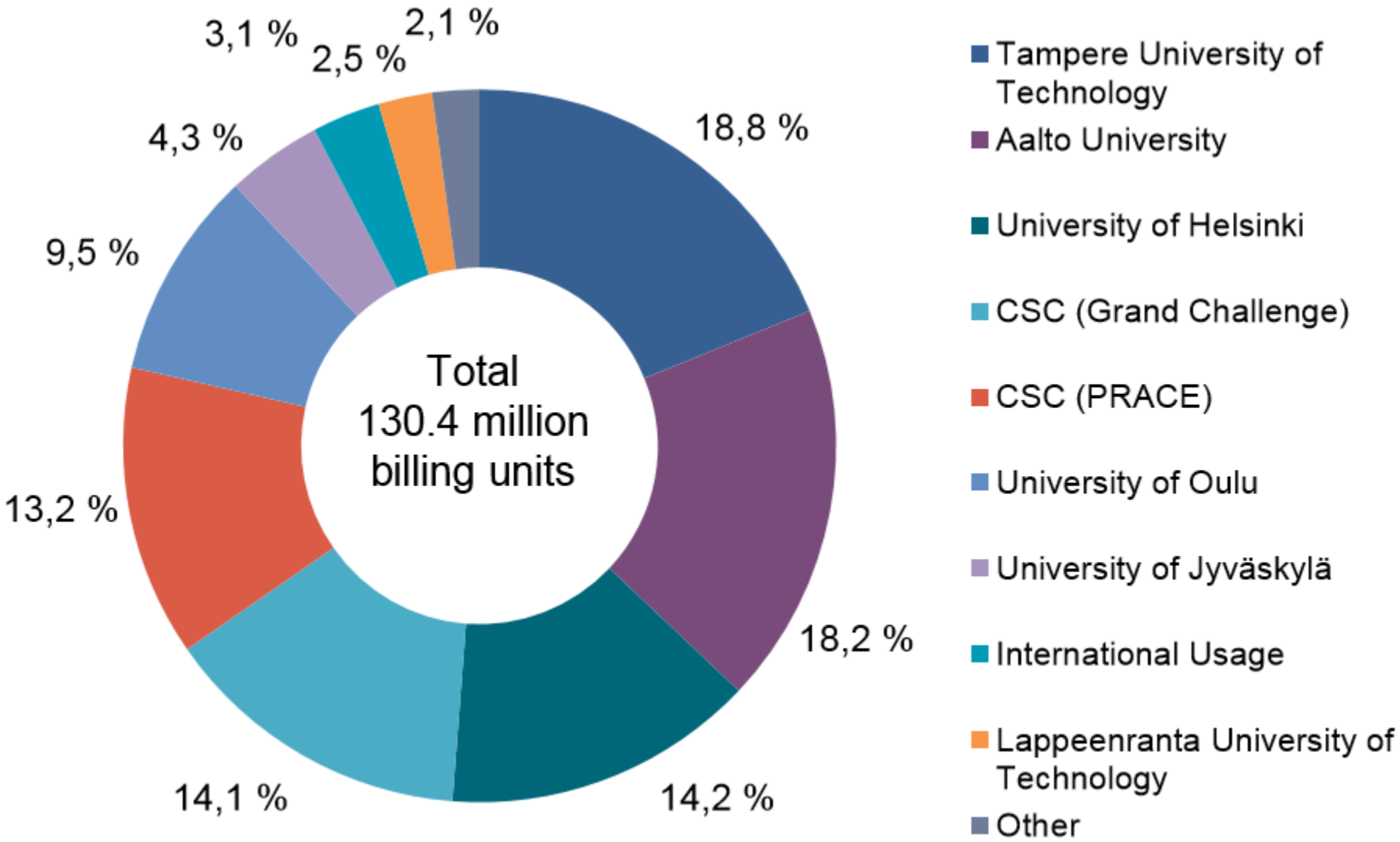
Users

- About 700 active computing projects
  - 3000 researchers use CSC's computing capacity
  - 4250 registered customers
- Haka-identity federation covers all universities and higher education institutes (287 000 users)
- Funet - Finnish research and education network
  - Total of 370 000 end users

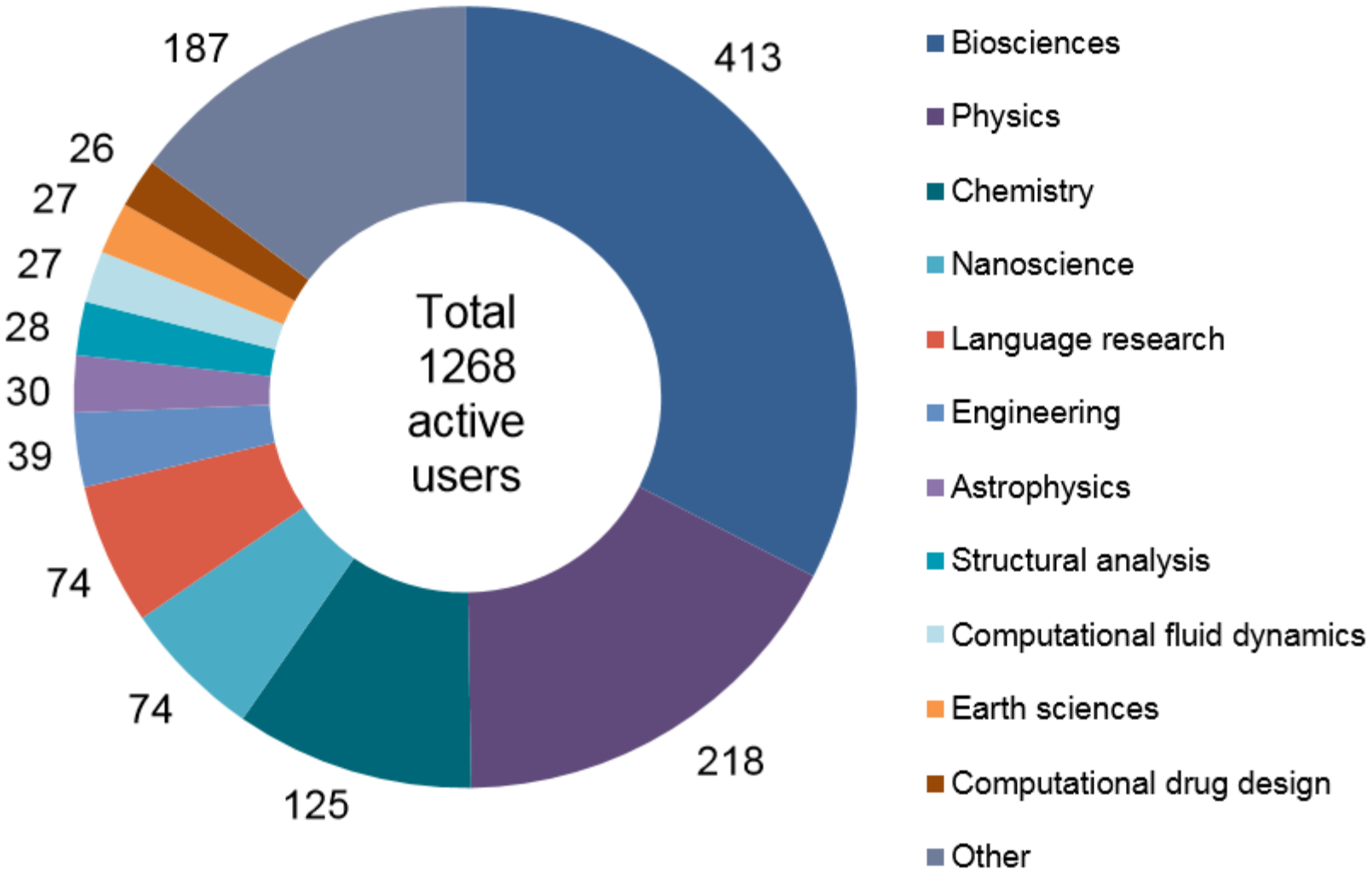
Users of computing resources by organization 1H2014



