EUROPEAN
SPALLATION
SOURCE

Research infrastructures as scientific cloud providers

ESS as a case study

PRESENTED BY JESPER RUDE SELKNÆS

2022-10-11

Agenda

1 Introduction to ESS

2 The Data Management and Software Center

3 Concrete technologies used at DMSC

4 Summing up

European Spallation Source

What it's all about

To build and operate the world's most powerful neutron source

enabling scientific breakthroughs
in research related to
materials, energy, health and the environment,
addressing some of the most important
societal challenges of our time

A coalition of 13 European countries

Host countries

Sweden, Denmark

Construction
47.5%

Operations
15%

Budget for construction €1.84 billion
Estimated annual budget €140 million

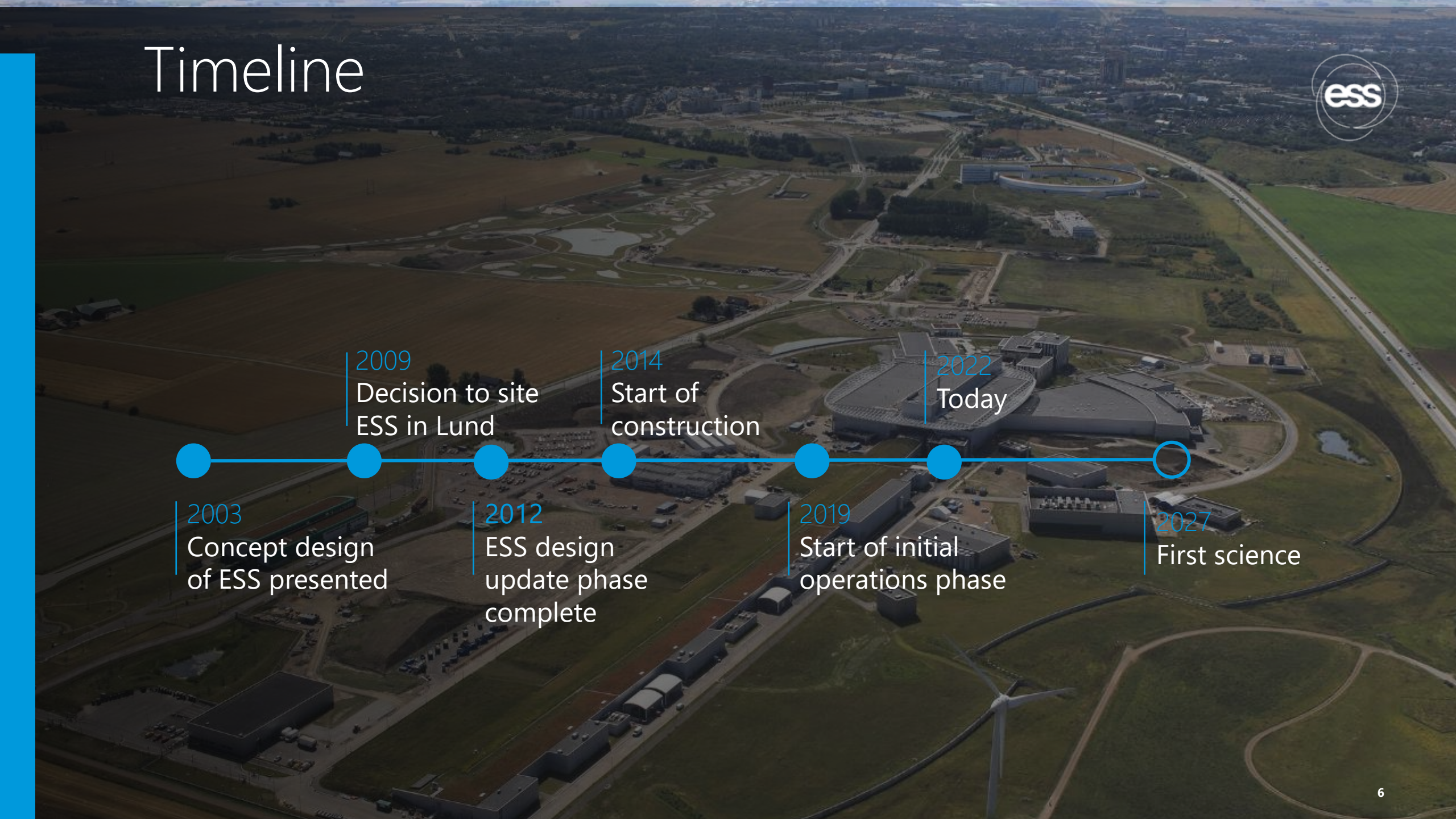
Non host member countries

Czech Republic, Estonia, France, Germany, Hungary, Italy, Norway, Poland, Spain, Switzerland, United Kingdom.

Construction
52.5%
(of which 70% is in-kind deliverables)

Operations
85%

Timeline


2003
Concept design of ESS presented
2009
Decision to site ESS in Lund
2012
ESS design update phase complete
2014
Start of construction
2019
Start of initial operations phase
2022
Today
2027
First science

