



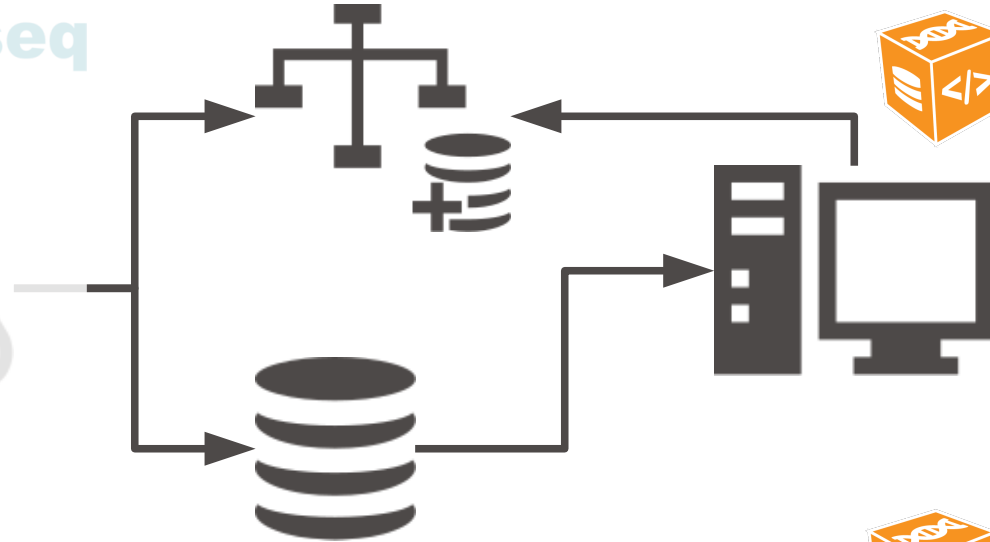
Data Storage Platforms



Korbinian Bösl
Data manager ELIXIR Norway
11 March 2020

NeLS

Norwegian e-Infrastructure for Life Sciences



sensitive
data



NeLS



Feide ola@uib.no



My Projects



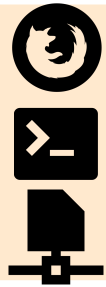
NeLS



Norwegian e-Infrastructure for Life Sciences

Data curation

NeLS



 **Feide** ola@uib.no
My Projects

NeLS

Norwegian e-Infrastructure for Life Sciences

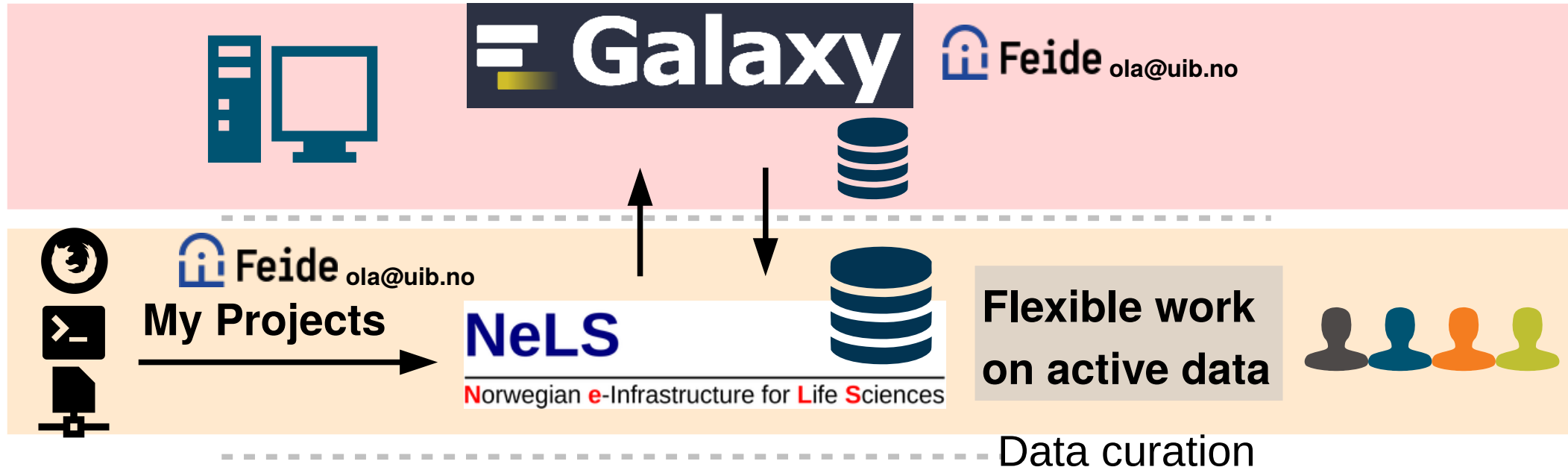