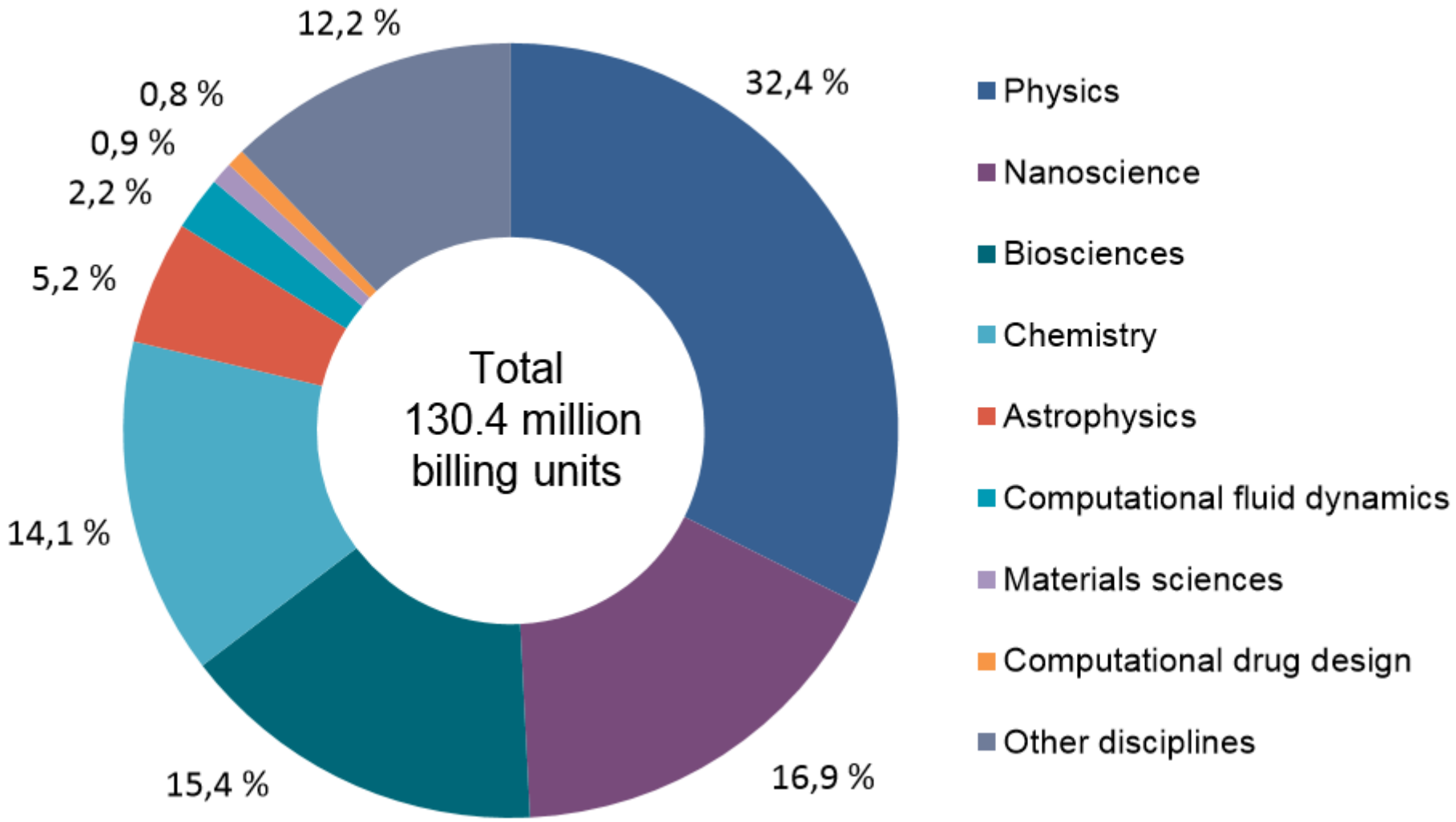
Computing usage by organization 1H2014



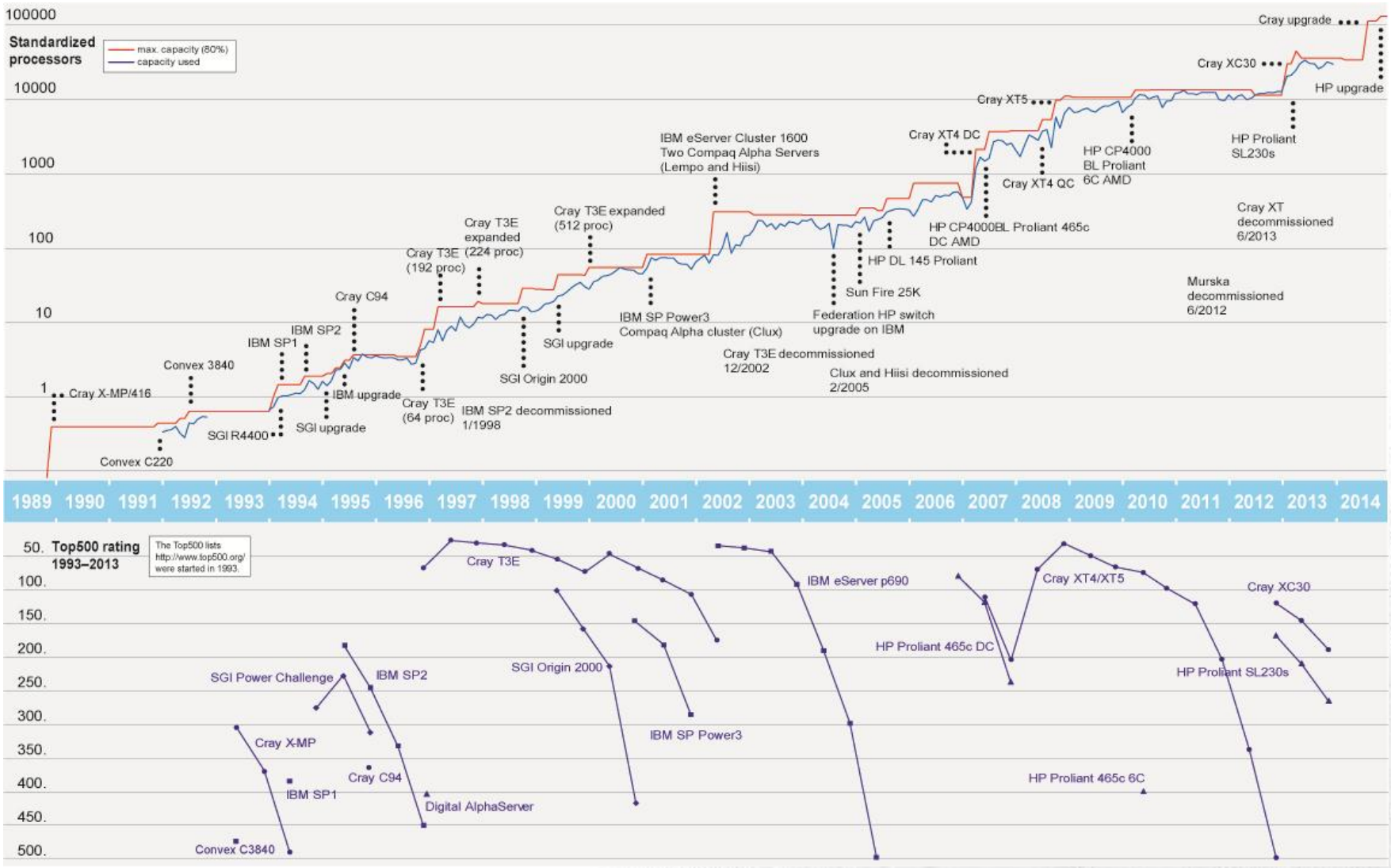
Users of computing resources by discipline 1H2014



Computing usage by discipline 1H2014

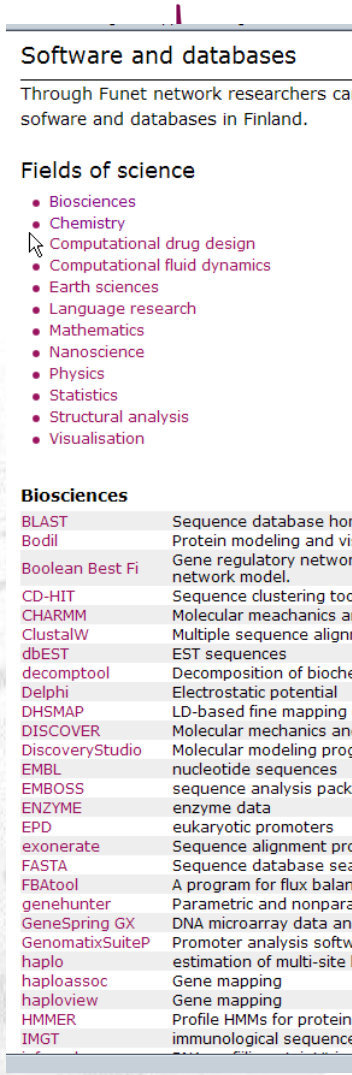


CSC Computing Capacity 1989–2014



Software and database offered by CSC

- Large selection (over 200) of software and database packages for research https://research.csc.fi/software-for-science
- Mainly for academic research in Finland
- Centralized national offering: software consortia, better licence prices, continuity, maintenance, training and support



Software and databases

Through Funet network researchers ca
sofware and databases in Finland.

