Sigma2s National e-Infrastructure for Research

Maria Francesca Iozzi, PhD, UNINETT/Sigma2

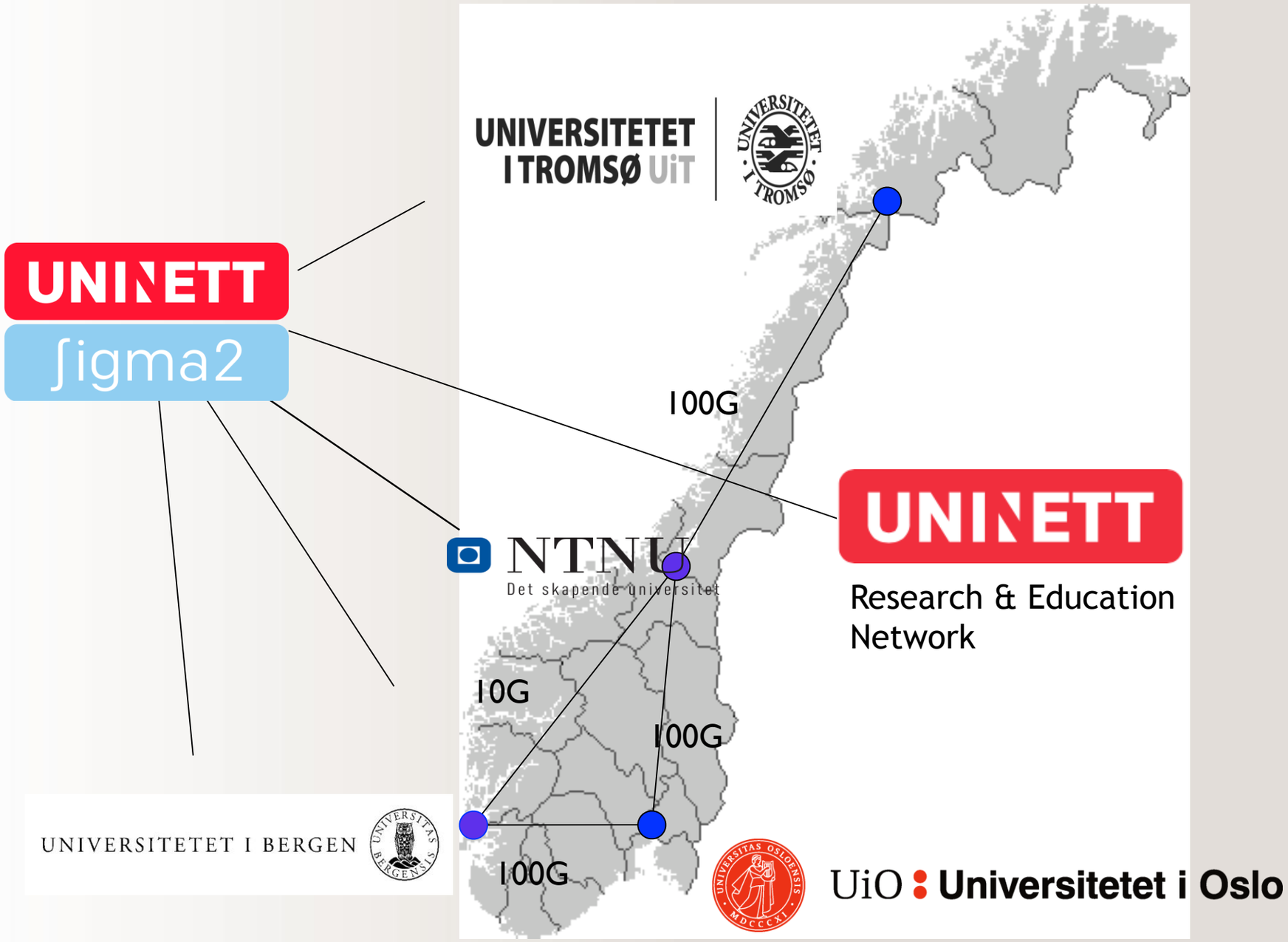
UNINETT Sigma2

Agenda

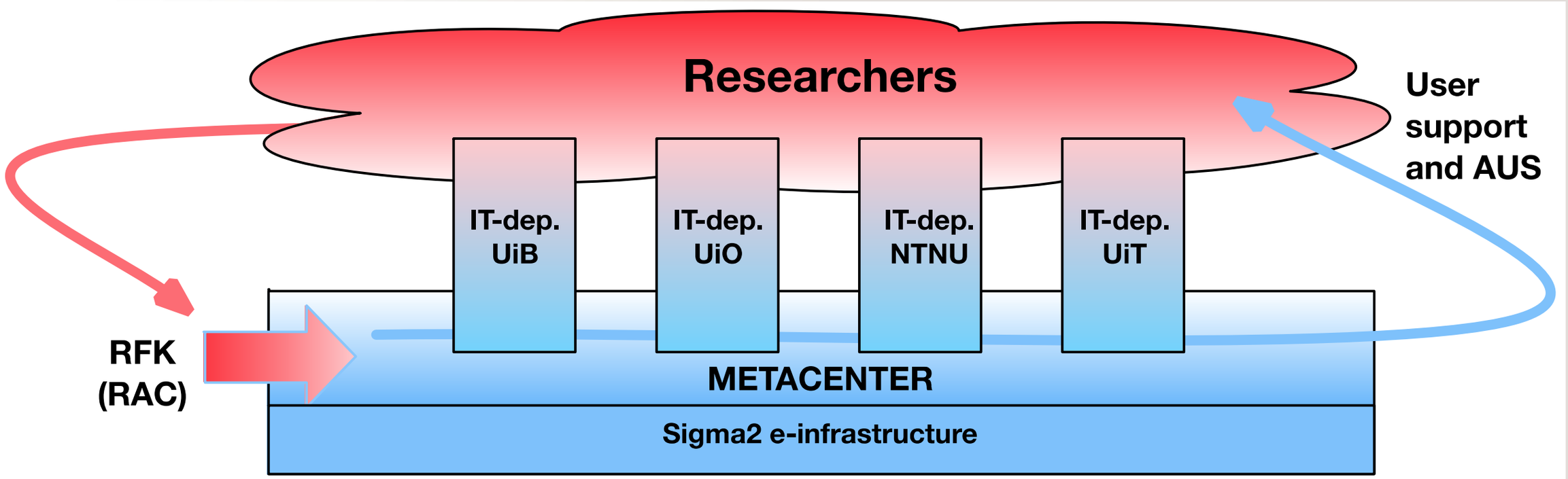
- About UNINETT Sigma2
- Sigma2 e-Infrastructure Services:
  - DMP
  - Storage
  - HPC and Data driven Computing
  - Archiving
- A future scenario for the Data Management?

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 4 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.**

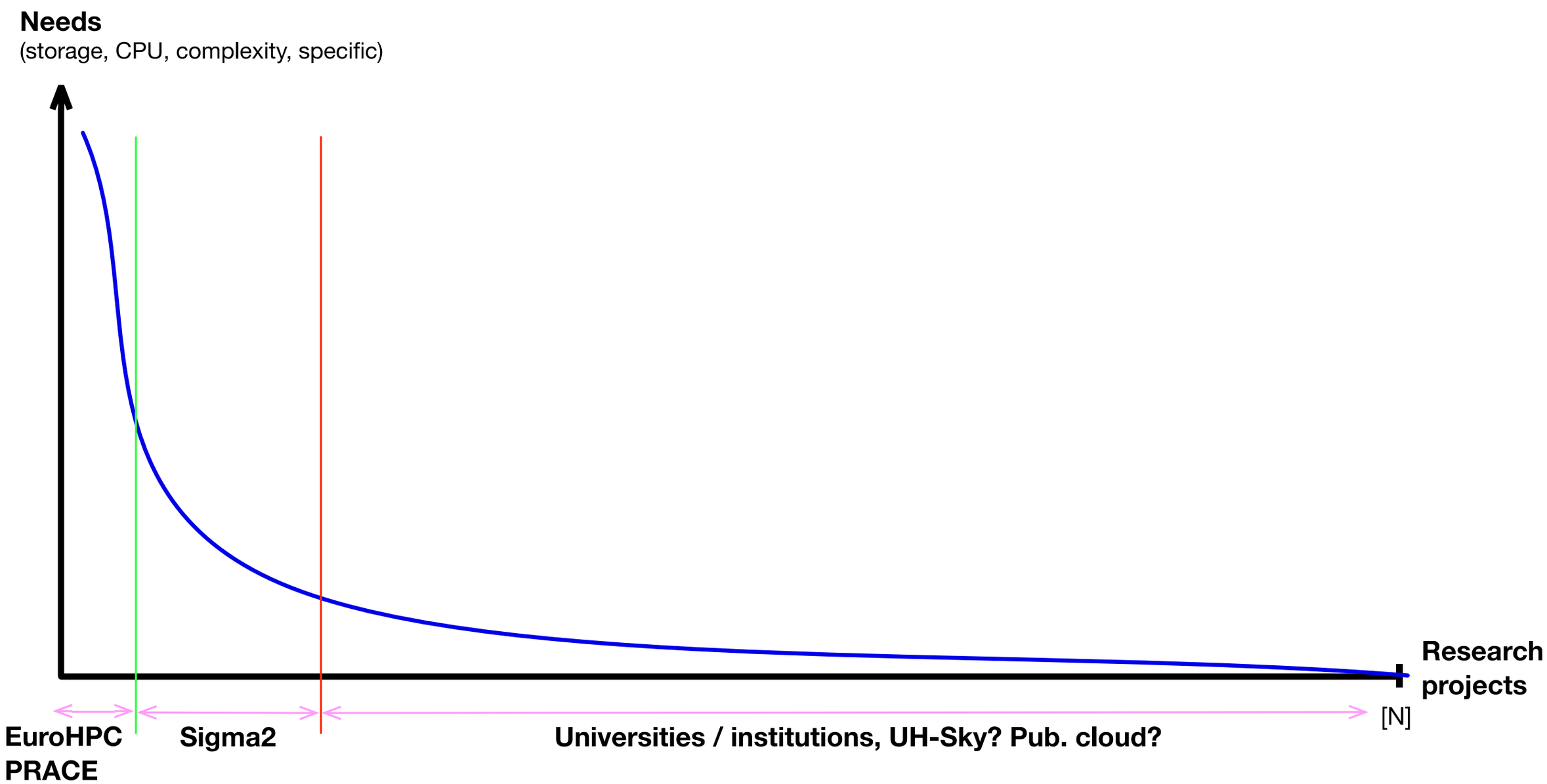

UNINETT
sigma2
UNIVERSITETET I TROMSØ UiT
100G
UNINETT
Research & Education Network
NTNU
Det skapende universitet
10G
100G
100G
UNIVERSITETET I BERGEN
UiO : Universitetet i Oslo

