E-infrastructure for research in Norway

UiB HPC course 2018.1

UNINETT Sigma2

Hans A. Eide, Phd. UNINETT Sigma2 AS

Agenda

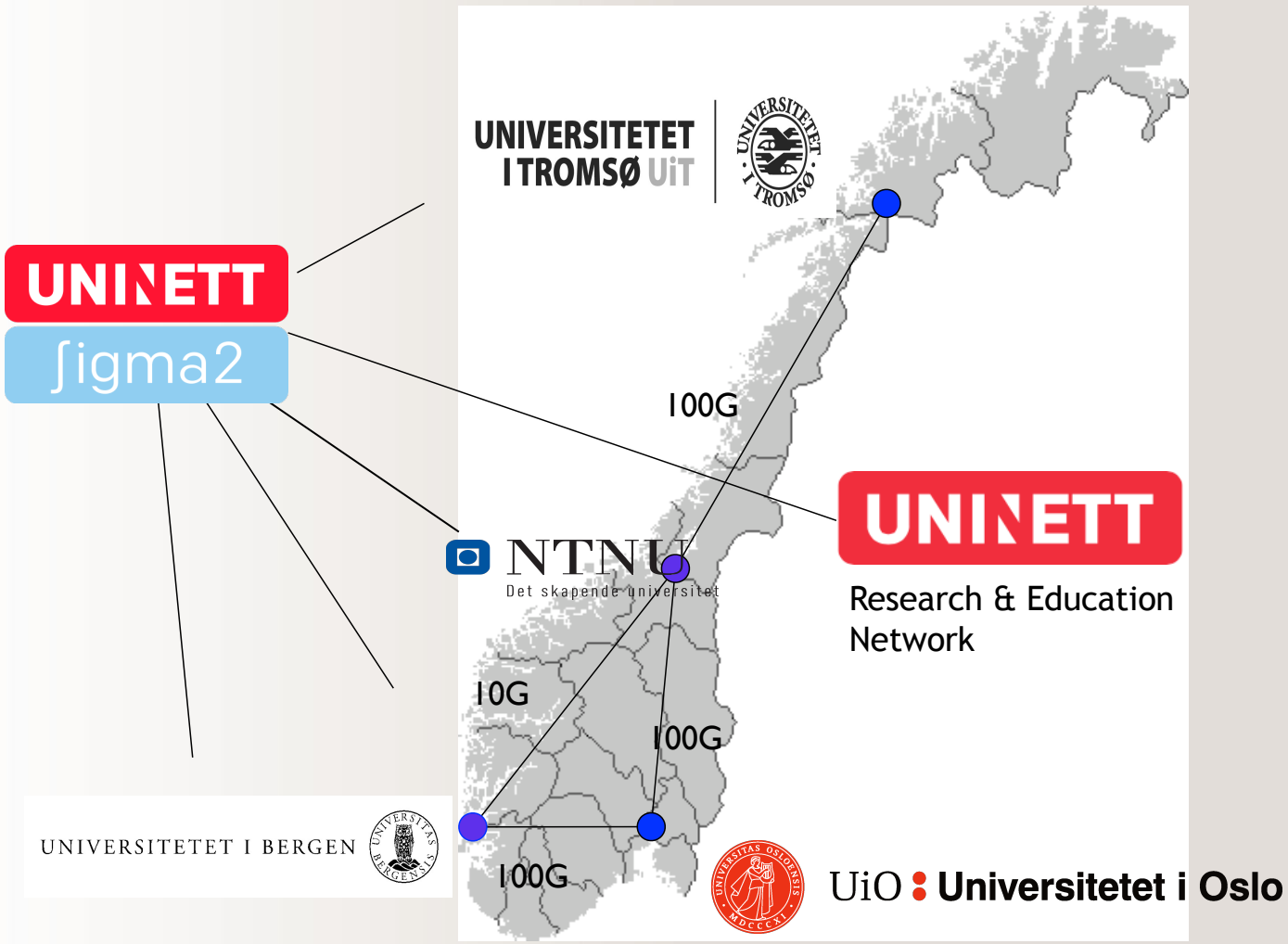
- About the national e-infrastructure, UNINETT Sigma2 and the Metacenter
- Core services
  - Computing
  - Storage and services
  - Getting access
  - Basic support and application support
  - Advanced User Support (AUS)

E-infrastructure: IT based infrastructure (networks, computers, data storage etc.), resources (software, tools, services etc.) and support that facilitate research, FAIR use of data and the collaboration among research communities.

- ! Not only hardware

National e-infrastructure - a very brief history

- From the beginning, it was always recognized that e-infrastructure, just like other research infrastructure, should be shared.
- Early on, research institutions competed for basically the same funding and established disconnected e-infrastructure resources.
- In the early 2000's, the need for coordination and sharing lead to the establishment of UNINETT Sigma and the Metacenter. Universities still competed for the same funding and had their own hardware resources, no common strategy.
- **In December 2014, the four major universities (UiB, UiO, UiT, NTNU) and the Research Council of Norway (RCN) decided to establish UNINETT Sigma2 and collectively operate the national e-infrastructure.**

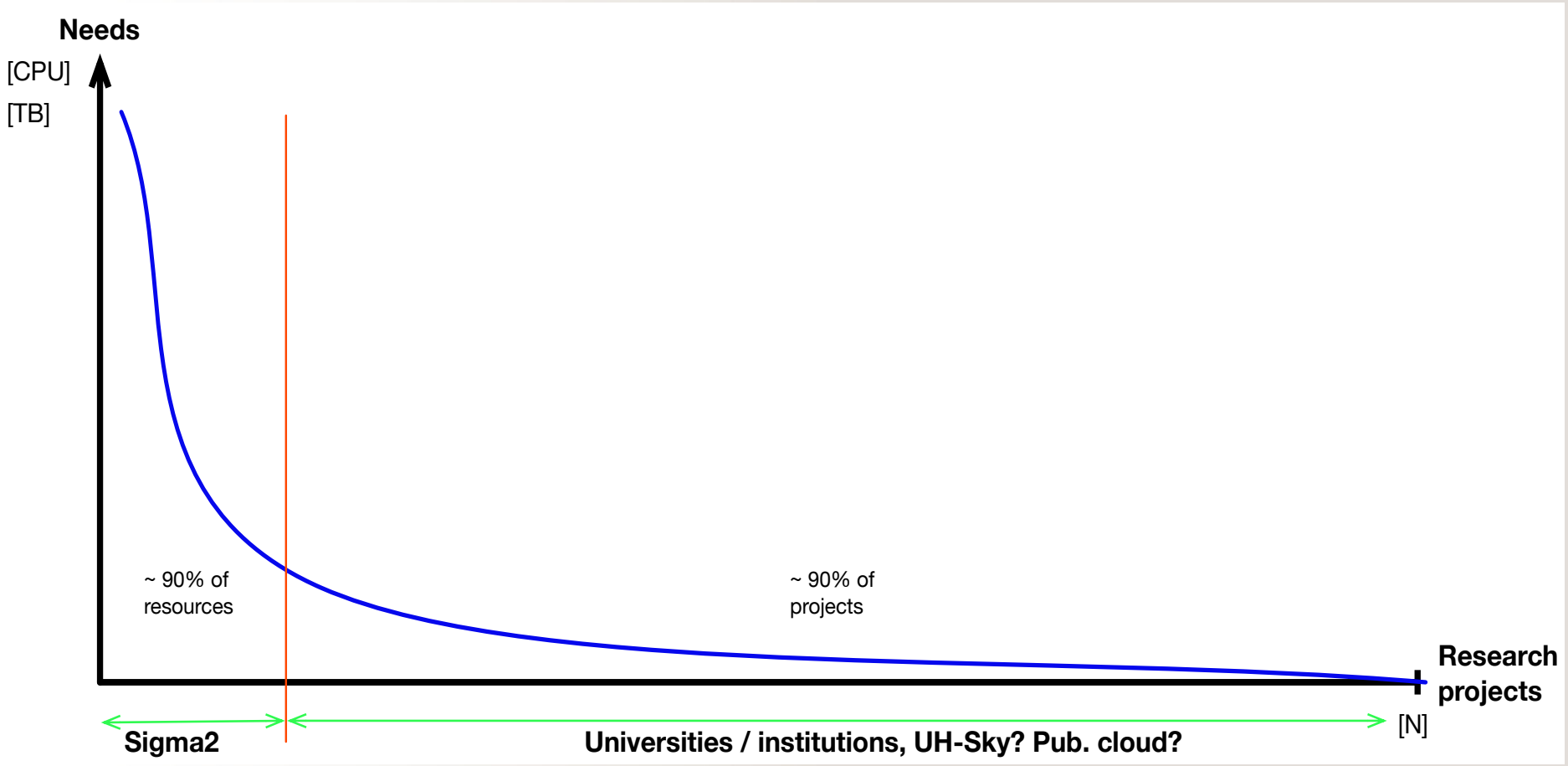


UNINETT Sigma2

Sigma2 - 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)

Local vs. national e-infrastructure

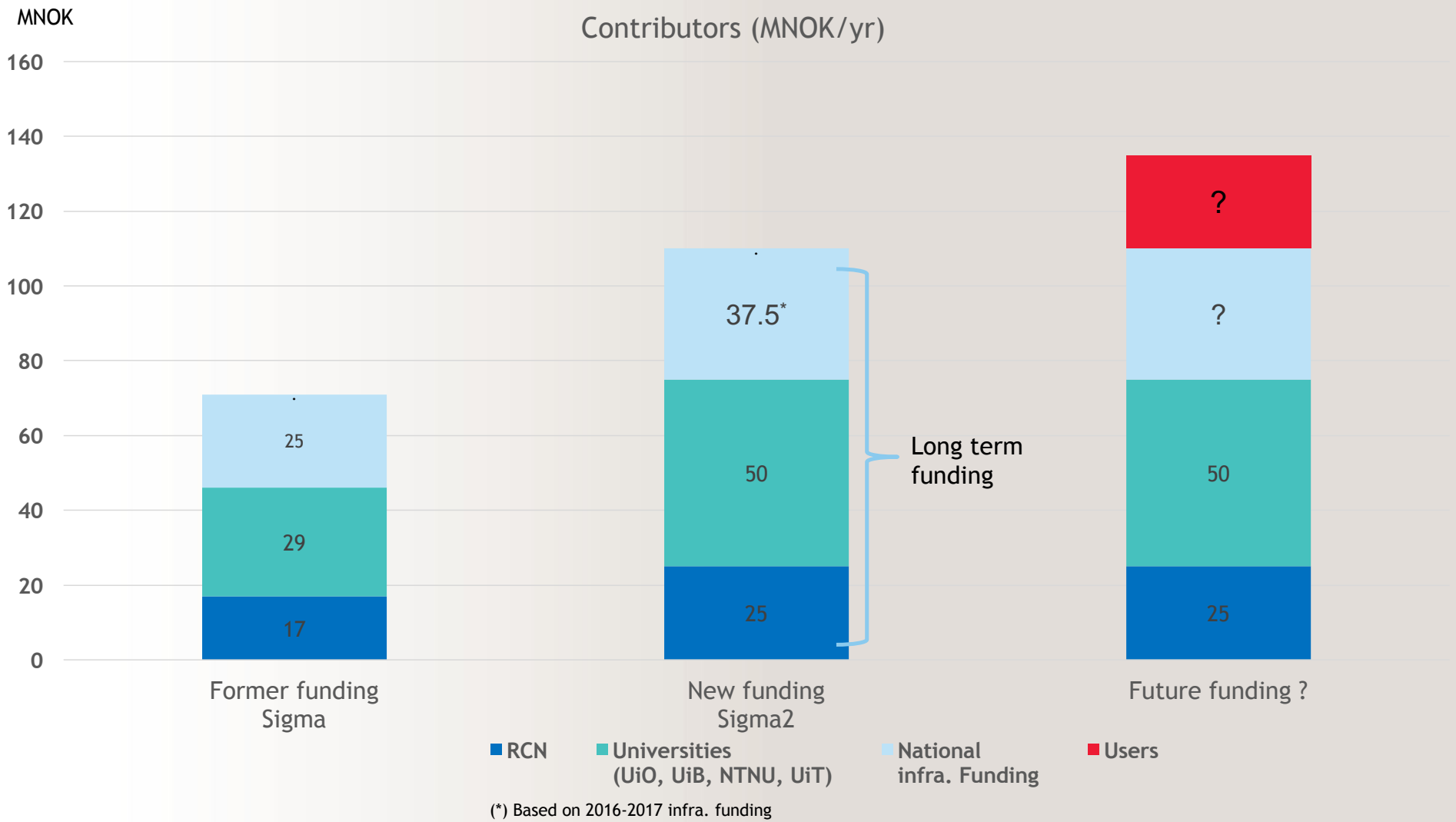


International cooperation

- High Performance Computing (HPC)
  - PRACE (Partnership for Advanced Computing in Europe)
  - PRACE DECI (Distributed European Computing Initiative)
- Storage
  - EUDAT / EOSC
- NeIC (Nordic e-Infrastructure Collaboration)
  - Code Refinery
  - Pool of Competences
  - Glenna Nordic Cloud project
  - Tryggve (sensitive data)
  - Nordic Tier-1 facility for WLCG (CERN)