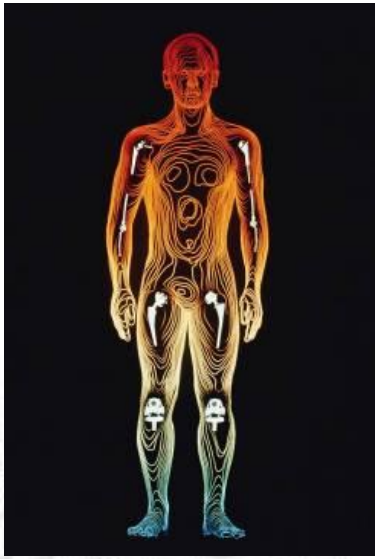
Fields of science

- Biosciences
- Chemistry
- Computational drug design
- Computational fluid dynamics
- Earth sciences
- Language research
- Mathematics
- Nanoscience
- Physics
- Statistics
- Structural analysis
- Visualisation

Biosciences

BLAST	Sequence database hor
Bodil	Protein modeling and vi
Boolean Best Fi	Gene regulatory networ network model.
CD-HIT	Sequence clustering too
CHARMM	Molecular meachanics a
ClustalW	Multiple sequence alignm
dbEST	EST sequences
decomptool	Decomposition of bioche
Delphi	Electrostatic potential
DHSMAP	LD-based fine mapping
DISCOVER	Molecular mechanics an
DiscoveryStudio	Molecular modeling prog
EMBL	nucleotide sequences
EMBOSS	sequence analysis pack
ENZYME	enzyme data
EPD	eukaryotic promoters
exonerate	Sequence alignment pro
FASTA	Sequence database sea
FBAtool	A program for flux balan
genehunter	Parametric and nonpara
GeneSpring GX	DNA microarray data an
GenomatixSuiteP	Promoter analysis softw
haplo	estimation of multi-site
haploassoc	Gene mapping
haploview	Gene mapping
HMMER	Profile HMMs for protein
IMGT	immunological sequence

New customers



Apply for CSC account:_ https://research.csc.fi/accounts-and-projects

- Most of CSC services are free for academic researchers, but usually a **CSC user account** is required.
  - Basic usage: register as CSC customer via SUI
  - Larger computing resources via an application form
- Benefits
  - A wide selection of scientific programs and databases available at CSC servers.
  - ICT resources and science-aware support (helpdesk@csc.fi)
  - Courses and events covering many areas are organized regularly.
  - Guide books and magazines in PDF.
  - CSC's research and development to improve services.
  - Networks bring together people with similar interests in science and technology.

HPC PHASE 2 RESOURCES

– SISU, TAITO
– BULL
– STORAGE: DDN (PHASE 3)

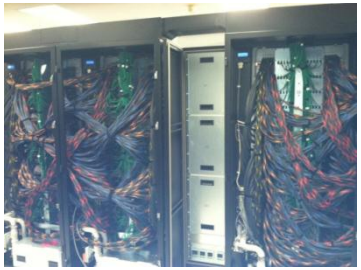
Sisu: Cray Supercomputer

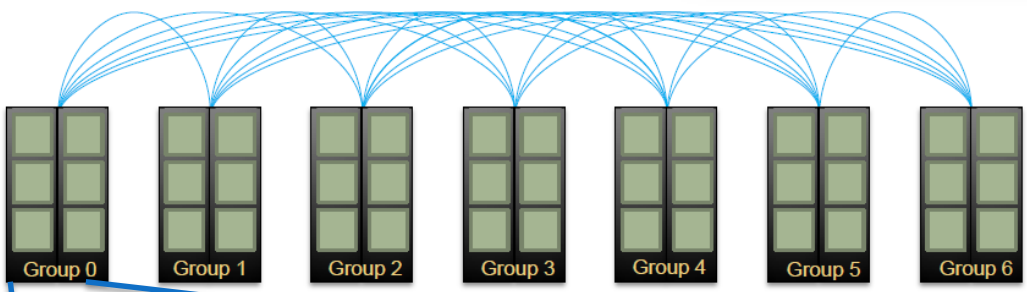
- Intel® Haswell® processor E5-2690 v3 product family
- Cray Aries Interconnect
- 40 512 cores
- 64 GB memory per node

Running on Sisu Phase 2

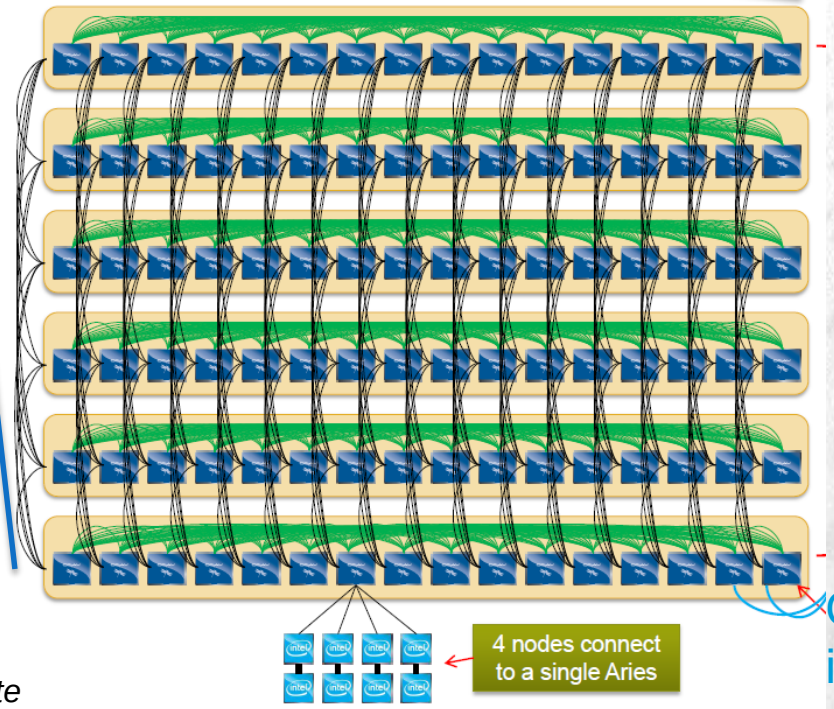
- Sisu guide
  - *https://research.csc.fi/sisu-user-guide*
- Phase 1 binaries may run off-hand, **CSC strongly advises to recompile the code**
- Login nodes based on Sandy Bridge (as they were in Phase 1)
  - Nonetheless anything running on ***login nodes*** needs to be recompiled (libraries/compilers' versions updated, e.g. gcc, some old ones removed)
- Scalability tests for more than 1008 cores
  - *https://research.csc.fi/sisu-scalability-tests*

Cray Dragonfly Topology





All-to-all network between groups



2 dimensional all-to-all network in a group

Optical uplinks to inter-group net

Source:
Robert Alverson, Cray
Hot Interconnects 2012 keynote

Sisu Phase 2 features

- AVX-2
  - May need to optimize for wider vectors' size
- DDR4
  - Higher bandwidth, lower power consumption
- Max job size increased
- Native SLURM on the way
  - We might be moving to it at some point

Taito: HP Supercluster

- Intel® Sandy Bridge & Intel® Haswell processors
- FDR InfiniBand interconnect
- ~18 000 cores
- Different memory per node sizes: 64, 128, 256 GB and 1.5 TB

Taito is a heterogeneous cluster

- Different jobs need different resources
- Bulk Sandy Bridge compute nodes
- Largemem Sandy Bridge compute nodes
- Hugemem Sandy Bridge compute nodes
- Bulk Haswell compute nodes

- Local */tmp* disk 2 TB on each node

→ reserve only what you need

One SLURM to serve them all…

- Do old applications run on new CPUs?
  - May run, CSC **recommends re-compiling**
  - Build your software for both (old and new) architecture
  - Gain depends on architecture
- <u>Batch job scripts need to be updated</u>
  - Number of cores per node: Phase 1: 16, Phase 2: 24
  - Memory changes
  - Instructions will be available through user guides
  - Partition CPU architecture can be specified

SLURM configuration: Fair usage

- SLURM uses fair share: the highest priority jobs go into execution next
  - Priority is decreased by the total amount of resources used in last 2 weeks per user
  - Priority is increased by time spent queueing
  - Backfiller will try to put small jobs into gaps due to current available resources and highest priority job
  - Jobs labeled ”Association limit” are not eligible to run (due to too many jobs in queue by the user)
- *Due to abuse, a maximum limit of jobs in queue now enforced*
- Chain jobs (--dependency –flag for SLURM) if you need long running time
- Don’t overallocate memory (add this command to your batch script `used_slurm_resources.bash` will print requests vs. used at stdout)
  - If you request a full node (-N 1), use –mem=55000 instead of –mem-per-core=something)
  - If you see abuse or think that the setup is unfair, contact *helpdesk@csc.fi*
- SUI has a monitoring tool for your jobs and used resources (*Services -> eServices -> My Project*)

How to prepare for Taito Phase 2?