The Metacenter

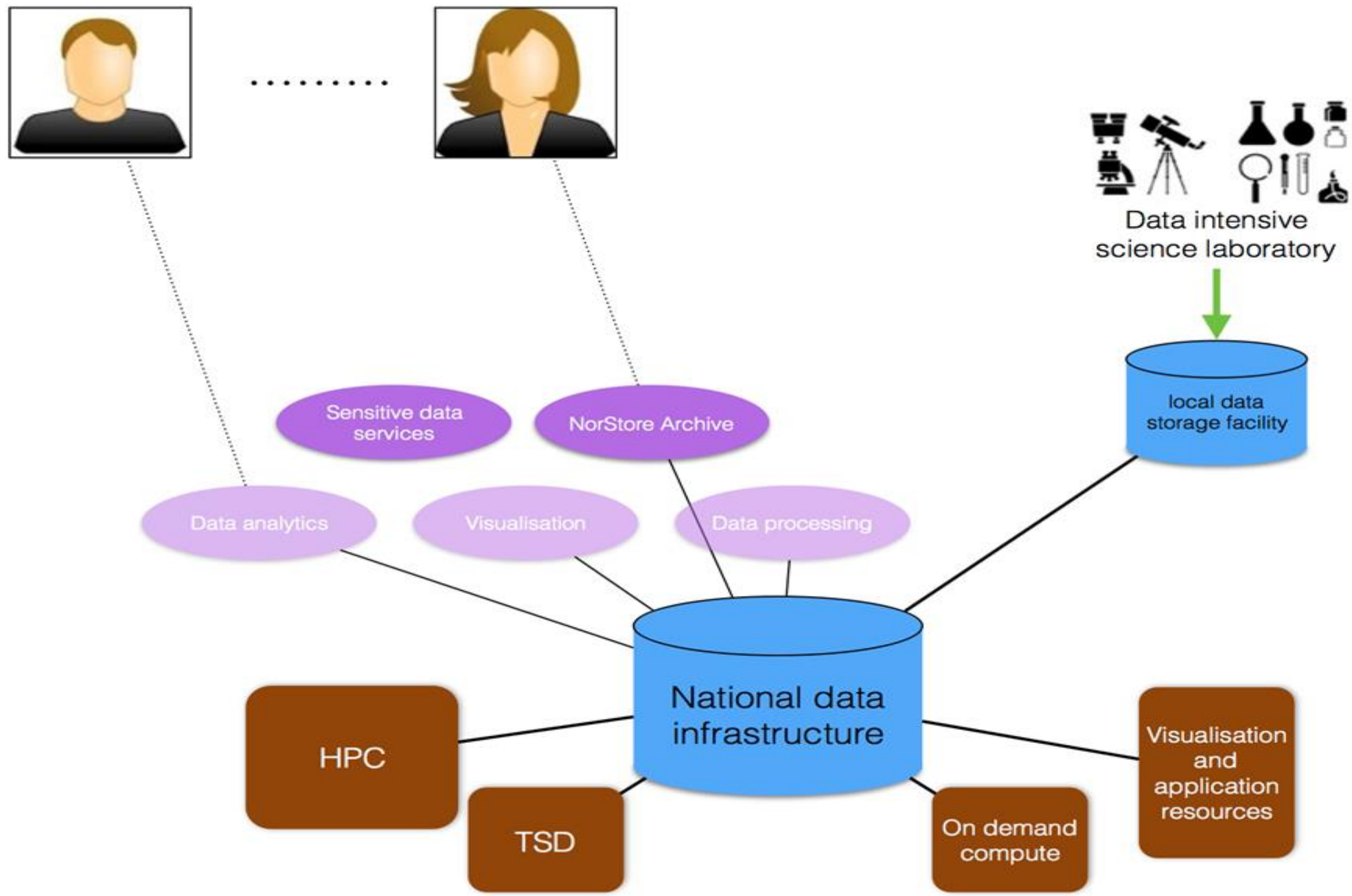


- National coordination and shared, consolidated resources have cost and efficiency advantages but creates a “distance” to the end-users (researchers)
- This is countered by keeping the support staff and competence near where the research is going on, at the universities
- Combined with a **data-centric**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

Who provides what?

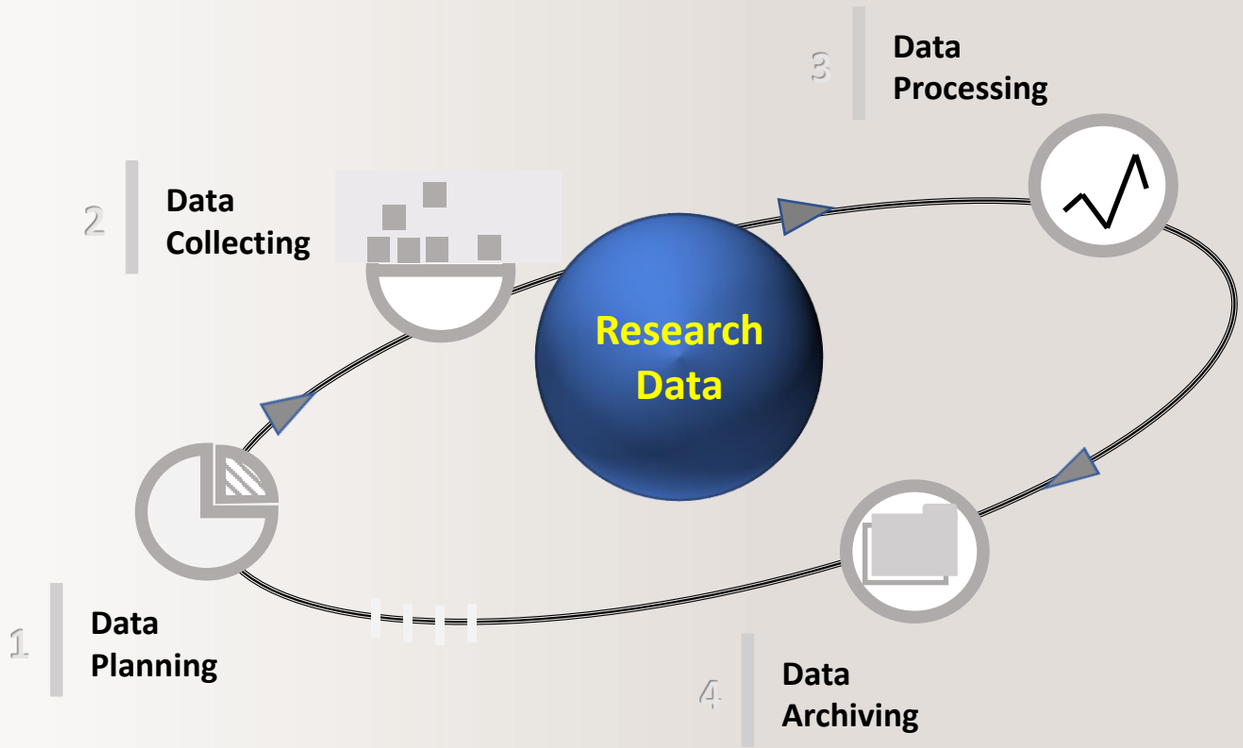


Data-centric architecture

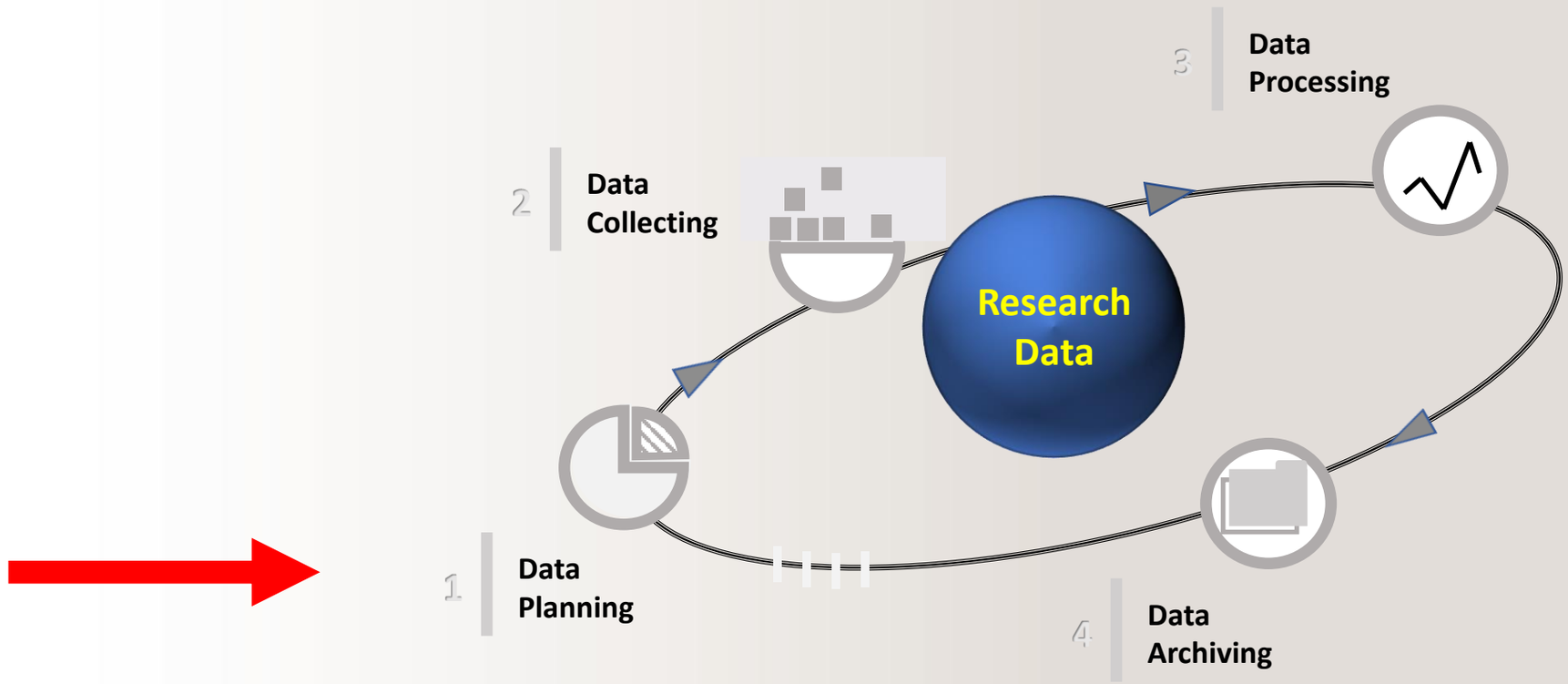


Sigma2 e-infrastructure services

The NIRD - National Infrastructure for Research Data - A complete and FAIR eco-system for research data