UNINETT Sigma2

Sigma2 financing



UNINETT Sigma2

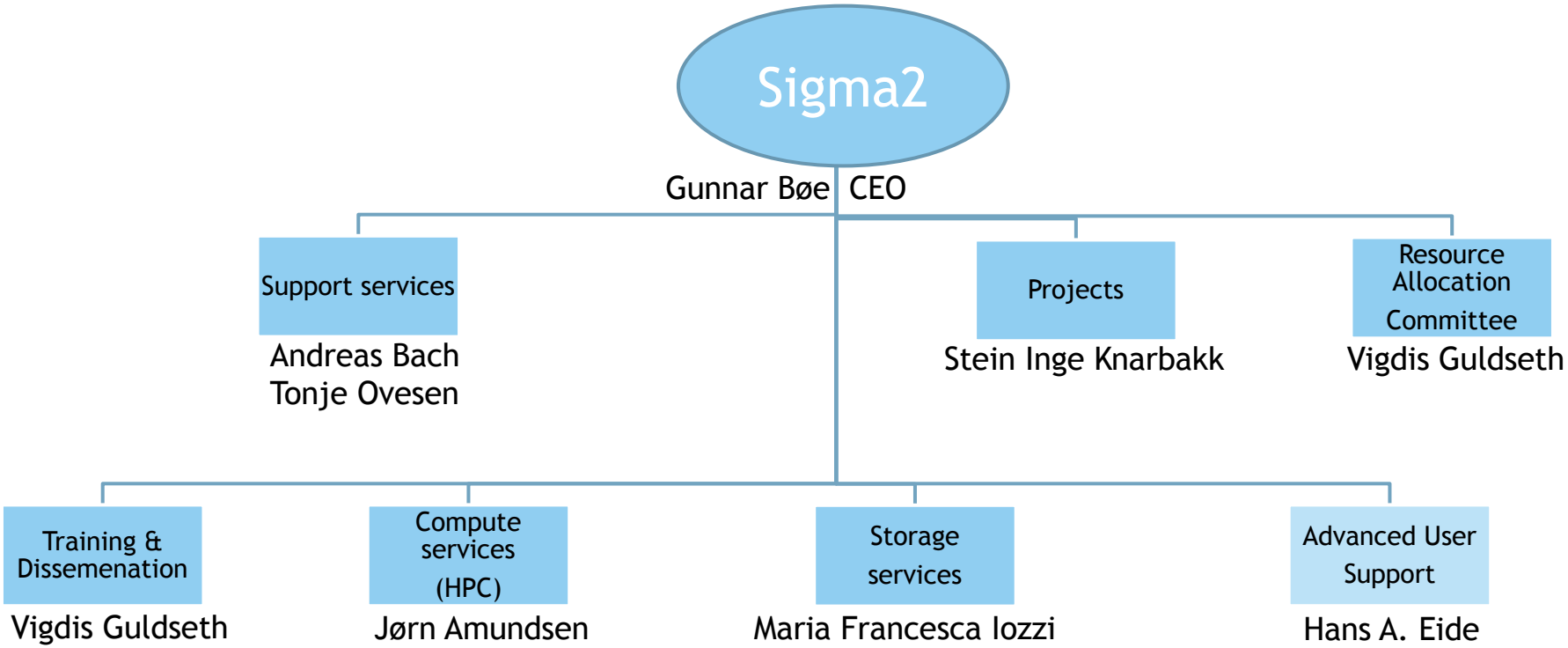
Sigma2 governance and management

- Sigma2 board members
  - Terese Løvås, **NTNU**, Professor, Department of Energy and Process Engineering
  - Nathalie Reuter, **UiB**, Professor, Department of Molecular Biology
  - Morten Dæhlen, **UiO**, Dean Faculty of Math. and Natural Sciences, Professor Mathematics
  - Kenneth Ruud, **UiT**, Prorector of Research, Professor of theoretical chemistry
  - Øyvind Hennestad, Corporate Laywer, **Sintef**
  - Juni Palmgren, **Karolinska institutet**, Professor, Department of Medical Epidemiology and Biostatistics
  - Roar Olsen, former Adm. Dir. **UNINETT**, Chairman
- Other stakeholders
  - The Research Council of Norway
  - The IT-directors (4 universities)
  - The Metacenter managers (4 universities)

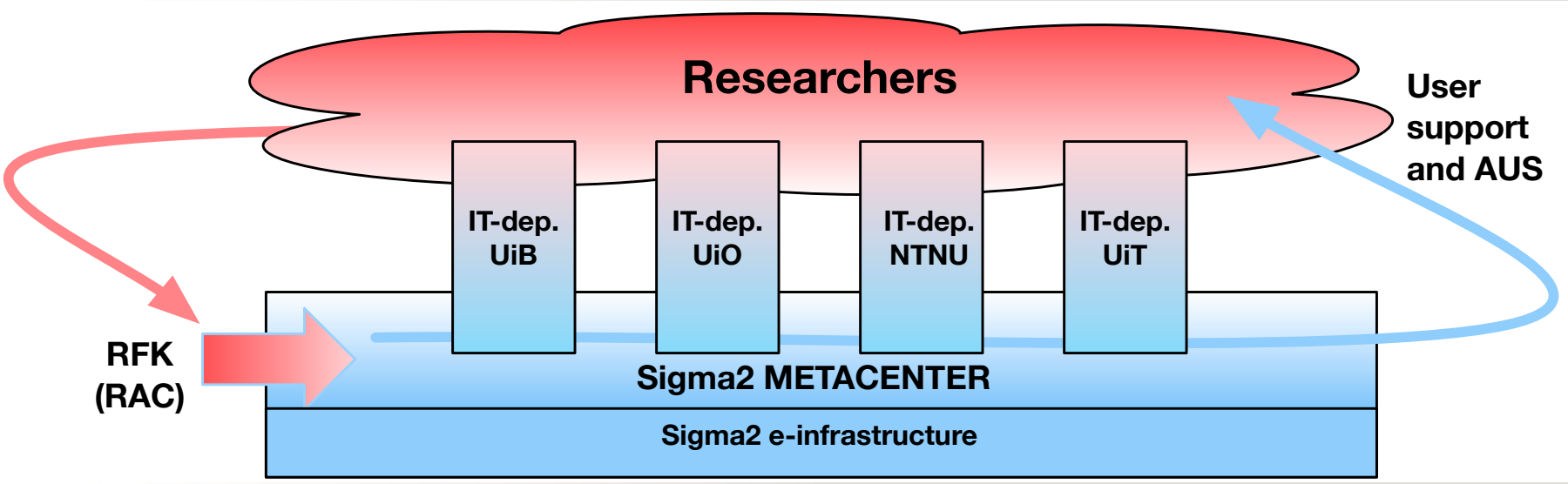
UNINETT Sigma2

- The national e-infrastructure for research and education
- Supports as of today ca. 1600 users, 400+ research projects
- Procurement, project lead, coordination, strategic responsibility
- 8 people employed (in Sigma2 itself)
- The Metacenter: ca. 35 FTEs over ca. 50 highly competent people employed at the IT-departments of the four partner universities

Sigma2 organization (focus areas)

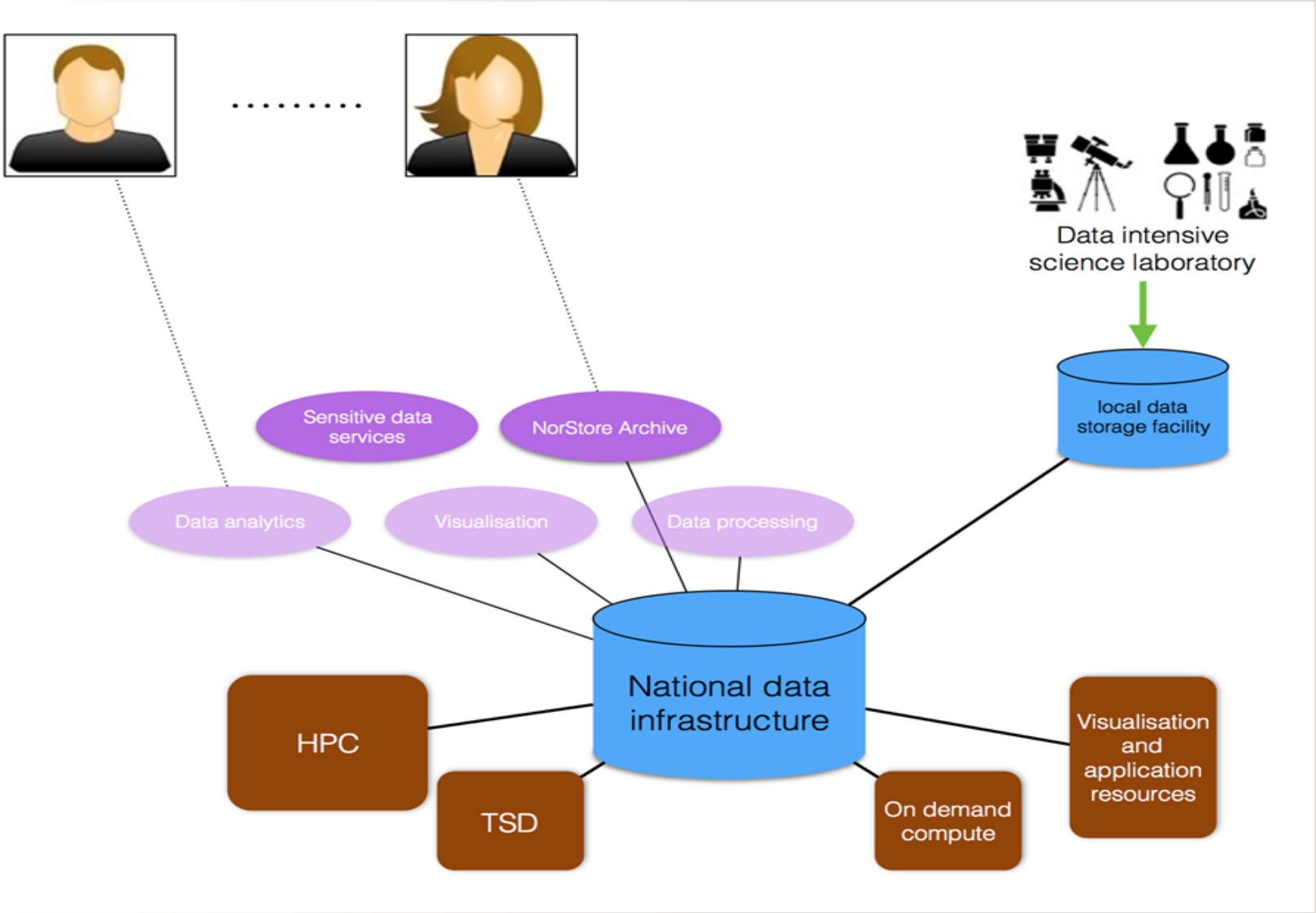


The Metacenter



- National coordination and shared, consolidated resources have cost and efficiency advantages but creates a “distance” to the end-users (researchers)
- This is countered 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

