

# Digitale forskningsdata i et nasjonalt perspektiv

UiB IT-forum

9. mai 2017

Gunnar Boe,  
Daglig leder

**UNINETT**

figma2



# Agenda

- Om UNINETT Sigma2
- Om forskningsdata
- Om oppgavene (nasjonalt vs lokalt)
- Om e-infrastrukturen

# About UNINETT Sigma2

- Established in December 2014 based on a decision from the 4 oldest universities and the Research Council of Norway
- A long-term model with 5+5 years and evaluation of the company after 5 years. (i.e. minimum 10 year lifetime for the company)
- Part of the UNINETT corporation, separate company
- Collaboration agreement with the 4 oldest universities incl. 50 MNOK yearly funding
- Contract with the Norwegian Research Council incl. 25 MNOK yearly funding
- Granted infrastructure funding (75.7 MNOK investment 2016-2017) from the Norwegian Research Council
- Operation and support contract with the 4 oldest universities
- Frame agreement with the universities for project work

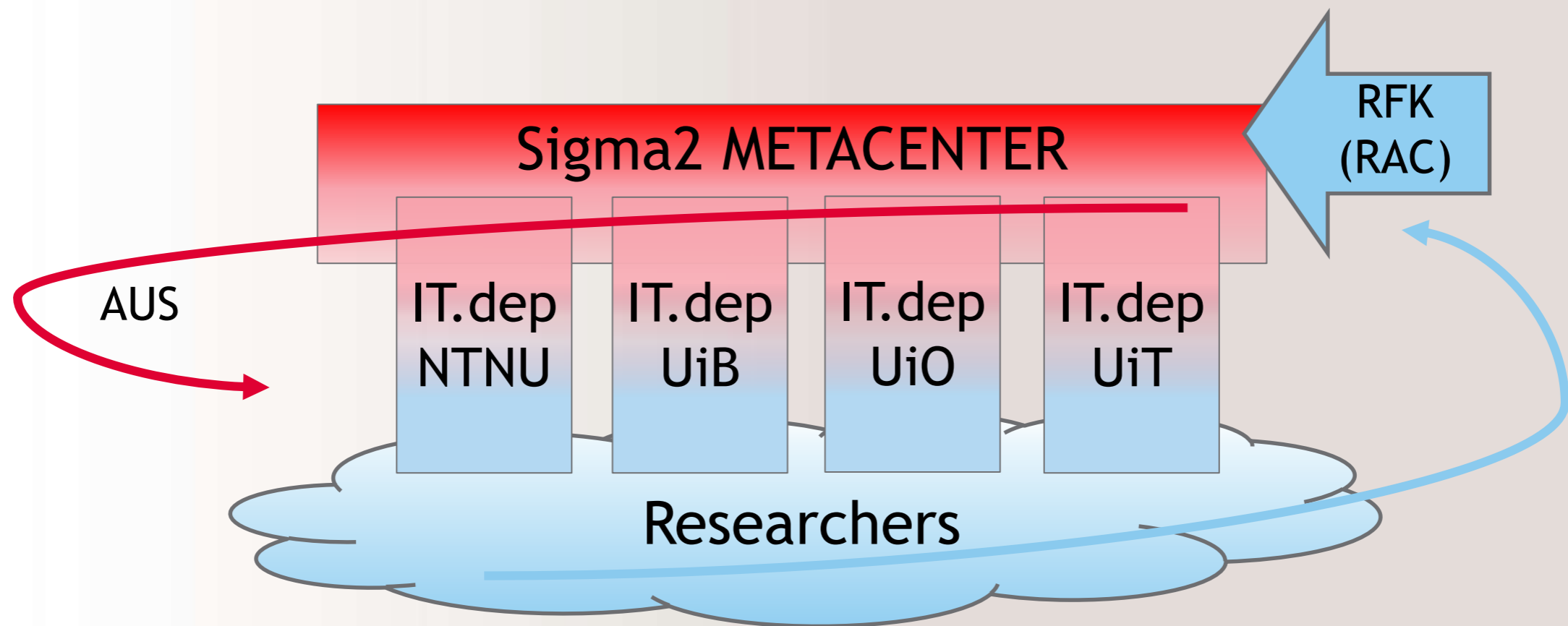
# «3. Det digitalt inviterende universitetet»



- «Tilrettelegge for å dele tjenester i sektoren, og ta et nasjonalt ansvar for digitale tjenester eller ressurser der det er ønskelig for å understøtte forskning og utdanning»

# The Metacenter

- National coordination and shared, consolidated resources have cost and efficiency advantages but creates a “distance” to the end-users (researchers)
- This is avoided by keeping the support staff and competence near where the research is going on, at the universities
- Combined with a data-centric architecture for the e-infrastructure, this model combines the advantages of the centralized model and the local model



# High level objectives

- Procure, operate and develop a critical national e-infrastructure
- Promote e-infrastructure to new research communities
- Lead and coordinate participation in international cooperation for e-infrastructure
- Provide an attractive and sustainable e-infrastructure for all research communities, with the following characteristics:
  - High reliability and availability
  - Cost effectiveness
  - Predictable access
  - Interoperability within the national e-infrastructure and between national and international infrastructures (e.g. PRACE, EUDAT)
- Provide services for data analytics of large datasets (Big Data)

# Research data

# Research Council Policy Objectives

- Improve quality in research through better opportunities to use previous work and combine data in new ways
- Transparency in research process and better opportunities to verify scientific results
- Increased collaboration and less duplication of research
- Increase innovation in business and public sector
- Efficiency improvement and better use of public funding

Forskningsrådet. Tilgjengeliggjøring av forskningsdata  
- Policy for Norges forskningsråd.  
Norges forskningsråd; 2014