**Flexible work
on active data**

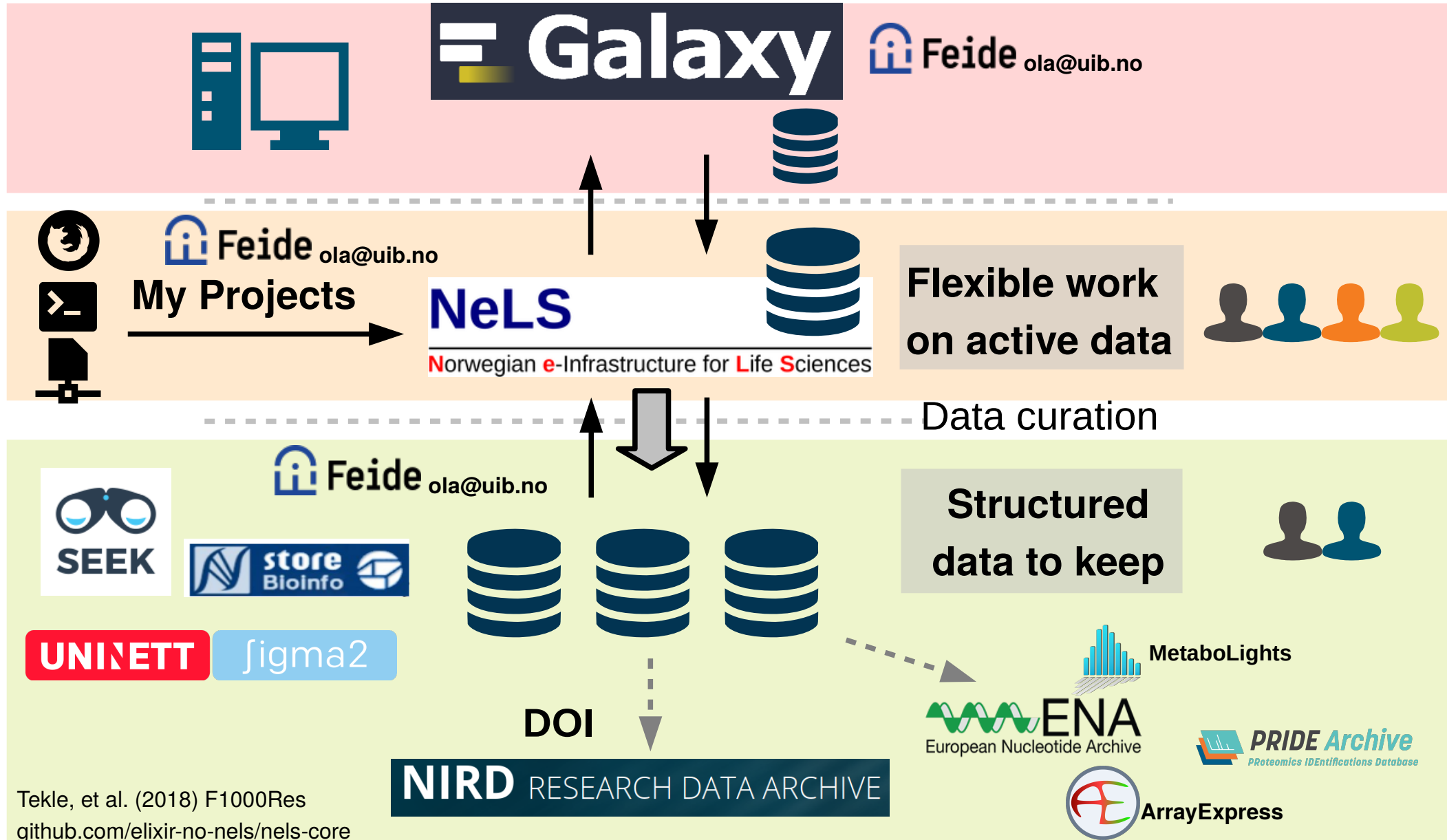


Data curation

NeLS



NeLS



Tekle, et al. (2018) F1000Res
github.com/elixir-no-nels/nels-core

NeLS

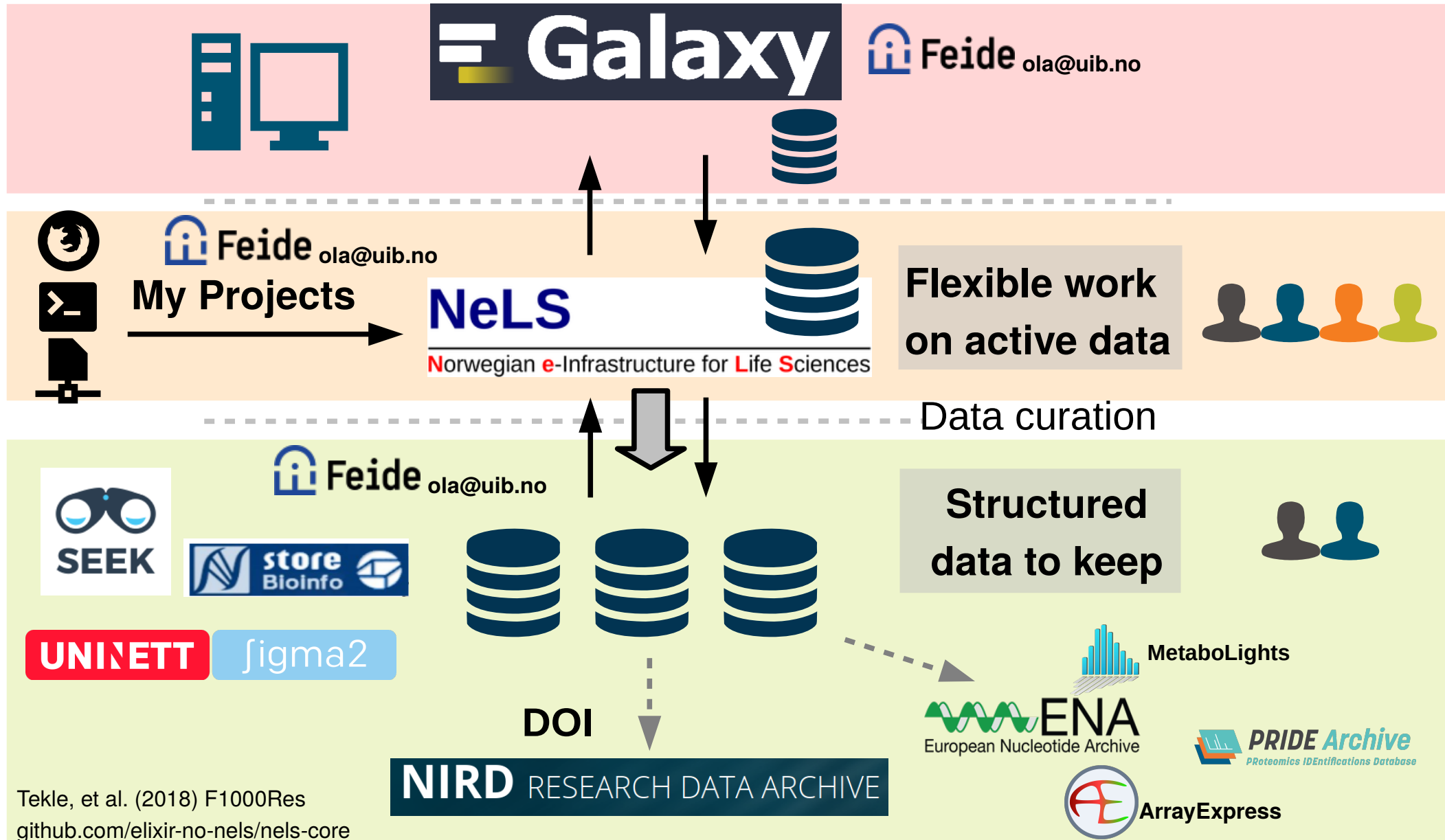
Days - weeks

 **nor seq**
Months

Years

Decades

<https://nels.bioinfo.no/>



NeLS - Accessibility

small datasets



ola@uib.no

NeLS - Accessibility

small datasets



ola@uib.no

larger datasets



ssh-key






git










sFTP

















Data Analysis in NeLS

Available Pipelines

Analysis Workflows   

      Search:



Category	Description	Galaxy Nodes
 DNA-Seq		
 Metagenome		
	<div><div>Taxonomic classification of metagenomic sequences (Paired-end)</div><div></div></div>	<div><div>Taxonomic classification of paired shotgun metagenomic sequences based on prediction of reads containing 16S rRNA</div><div> </div></div>
 RNA-Seq		  
	<div><div> Eukaryote RNA-Seq workflows</div><div> </div></div>	
	<div><div> LiceBase RNA-Seq workflows</div><div></div></div>	