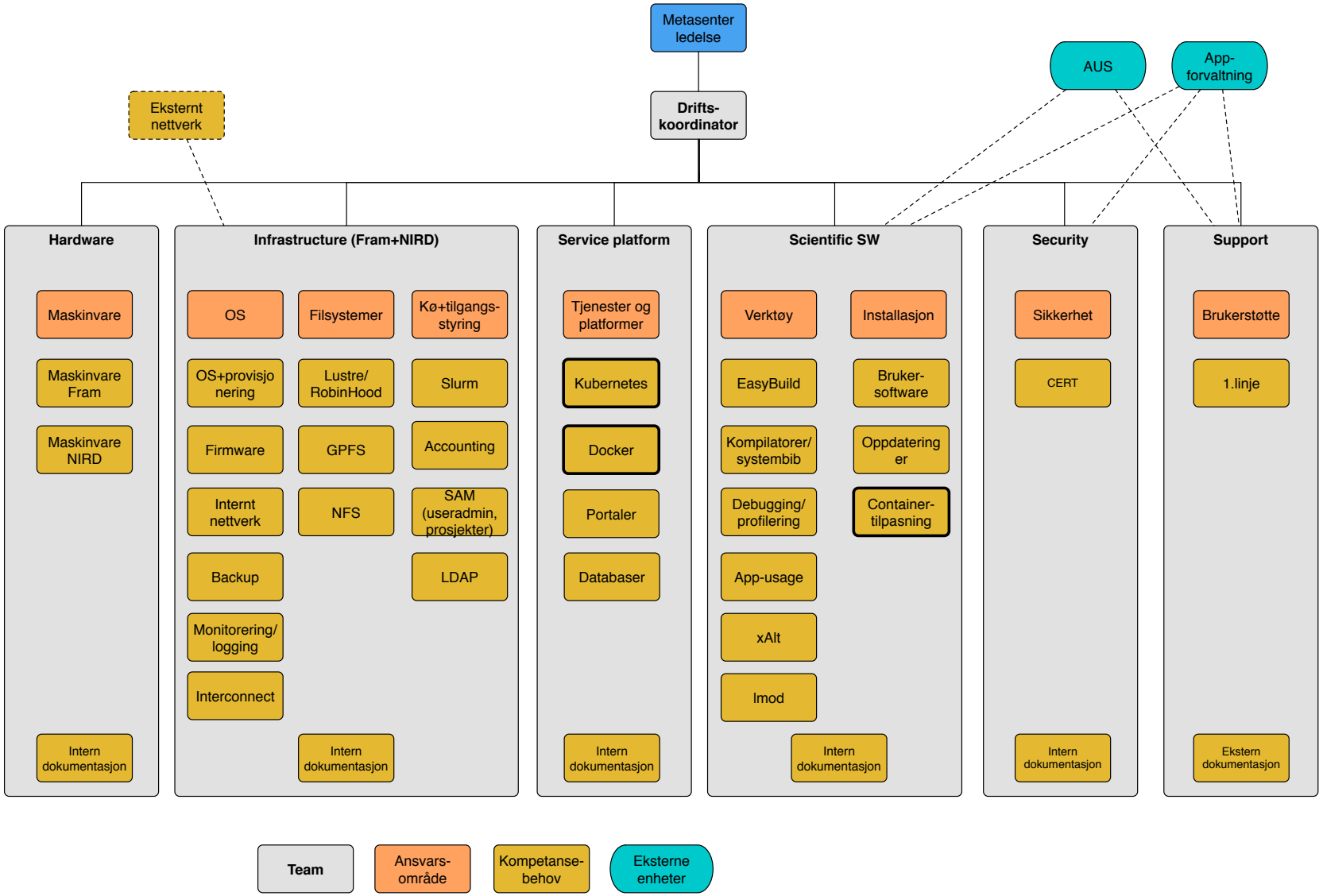
Data-centric architecture



Operations organization

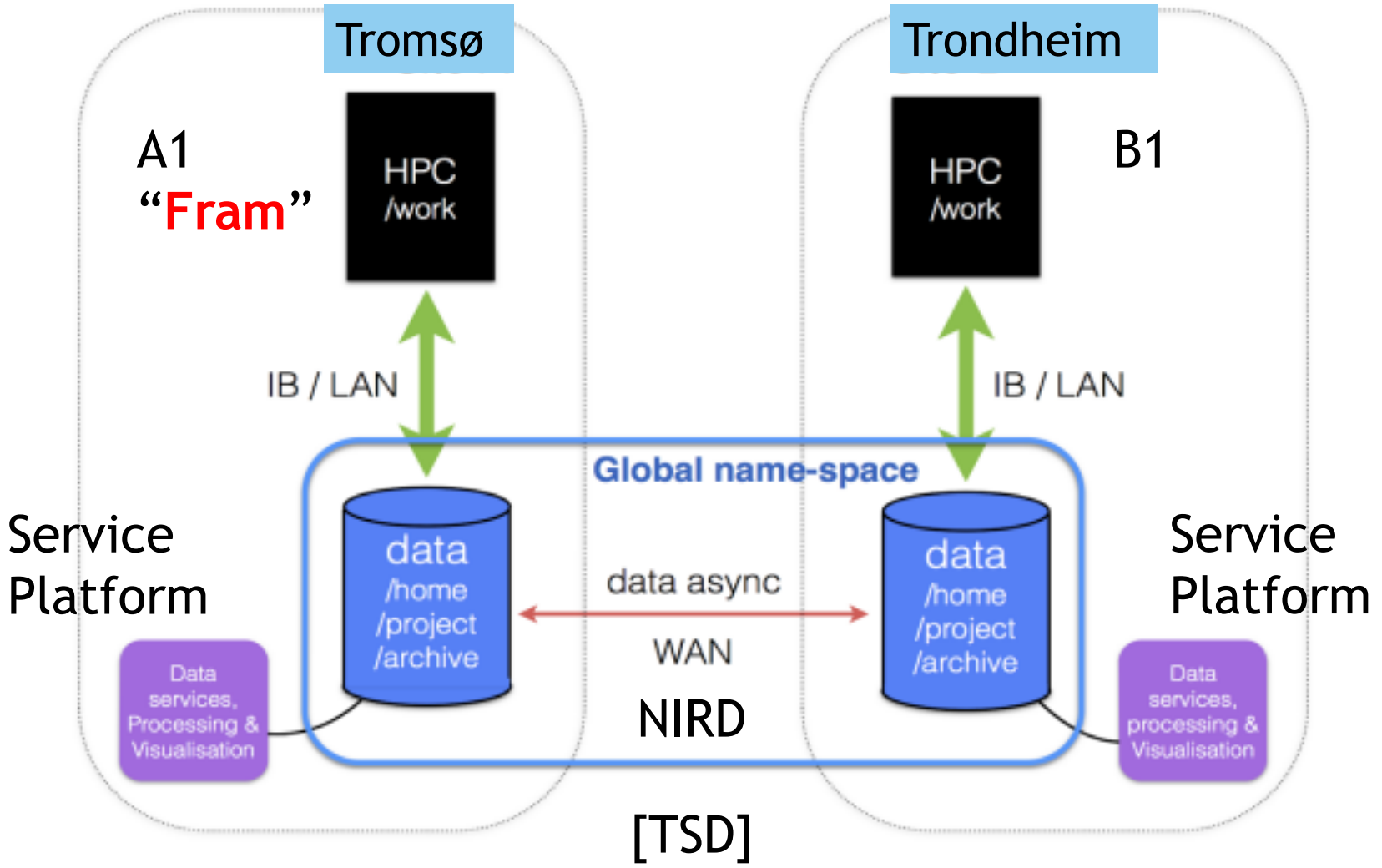
- Shared operations between the 4 partner universities
- Organization, staffing and agreement ”Drift og Brukerstøtte” established by the four partner universities in collaboration with Sigma2
- Agreement in place since 1 June 2017
- Area-specific teams with own team leaders
- Rotating first-line support team

Operations organization



Sigma2 core e-infrastructure services

Implementing the data-centric architecture



UNINETT Sigma2

Sigma2 core e-infrastructure services

- Computation
  - Compute cycles for computational research, including for sensitive data
- Storage
  - Data storage (archive and project), including for sensitive data
  - Data management planning (DMP)
  - Service platform (visualization, data-analytics, discipline and project specific services)
- Basic user support
  - Basic tech support through a ticket-based support service
  - Training
- Advanced user support (AUS)

Sigma2 core e-infrastructure services

- Computation

Computing (HPC) - past to future

- The past
  - Load is serviced by Abel, Stallo, Hexagon and Vilje
  - A virtual organization (the Metacenter), but …
  - Independent systems, independent software stack, independent storage and independent systems administration
- The future
  - Moving HPC from a 4-system to a 2-system IS a with 2-year leap-frogged installation across a 4-year lifetime
  - Data-centric model with close connection between HPC and storage (NIRD)
  - Two compute platforms: HPC and Service platform
  - Common operations and SW stack, based on EasyBuild and Slurm

Computing (HPC) - past to future

- The future (continued)
  - HPC platform manycore nodes with fast IC
  - Service platform (SP) for 1-node jobs not needing fast IC
  - GPUs planned for the SP (8 nodes with 2 CPUs and 2 GPUs each)
- The present
  - We are in between two train stops!
  - **Fram** as part of the new IS, with Abel and Stallo from the old IS in service until end 2018/beginning 2019 until next system, ‘B1’
  - Might experience pressure on compute and storage resources along this path
  - Might be mitigating actions in between, e.g. Vilje operated throughout 2017 and possibly beginning of 2018

High Performance Computing (HPC) resources

System	Sigma2 capacity (MCPUhrs/yr)	Tot. performance (TFLOP/s)	Deployed
Hexagon	102.1	109	4/2012
Abel	75.9	182	10/2012
Vilje	113.0	312	10/2012
Stallo	120.4	~291	10/2012 (+ utv.)
Colossus*	<13	~30	4/2014
Sum	322.1	894	
Fram	279.2	1071	10/2017
”B1”	?	?	(4Q/2018)
“HTC** platform”	?	?	(2H2018)

(*) For sensitive data, part of TSD

(**) HTC = High Throughput Computing / cloud platform

UNINETT Sigma2

Computing (HPC)

Hardware

- From 1 April 2018 will compute load be serviced by Abel, Stallo and **Fram**
- Access to compute time on Colossus (TSD) available also from Sigma2
- Accelerators, GPUs and Xeon Phis, currently available on Abel
- GPUs (Volta) will be available on the service platform

Software platform

- A common software platform based on EasyBuild and Lmod is provided on Fram
- Toolchain based on Intel compilers and Intel MPI, but GCC, PGI and HPCX OpenMPI also provided
- TotalView and Performance Reporter main debug and profiling tools
- User EasyBuild module builds is also provided
- The software platform which is developed on Fram will be eventually be introduced on Abel and Stallo later