- Porting strategy
  - Getting started document and a User Guide for Sisu prepared
  - Compilers, libraries, flags, …
  - Preliminary performance data
  - Add **AVX-2 flag** when compiling your code
  - CSC ports and optimizes a number of applications for the new architectures
  - Consider running testing your code on Sisu, which has Haswell CPUs

Status of Phase 2 Sisu and Taito

- *Sisu:* available since ***9.9.2014***
- *Taito:* general availability planned in ***Q4 2014***

Bull

- Official opening on ***1.10.2014***
- ***Accelerators and coprocessors***
  - 38 NVIDIA K40 nodes (76 gpus)
    - 12 GB memory per card
  - 45 Intel Xeon Phi nodes (90 Xeon Phis)
    - 16 GB memory per card
  - Energy efficient CPU's

How to access Bull

- **Accessing the resources**
  - **Intel Xeon Phi: ssh taito-mic.csc.fi (TBC)**
  - **NVIDIA K40: ssh taito-gpu.csc.fi**

Fast and large storage: DDN Phase 3

- HPC storage used by Sisu and Taito
- System size increased to ~4 PB
  - About 1.9 PB added to the current configuration in early October 2014
  - Aggregate bandwidth > 80 GB/s (currently ~48 GB/s)
- Available together with Phase2 supercomputers

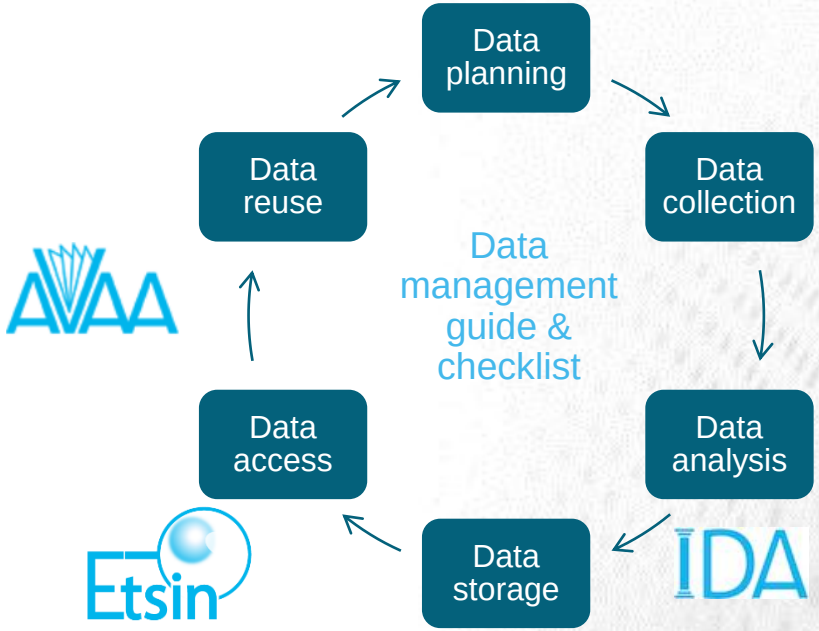
Disks in total

- *4.0 PB on DDN*
  - $HOME directory (on Lustre)
  - $WRKDIR (*not backed up*), soft quota 5 TB / user
  - Up to 100 TB / project
- *HPC Archive*
  - 2 TB / user, common between Sisu and Taito
- *3 PB disk space through TTA/IDA*
  - 1 PB for Universities
  - 1 PB for Finnish Academy (SA)
  - 1 PB to be shared between SA and ESFRI
  - more could be requested
- 1.1 PB cloud NFS (Netapp) for virtual machines of IaaS customers funded by ELIXIR Finland
- */tmp* on Sisu and Taito (around 1.8 TB) to be used for *compiling* codes on login nodes

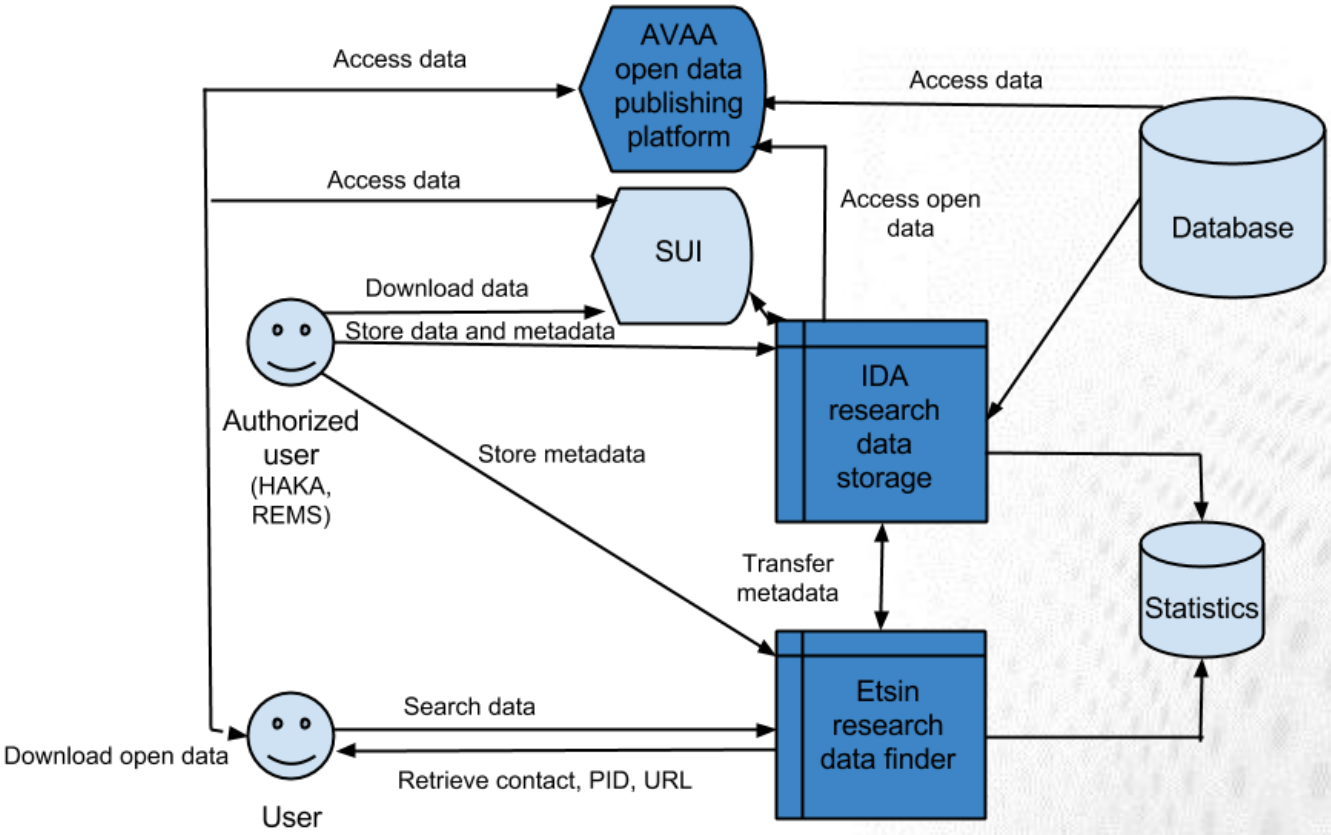
The Open Science and Research Initiative

- **The Ministry of Education and Culture has launched the Open Science and Research Initiative, for the promotion of information availability and open science for years 2014–2017.**
- Goal to make Finland a leading country in openness of science and research
- The initiative will promote the reliability, openness and societal impact of science and research
- Scope of the initiative includes publications, research data, and methods
- The aim is to provide researchers with practical knowledge in how they as individuals can implement open science
- Several services provided by the Ministry to researchers