NeLS - Getting Access

NeLS

Norwegian e-Infrastructure for Life Sciences

Typically > 1TB data

Free for regular sized projects

Allocation through Helpdesk: contact@bioinfo.no

Direct upload through Nourseq facilities

<https://nels.bioinfo.no/pages/user-terms.xhtml>



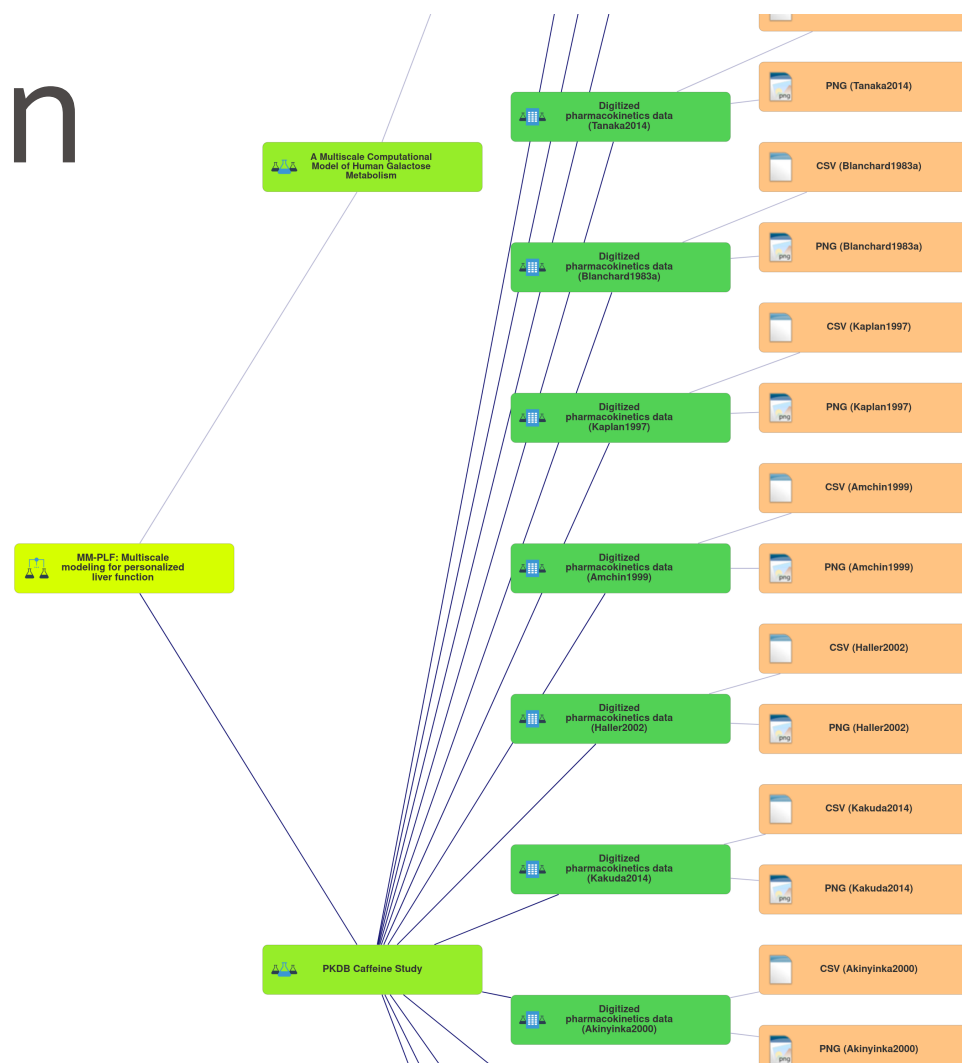
SEEK





SEEK

Investigation Study Assay



RightField



File Edit Sheet Help

	A	B	C	D	E	F
1	# This is an excel template...					
2	# Use this template for ...					
3	# Click the Metadata Ex...					
4	# Field names (in blu...					
5	# CLICK HERE for the F...					
6						
7	SERIES					
8	# This section describes ...					
9						
10	title					
11	summary					
12	summary					
13	overall design					
14	contributor					
15	contributor (SEEK ID)					
16	SEEK Project	Project				
17	Experiment Class (a...	transcriptomics				
18	Experiment Design t...	ExperimentDesignT...				
19	Technology type	microarray				
20	quality control type	QualityControlDesc...				
21						
22	SAMPLES					
23	# The Sample name...					
24	# CLICK HERE to find t...					
25						
26	Sample name	title	CEL file	source name	organism	characteristics:...