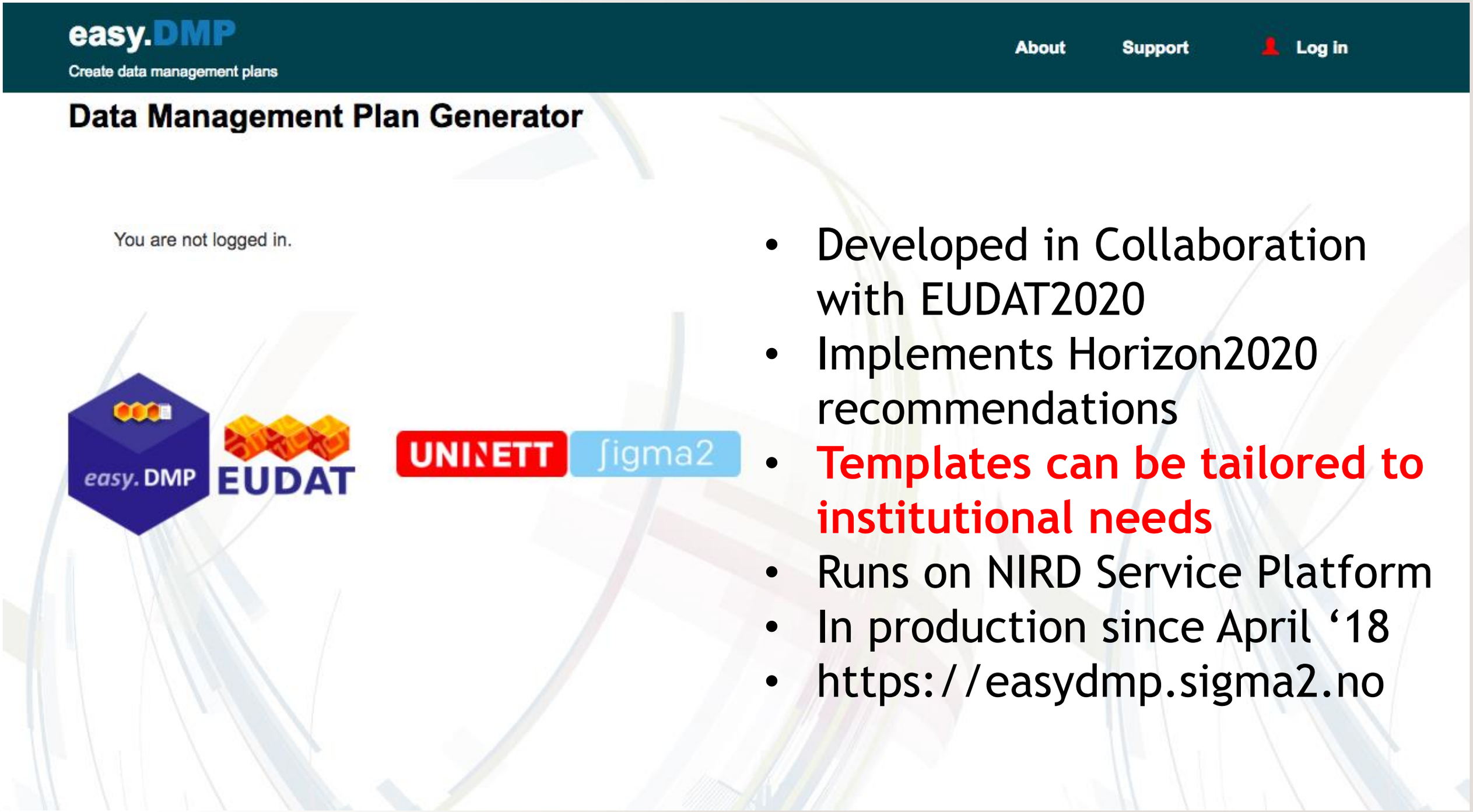
Data Planning: easyDMP



Motivation for easyDMP

- Data management guidelines, are by definition not specific.
- Researchers need help understanding and implementing the guidelines.
- There was a need for a tool that breaks down the guidelines and support institutional/community specific guidelines
- At the same time the tool can fulfil a self-learning role.

easyDMP



- Developed in Collaboration with EUDAT2020
- Implements Horizon2020 recommendations
- **Templates can be tailored to institutional needs**
- Runs on NIRD Service Platform
- In production since April ‘18
- https://easydmp.sigma2.no

easy.DMP
Create data management plans

Your plans | Create a new plan | About | Support | Log out

Test1000

Go to plan summary

Your answers are saved every time you press "prev" or "next"

1 Data Summary | 2 FAIR data | 3 Allocation of resources | 4 Data Security | 5 Ethical aspects | 6 Other

2.1 Making data findable, including provisions for metadata

Q2.1.1 Will you use metadata to describe the data?*

A structured description of your data will help you and your colleagues to use your data. Describing what the fields are, what's in the files and the structure of your data. You should think about using a metadata scheme that your community uses.

- (•) **Yes**
- ( ) **No**

More information

If you need to go more in depth, do so here. This will be shown in the generated text

Prev | Question 1/24 | Next

DMP Institutional Templates (II)

Test-psi-new

Go to plan summary

Your answers are saved every time you press "prev" or "next"

7 Hvordan sikrer prosjektet at personidentifiserbare data ikke lagres andre steder enn TSD?*

Kun PI har rettigheter til å eksportere dataene ut fra TSD

Shorter/simpler questionnaire with pre-filled answers.

Prev

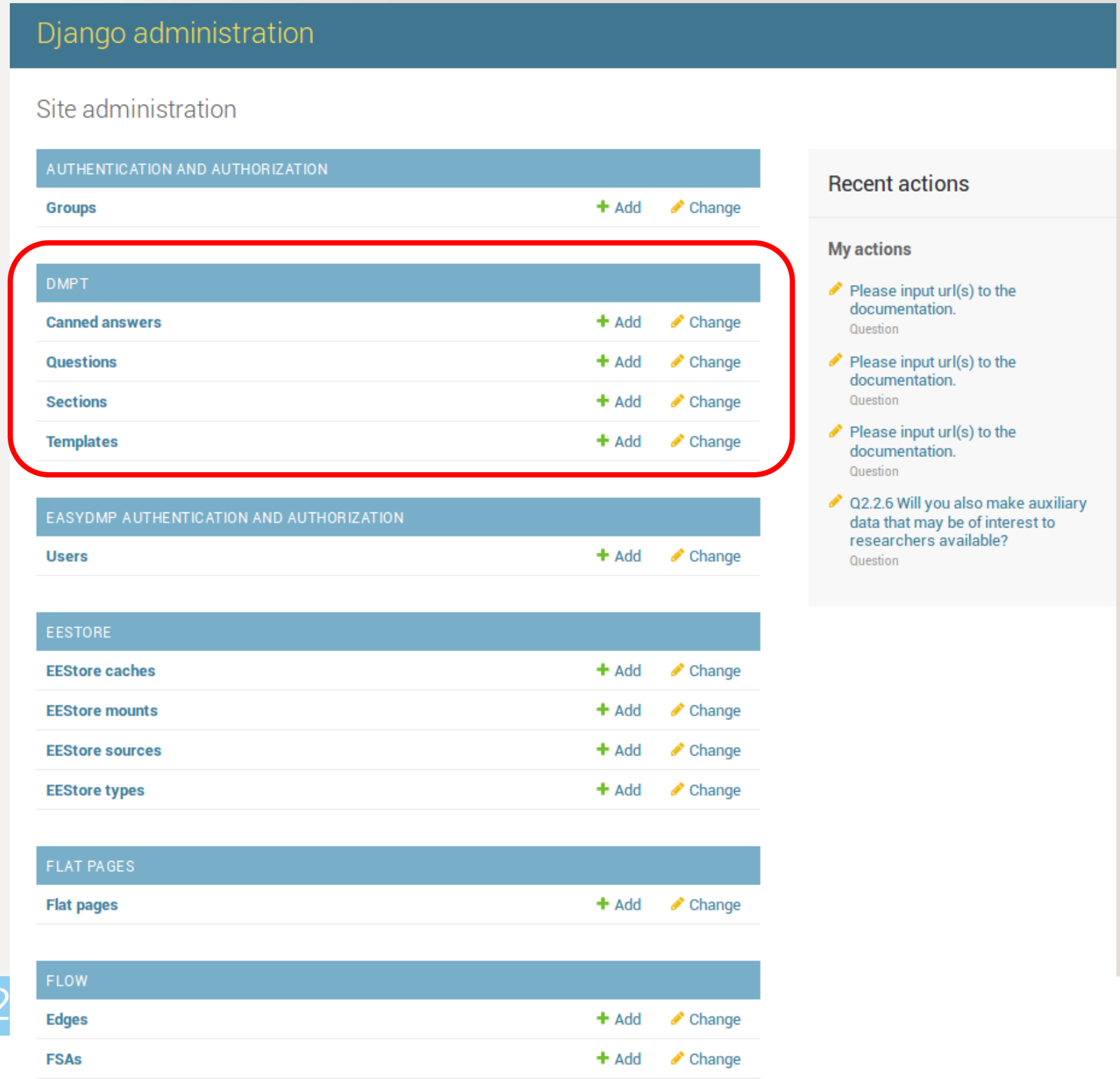
Question 7/8

Next

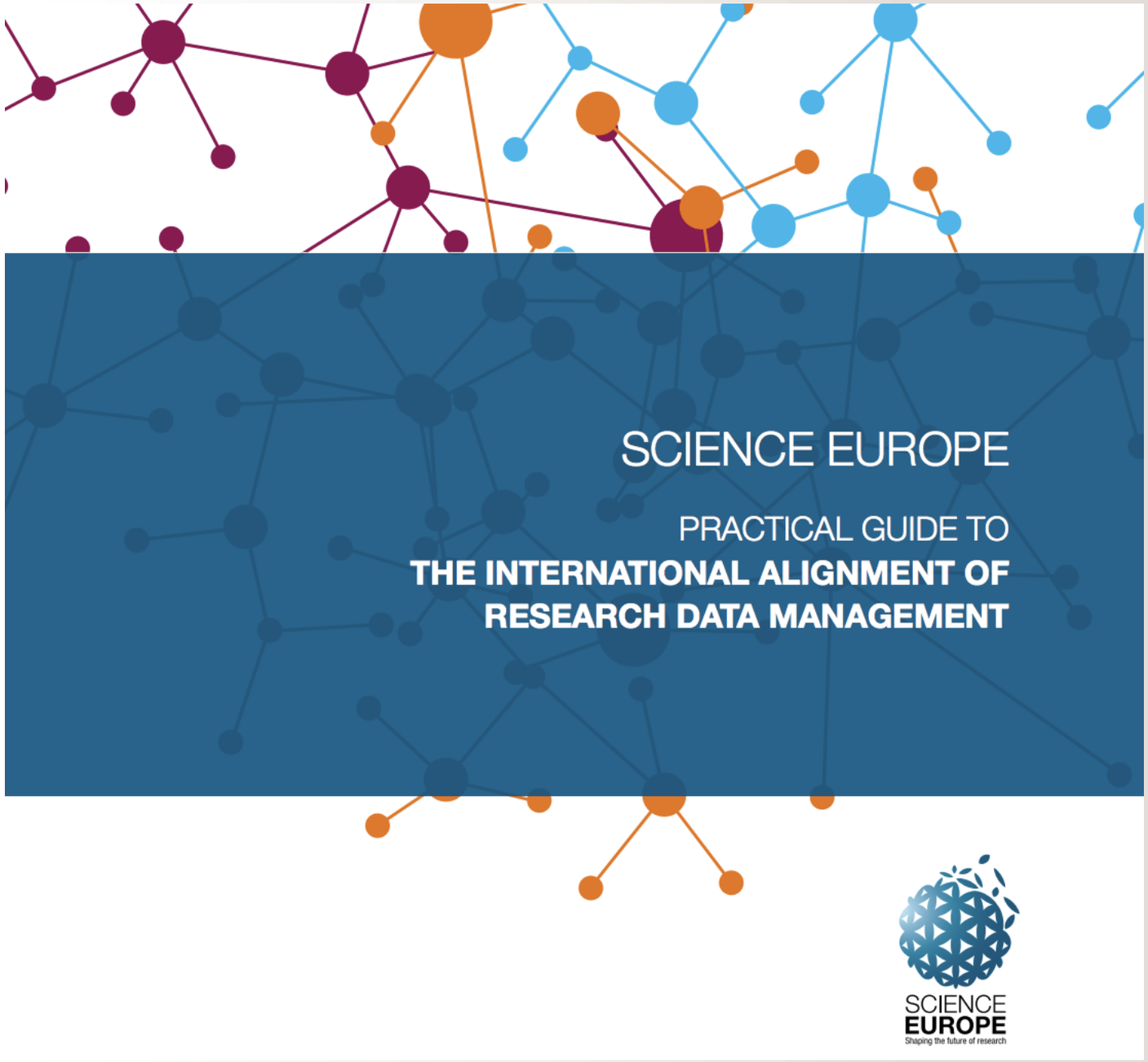
UNINETT sigma2

Customized DMP Templates

There is an admin interface that allows admins to create new templates or update existing templates.