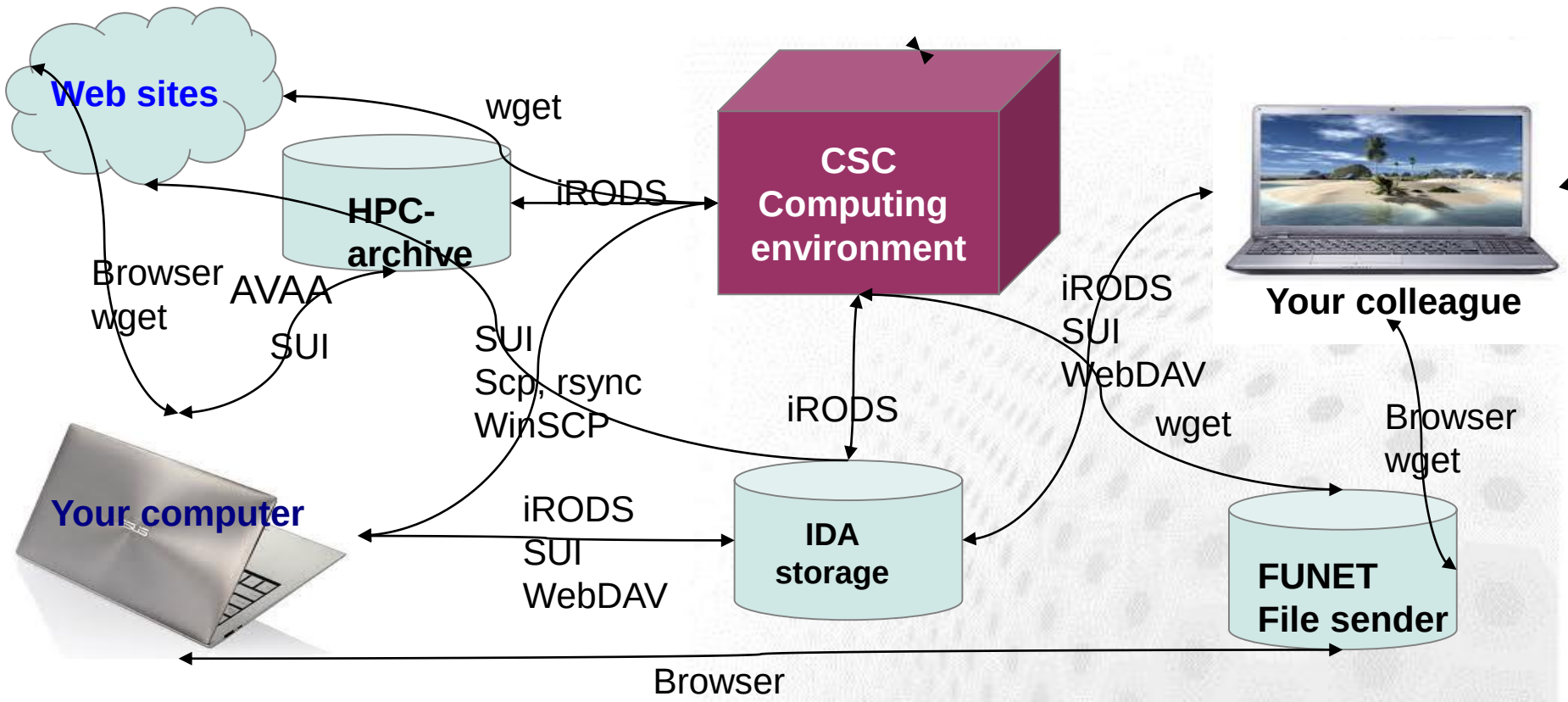
Research data lifecycle and open science services produced by CSC



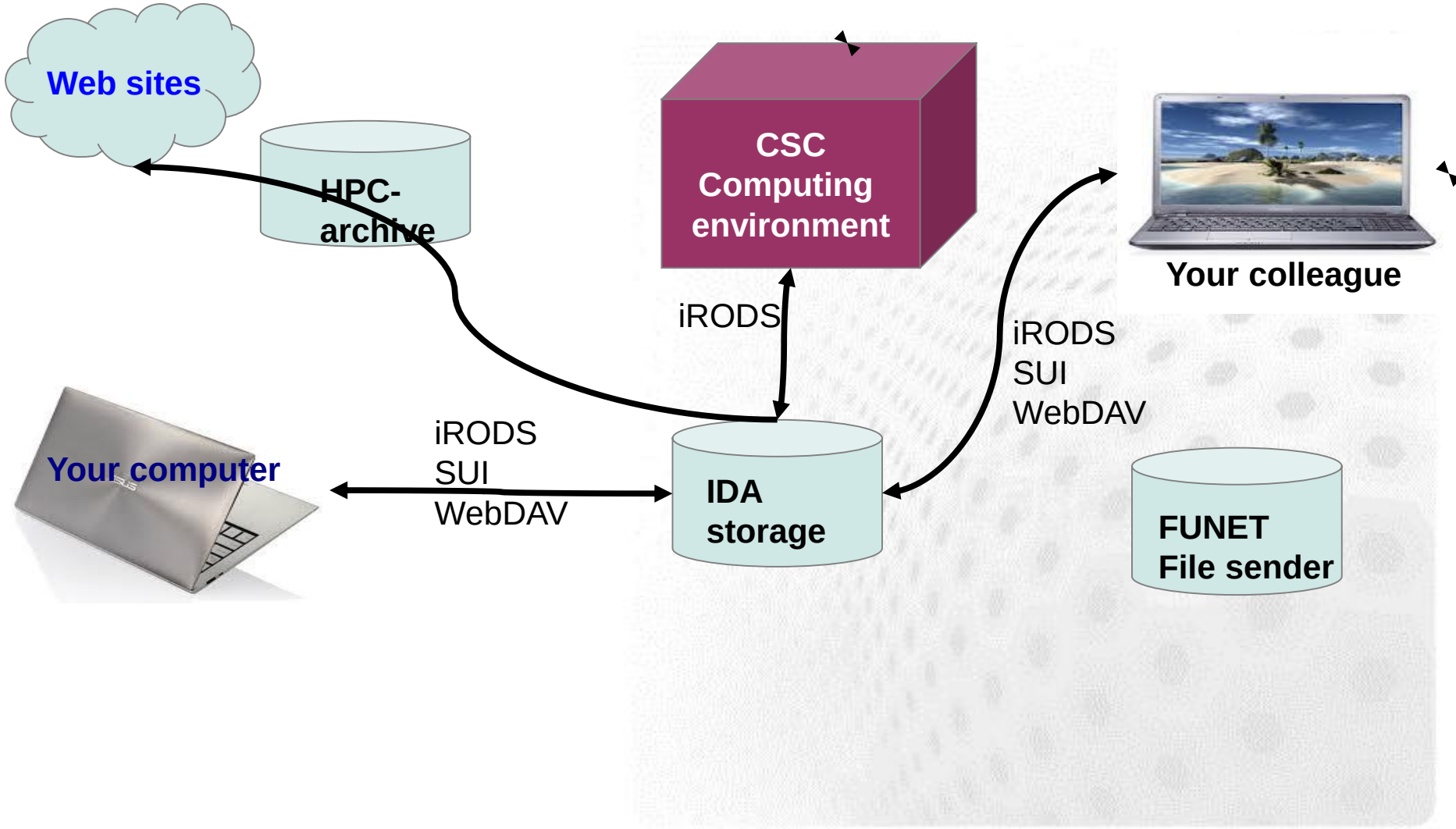
Use of open science services



Moving data to and from CSC



IDA storage service



IDA storage service

- iRODS-based storage system for storing, archiving and sharing research data
- Designed for stable data sets, not data analysis (low speed / high safety and availability)
- Storage quotas are granted by the universities
  - JyU quota 210 TB, 119 TB allocated to projects, 30 TB saved data
  - Contact person at JYU: Ilari Korhonen, puh. 0401877986, ilari.x.korhonen@jyu.fi
- Project based usage with personal accounts
- Interfaces: iRODS commands, WWW (SUI), WebDav network disks
- Linked to other ATT services:
  - Metadata may be stored via Etsin and data can be published through AVAA
- https://www.tdata.fi/ida

IDA

- Part of ATT
- Quotas granted by universities and Academy of Finland
- Several interfaces (WWW/SUI, network disk, i-commands)
- Internet accessible
- Project based structure
- Data can be made public through AVAA