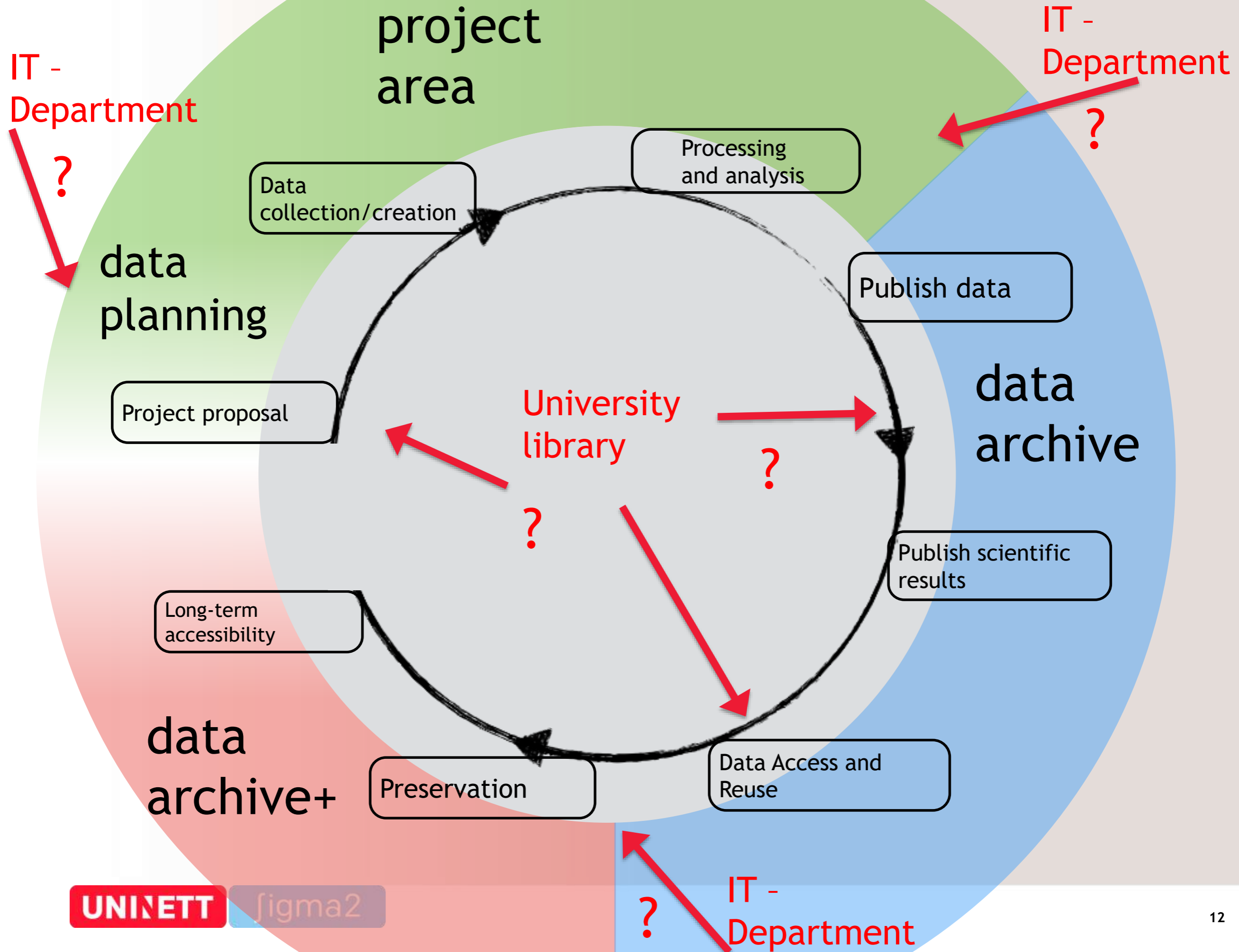




# «1. En gjennomgående digitalisert infrastruktur»



- «Bygge sammenhengende verdikjeder fra data-innsamling, via lagring, tilgjengeliggjøring og prosessering til publisering og etterprøving»



# The actors... who provides what

- International level
- National level
- University/institutional level
- Departments / Faculties
- Institute or research group



# National e-infrastructure level

- The global view, Interfacing with international services/e-infrastructures
- Generic services shared by many
- Economy of scale
- Providing services for publicly funded (RCN) research and enabling interaction between various stakeholders
- Competence

# University/institutional level

- Special local needs, Specific for the university
- Integration with local services
- Connect and promote data to higher level repositories
- Data curation best done locally?

# Services

# Sigma2 e-infrastructure services 1 / 2

## ➤ Computation

- Compute cycles for computational research

## ➤ Storage

- Data management planning
- Data storage, including Sensitive data
- (Visualization, Data-analytics)

## ➤ Basic user support

- Basic tech support through a ticket-based support service
- Training

## ➤ Advanced user support

# Advanced User Support (AUS)

- 1) Project based AUS:
  - can be the sole initiative of a researcher or a science area
  - granted by RFK with 2-3 PMs spent over a maximum of 6 months.
- 2) Discipline specific AUS
  - initiated by Sigma2 in cooperation with a science discipline
  - can have allocations of more than 12 PMs spent over a maximum for 2 years
  - joint funding



# Advanced User Support (AUS)

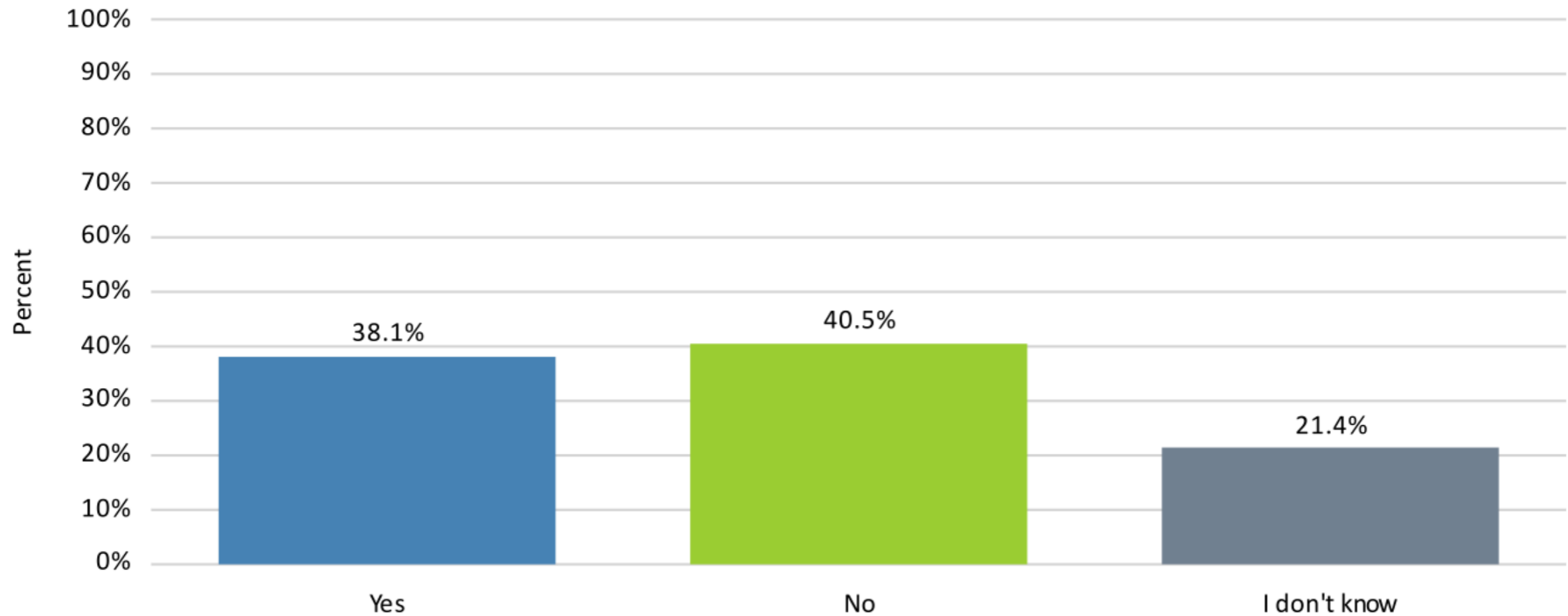
For the storage services, project based advanced user support aims at:

- assist researchers to create data plans
- implementing best practices for collecting and handling data
- identifying or defining meta-data schema
- identifying suitable storage formats
- identifying dedicated or specialised tools to help access or visualize data, utilise the facilities better

# Data management plans?

## 24. Does your project have a data management plan (DMP)?

NorStore project managers



# Data 'policy' for Research data

UNINETT

sigma2

DMP

Home

Your plans

Create a new plan

Log out

## Data Management Plan generator

### Data Management

The aim of this section is to help you to identify and take into account all the factors that can influence the management of your data during its lifetime. To create an effective data management plan you will need to identify where the data come from, where and how it will be stored, and what will be needed to access the data.

Q1.1 Q1.2 Q1.3 Q1.4 Q1.5 Q1.6 Q1.7 Q1.8 Q1.9 Q1.10 **Q1.11**

37% complete

Q: Who is responsible for managing the data?\*

- Don't Know
- Principal Investigator
- Data Manager

Notes

<http://sigma-dmp.paas.uninett.no>

UNINETT

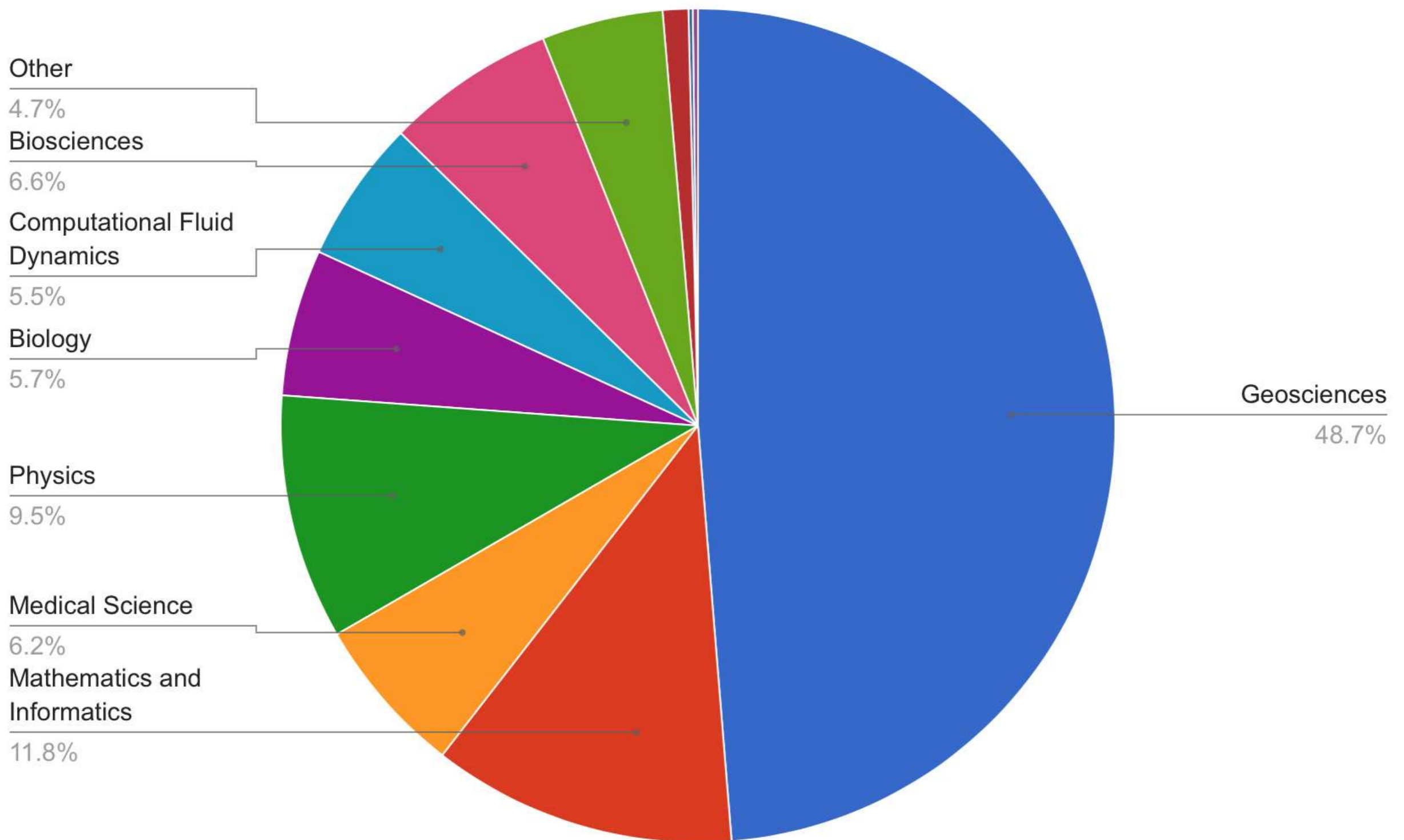
sigma2

# Sigma2 e-infrastructure services 2/2

## Specific services supported on NorStore resource

Service	Project/community	Reference	Contact
BioGateway	Biology	<a href="#">Semantic systems biology</a>	Martin Kuiper
NorMAP THREDDS	Climate, wind energy	<a href="http://normap.norstore.uio.no">normap.norstore.uio.no</a>	<a href="mailto:support@norstore.no">support@norstore.no</a>
StoreBioInfo Portal	Bio-informatics	<a href="http://storebioinfo.norstore.no">storebioinfo.norstore.no</a>	Kjell Petersen
Earth Systems Grid	Climate	<a href="#">ESG data node</a>	Mats Bentsen
ELMCIP	Humanities	<a href="#">ELMCIP Knowledge Base</a>	Scott Rettberg
LTR	Humanities	<a href="http://WEBDAV.ltr.norstore.uio.no">WEBDAV ltr.norstore.uio.no</a>	Stephan Oepen
z9	Mediacal imaging	<a href="http://d9.norstore.uio.no">d9.norstore.uio.no</a>	Jonas Ødegaard
UniKode	Climate	<a href="http://unikode.norstore.no">unikode.norstore.no</a>	Martin King