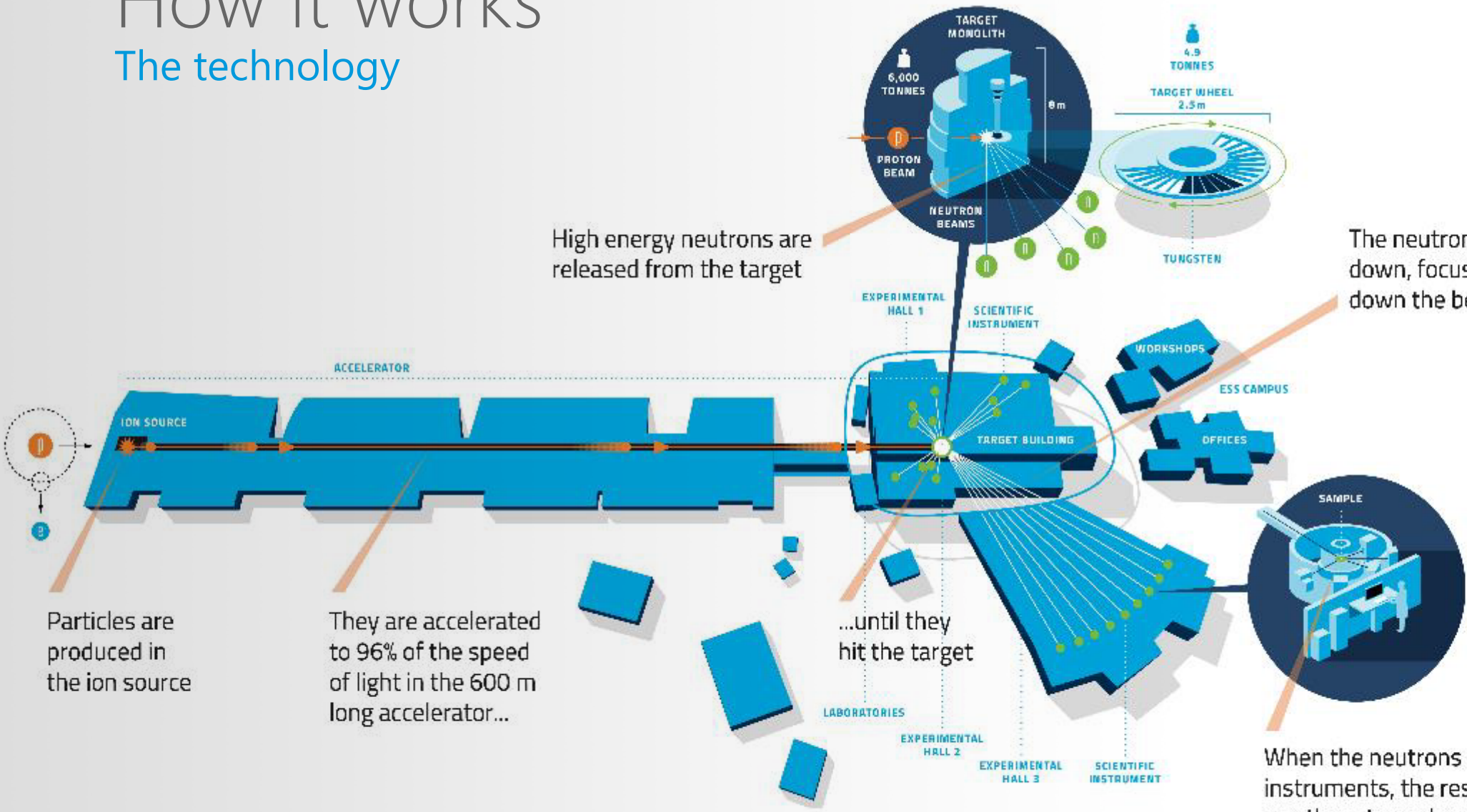
Work in progress

From green field to a state of the art RI

- Construction on the ESS site began in 2014
- Accelerator + target & first neutrons in 2023
- Instrument rollout 2024 – 2027 (15 instruments)

How it works

The technology



Neutron scattering simplified

Labs prepare the samples locally

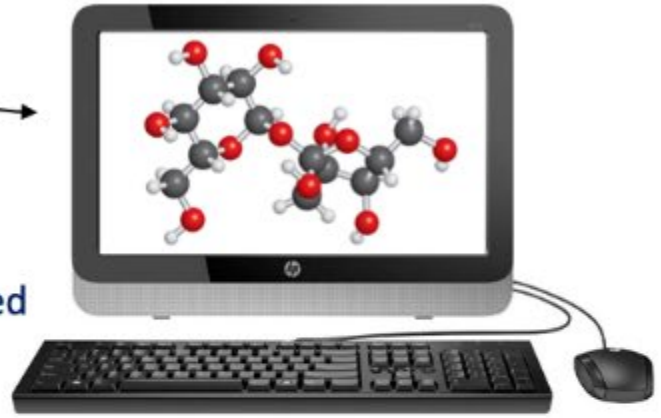
The neutrons are accelerated towards the sample

The neutrons hit the sample and the nuclei are 'scattered' off inside

The **pattern** of the scattered neutrons is detected

The **flight time** of the scattered neutrons is detected

Computational analysis translates the detected pattern into information on what's in the sample



Neutrons tell us:
Where the atoms are (pattern)
What they are doing (flight time)

Sample environment to vary temperature, apply magnetic fields, pressure, humidity, in-operando processes like operating battery, welding, forging, etc.

ESS will provide support in the form of data analysis packages, user support, data storage, etc.