27	SAMPLE 1				organism	
28	SAMPLE 2				organism	
29	SAMPLE 3				organism	
30	SAMPLE 4				organism	
31	SAMPLE 5				organism	
32	SAMPLE 6				organism	
33	SAMPLE 7				organism	
34	SAMPLE 8				organism	
35	SAMPLE 9				organism	
36	SAMPLE X				organism	
37						
38						
39	PROTOCOLS					
40	# This section includes pr...					
41	# Protocols which are ap...					
42						
43	growth protocol					
44	treatment protocol					
45	extract protocol					
46	label protocol					

Selected cells: B17:B17

ONTOLOGY HIERARCHIES

MGEDOntology.owl x

JERMOntology x

- ExperimentClassType
 - fluxomics
 - genomics
 - interactomics
 - metabolomics
 - proteomics
 - reactomics
 - single_cell
 - transcriptomics
- InformaticsAnalysisType
 - ModelAnalysisType
- CultureGrowth
 - FactorsStudied
 - concentration
 - expression

TYPE OF ALLOWED VALUES

☐ Free text

☐ Direct subclasses

☒ Subclasses

☐ Instances

☐ Direct instances

ALLOWED VALUES

- Comparative genomic hybridization
- RNAi
- gene expression profiling
- methylation profiling
- microRNA profiling
- tiling path

Apply

Metadata Template Matrix Template Metadata Example Matrix Example 1 Matrix Example 2