Data-analytics (Big data) and HTC

- Low demand for analytics and machine learning so far
- Abel handles most of the HTC (esp. life science)
- Analytics technology already in use for other services from UNINETT
  - Pilot (Spark) in cooperation with
    - St.Olav hospital/NTNU (Protein and Genomic analysis)
  - Other use cases:
    - Computational Linguistics (common Crawl dataset (500 TB))
    - Fish genomics
    - EISCAT data
- A new platform for this type of compute needs is being built in connection with the NIRD storage infrastructure, possibly also for TSD late 2018

Sigma2 core e-infrastructure services

- Storage

Storage infrastructure

- Research data archive
- Project storage (minimum 10 TB)

- Norstore is replaced by NIRD – National Infrastructure for Research Data

System	Capacity [PB]	Deployed	Location
Norstore	3.7	1/2013	Oslo (+Tromsø)
NIRD	5.6	9/2017	Tromsø + Trondheim
(NIRD exp.)	~10?	(2/2018)	

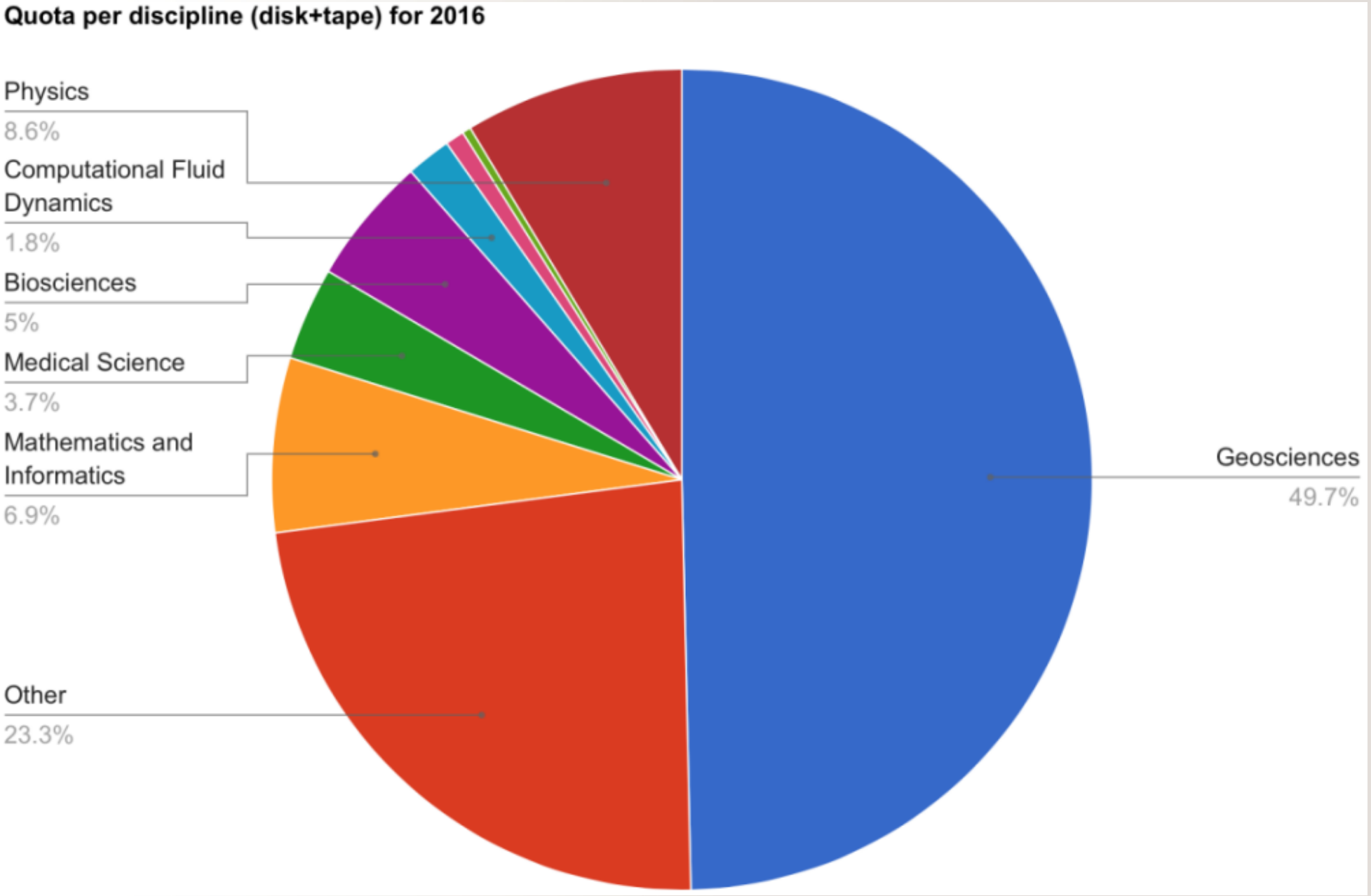
Storage - NIRD

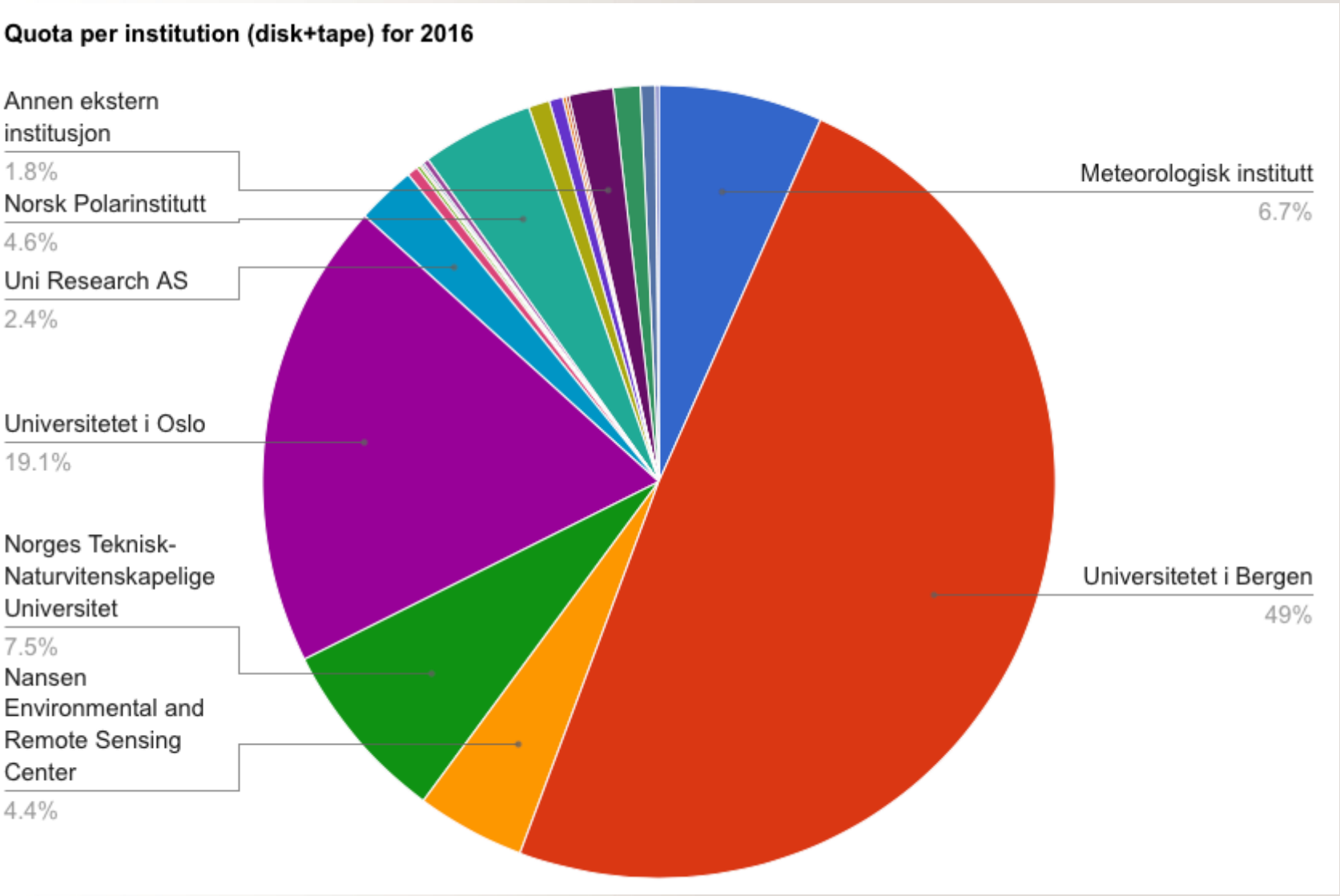
- NIRD = National Infrastructure for Research Data
  - 6+ PB of disk (Tromsø and Trondheim) as of fall ‘17.
  - GPFS 4.2 parallel filesystem
  - Login node “login.nird.sigma2.no” with ssh access
  - Data in /projects/NsxxxxK (NIRD projects) symlinked to /nird/projects/nird on Fram
  - Data in /nird/projects/fram/nnxxxxk (Notur projects on Fram)
  - $HOME on Fram
  - Get quota usage with dusage -p NsxxxxK [or nnxxxxk]

```
$ dusage -p NS2345K
================================================================
Project Account Resource
NS2345K $PROJECT nird
NS2345K $PROJECT nird
NS2345K $PROJECT replica Disk 919.961TB 1000TB 1000TB NS2345K $PROJECT
replica Files 20026994 None None
----------------------------------------------------------------
```

Storage – NIRD (cont.)

- Geo-replication from TOS to TRD (~realtime)
- Daily snapshots (daily for last 7 days and weekly for last 6 weeks). Check /projects/NsxxxxK/.snapshots and /nird/projects/fram/nnxxxxk/.snapshots
- Home dirs are NFS mounted on Fram → do not run demanding jobs directly from here but first copy data to /cluster/work and start jobs from there


Quota per discipline (disk+tape) for 2016
Physics
8.6%
Computational Fluid Dynamics
1.8%
Biosciences
5%
Medical Science
3.7%
Mathematics and Informatics
6.9%
Other
23.3%
Geosciences
49.7%