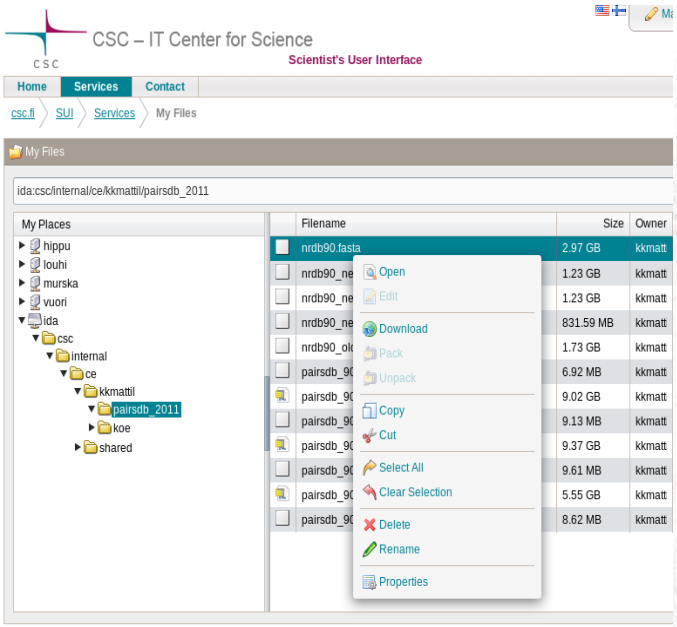
HPC-archive

- Part of CSC computing environment
- 2 TB default quotas for CSC users
- Usage with i-commands
- Visible only to CSC environment
- Personal storage area
- Replaced the old $ARCHIVE service

iCommands

- `iput` *`file`* — move file to IDA
- `iget` *`file`* — retrieve file from IDA
- `ils` — list the current IDA directory
- `icd` *`dir`* — change the IDA directory
- `irm` *`file`* — remove file from IDA
- `imv` *`file file`* — move file inside IDA
- `imeta` *`command`* — view and edit metadata
- `irsync` — synchronize the local copy with the copy in IDA
- `imkdir` — create a directory to IDA
- `iinit` — Initialize your IDA account

IDA in Scientist's User Interface



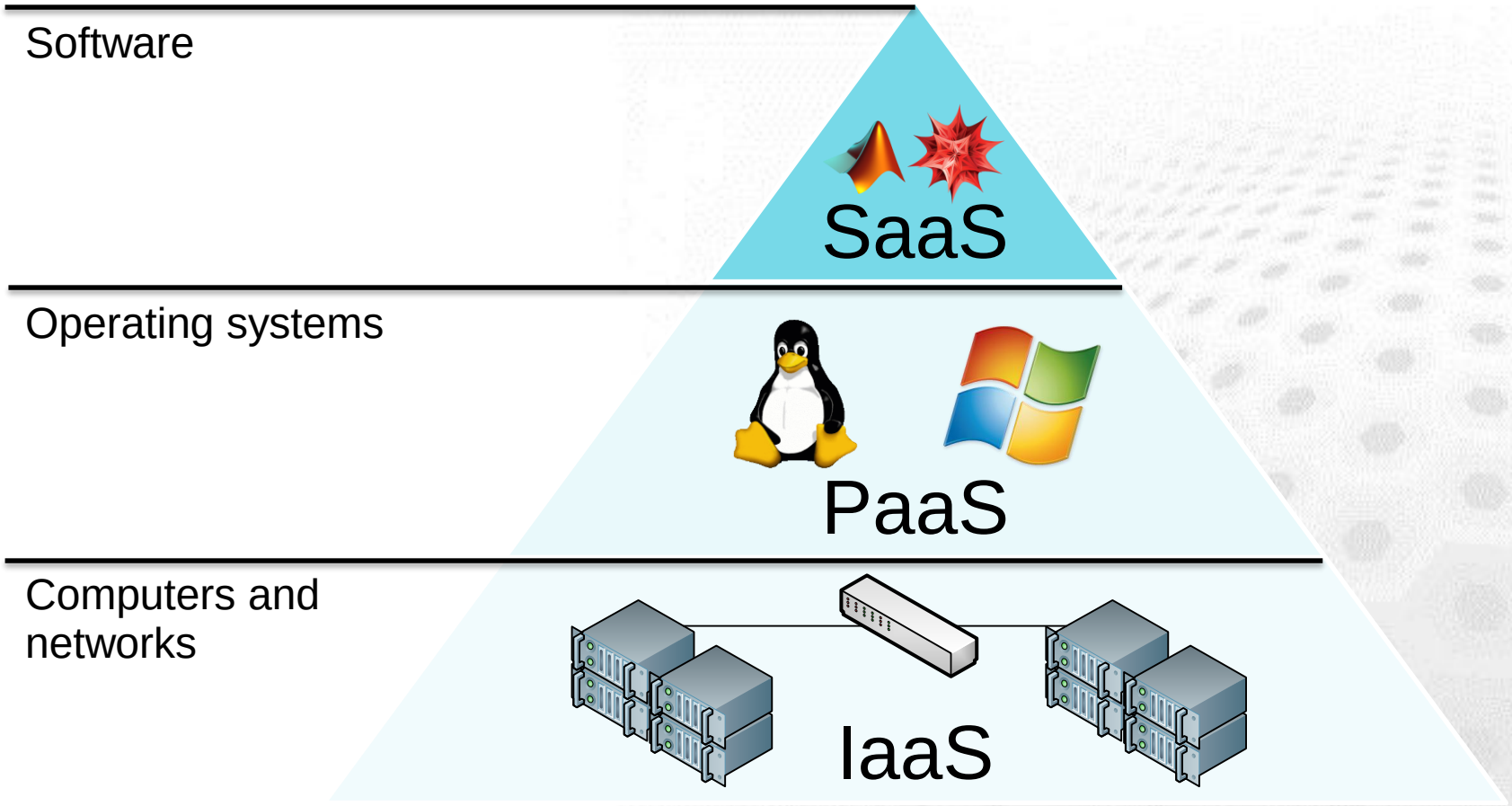
EUDAT and Research Data Alliance – Open Data and Services for Research

- EUDAT is a pan-European initiative building a sustainable **cross-disciplinary** and **cross-national** data infrastructure for research data. EUDAT provides services for accessing and preserving research data. www.eudat.eu

- Research Data Alliance Europe (RDA) is building the social and technical bridges that enable **global open sharing of data**. Researchers, scientists, data practitioners from around the world are invited to work together to achieve the vision. Finland is the most active from the Nordic countries. Join us at: http://europe.rd-alliance.org/

- CSC coordinates both projects.

Cloud computing:
three service models



cPouta – computing in the Cloud

- cPouta
  - Virtual machines on demand
  - Taito hardware
  - Dedicated resources (HPC focus)
- More freedom
- More responsibility

cPouta on Taito

Taito cluster:

two types of nodes, HPC and cloud

HPC node

HPC node

Cloud node

Cloud node

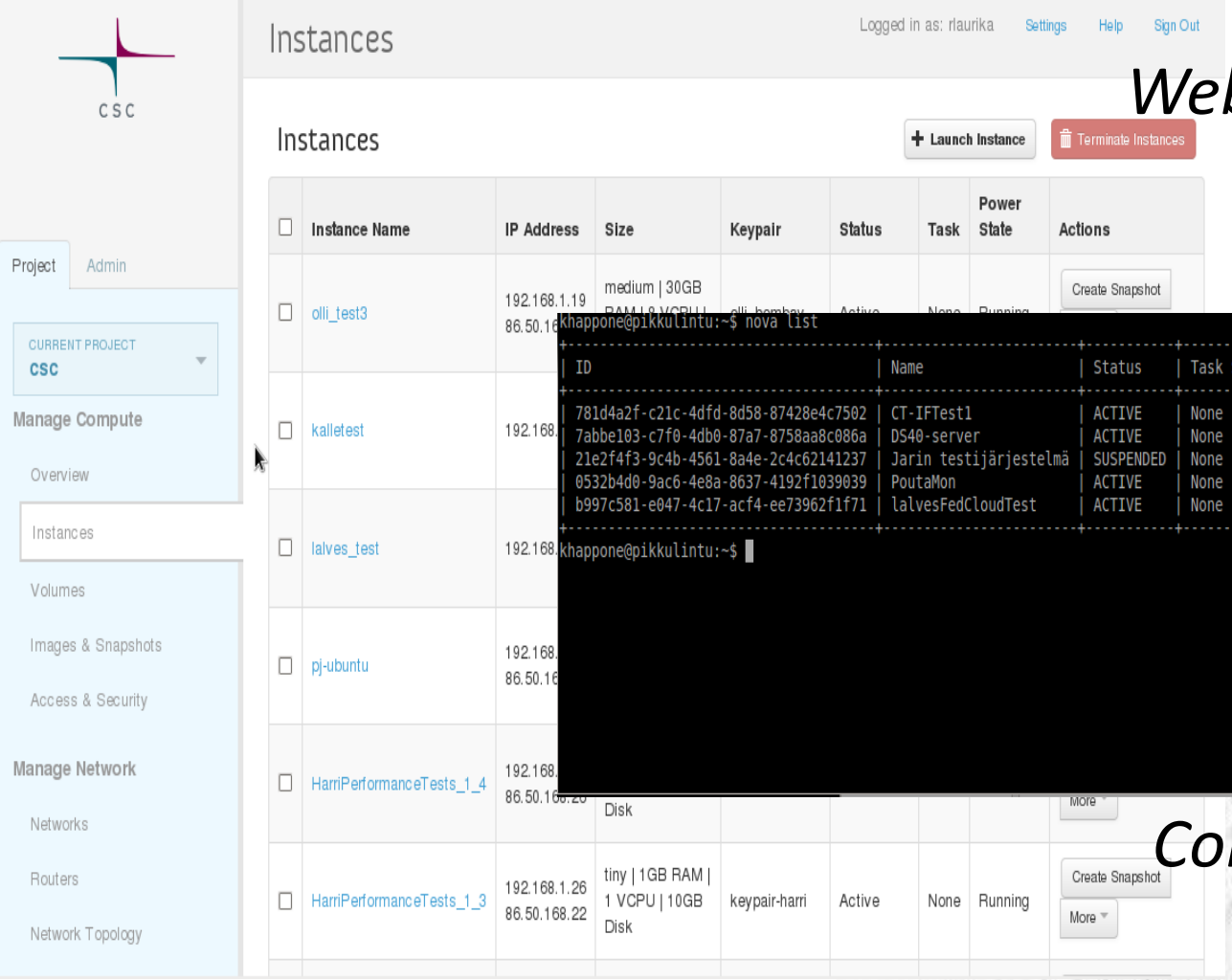
Host OS: RHEL

Virtual machine

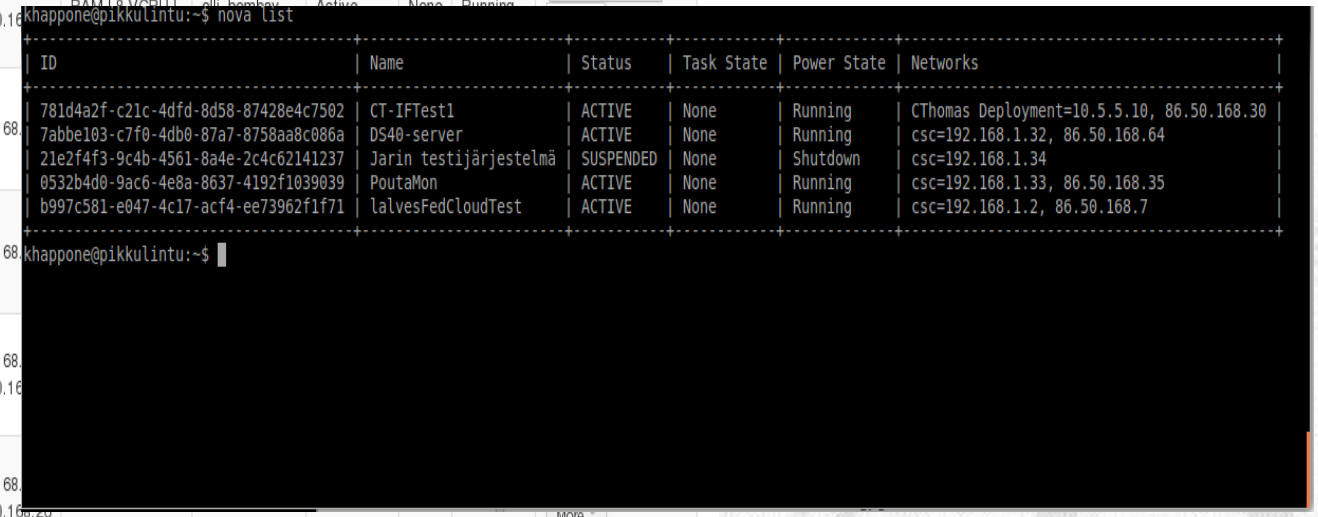
- Guest OS: Ubuntu

Virtual machine

- Guest OS: Windows



Web interface



```
khappone@pikkulintu:~$ nova list
+--------------------------------------+-----------------------+-----------+------------+-------------+-----------------------------------------+
| ID                                   | Name                  | Status    | Task State | Power State | Networks                                |
+--------------------------------------+-----------------------+-----------+------------+-------------+-----------------------------------------+
| 781d4a2f-c21c-4dfd-8d58-87428e4c7502 | CT-IFTest1            | ACTIVE    | None       | Running     | CThomas Deployment=10.5.5.10, 86.50.168.30 |
| 7abbe103-c7f0-4db0-87a7-8758aa8c086a | DS40-server           | ACTIVE    | None       | Running     | csc=192.168.1.32, 86.50.168.64          |
| 21e2f4f3-9c4b-4561-8a4e-2c4c62141237 | Jarin testijärjestelmä | SUSPENDED | None       | Shutdown    | csc=192.168.1.34                        |
| 0532b4d0-9ac6-4e8a-8637-4192f1039039 | PoutaMon              | ACTIVE    | None       | Running     | csc=192.168.1.33, 86.50.168.35          |
| b997c581-e047-4c17-acf4-ee73962f1f71 | lalvesFedCloudTest    | ACTIVE    | None       | Running     | csc=192.168.1.2, 86.50.168.7            |
+--------------------------------------+-----------------------+-----------+------------+-------------+-----------------------------------------+
khappone@pikkulintu:~$
```

Command line tools

```
https://pouta.csc.fi:8777/v2/csc/servers/0532b4d0-9ac6-4e8a-8637-4192f1039039
https://pouta.csc.fi:8777/v2/csc/flavors/1a0f1143-47b5-4e8a-abda-eba52ae3c5b9
https://pouta.csc.fi:8777/v2/csc/images/
```

REST API

Cloud service development in 2014

- **Pouta** (virtualisation) = CSC cloud service

  https://research.csc.fi/cloud-computing

  - In Production ”Amazon-type” Pouta for Research Communities and Organisations
    - Anyone can apply access
  - In development: Enterprise i.e. Biomedinfa –type virtual hosting to collaborate on organisational ICT capacities.
  - Development focus: Security features to support biobank (secure) data handling

- **Data** replication of key EMBL-EBI datasets to CSC
  - E.g. Computational access to local Ensembl from virtual machnies

cPouta's use cases

- Enhanced security – isolated virtual machines
- Advanced users – able to manage servers
- Difficult workflows – can't run on Taito
- Complex software stacks
- Ready made virtual machine images
- Deploying tools with web interfaces
- ”We need root access”

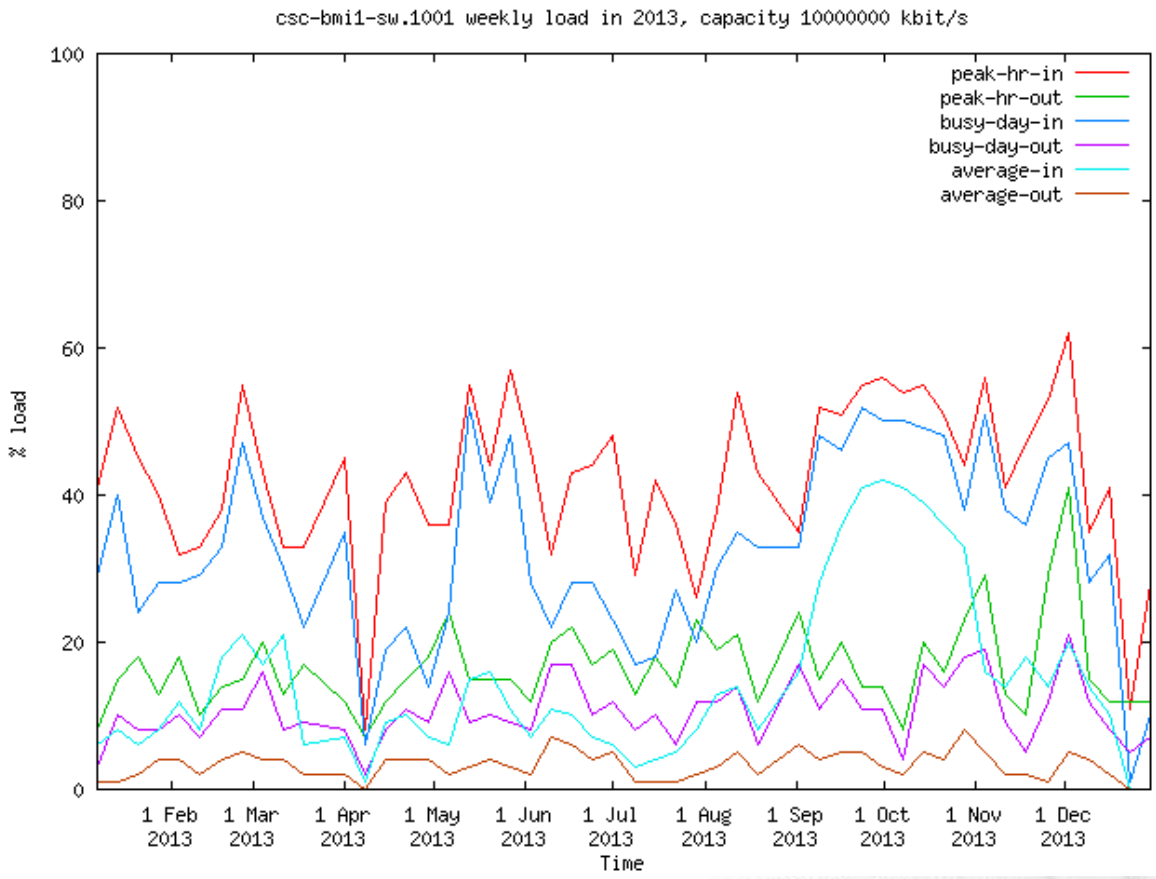
If you can run on Taito – run on Taito
If not – Pouta might be for you