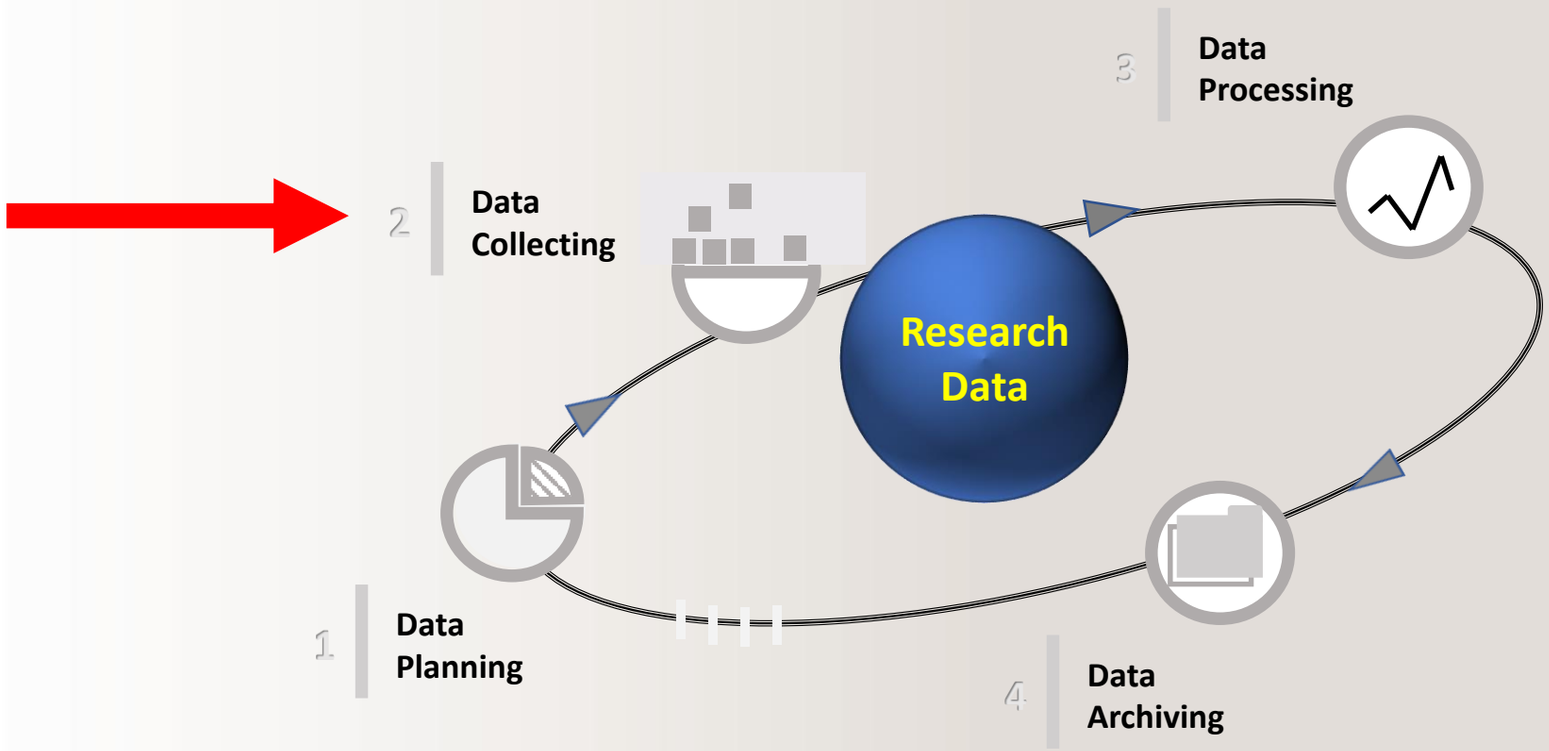
UNINETT Sigma2



- The Norwegian Research Council has adopted the Science Europe Template –

- The template will be soon available in easyDMP

Data Collection: Project Storage

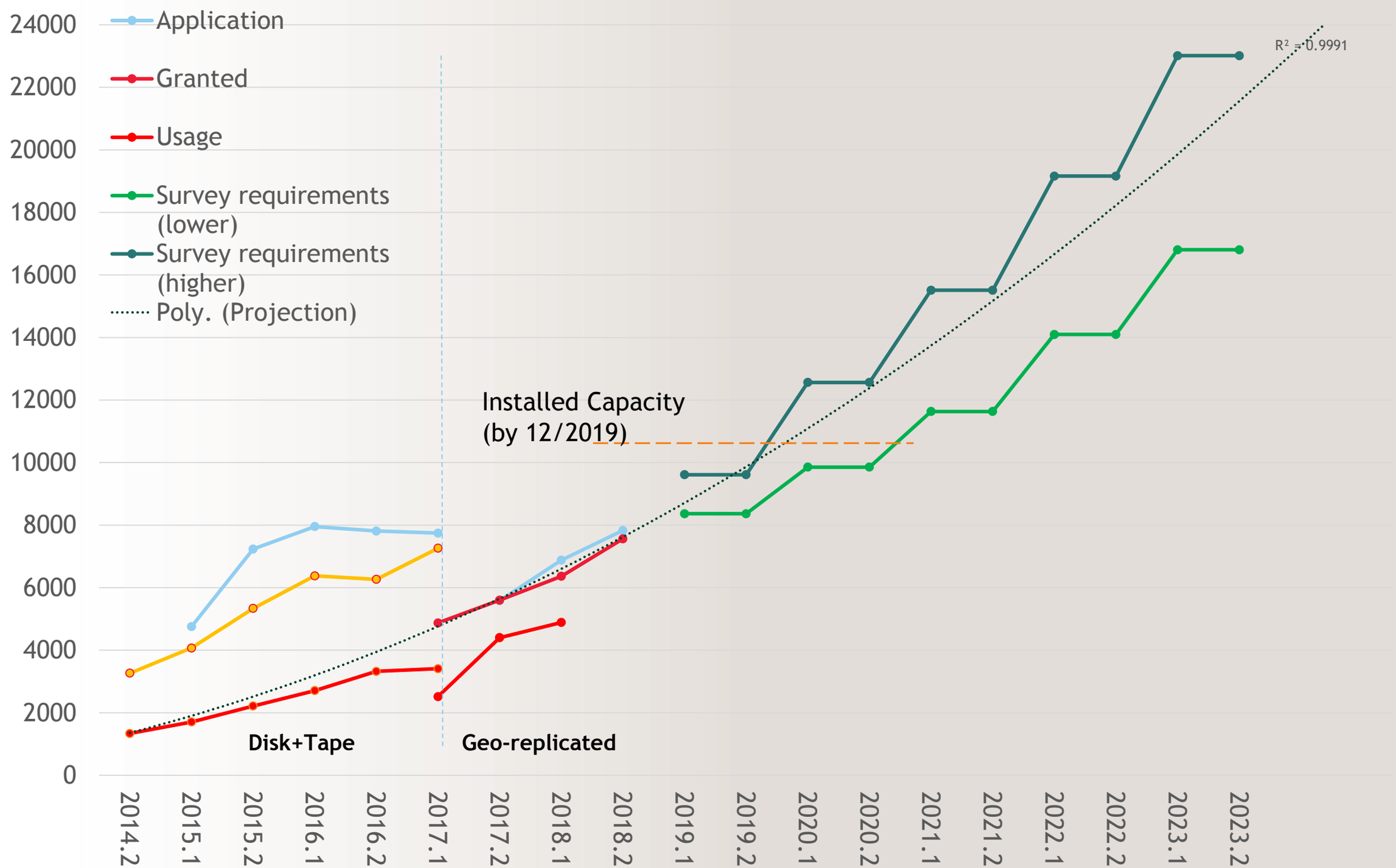


NIRD Storage

	Capacity	Allocated	Usage
NIRD (09.2017)	4807	5602	4403
NIRD (10.2018)	+/- 7400	+/- 8200	
NIRD (03.2019)	+/- 9200	?	