## Quota per discipline (disk+tape) for 2017



### Quota per institution (disk+tape) for 2017

UiT - Norges arktiske universitet

1.4%

Uni Research AS

1.9%

Universitetet i Oslo

18.5%

Norges Teknisk-Naturvitenskapelige Universitet

5.6%

Nansen Environmental and Remote Sensing Center

5.2%

Meteorologisk institutt

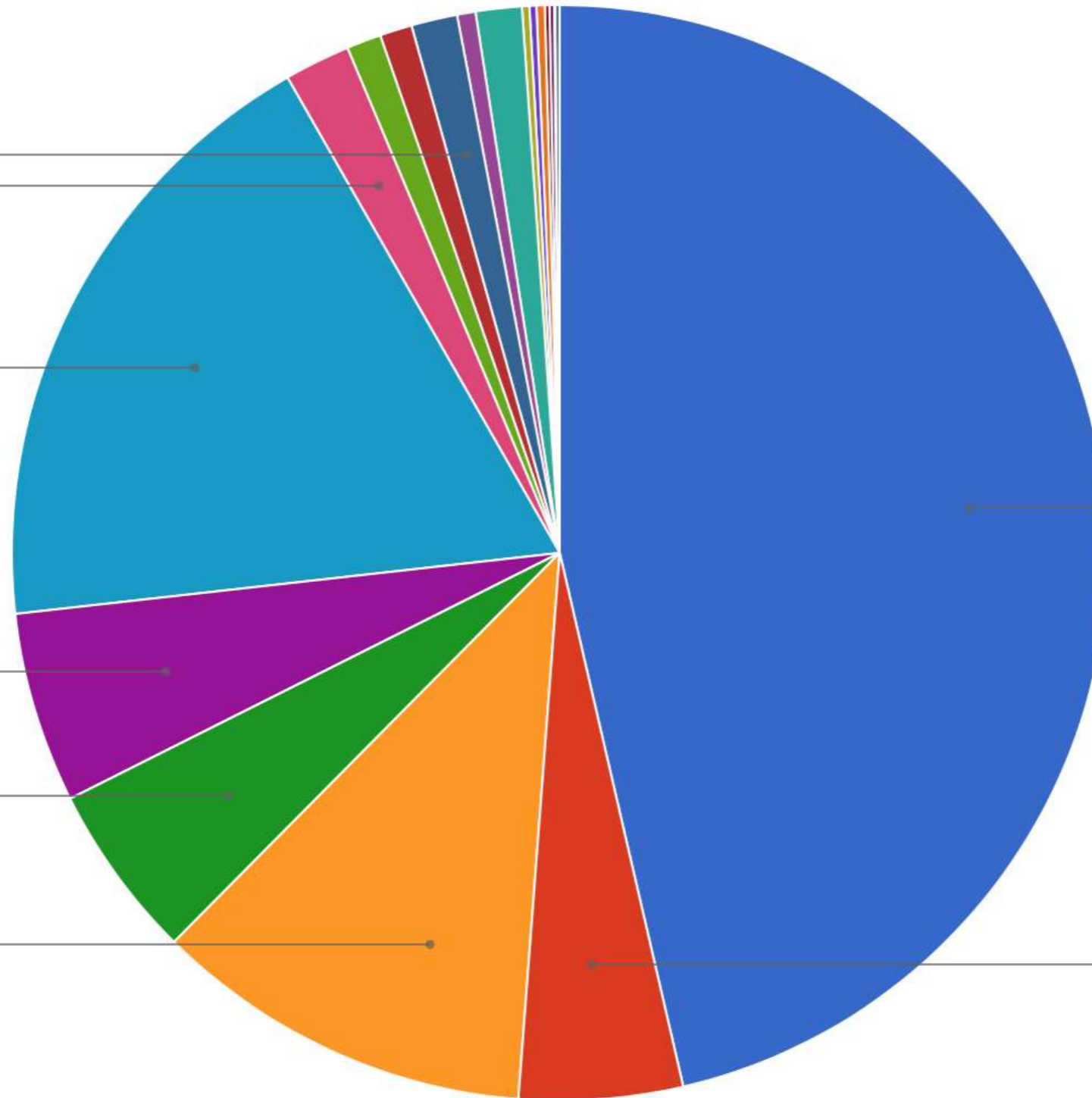
11.2%

Universitetet i Bergen

46.4%

Norsk Polarinstitut

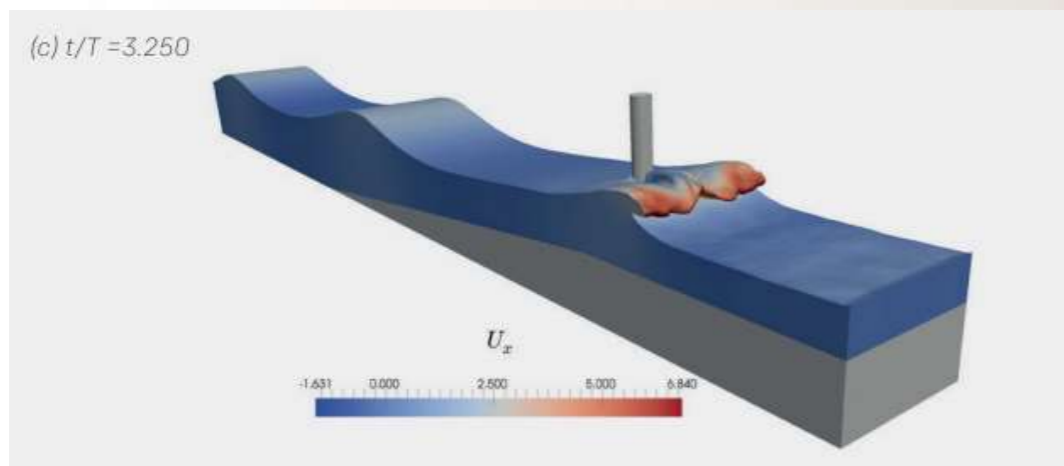
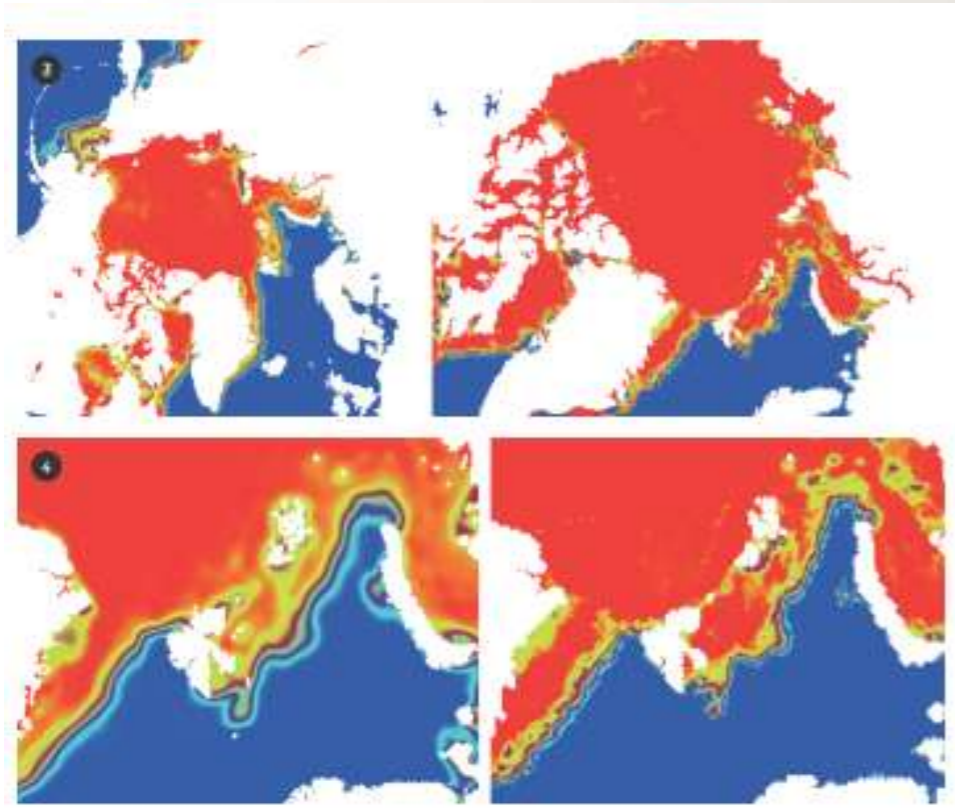
4.8%



# Data intensive Science Disciplines

- Climate (IPCC production, ESGC data node, HPC intensive data)
  - large datasets, avoid moving data, scalability, data longevity and integrity
- Neuroscience (HumanBrain, Kavli Inst., INCF)
  - sensitive data, raw sensor data, data mgmt tool