- Pouta user guide: *https://research.csc.fi/pouta-user-guide*

ePouta

- Renewing the cloud cluster equipment in Espoo in 2015
  - Changes to OpenStack cloud middleware (autumn 2014)
  - Focus on secure computing and service for organisations
  - Idea: seamless scaling of local resources using a trusted compute center (in Finland)
  - Requires local IT admin contact
  - Funding model and resource allocation policy is still under debate, supported by ELIXIR Finland

CSC – Meilahti genomics IaaS data traffic 2013



5.8 PB in

1.4 PB out

Avg. 221 MB/s 24 hours a day all year round

Summary of cloud resource @ CSC

- Pouta cloud services in production

  https://research.csc.fi/cloud-computing

  - Support researchers (”Communities”) and of organisations (”Enterprise”) to integrate to virtualised capacities on the CSC cloud platform
  - Cloud NFS net storage to 1.1 PB
  - Active replication of key biological datasets
  - Further information: contact@csc.fi

Grid computing with Finnish Grid Infrastructure (FGI)

ARC Grid Monitor

2014-05-27 CEST 12:45:37

Processes: Grid Local

Country	Site	CPUs	Load (processes: Grid+local)	Queueing
	Aesyle (FGI)	72	0+35	0+0
	Alcyone (CMS)	892	156+312	1040+0
	Alcyone (FGI)	892	6+461	19+0
	Asterope (FGI)	192	84+0	10+1
	Celaeno (FGI)	448	172+0	9+0
	Electra (FGI)	672	0+478	0+0
Finland	Jade (HIP)	768	227+541	25+49
	Maia (FGI)	768	360+408	14+0
	Merope (FGI)	1612	0+1319	14+0
	Pleione (FGI)	288	144+0	13+0
	Taygeta (FGI)	360	42+174	15+0
	Triton (FGI)	6972	182+0	2+0
	Usva (CSC/FGI/test)	144	12+0	0+0
TOTAL	*13 sites*	*14080*	*1385 + 3728*	*1161 + 50*

ALL

FGI

- In grid computing you can use several computing clusters to run your jobs
- Grids suits well for array job like tasks where you need to run a large amount of independent sub-jobs
- You can also use FGI to bring cluster computing to your local desktop

- FGI: 12 computing clusters, about ***10 000*** computing cores
- Software: Run Time Environment include applications from all fields, e.g., bioinformatics, chemistry, physics:
  - *https://confluence.csc.fi/display/fgi/Runtime+Environments*

Using grid

- The jobs are submitted using the ARC middleware (http://www.nordugrid.org/arc/)
  - Using ARC resembles submitting batch jobs in Taito or Sisu
- ARC is installed in Hippu and Taito, but you can install it to your local machine too.
  - Setup command in Hippu:
    - module load nordugrid-arc
  - Basic ARC commands:
    - **arcproxy** (Set up grid proxy certificate for 12 h)
    - **arcsub *job.xrsl*** (Submit job described in file job.xrsl)
    - **arcstat -a** (Show the status of all grid jobs)
    - **arcget job_id** (Retrieve the results of a finished grid job)
    - **arckill job_id** (kill the given grid job)
    - **arcclean -a** (remove job related data from the grid)

Sample ARC job description file

```
&
(executable=runbwa.sh)
(jobname=bwa_1)
(stdout=std.out)
(stderr=std.err)
(gmlog=gridlog_1)
(walltime=24h)
(memory=8000)
(disk=4000)
(runtimeenvironment>="APPS/BIO/BWA_0.6.1")
(inputfiles=
( "query.fastq" "query.fastq" )
( "genome.fa" "genome.fa" )
)
(outputfiles=
  ( "output.sam" "output.sam" )
)
```

Getting started with FGI-Grid

1. Apply for a grid certificate from TERENA (a kind of grid passport)
2. Join the FGI VO (Access to the resources)
3. Install the certificate to Scientists' User Interface and Hippu.
4. Install ARC client to your local Mac or Linux machine for local use)
5. Instructions*: http://research.csc.fi/fgi-preparatory-steps*

Please ask help to get started: helpdesk@csc.fi

FGI user guide: *http://research.csc.fi/fgi-user-guide*

Courses

- ***Sisu*** Phase 2 workshop
  - *4.-6.11.2014*
- ***Taito*** Phase 2 workshop
  - *Spring 2015*
- **CSC courses**: *http://www.csc.fi/courses*
  - Introduction to Linux and Using CSC Environment Efficiently + Pouta training *20.-22.10.2014*
  - CSC HPC Summer School
  - Spring, Autumn, Winter Schools
  - Parallel Programming

Grand Challenges

- Normal GC *(call in half a year / year intervals)*
  - New CSC resources available for a year
  - No limit for number of cores
  - *Next call beginning of 2015*
- Remember also PRACE/DECI calls
  - CSC supports the technical aspects of the applications

CSC Phase2 resources' summary

- ***Sisu*** *supercomputer*
  - General availability since **9.9.2014**
- ***Taito*** *supercluster*
  - Installation planned in *Q4 2014*
  - Part of Taito used for *Pouta Cloud*
- ***Bull*** system
  - General availability since **1.10.2014**
  - *45 nodes* with *2 Intel Xeon Phi coprocessors* each
  - *38 nodes* with *2 NVIDIA Tesla K40 accelerators* each
- ***DDN*** HPC storage system
  - Totaling *4 PB of fast parallel storage*

Feedback form and Round robin

Feedback form

- *http://bit.ly/GUT14_JY*
  - *Direct link: https://www.webropolsurveys.com/S/B8235DFD6E515C75.par*

- *(link also on the seminar home page www.csc.fi →)*

Round robin

- *What are your needs for your research?*
  - How CSC can help?
  - Special libraries/tools?

- How much data you produce that needs processing?
- Courses/training?
- Queue length: 3 (Sisu) / 14 (Taito) days enough?
  - Codes that can't checkpoint?
- Is memory an issue for you?
  - 1.5 TB/nodes usage policy?
- Do you need to move a lot of files? (from where?)
- Interested in GPGPU/MICs? Which code?

- Vesa Apaja
- Seyed Mohammad Mehdi Tabatabaei
- Riku Tuovinen
- Iballa Burunat
- Heikki Mäntysaari
- Ilari Korhonen
- Esa Haapaniemi
- Sergey Chernov
- Tinkle Chugh
- Maryam Ghalibaf
- Jen Pickett
- Juho Kopra
- Piret Avila
- Jia Liu
- Miika Nurminen
- Qingyun Fan
- Yegor Chechurin
- Elina Kalenius
- Anbu Poosakkannu
- Niko Säkkinen
- Hannu Vuori
- Roshan Budhathoki
- Anu Mikkonen
- Hoda Noroozi
- Karoliina Honkala
- Pekka Koskinen
- Riaz Mondal
- Muhammad Asghar
- Michael Cochez
- Pekka Wartiainen
- Sanna Rauhamäki
- Sanna Niinivehmas
- Sakari Lätti
- Kari Aliranta
- Mira Kuusisto