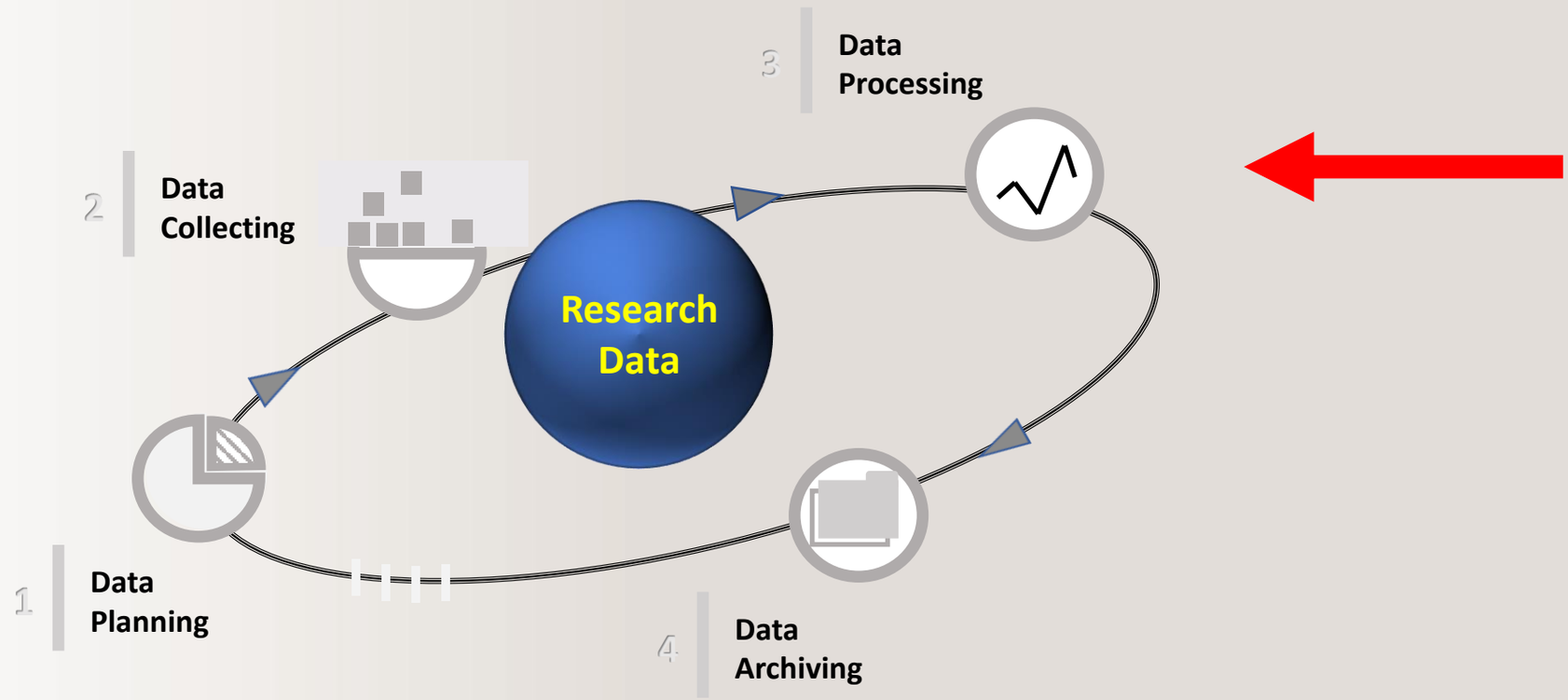
- All storage is disk-based with object storage interface (OpenStack Swift, S3)
- Geo-replication (async) + snapshots (daily)
- Storage divided into various areas covering data lifecycle: Project working area, data archive and backups

NIRD Storage - Trends

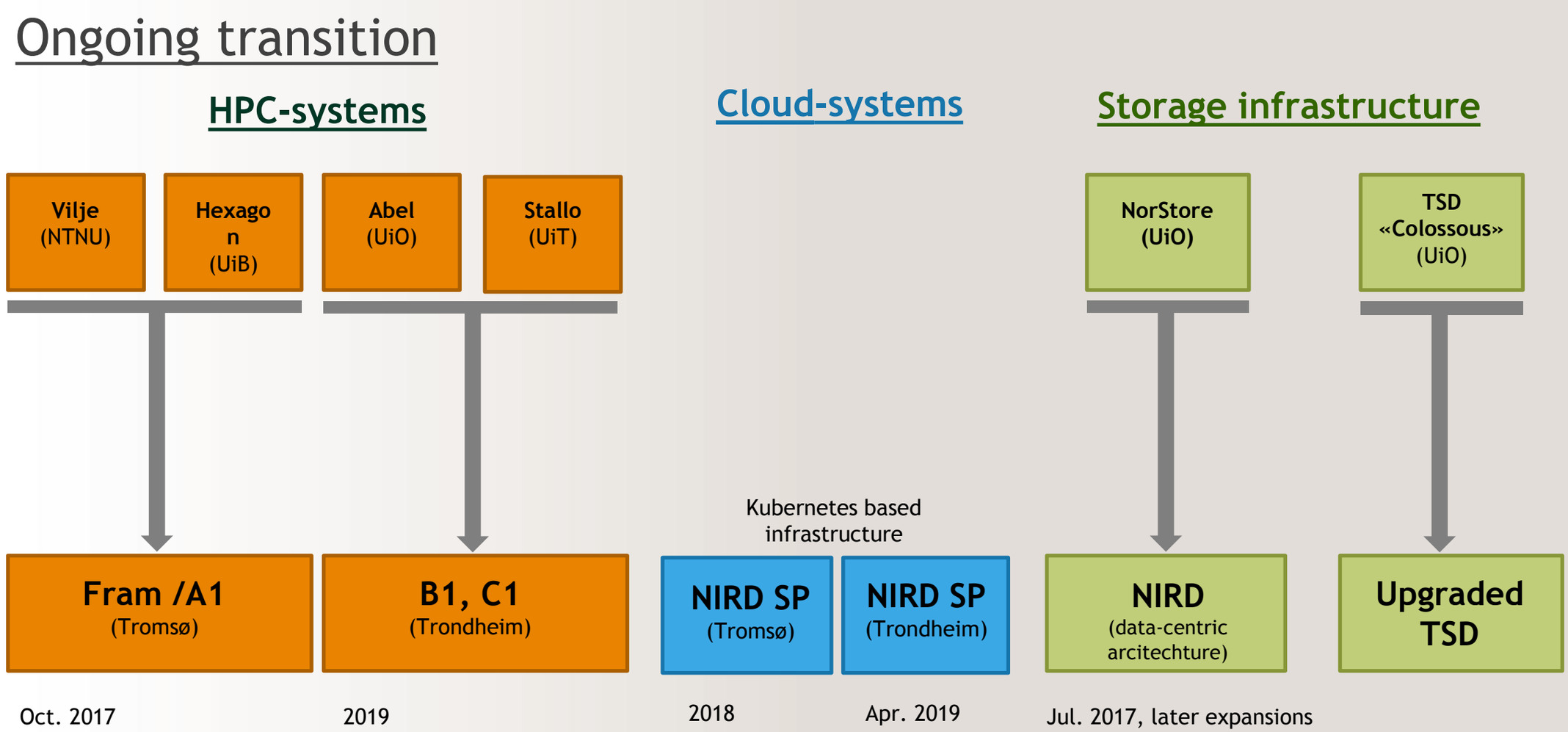


UNINETT Sigma2

Data Processing



Computing Services - Overview



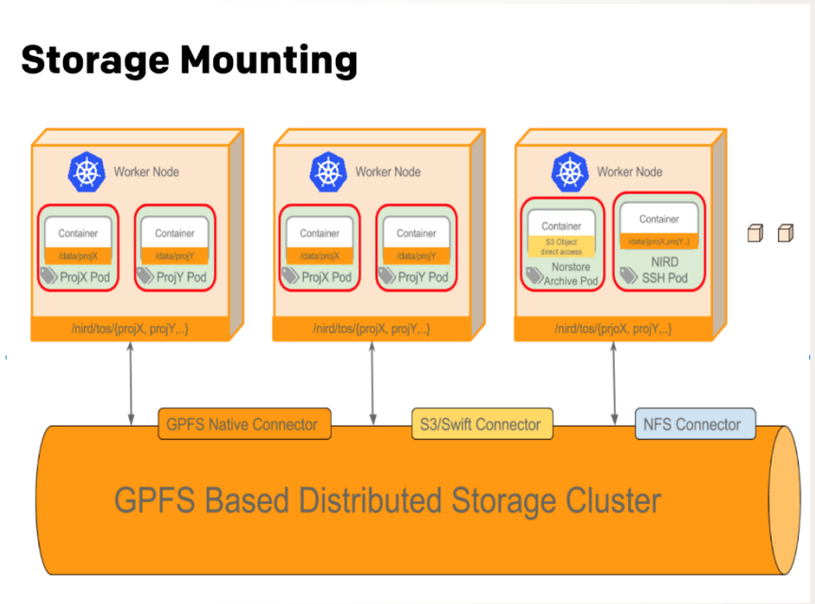
High Performance Computing (HPC) resources

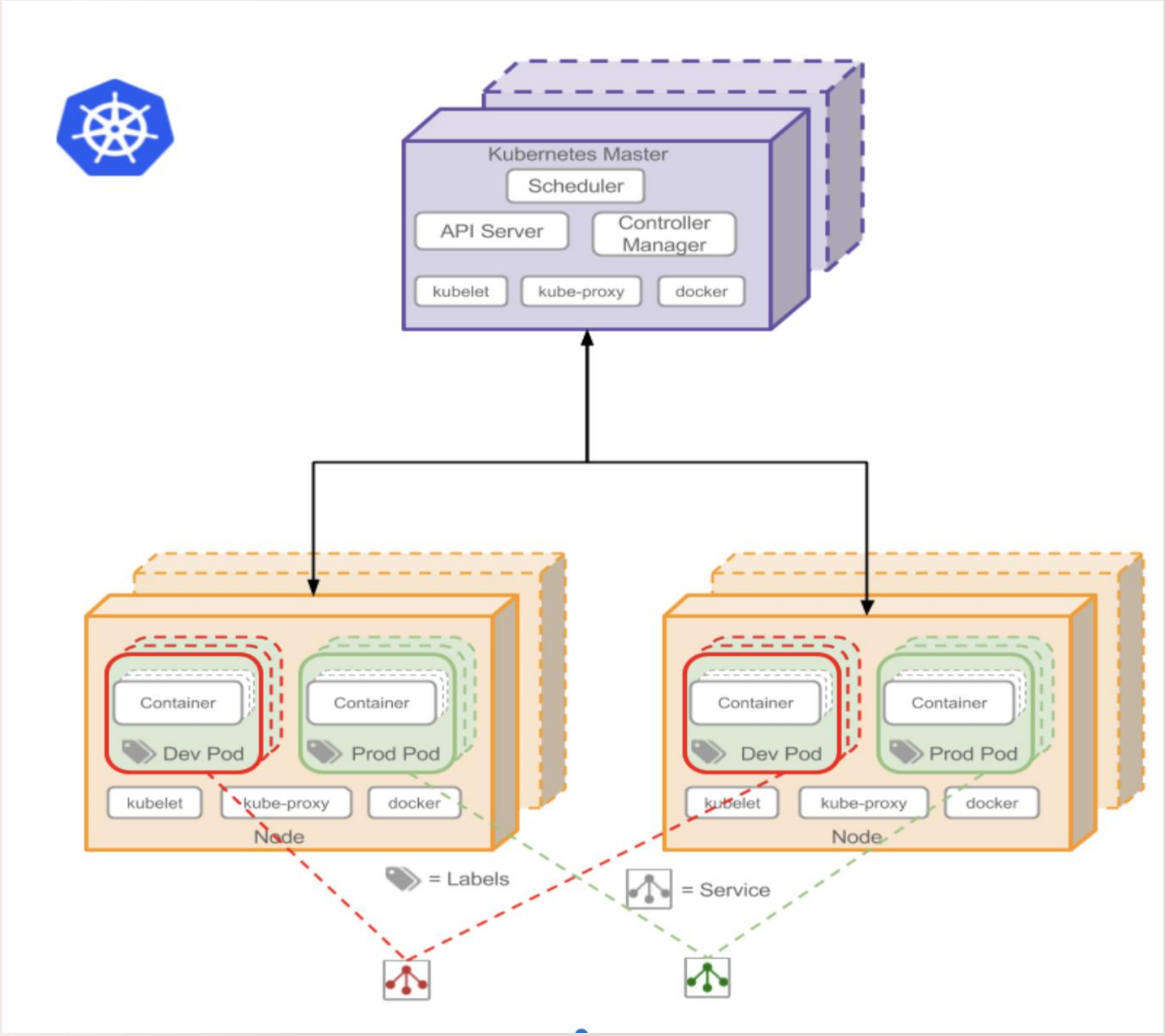
System	2018.2 capacity [MCPUh]	RFK Allocation (2018-09-13)	Deployed
Abel	38.0	35.6	10/2012
Stallo	59.8	56.6	10/2012 + ext.
Fram	129.2	123.0	11/2017
Vilje	98.1	97.1	10/2012
Sum*	**325.1**	**312.3**	(96%)
B1	?		(2019)
Saga	?		(2019)
HTC** platform	?		(2018)