NeLS

Norwegian e-Infrastructure for Life Sciences



Integration



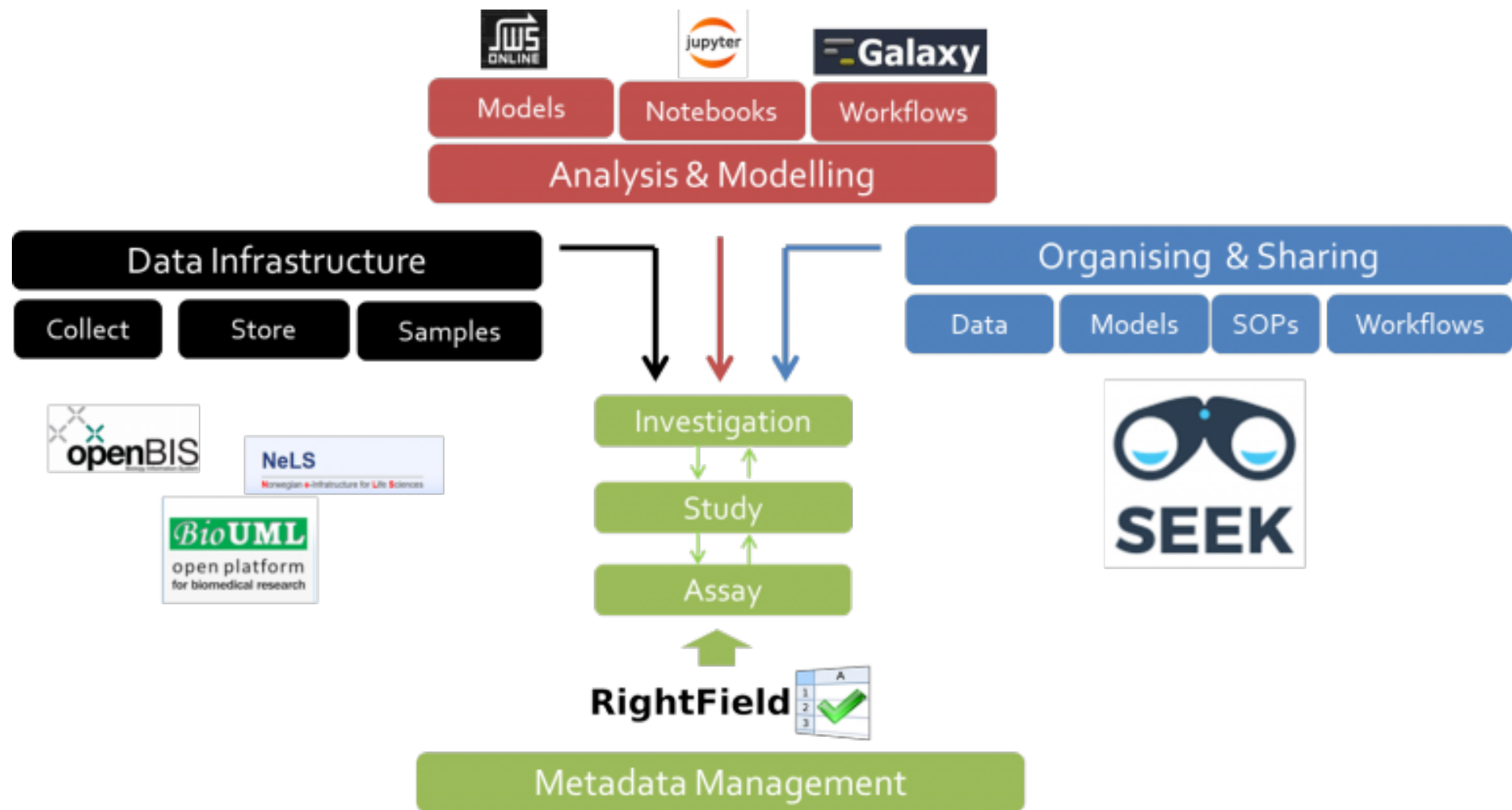
Reference data in NeLS in SEEK



Access data in NeLS through SEEK



Save Metadata to SEEK through NeLS



National HPC infrastructure

UNINETT

sigma2



Advanced Storage Needs

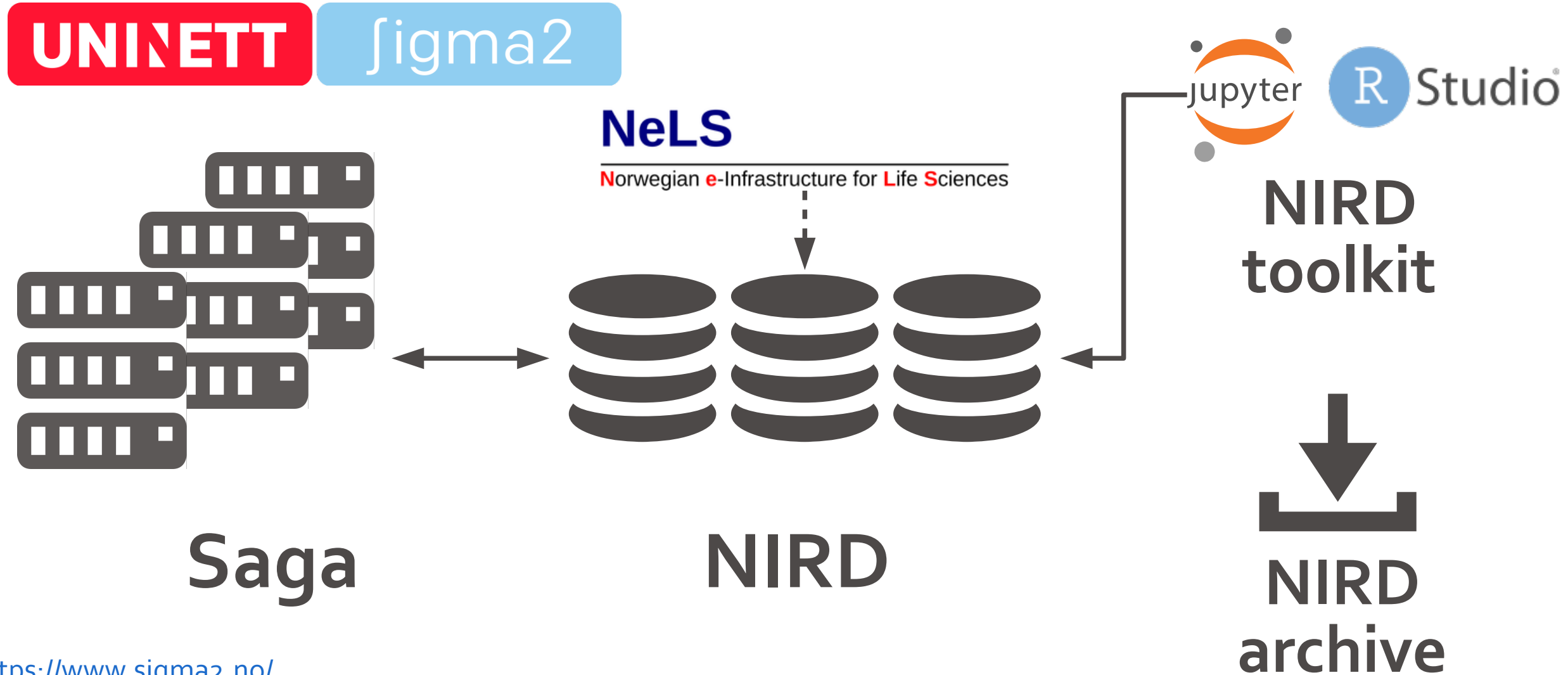


Advanced Computing Resources



Free Analysis Toolkit (Rstudio, Jupyternotebook)

National HPC infrastructure



HPC infrastructure

UNINETT

sigma2