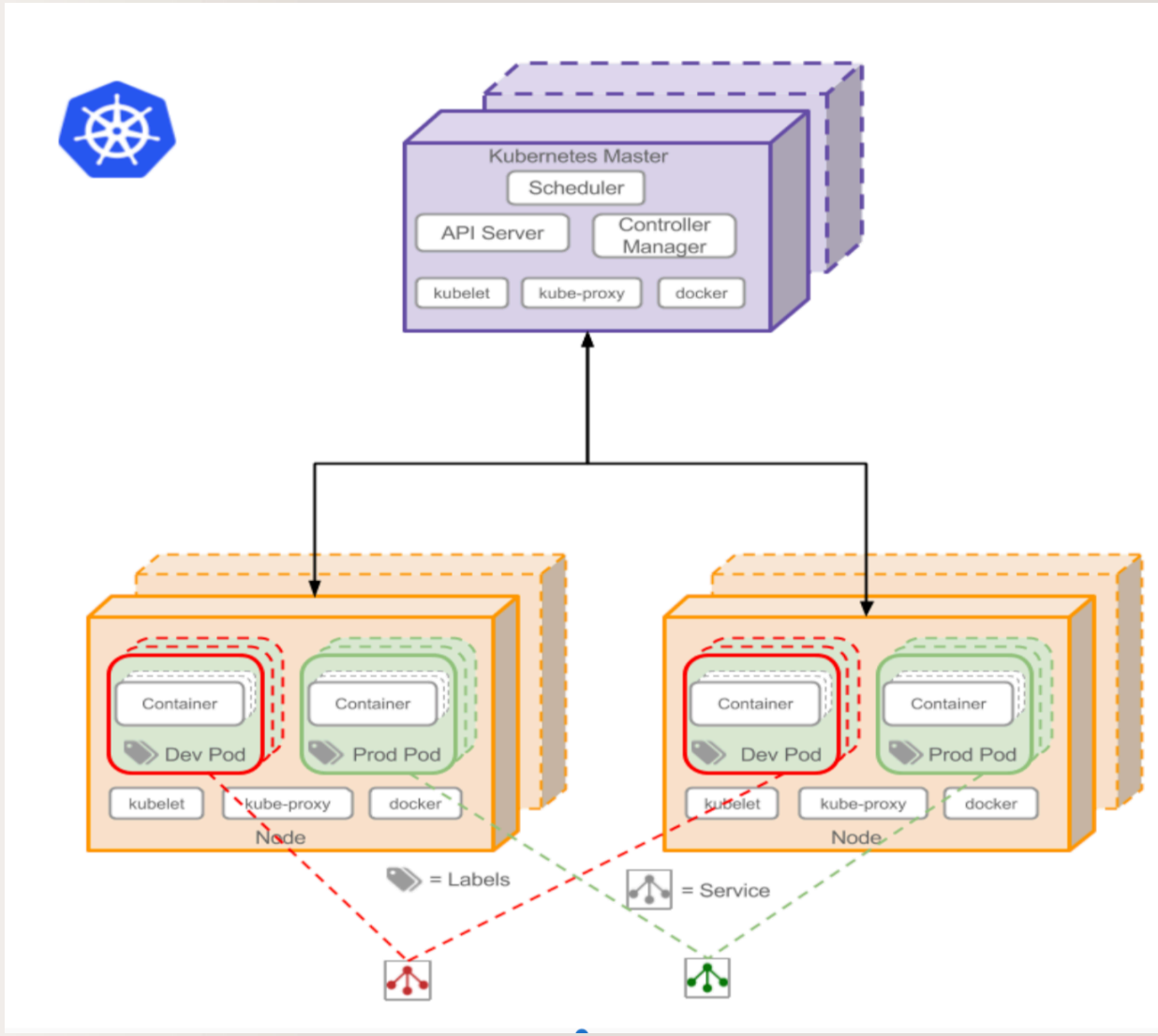

Quota per institution (disk+tape) for 2016
Annen ekstern institusjon
1.8%
Norsk Polarinstitutt
4.6%
Uni Research AS
2.4%
Universitetet i Oslo
19.1%
Norges Teknisk-Naturvitenskapelige Universitet
7.5%
Nansen Environmental and Remote Sensing Center
4.4%
Meteorologisk institutt
6.7%
Universitetet i Bergen
49%

NIRD Service Platform

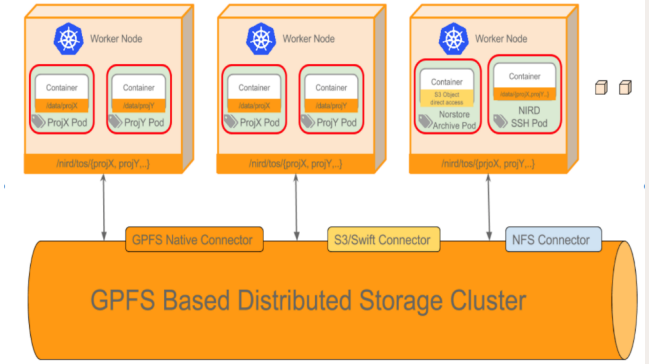
- Bring compute to the data, not the other way around (data-centric architecture, sits “on top of” NIRD)
- Powerful compute nodes and virtualization technology (Kubernetes, Docker containers) for on-demand tasks and fast service deployment

NIRD Service Platform (SP)

A Kubernetes-platform running on computing nodes that access the NIRD distributed storage. Services run in Docker containers.

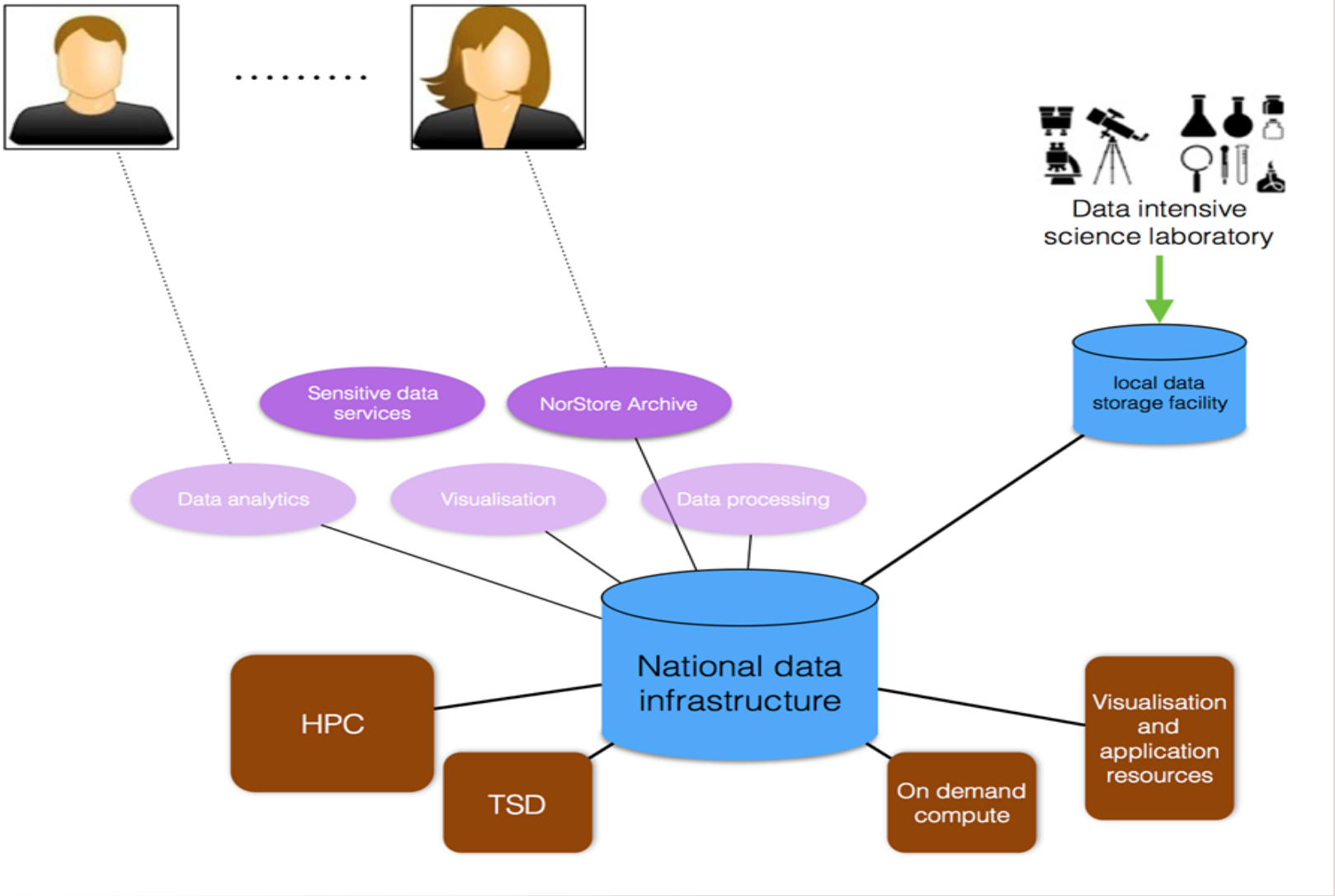


Storage Mounting



UNINETT Sigma2

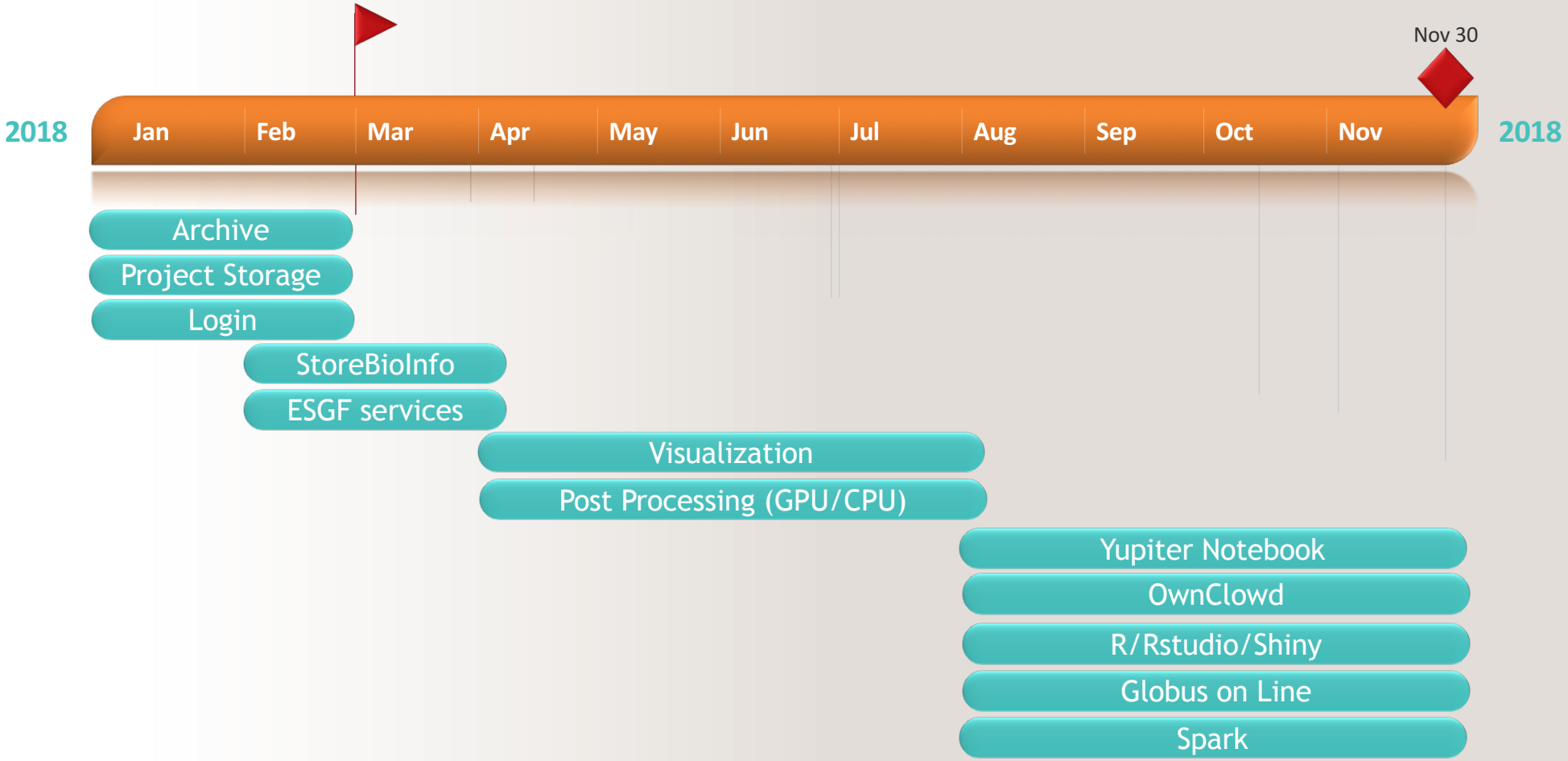
Remember the data-centric architecture



Strength of the Service Platform (SP)