(*) Not including approx. 11M from late applications
(**) HTC = High Throughput Computing / cloud platform

Data Processing: NIRD Service Platform

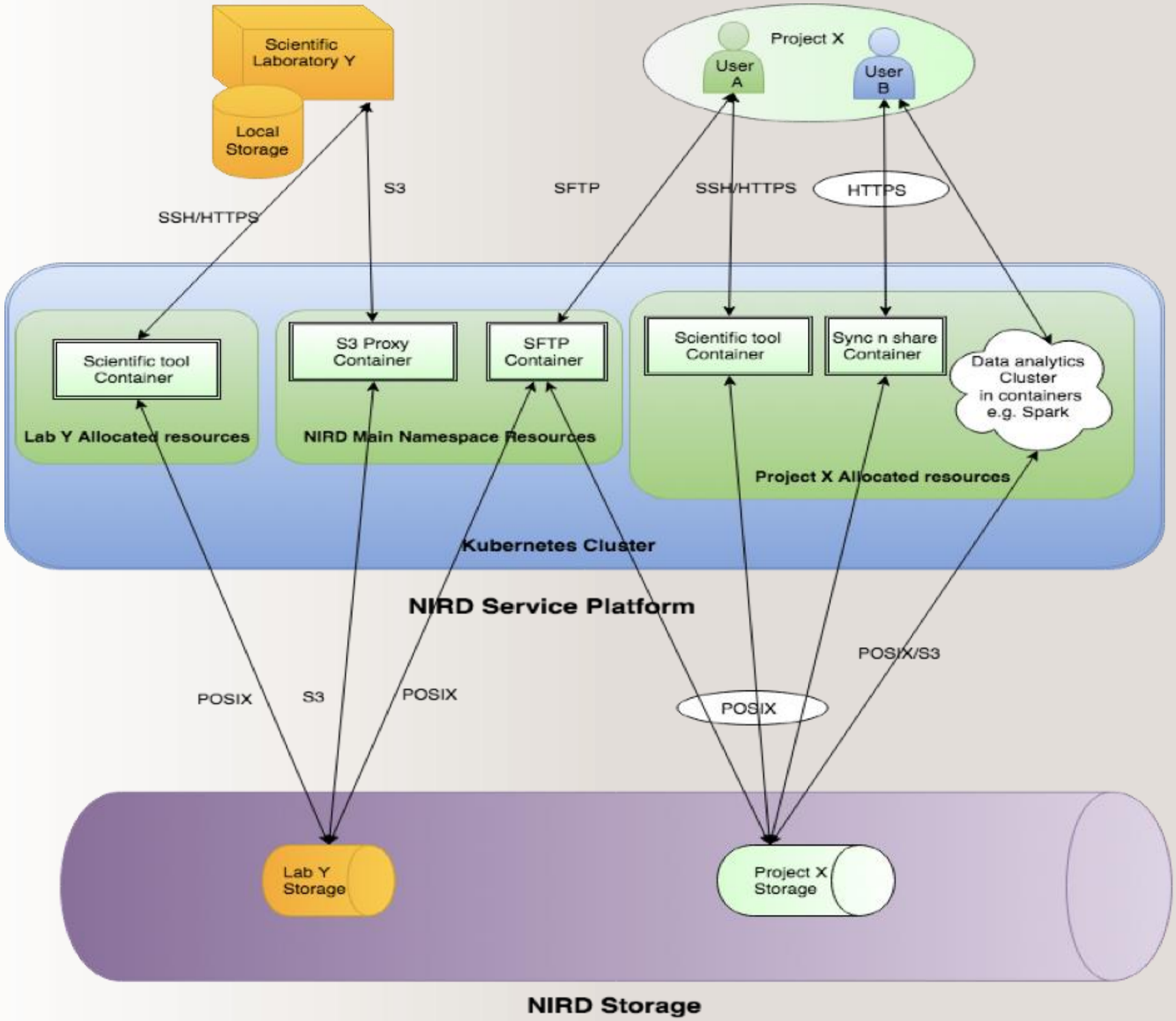
A **kubernetes-platform** running on computing nodes that mount the GPFS distributed **storage**. Services run into **docker containers.**



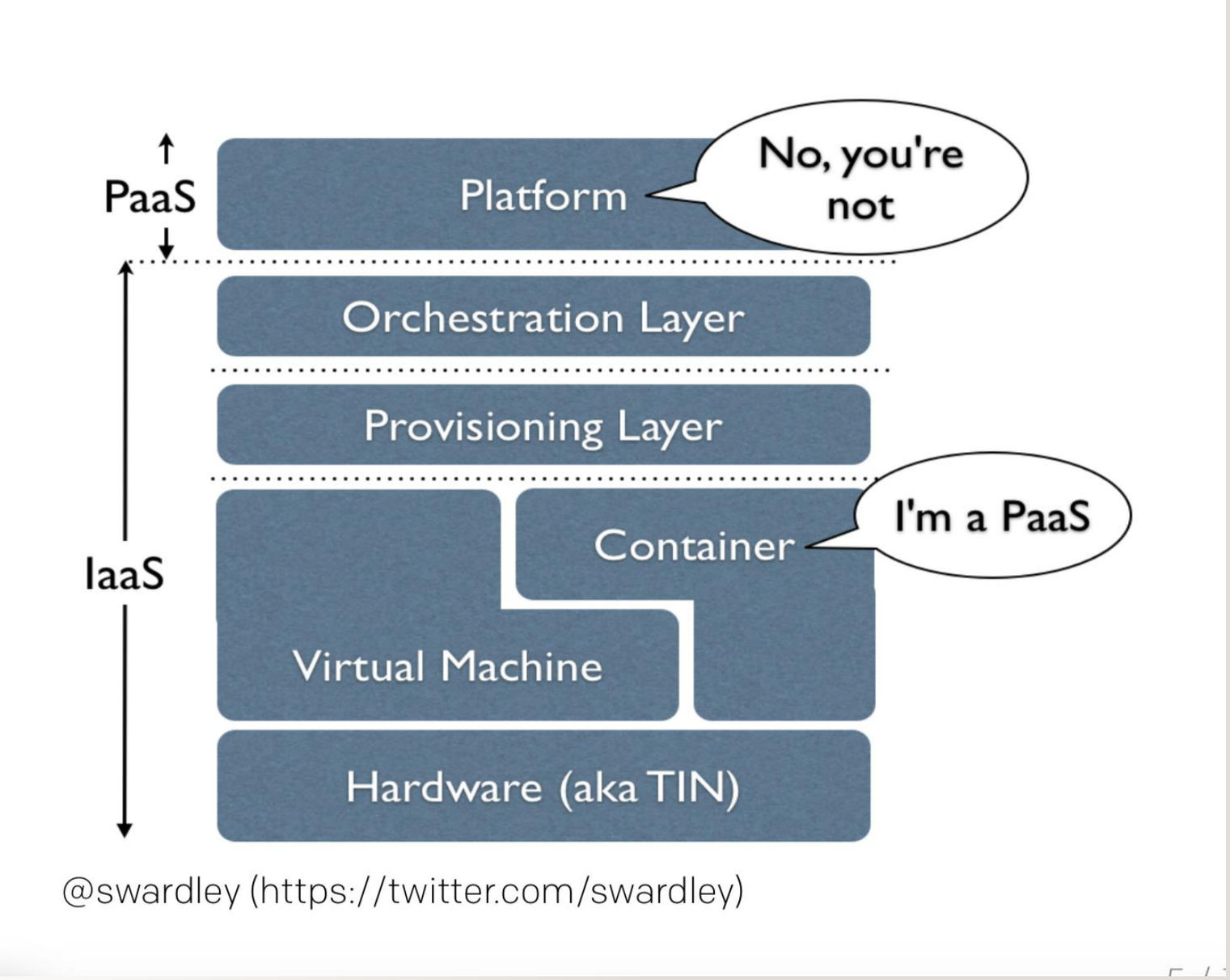


UNINETT ∫igma2

NIRD Service Platform

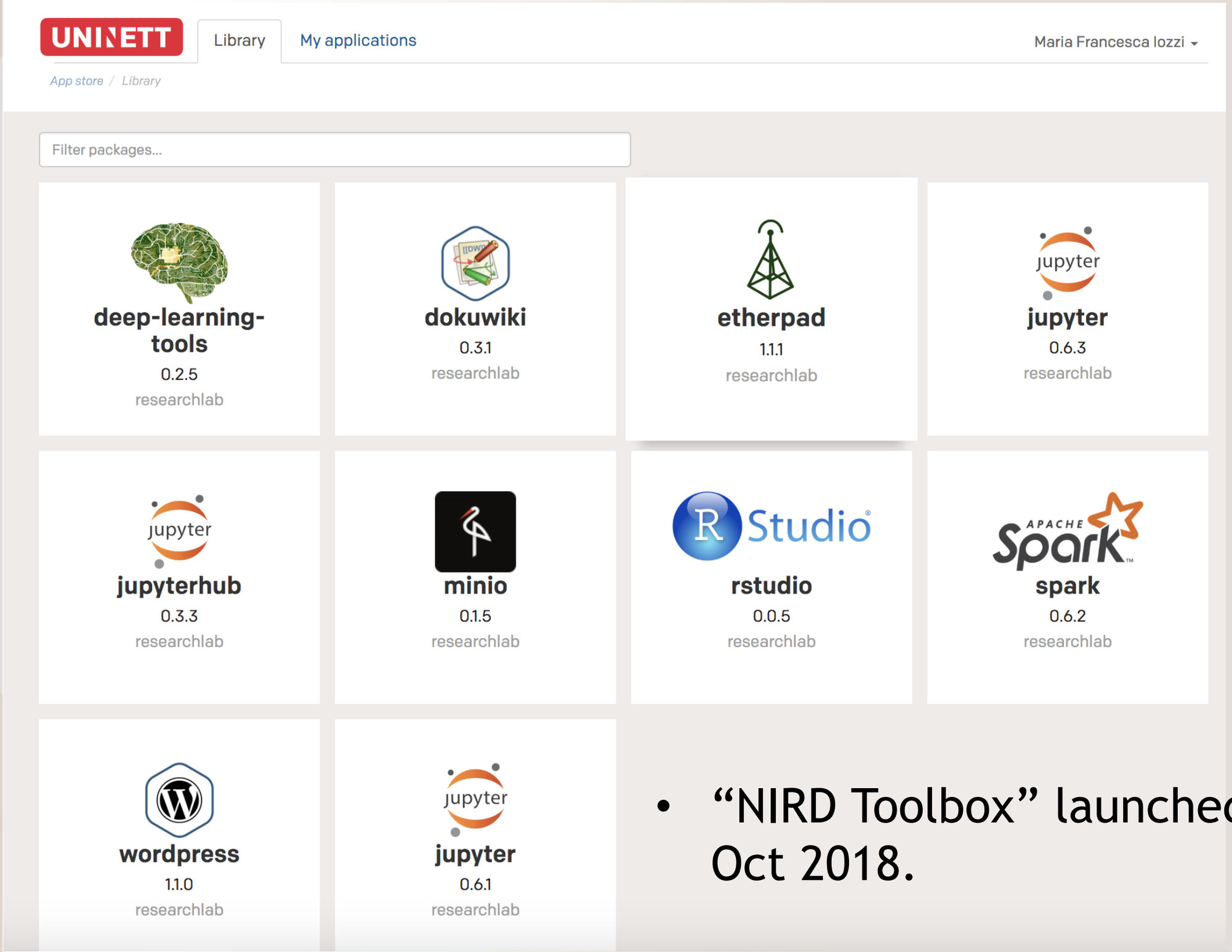


PaaS or NOT.PaaS?