- ELIXIR.NO (next generation sequencing, analysis/processing, sharing/archiving, data product delivery)
  - portals, AAI, work flow mgmt, access to tools
- CLARINO (structured data, corpus)
  - AAI, data access, DOIs, centralising HPC+data
- Biodiversity (GBIF, LifeWatch)
  - portals, access/sharing, metadata, own PIDs, Biobanks)
- Marine environment (sensor collection, basic service needs) ...
- EPOS (implementation phase, sensor collection) ...

# Examples of projects



Pictures from  
META 1/2015



# The infrastructure: A new architecture

- National Infrastructure for Research Data (NIRD)

## «2. Brukerorienterte digitale tjenester»



- «Ta utgangspunkt i behov, brukervennlighet og brukeropplevelser i realisering av digitale tjenester»

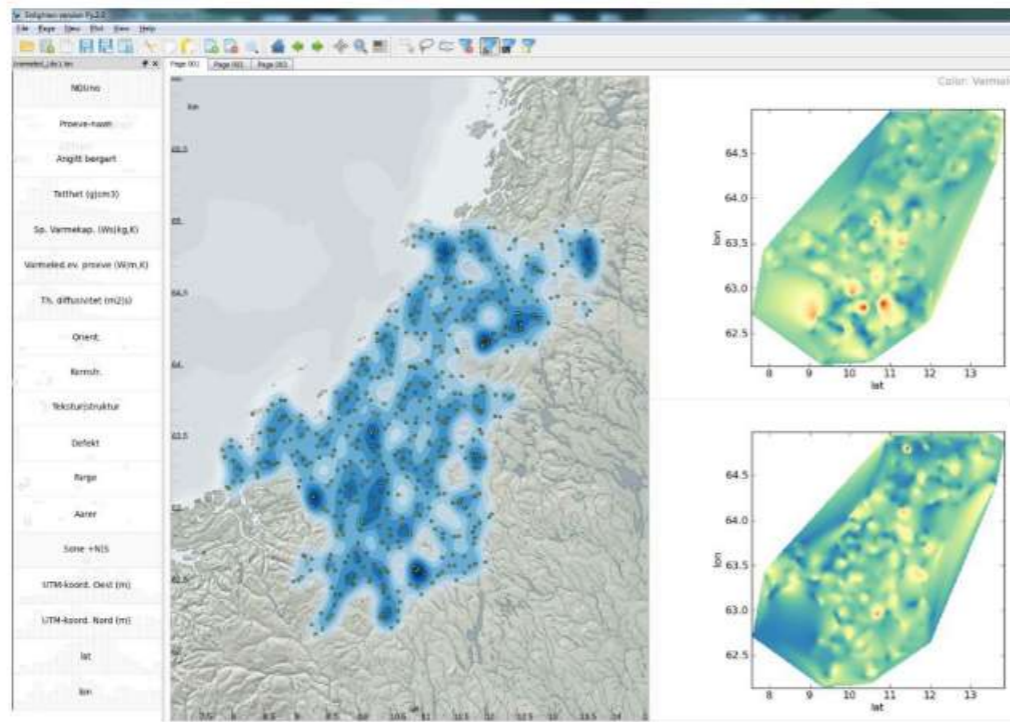
# What researchers requests:

## Software requirements

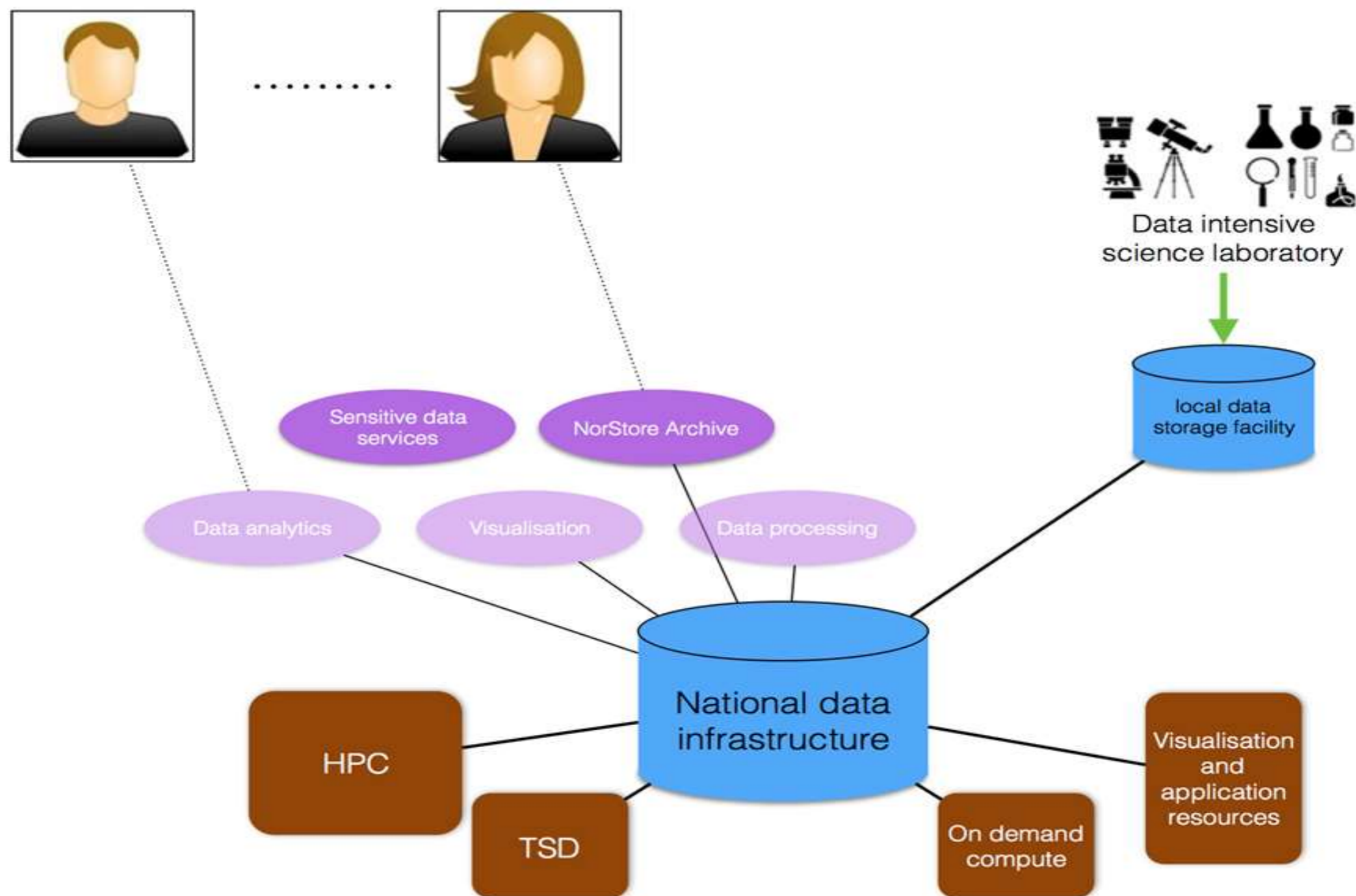
- Jupyter notebook
- Jupyterhub
- Python, scientific stack ~ Anaconda
- Docker
- Enlighten (server)
- Enlighten web (client)

## Processing and visualization software

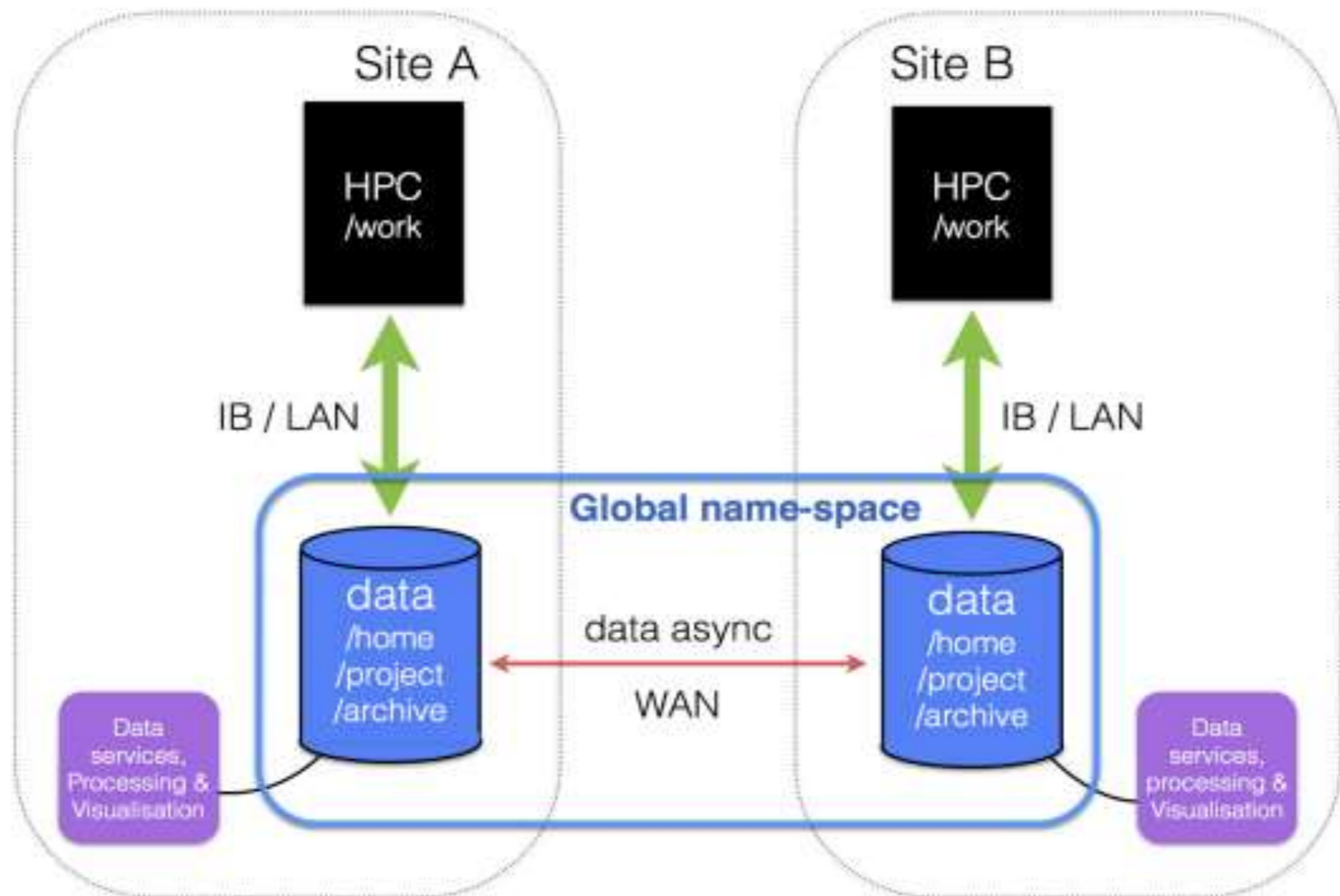
- (as cloud services or available for download)
  - Visualization – ICS-D functionality
- SEISAN, Earthquake analysis software (UiB)  
NORSAR 3D, 3D modelling tool