- **Flexible and versatile**: SP can host any “dockerized” service
- **Cost-effective**: SP computing resources can be use for “dockerized” jobs or traditional HPC jobs (single threaded or OpenMP jobs)
- **Customizable**: researchers can run their own service (web service, computing workflows etc…) provided that it is “dockerized”
- **GPUs for visualization** and **GPU/CPU computing** (machine learning)

Services Deployment Roadmap

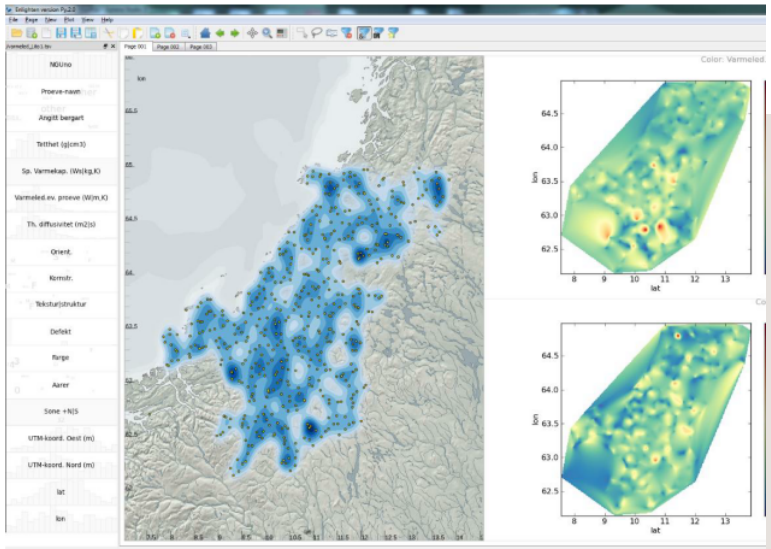


UNINETT Sigma2

What researchers request (example):

Software requirements

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



Processing and visualization software

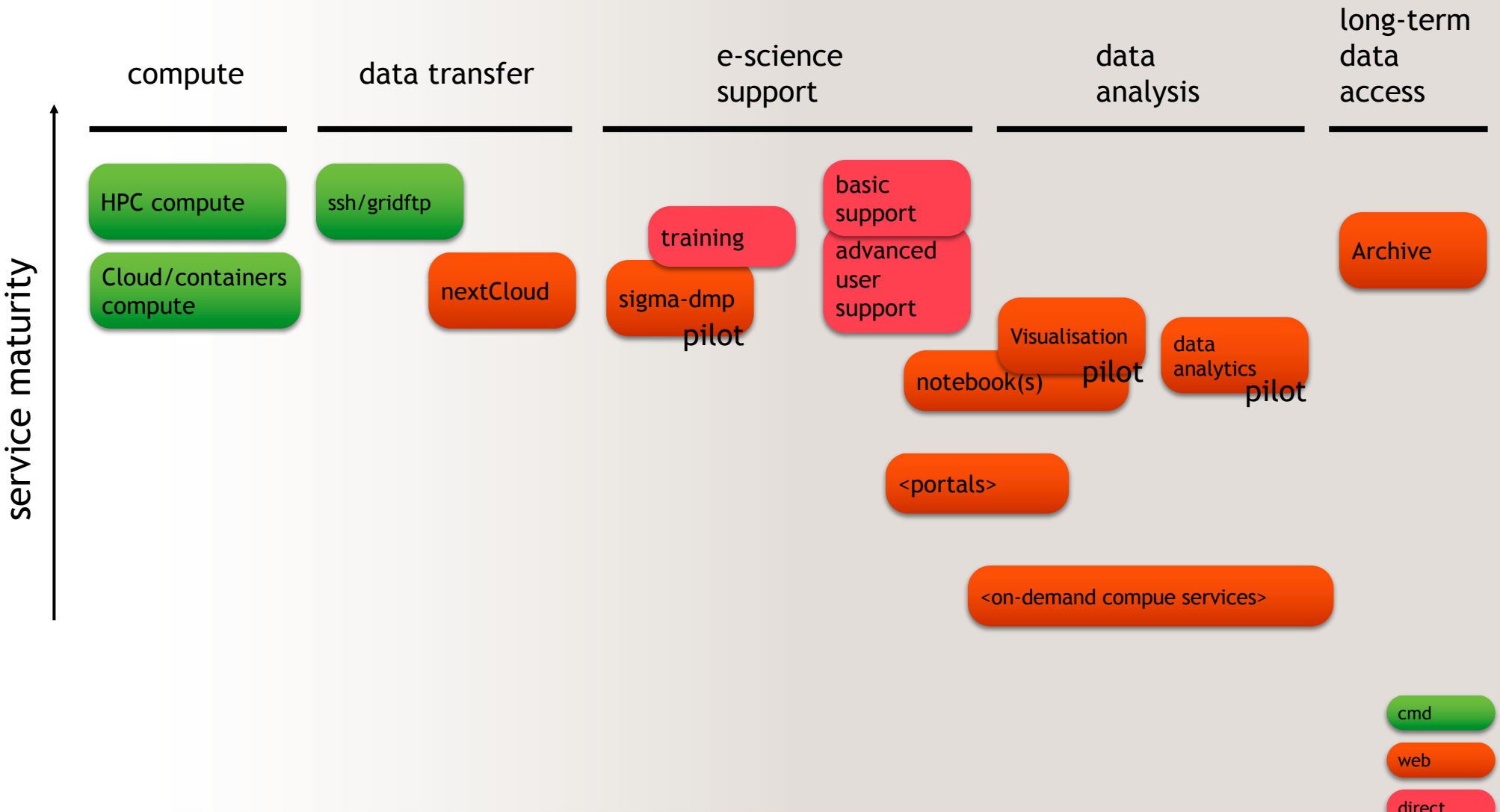
- (as cloud services or available for download)
- Visualization – ICS-D functionality
- SEISAN, Earthquake analysis software (UiB)
- NORSAR 3D, 3D modelling tool

Sigma2 community services

Specific services supported on NorStore resource

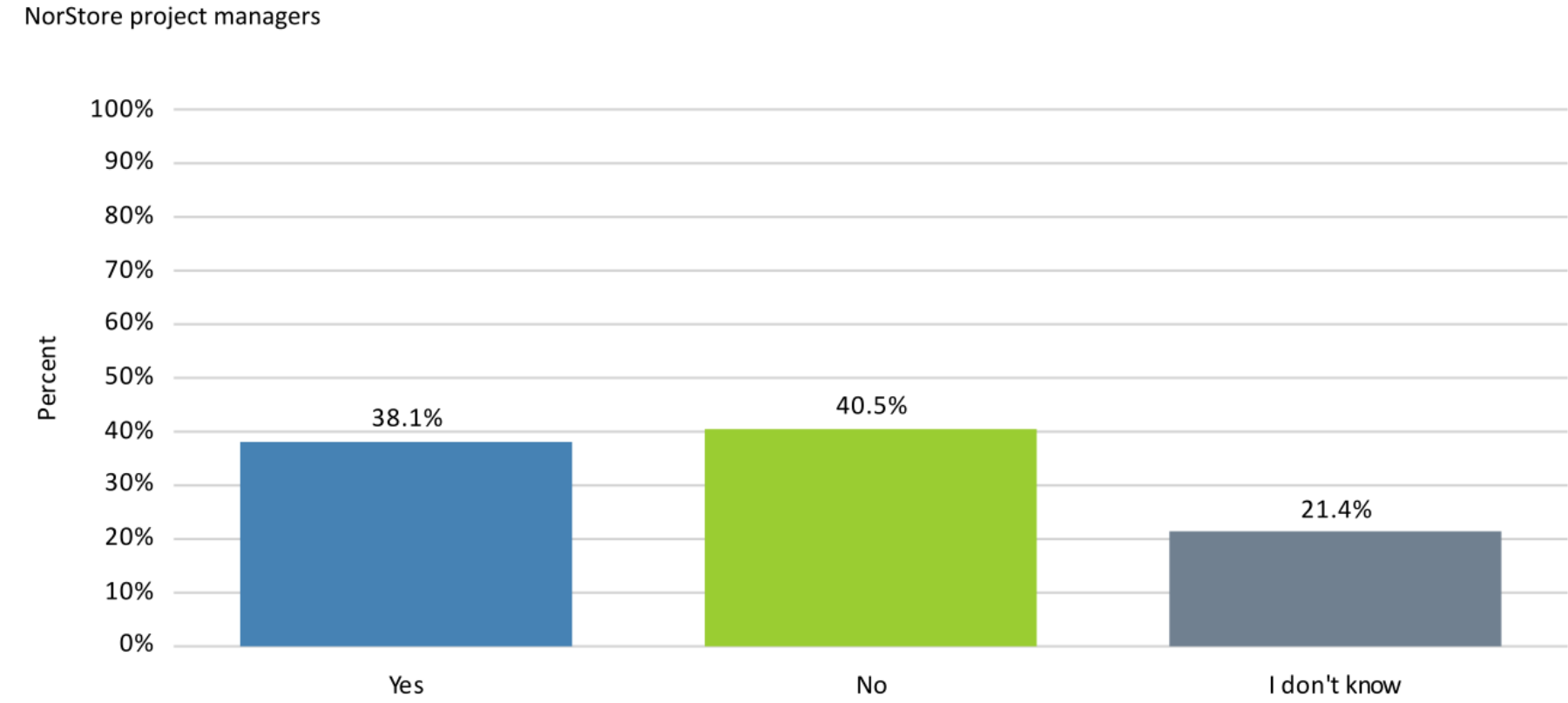
Service	Project/community	Reference	Contact
BioGateway	Biology	Semantic systems biology	Martin Kuiper
NorMAP THREDDS	Climate, wind energy	normap.norstore.uio.no	support@norstore.no
StoreBioInfo Portal	Bio-informaticas	storebioinfo.norstore.no	Kjell Petersen
Earth Systems Grid	Climate	ESG data node	Mats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