A PaaS needs a resorse manager!

NIRD Toolbox



- “NIRD Toolbox” launched by end of Oct 2018.

UNINETT Sigma2

Services for sensitive research data

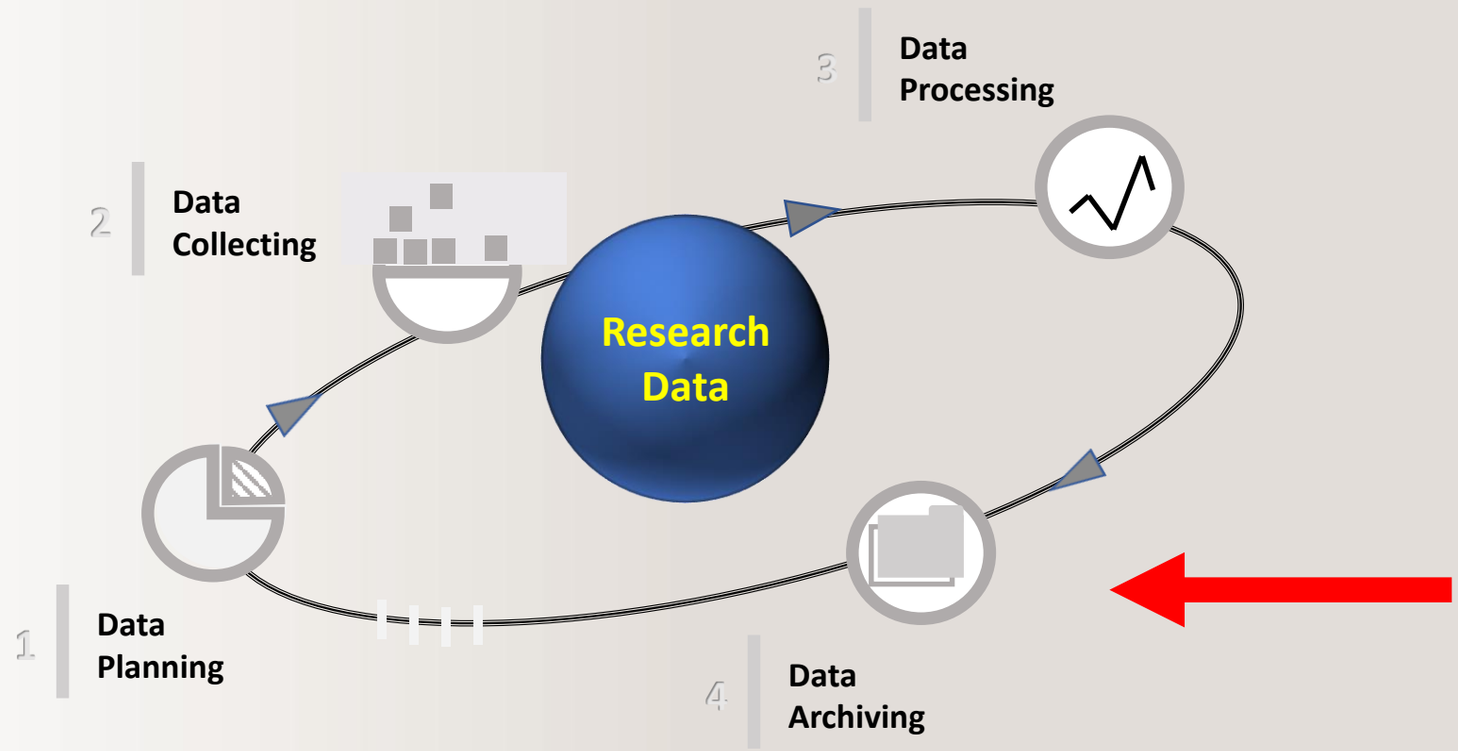
- In early 2018, Sigma2 and UiO/USIT has signed a contract to make the TSD part of the national e-infrastructure

- As a result of this contract, the uptake of the TSD will be facilitate for the projects that:
  - Need HPC compute resources
  - Need 10 TB or more storage capacity

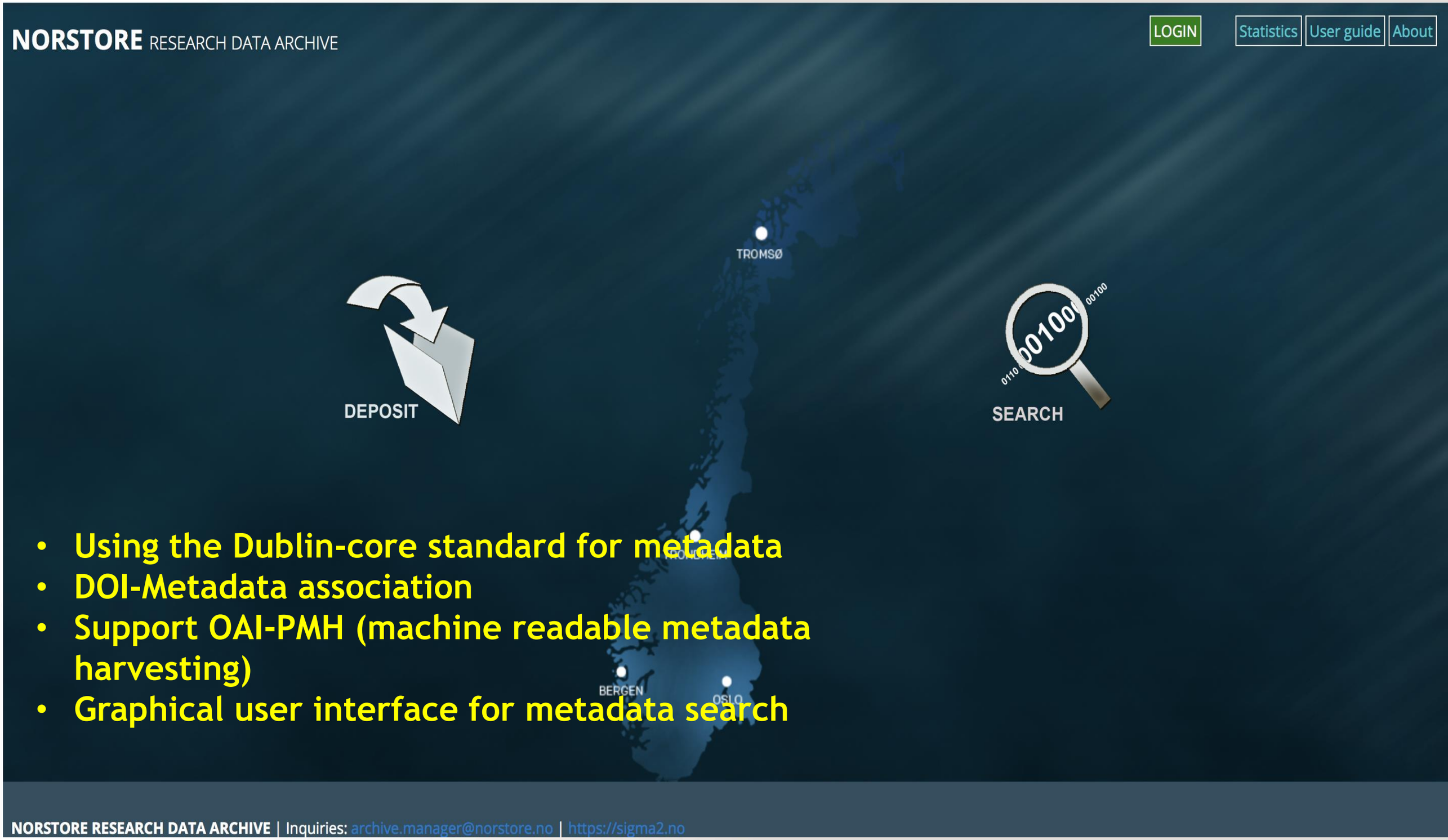
- The costs associated with this services will be then lowered if allocated through Sigma2

- The application process will be the same as for the other services in the SIgma2's infrastructure

Data Archiving: Research Data Archive



NIRD Archive



- Using the Dublin-core standard for metadata
- DOI-Metadata association
- Support OAI-PMH (machine readable metadata harvesting)
- Graphical user interface for metadata search