Facts about ESS

5 MW particle accelerator

2 MW at start

15 instruments

next step is 22

3 000 guest scientists visiting yearly

to conduct experiments

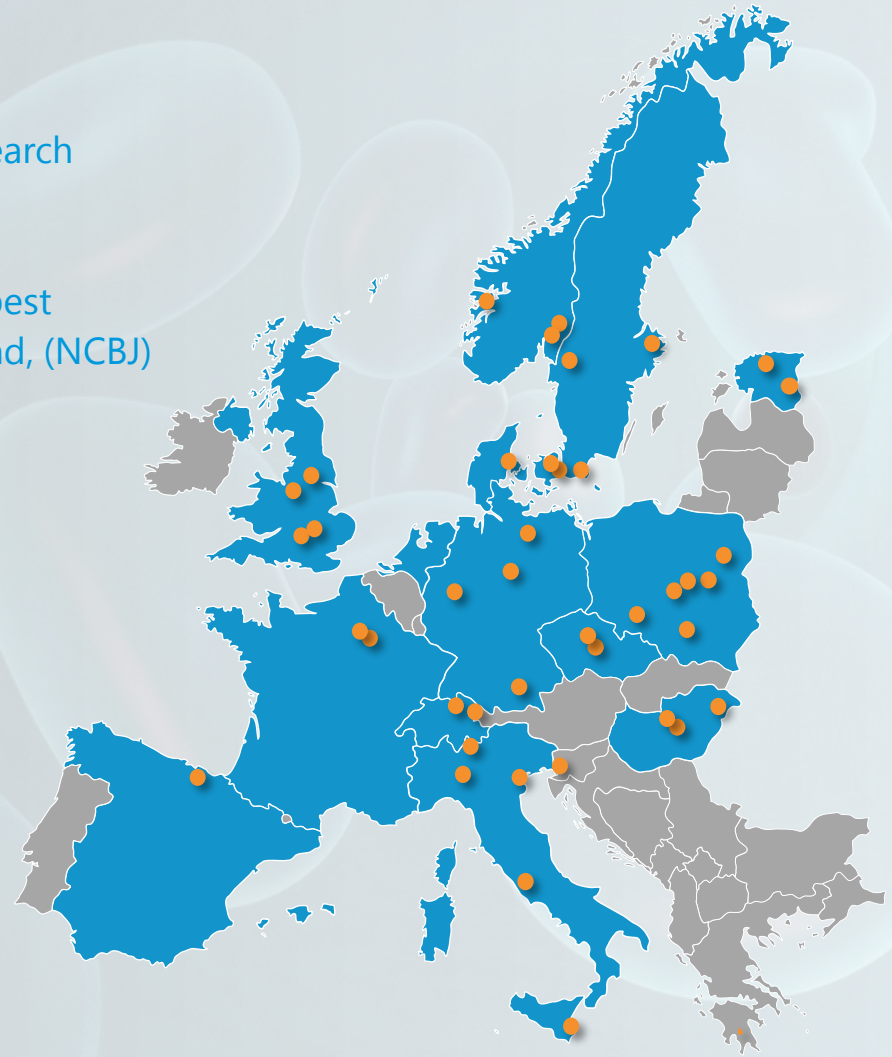
800 experiments per year

BREEAM Renewable energy & waste heat recovery

Unique international project

How will it be built?

Aarhus University
Atomki - Institute for Nuclear Research
Bergen University
CEA Saclay, Paris
Centre for Energy Research, Budapest
Centre for Nuclear Research, Poland, (NCBJ)
CNR, Rome
CNRS Orsay, Paris
Cockcroft Institute, Daresbury
Elettra – Sincrotrone Trieste
ESS Bilbao
Forschungszentrum Jülich
Helmholtz-Zentrum Geesthacht
Huddersfield University
IFJ PAN, Krakow
INFN, Catania
INFN, Legnaro
INFN, Milan
Institute for Energy Research (IFE)
Rutherford-Appleton Laboratory, Oxford(ISIS)
Copenhagen University
Laboratoire Léon Brillouin (CEA/CNRS/LLB)
Lund University
Nuclear Physics Institute of the ASCR
Oslo University
Paul Scherrer Institute (PSI)
Polish Electronic Group (PEG)
Roskilde University
Tallinn Technical University
Technical University of Denmark
Technical University Munich
Science and Technology Facilities Council
UKAEA Culham
University of Tartu
Uppsala University
WIGNER Research Centre for Physics
Wroclaw University of Technology
Warsaw University of Technology
Zurich University of Applied Sciences (ZHAW)



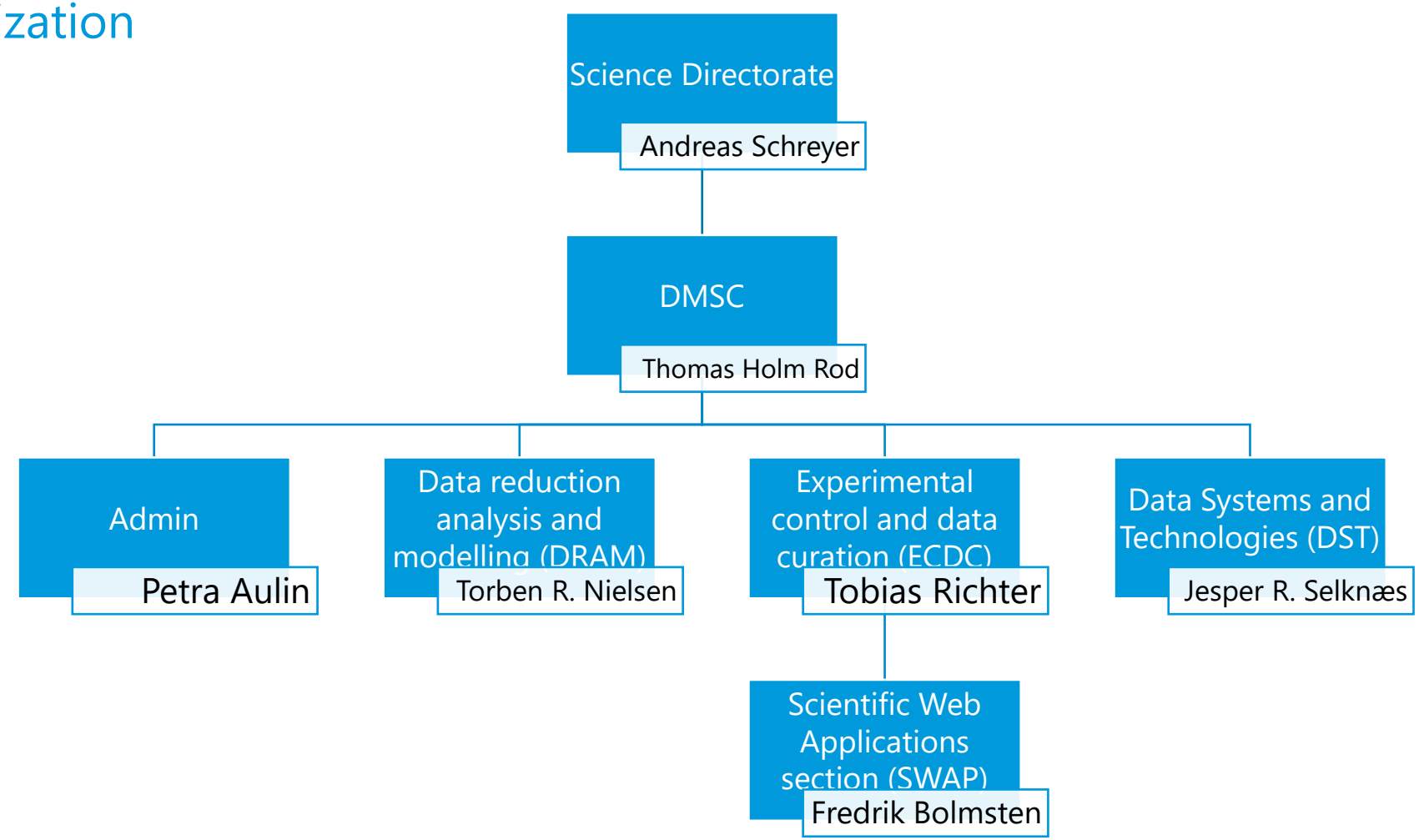
Data Management and Software Center

Enabling science

DMSC mission:
To use the techniques and methods of scientific computing
to facilitate, enable and advance the scientific research to be carried out using the neutron beam instruments at the
European Spallation Source.