Services

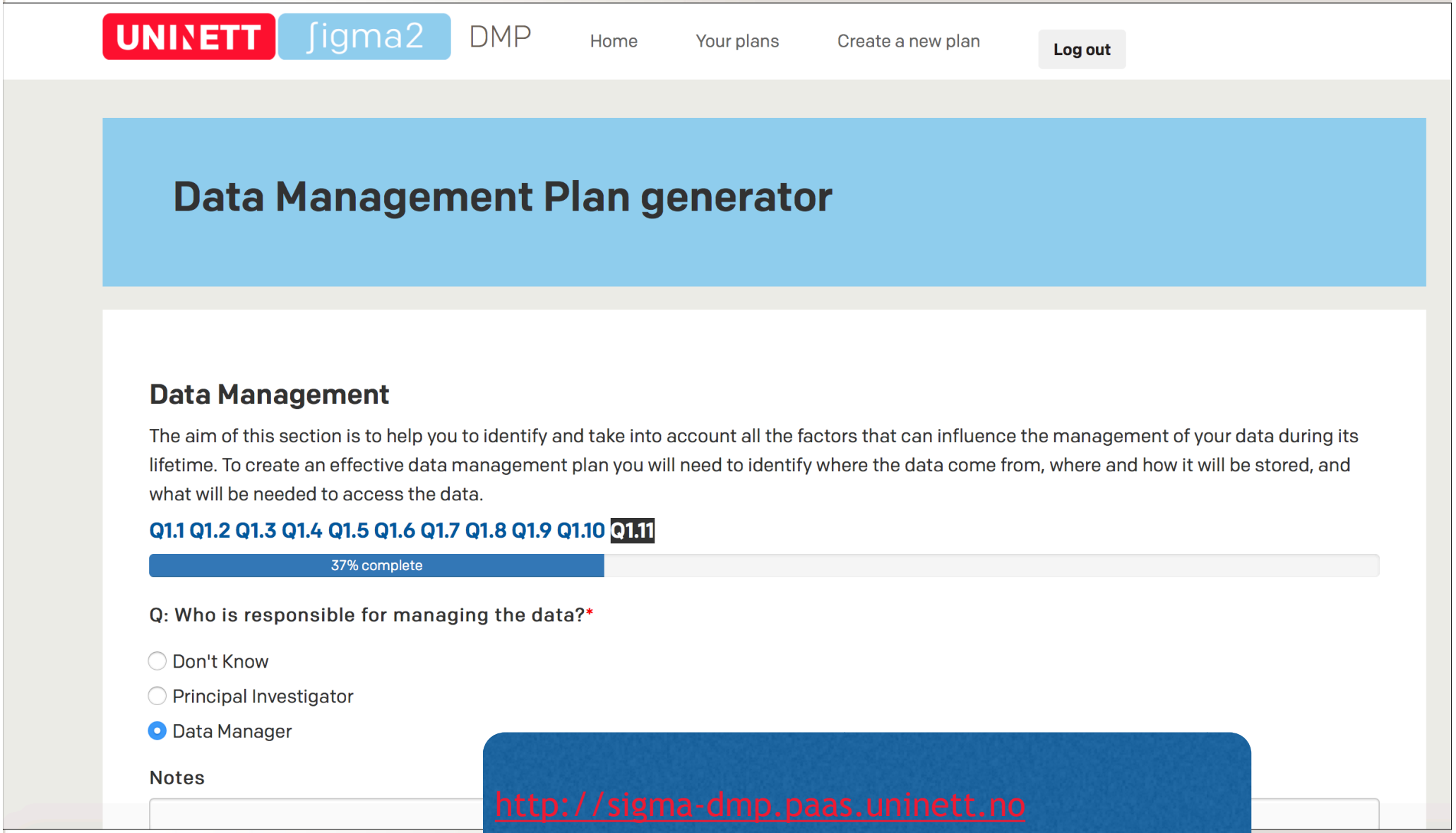

compute
data transfer
e-science support
data analysis
long-term data access
service maturity
HPC compute
Cloud/containers compute
ssh/gridftp
nextCloud
training
sigma-dmp
pilot
basic support
advanced user support
Visualisation
pilot
notebook(s)
data analytics
pilot
Archive
<portals>
<on-demand compue services>
cmd
web
direct

Data management plans?

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



Data ‘policy’ for Research data



Services for sensitive research data

- Data that can be related to human subjects is by law/nature sensitive*, and the importance and prevalence of this type of data in research is rapidly increasing as it relates to health and other societal issues of high impact and visibility.

- Our ability to do research involving sensitive data is dependent on e-infrastructure that can protect the data according to laws and regulations while at the same time providing access and resources according to the needs of the researchers.

- UiO/USIT, together with Sigma/Sigma2 and others, have collaborated on establishing a secure e-infrastructure to provide services for sensitive data. The resulting ”TSD” is a national platform for all types of research involving sensitive data.

(*) PERSONAL DATA REVEALING INFORMATION REGARDING RACIAL OR ETHNIC ORIGIN, POLITICAL OPINIONS, RELIGIOUS OR PHILOSOPHICAL BELIEFS, TRADE-UNION MEMBERSHIP, DATA CONCERNING HEALTH, SEX LIFE.

UNINETT Sigma2

Getting access to the national e-infrastructure

By application

- Calls twice a year (Jan/Feb, Aug/Sep):
  - **https://www.metacenter.no/mas/application/project/**

Right away

- Small and exploratory needs (Fram only)
  - https://www.metacenter.no/mas/application/project/
  - If in doubt: sigma2@uninett.no

- See https://www.sigma2.no/content/apply-e-infrastructure-resources

Resource allocation

- Resources made available to all research carried out under the auspices of Norwegian research institutions
- Controlled by the Resource Allocation Committee (RFK)
- Applications are assessed on the basis of the project's scientific quality
- 2 calls every year for major applications (minor applications throughout the year)

RFK working group (RFK-wg)

- One technical person with HPC knowledge from each university, and one person from each site hosting storage
- Reads application requirements and proposes quotas and host assignments to RFK
- Used as primary source for allocation advices on extra allocations and new projects during allocation periods
- 2017 staffing: Steinar Trædal-Henden, Lorand Szentannai, Henrik Nagel, Ole Widar Saastad (all HPC), Thierry Toutain (storage)

Sigma2 core e-infrastructure services

- Basic user support

Help!

Technical support

- User documentation:
  - **https://www.sigma2.no/content/support-e-infrastructure-users**

- **All** support requests: **support@metacenter.no**
  - Applications for compute and storage resources go to sigma2@uninett.no

Application management and support

- Organization for strategic management of applications
- Connection between scientific groups and the operations org.
- Looks at application usage on the resources, which applications should be added, which should be phased out
- Success depends on good relationship with relevant scientific groups
- Will be organized in application areas
  - Chemistry / Matr. Science, Data analysis, Life Sciences, Geo. Physics / Earth Science (Climate), CFD, Performance Mon. / SW Development / Misc
- Startup – RSN (spring ‘18)

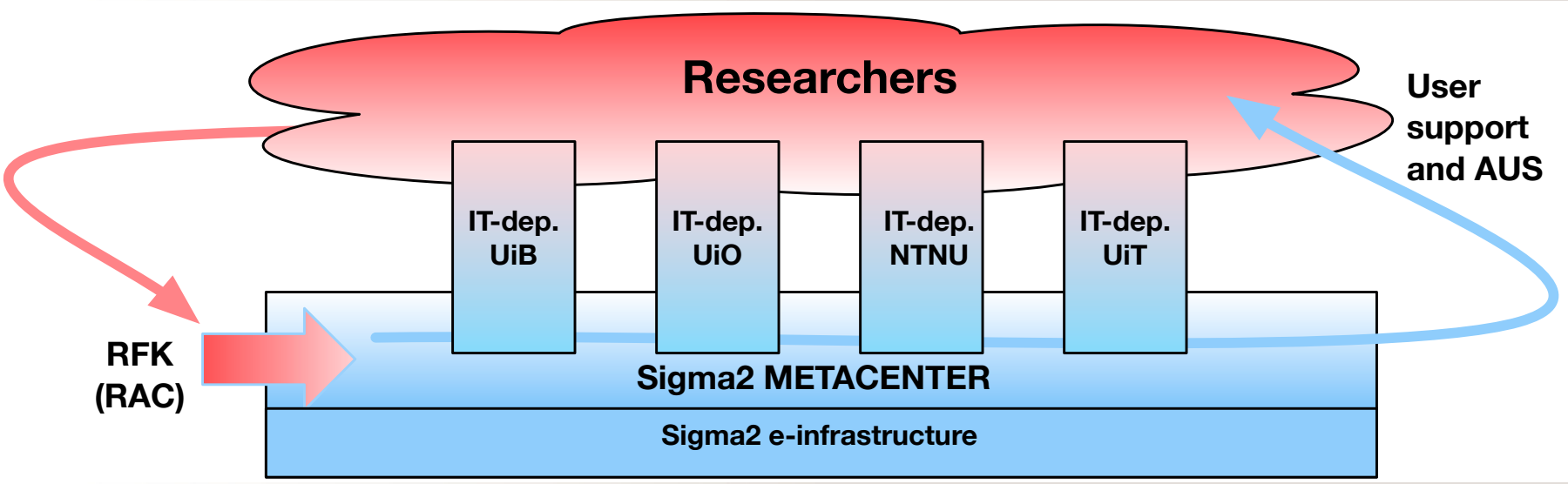
Sigma2 core e-infrastructure services

- Advanced user support (AUS)

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, continuous applications
- 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

The Metacenter- assistance is near!



- National coordination and shared, consolidated resources have cost and efficiency advantages but creates a “distance” to the end-users (researchers)
- This is countered 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

Advanced User Support (AUS) (cont.)

For the HPC services, project based advanced user support aims at helping scientists to improve or extend the performance and capabilities of their applications. This can be in a number of ways, including:

- code parallelization
- code porting
- code profiling, optimization, benchmarking
- improving user-interfaces
- software development

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

AUS example: OILCOM (Skogen/IMR)

cores [0,16]. In parentheses, speedup wrt. the original, non-optimized code is shown, where the iteration time was 10 seconds. The maximum speedup obtained was $\sim$ 18 when running on 32 cores.

	iteration time [s] (speedup)	
ncores	1 socket	2 sockets
1	4.6 (2.2)	-
2	2.4 (4.2)	2.7 (3.7)
4	1.3 (7.7)	1.4 (7.1)
8	0.84 (11.9)	0.85 (11.8)
16	0.69 (14.5)	0.64 (15.6)
32	-	0.56 (17.9)

Advanced User Support (AUS)

- How to apply for AUS:
  - At any time, contact sigma2@uninett.no or start from https://www.sigma2.no/content/advanced-user-support-0
  - Small AUS projects might be granted within a week, larger projects (e.g. discipline speciffic AUS) might need longer time

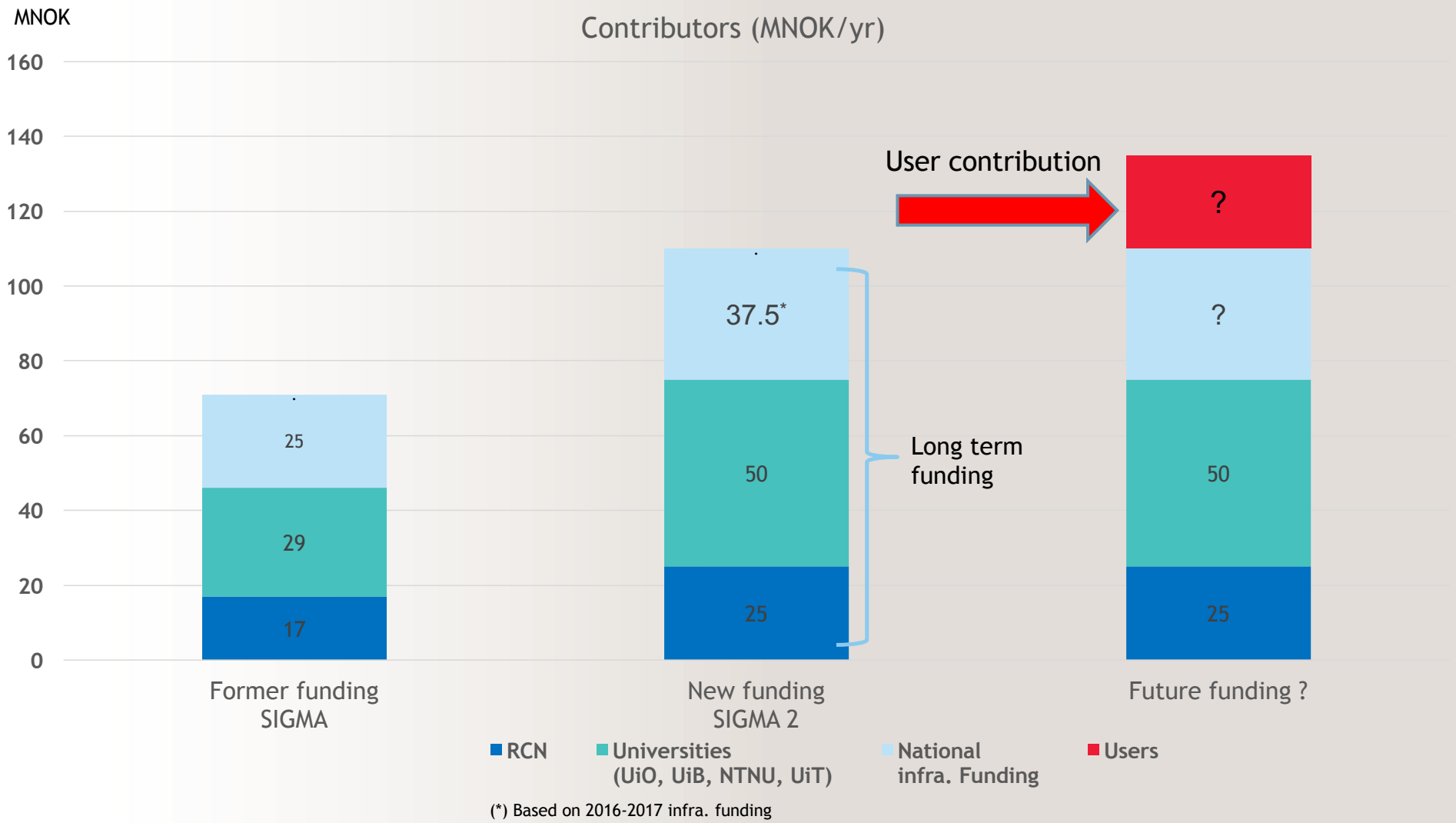
www.sigma2.no

Backup slides

Contribution (payment) model

- Are all these things free for users??