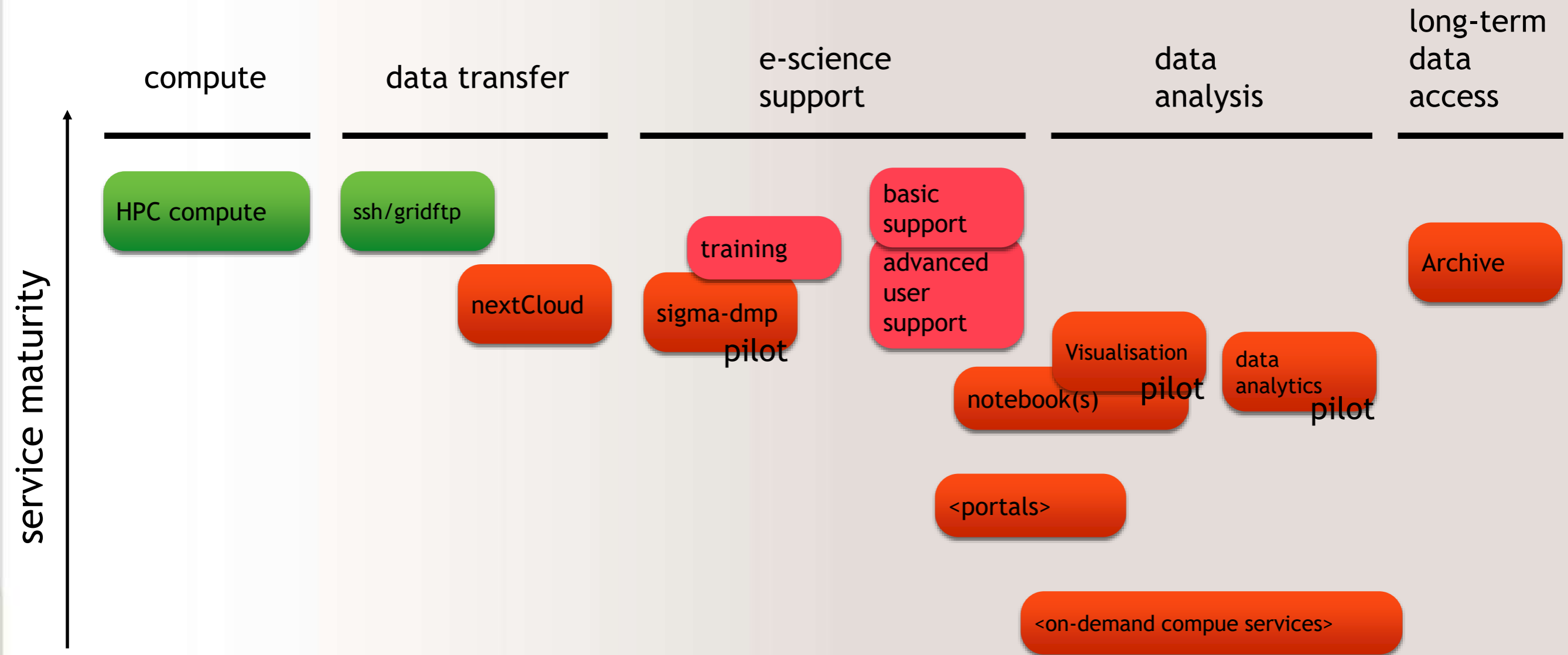
# Data-centric architecture



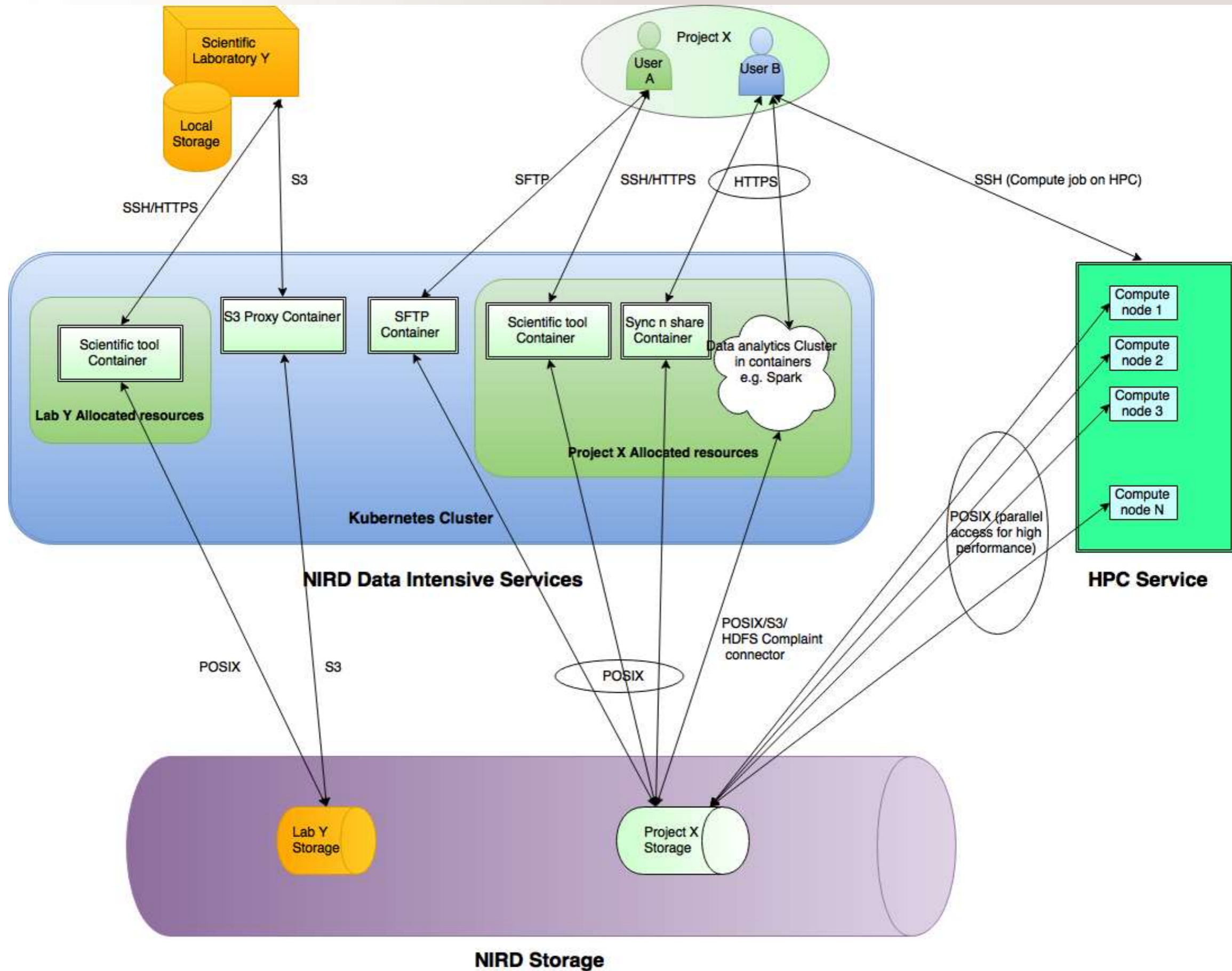
# Distributed data centre model