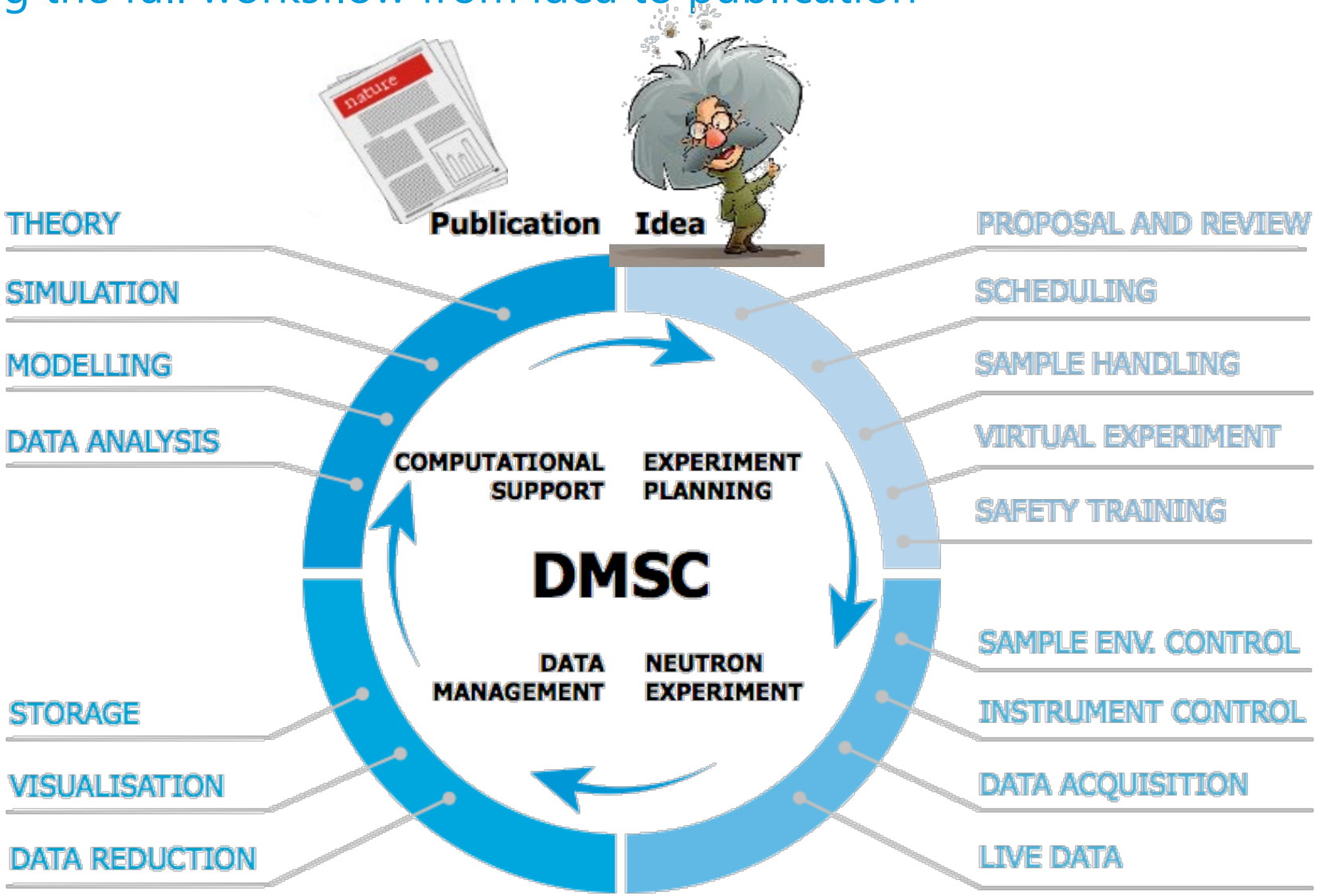
Data Management and Software Center

Organization



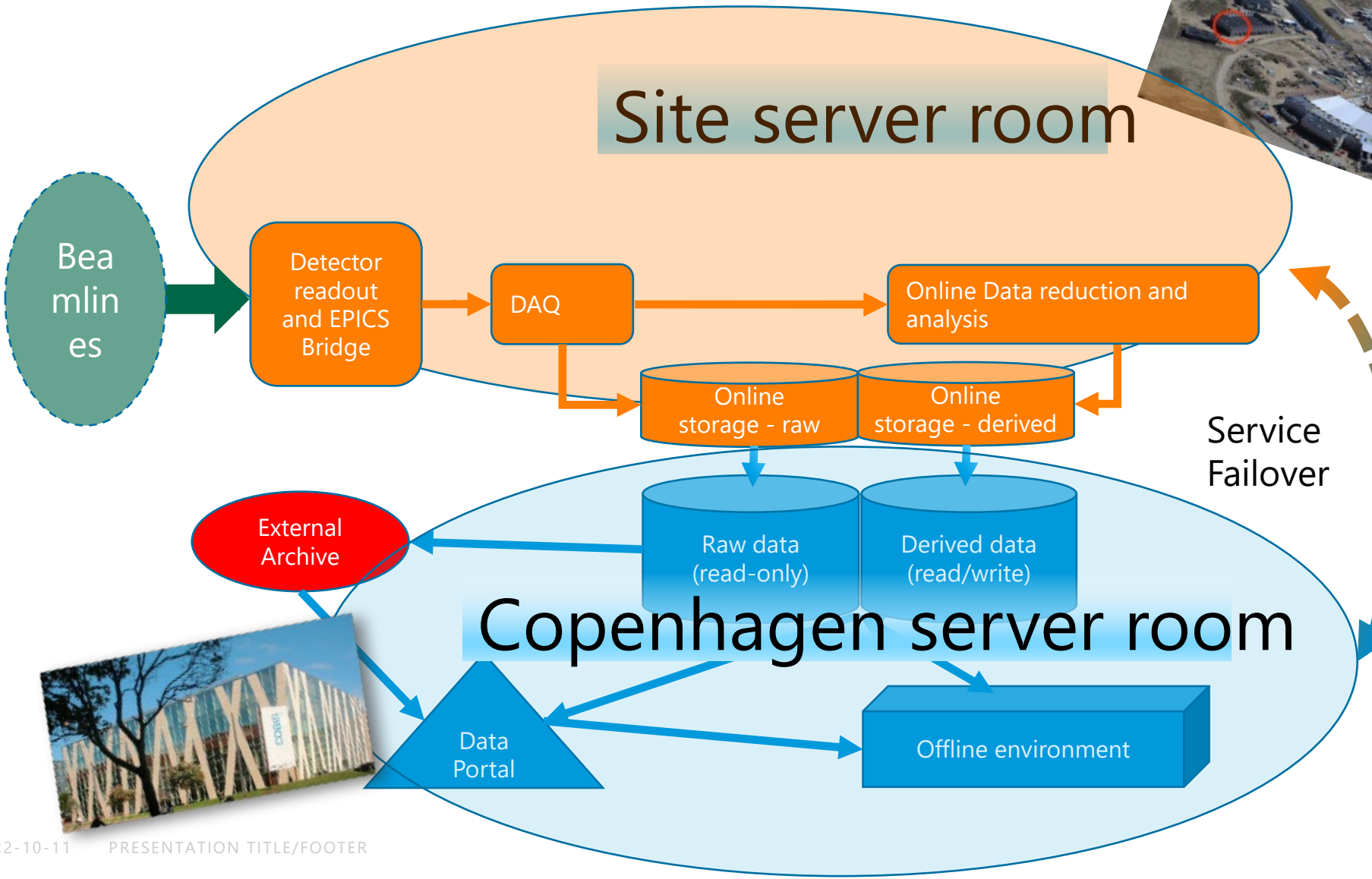
Data Management and Software Center

Supporting the full worksflow from idea to publication



The ESS data pipeline

From detectors to publication



ESS as a scientific cloud provider

Bringing the results to the end-user

Boundary conditions for ESS

- ESS adheres to the FAIR principles for scientific data
- ESS must ensure a good and smooth user experience for users conducting experiments.
- Data sets at ESS ranges from gigabytes to double digit terabytes per experiment plus derived data!

Ambition for how to work with data at ESS:

- Users are allowed to download data and use it in accordance with the ESS DMP but…
- We would rather that users find DMSC service the better alternative and that DMSC services will always be the preferred choice.
- Users will thereby work in ”the cloud” and ESS can leverage resources and combined data sets from other facilities in the *OSC’s
- Making ESS a cloud provider in the Open Science Clouds.

Concrete technologies used at DMSC

Bringing the results to the end-user

- **ESS is part of the PaNOSC project (Photon and Neutron Open Science Cloud) collaborating on common solutions for:**
  - E-learning platform supporting courses and virtual experiments (pan-training.eu)
  - Data catalogue (SciCAT)
  - VISA (Virtual Infrastructure for Scientific (data)analysis) providing a front-end to users for data reduction and analysis. VISA is developed by ILL.