Sigma2 financing



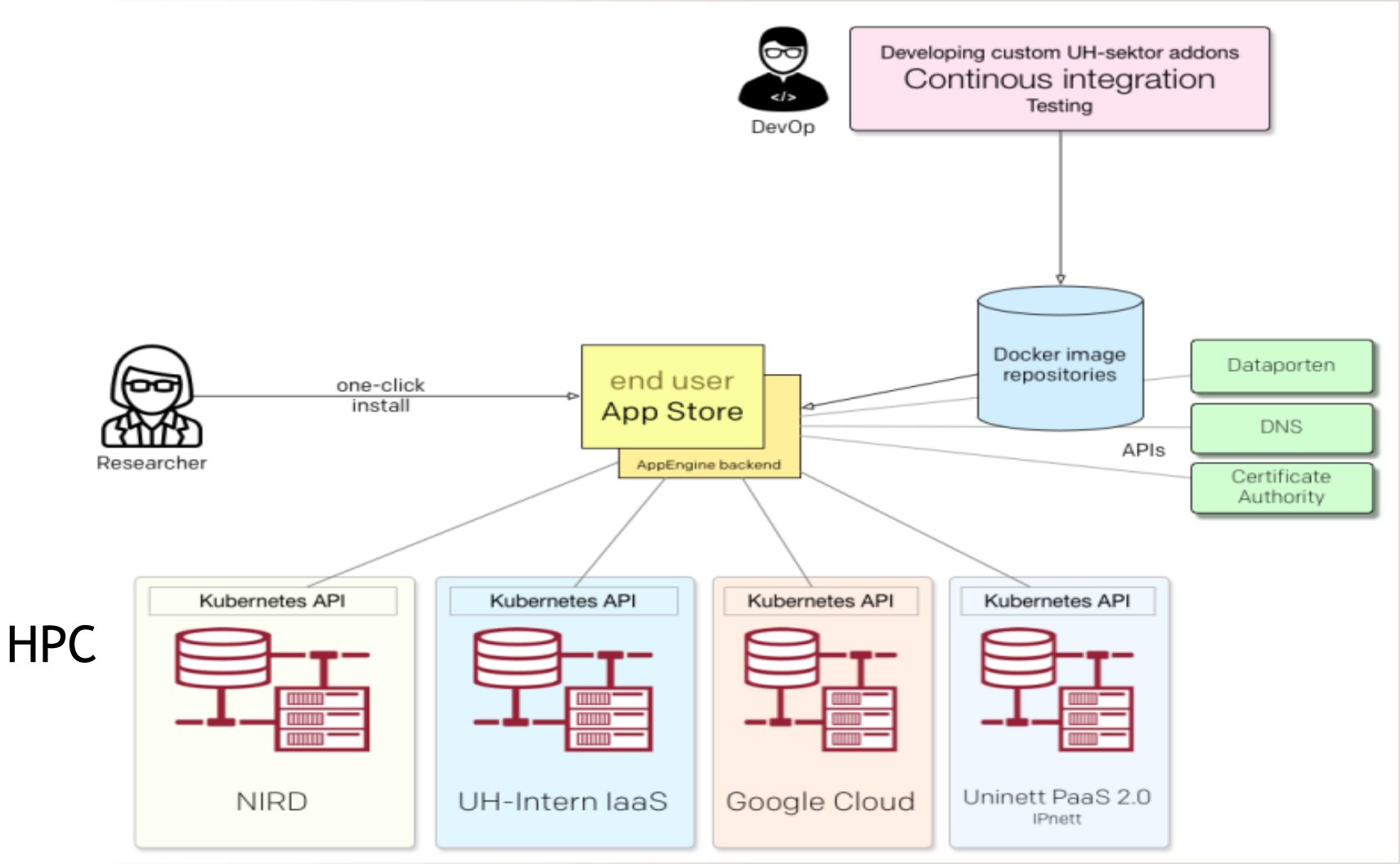
(*) Based on 2016-2017 infra. funding

UNINETT Sigma2

Contribution model: General principles

- Research data: All projects get X TB storage for **free*** on project area. Archiving research data is **free**.
- Compute resources: **free***.
- **(*)** Three exceptions:
  - A) Commercial research and industry
  - B) Large projects with EU or RCN funding
    Suggested definition of «large» 20 largest projects. (i.e. well above 4 million CPU hours pr. year)
  - C) Non-commercial projects needing **dedicated** or **special resources**
- This model is planned to be introduced during 2017/2018 so that existing research projects will get a reasonable time to adapt to these new rules and make provisions for this in their future applications for funding. (I.e. only projects with new funding from NRC/EU after 2017 where NRC has required budgeting for e-infrastructure resources. So far this is only valid for SFF and INFRA applications.)

A future common architecture?



Next round of procurements

High Performance Computing (HPC) resources

System	Sigma2 capacity (MCPUhrs/yr)	Tot. performance (TFLOP/s)	Deployed
Hexagon	12.8	109	4/2012
Abel	75.9	182	10/2012
Vilje	113.0	312	10/2012
Stallo	120.4	~291	10/2012 (+ utv.)
Colossus*	<13	~30	4/2014
Sum	322.1	894	
Fram	279.2	1071	10/2017
”B1”	?	?	(10/2018)
“HTC** platform”	?	?	(2H2018)

(*) For sensitive data, part of TSD

(**) HTC = High Throughput Computing / cloud platform

UNINETT Sigma2

Storage infrastructure

- Norstore is replaced by NIRD – National Infrastructure for Research Data

System	Capacity [PB]	Deployed	Location
Norstore	3.7	1/2013	Oslo (+Tromsø)
NIRD	5.6	9/2017	Tromsø + Trondheim
(NIRD exp.)	~10?	(2/2018)	

  - Research data archive
  - Project storage

Neste runde anskaffelser

- Søknad sendt NFR infrastrukturutlysning 2016 (INFRA 2016)
- Forarbeid inkluderte grundige undersøkelser av brukerbehov
  - Type ressurser
  - Kapasitet
- Søknad om 143 MNOK (Sigma2) + 40 MNOK (NeIC), totalt 183 MNOK
- Innvilget, men kutt på 32% (ned til totalt 125 MNOK)
- I forhandlinger med NFR, håper på litt bedring men ikke mye
- Endelig resultat i desember
- FOR-ANS-18 påbegynt 2H 2017

Capacity, usage and needs

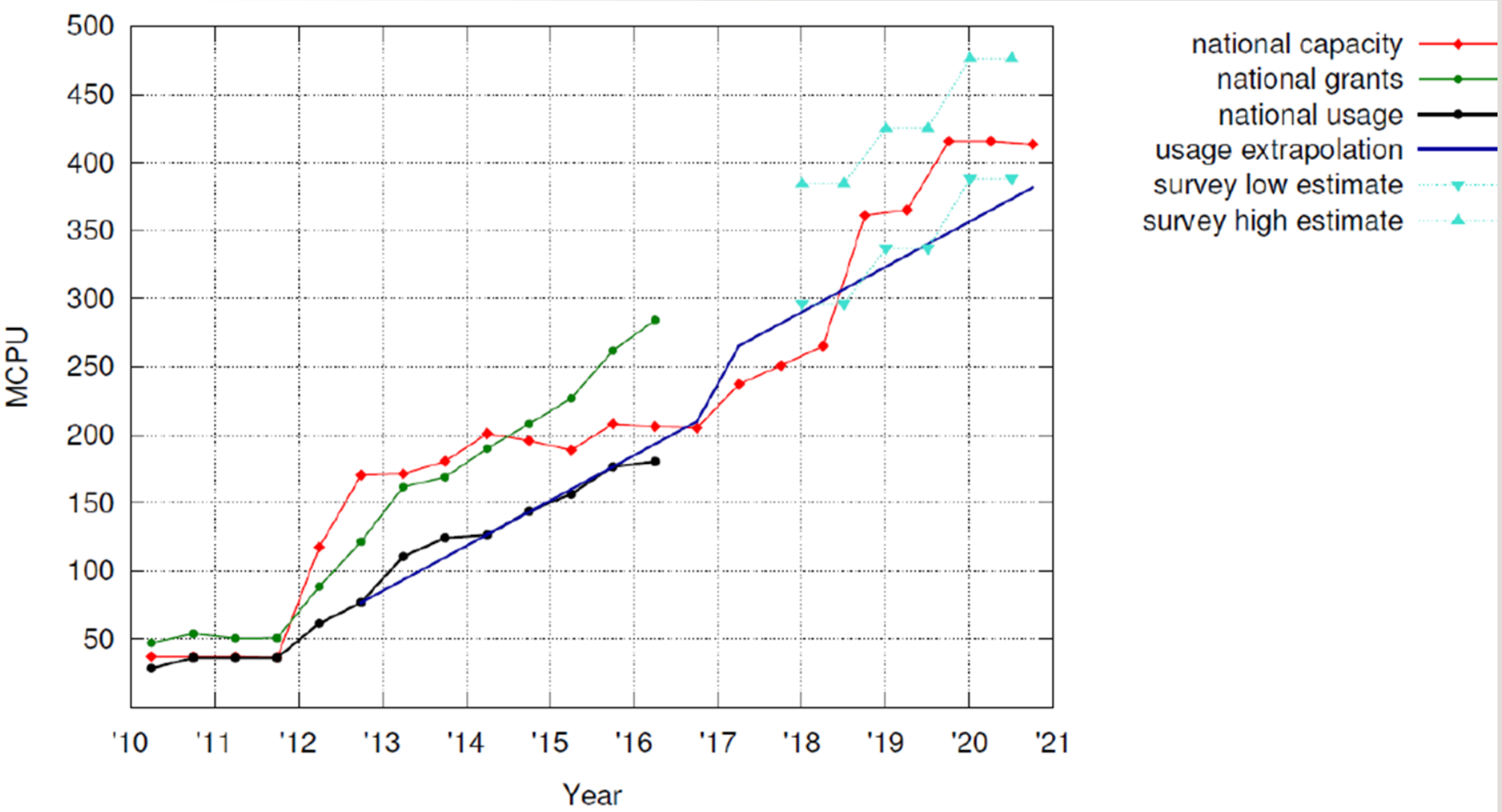


Figure 3.2 Historical national allocations together with usage extrapolations from historical data and user survey estimates.

Level of natl. compute capacity compared to other countries