UNINETT Sigma2

Metadata – essence for research data

- A must-have for credible research data

nometadata.org

Data Management - National Collaborations

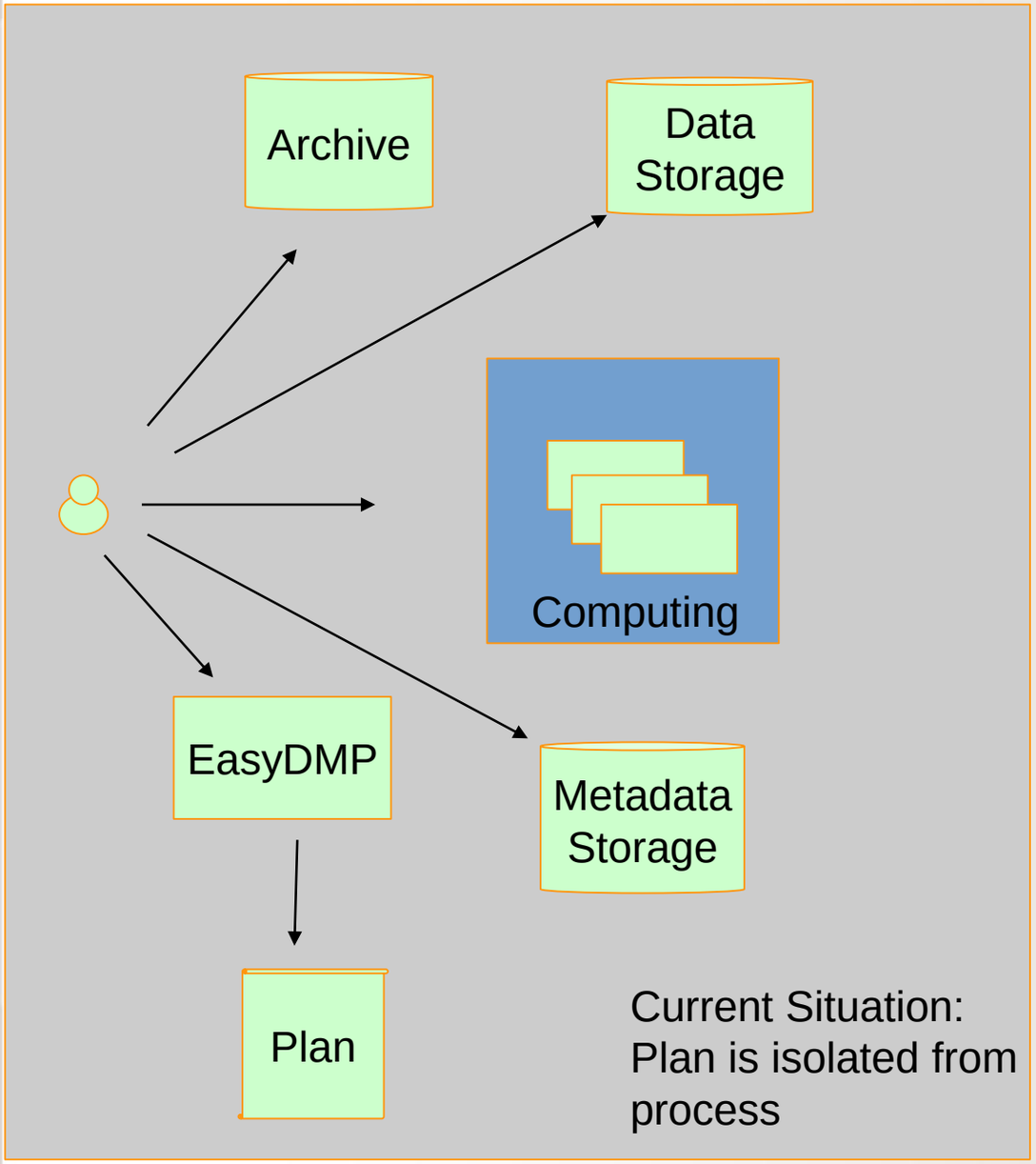
- Starting a tighter collaboration with NSD to develop a national DMP Tool
- A national reference group for DMP tool soon established
- Proposal for a Norwegian RDA node under developpement – will be sent in April 2019 → Forum for user driven discussion on data management

Data Management - Nordics Collaborations

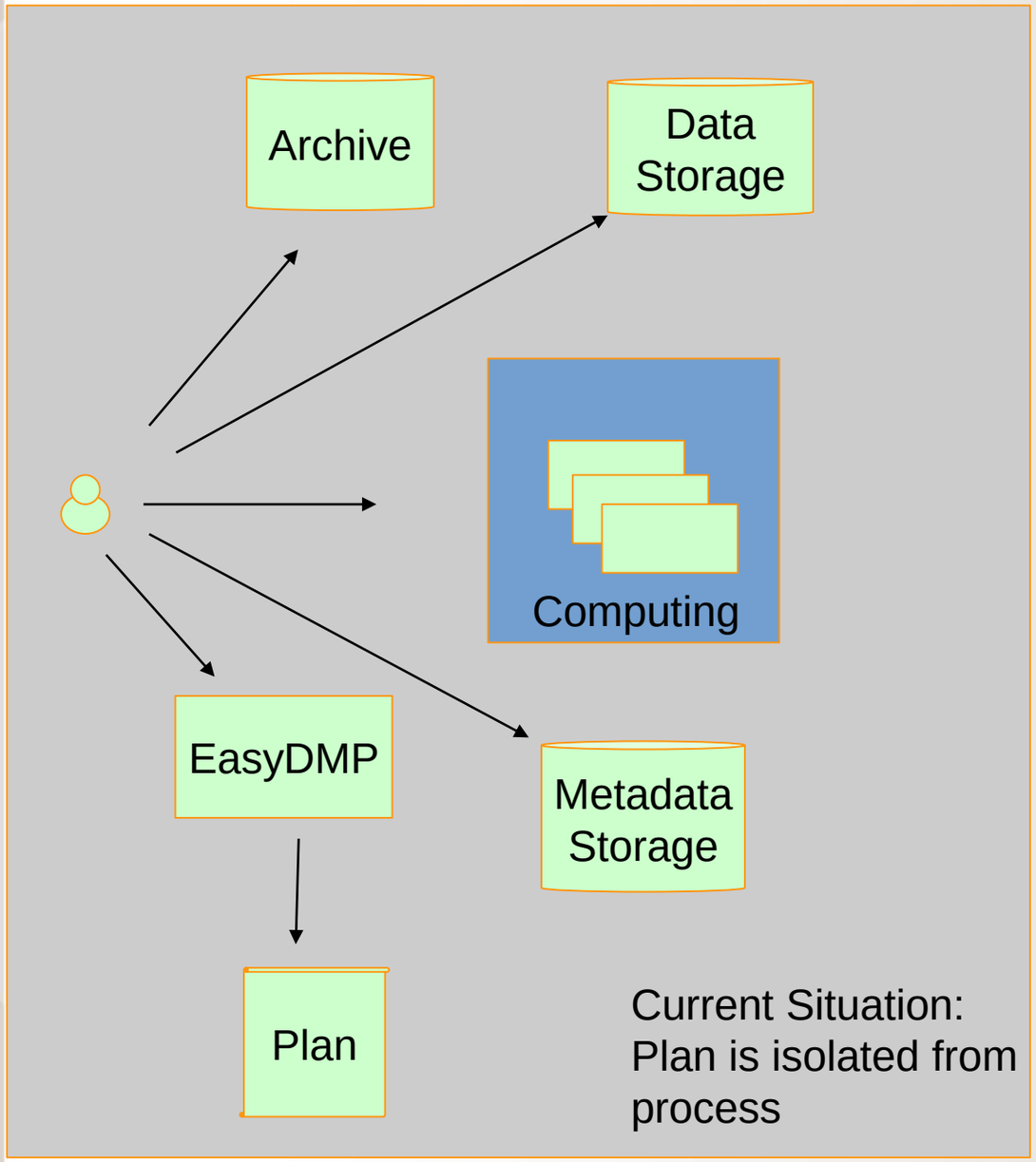
EOSC Nordic project granted by the EOSCINFRA-05, starting in May-Aug 2019:

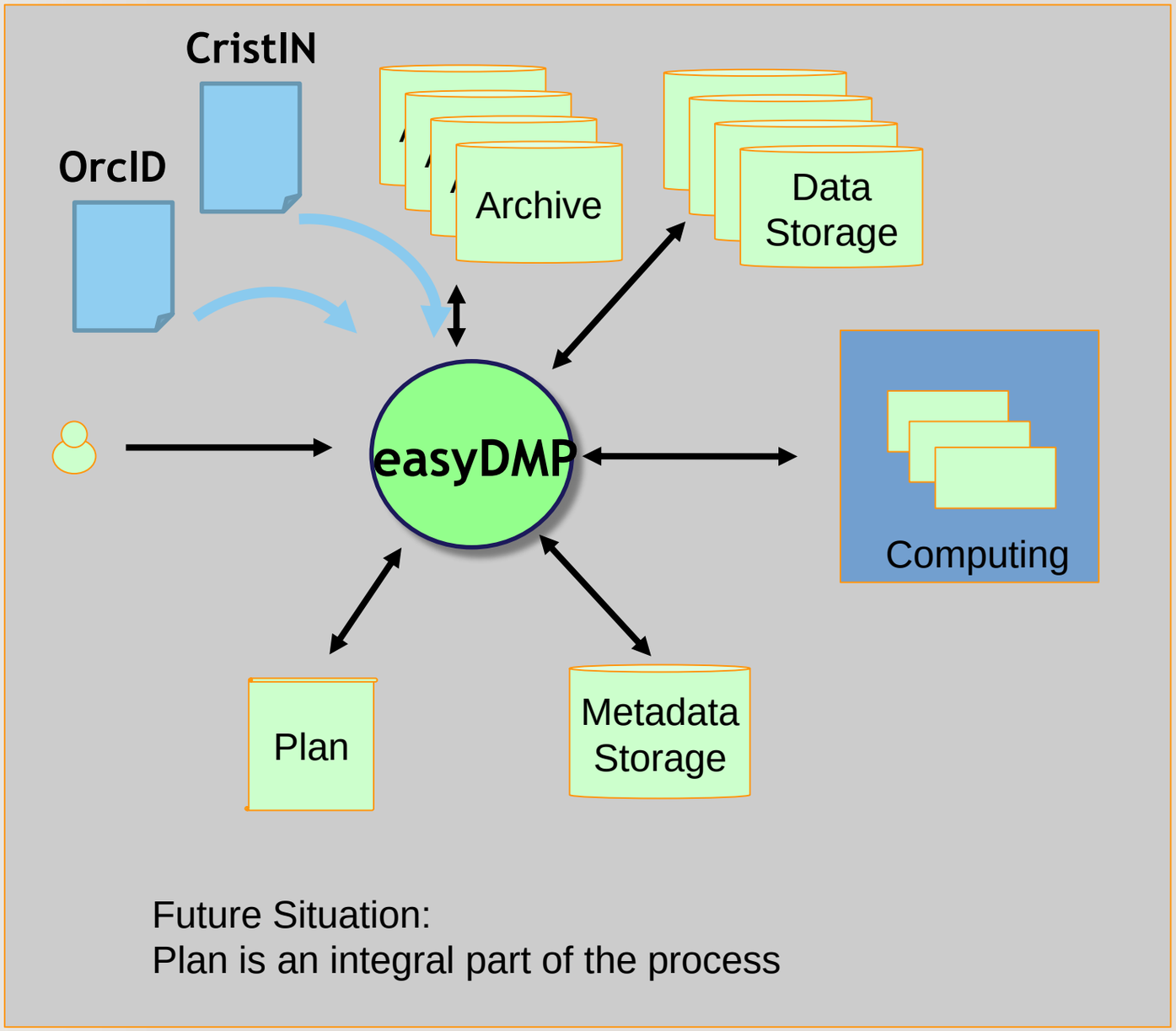
- Circa 6Meuro over 3 years – All Nordics countries involved
- Sigma2 leading WP5 (2MEuro) – ”Nordics Demonstrators”
- T5.3 about Nordic distributed data management workflows
- Actionable DMP tool (Collaboration with GoFAIR)

Currently …


Archive
Data Storage
Computing
EasyDMP
Metadata Storage
Plan
Current Situation:
Plan is isolated from process

A possible future scenario?





UNINETT ∫igma2