  - Integration with EOSC, including making ESS services part of the EOSC service catalogue, federated AII, long term data archival and data sharing.

- **Virtualization and containerization**
  - OpenStack, OpenShift
  - Kubernetes, Docker etc.
- **HPC and backend storage**
  - Classic HPC using SLURM plus Jupyter frontends.
  - SpectrumScale (GPFS), ZFS (in the longer term object storage is likely to be used along side block storage)

DMSC data center consists of

Hardware

The Datacenter consists of:

- A HPC cluster consisting of 95 nodes and 2656 cores+ 2x A100 GPU cards.
- Storage systems:
  - ZFS based home directories – 80TB total
  - ZFS based storage for project data – 320TB total.
  - Lustre for online scratch – 66TB fast storage. To be sunset in 2022.
  - Lenovo based GPFS SpectrumScale 2 x 850TB
  - A 120TB ZFS based backup system at I2 (DeIC) in Lyngby DK. To be sunset in 2022.
- A 3 host 1.1TB RAM virtualization system. Including 2 x 10TB SAS based storage systems
- 10 nodes (32 to 96 cores each) for virtualization being deployed on OpenStack in both sites.
- Mac OSX servers (1 MacPro server and 10 Mac Minis)
- 1GBE, 10GBE and 100GBE Ethernet
- QDR and HDR InfiniBand for cluster- and storage interconnect. QDR to be sunset.
- Total of 46 servers in Lund for event formation, KAFKA cluster and filewriters.
- 10GBE MPLS based site-to-site link connecting COBIS and H01

DMSC data center consists of

Software and services

DST is managing the following systems and services:

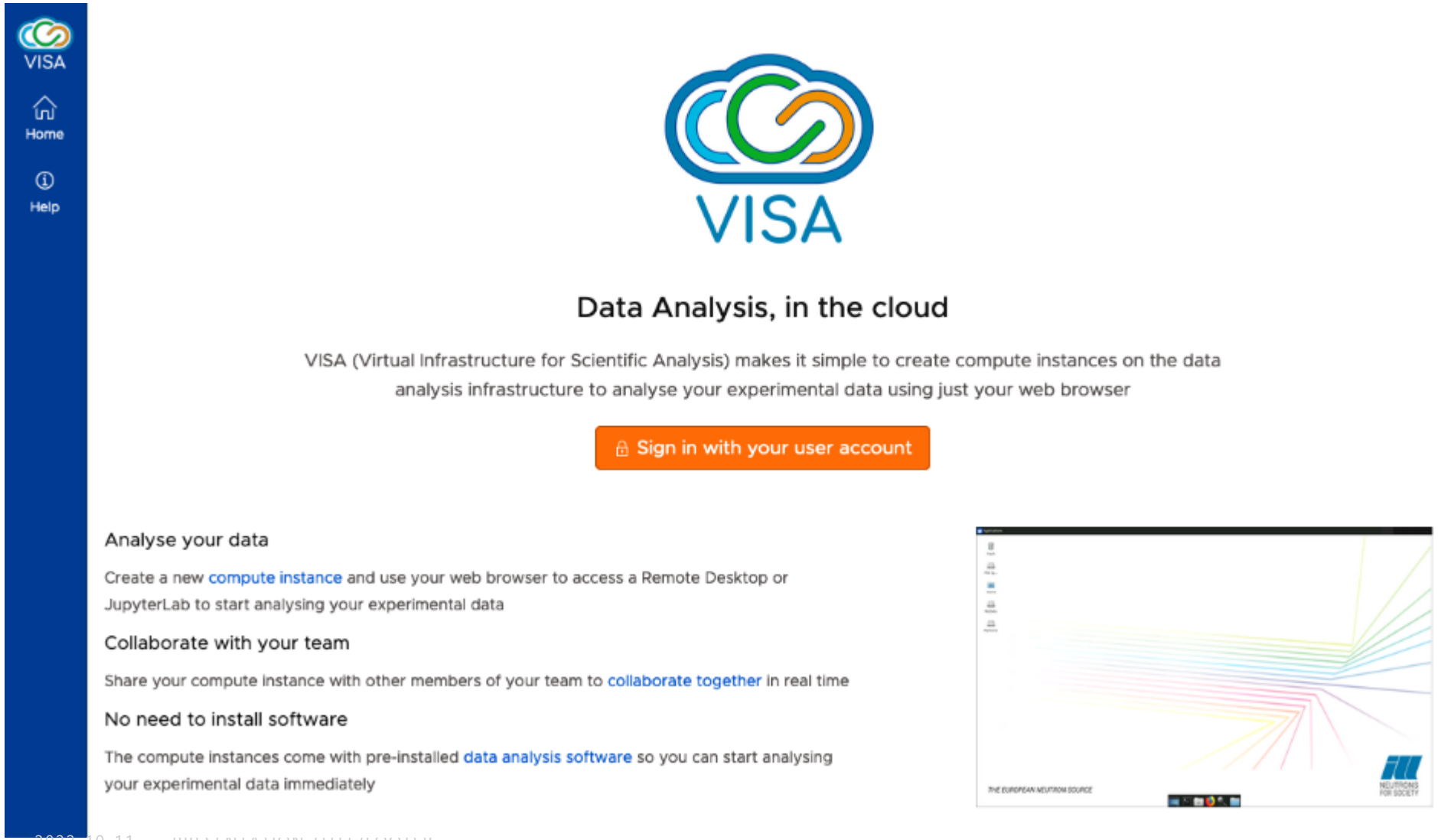
- Gitlab Community
- Jenkins CI/CD
- SLURM for cluster management
- Jupyter notebooks and Jupyter hub
- Mediawiki servers
- OwnCloud available for ESS as a whole – 17TB data
- Foreman/Puppet allowing dev. Teams to spin up machines for dev and test.
- VISA – current in a proof of concept setup
- Build servers for software projects
- Kubernetes

Infrastructure systems:

- OpenLDAP
- DNS
- OSSEC, Icinga, Graylog, GLPI for monitoring and asset management
- Yum and apt repository servers
- Etc.