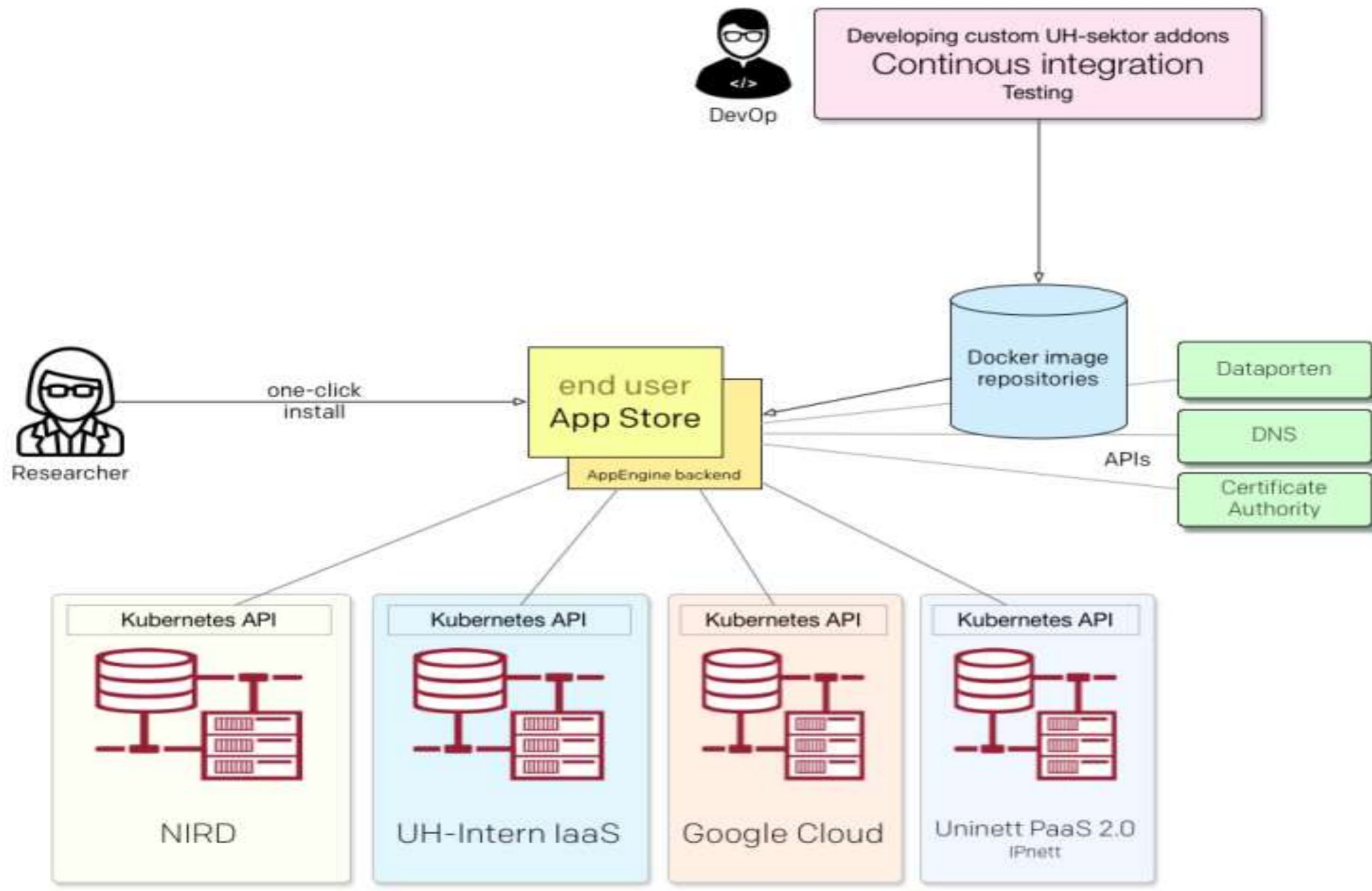
# Services



service maturity ↑



# A future common architecture?





[www.sigma2.no](http://www.sigma2.no)