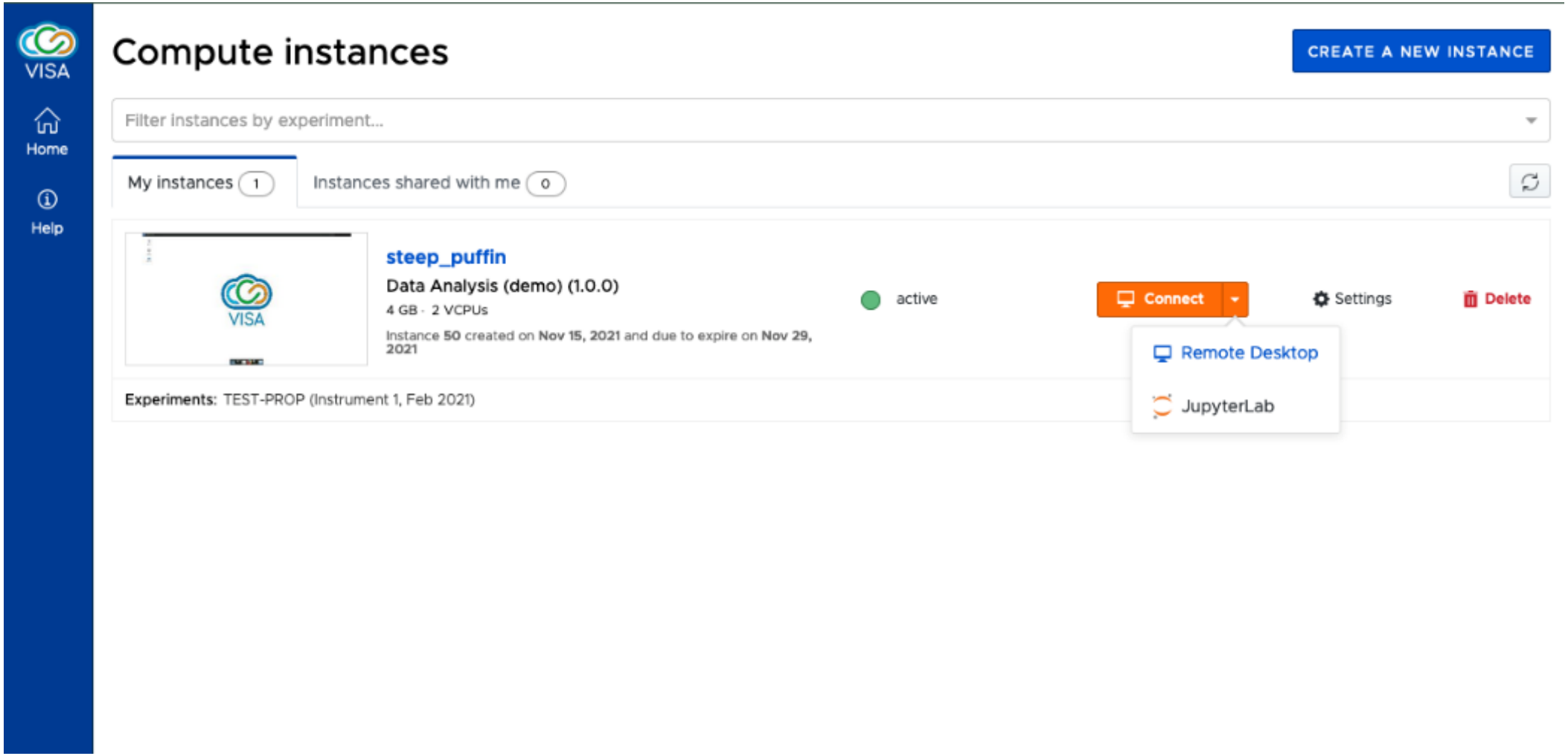
VISA

Landing page



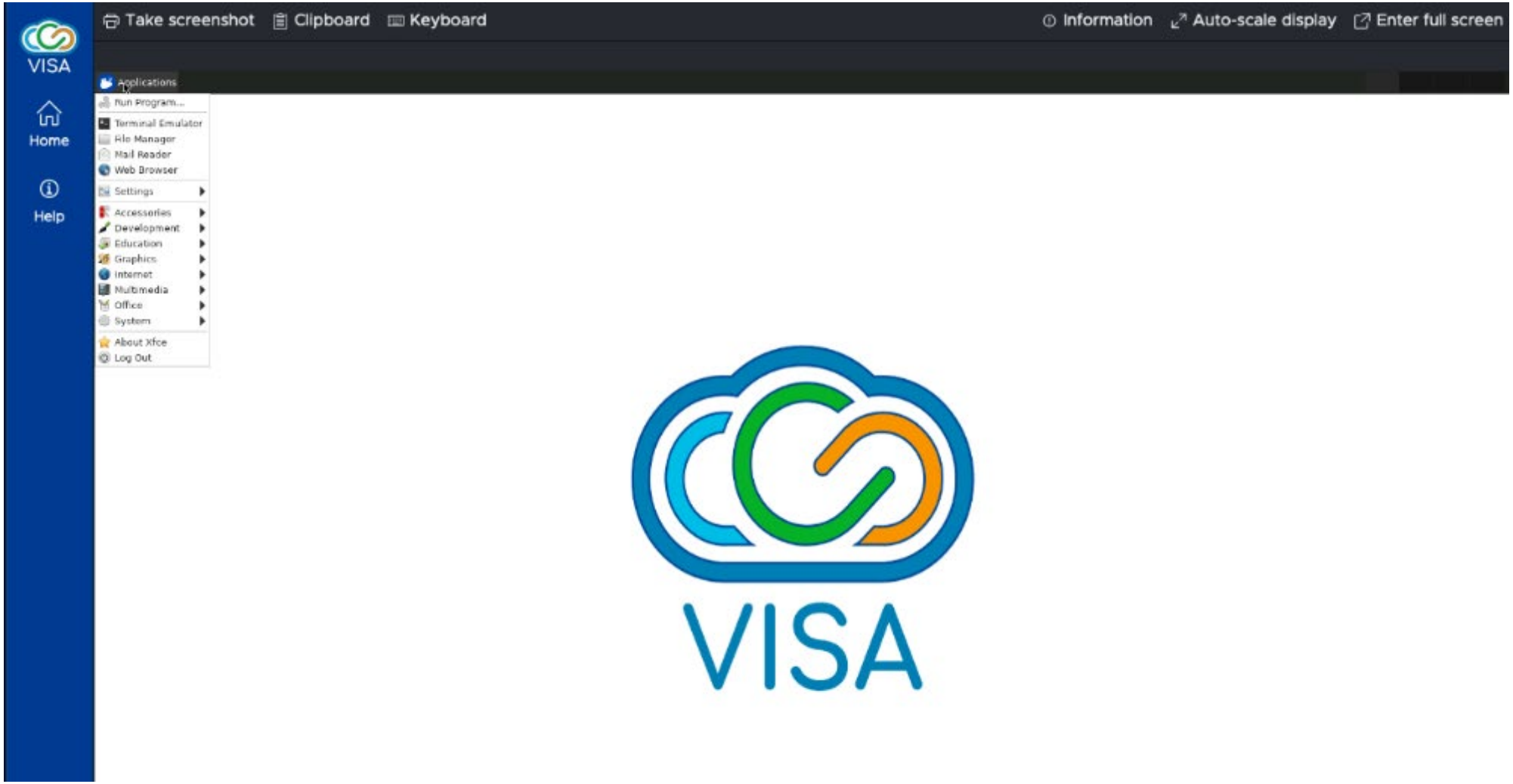
VISA

Users instances



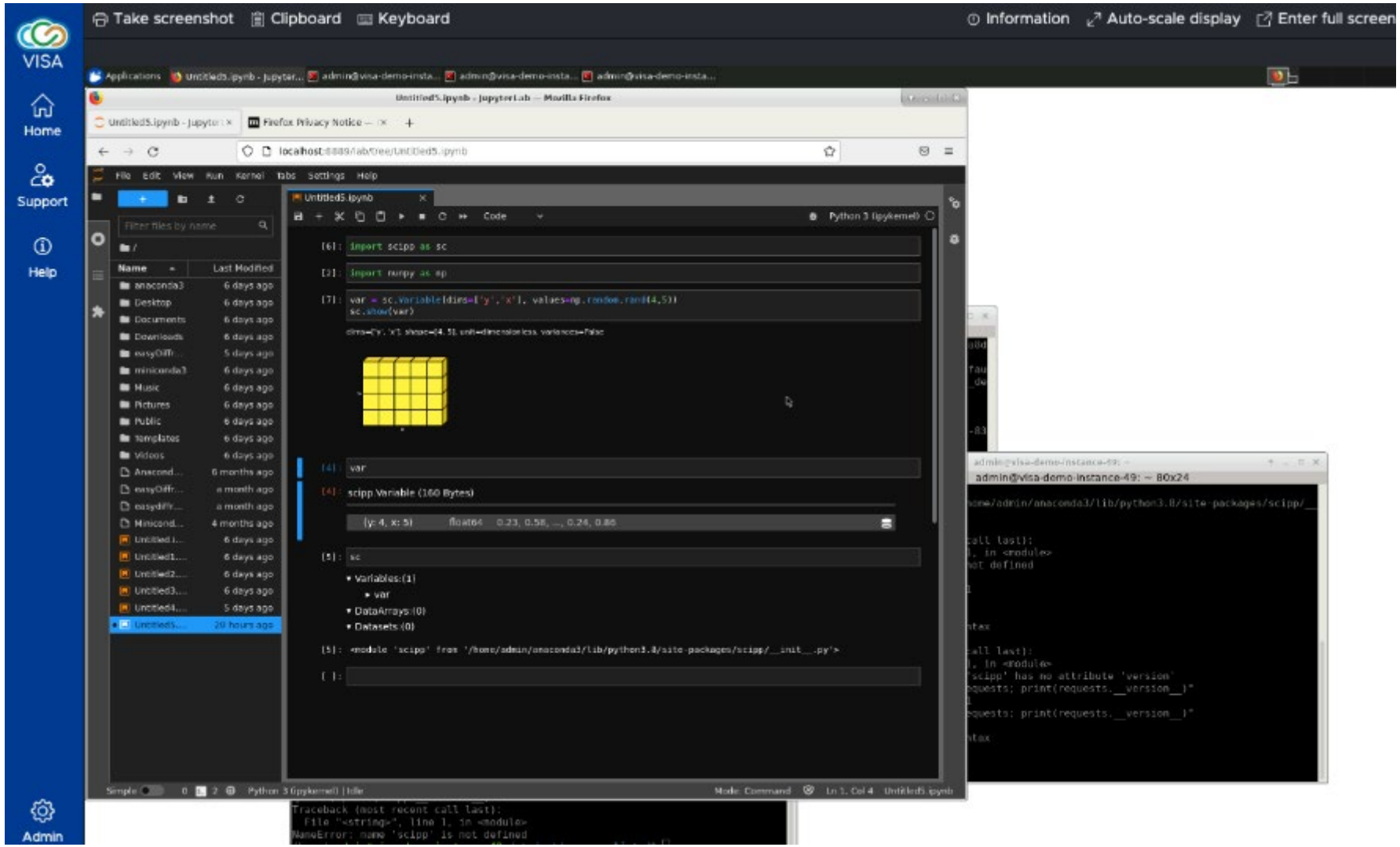
VISA

Remote desktop connection



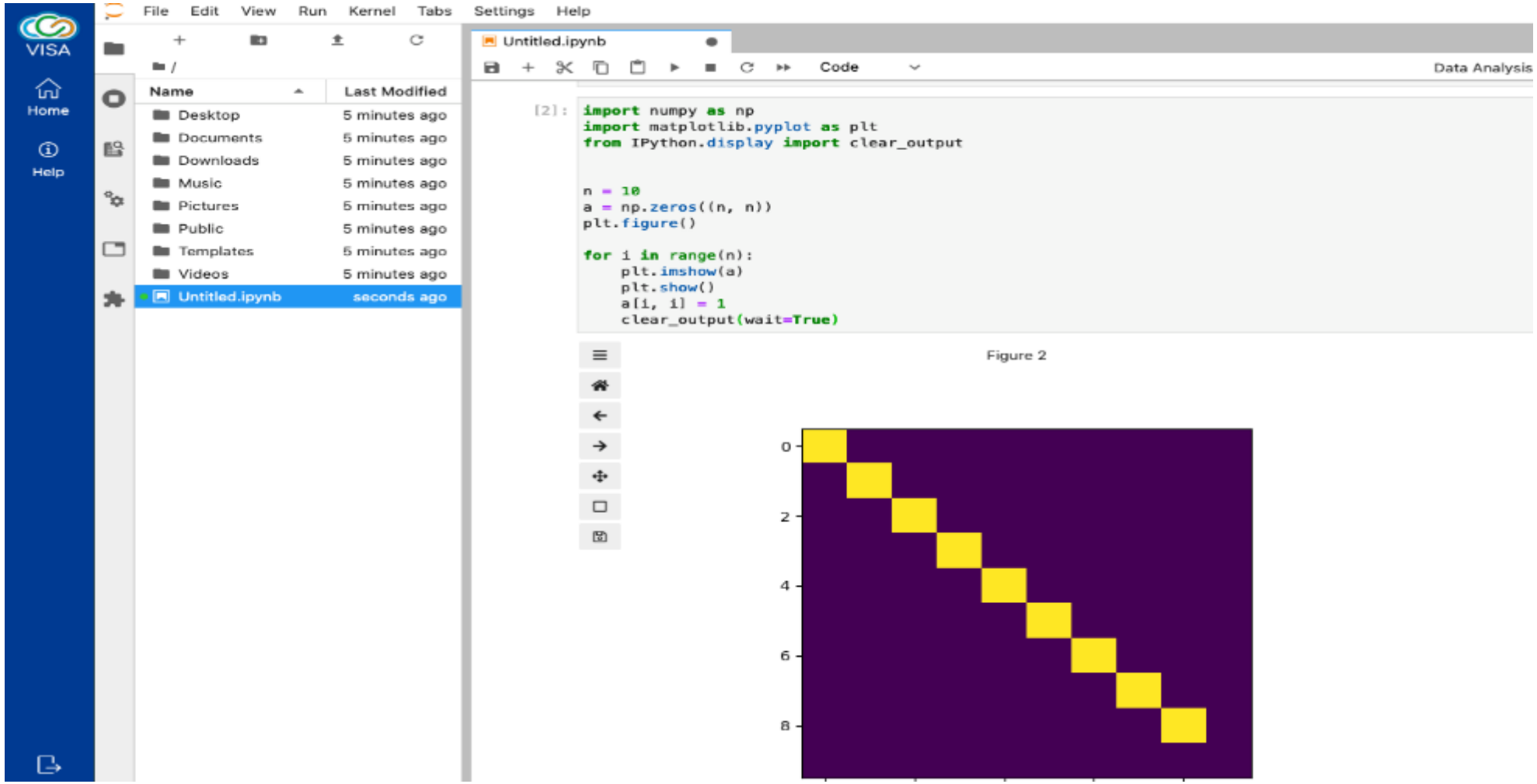
VISA

Remote desktop connection



VISA

JupyterHub connection



Summing up

ESS as a cloud service for neutron scattering research

- Data sets are too big and too expensive to be transported by portable storage devices.
- Data analysis requires considerable computational resources.
- ESS will strive to facilitate good and efficient data analysis by
  - Ensuring that the relevant software packages are developed and maintained.
  - That computational resources are available in a convenient but secure way.
  - That scientists does not have to care about the "how's" but only the "what's" – it should not be up to a scientist to find out how a computational run can be done – only which runs is relevant.
  - Working with ESS data should be doable from any location at any time.
  - The tax payers should get at much science out of an experiment as possible.
- **The purpose of DMSC is to ensure this – hence ESS is dedicated to provide the above.**

Finish presentation