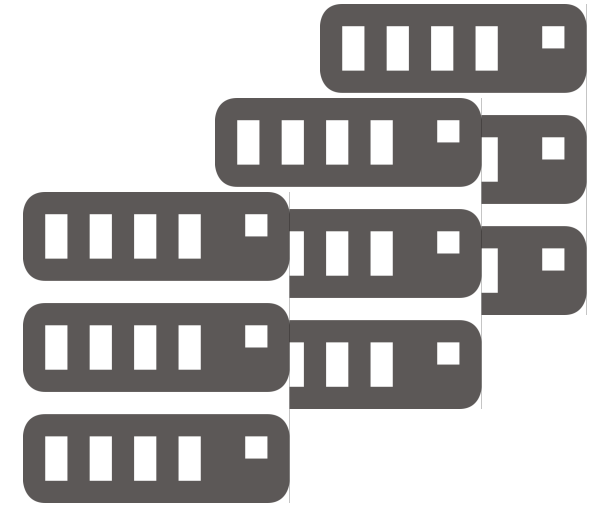
Saga

244 nodes – 9824 cores – 32 GPUs

Total 75TB RAM – 186-3066GB/node

89TB + 60TB NVMe+SSD

1PB file system



Storage infrastructure - NIRD

UNINETT

sigma2

Geo replicated

Nightly snapshots

12 PB (→ 36 PB next 4 yrs)



Sigma2 - Getting Access



Bi-annual calls (January & July) – Reapplication (!)

User contribution model (projects > 15 mill. NOK)

Application for small grants continually

TSD – National sensitive data service



Maximum classification



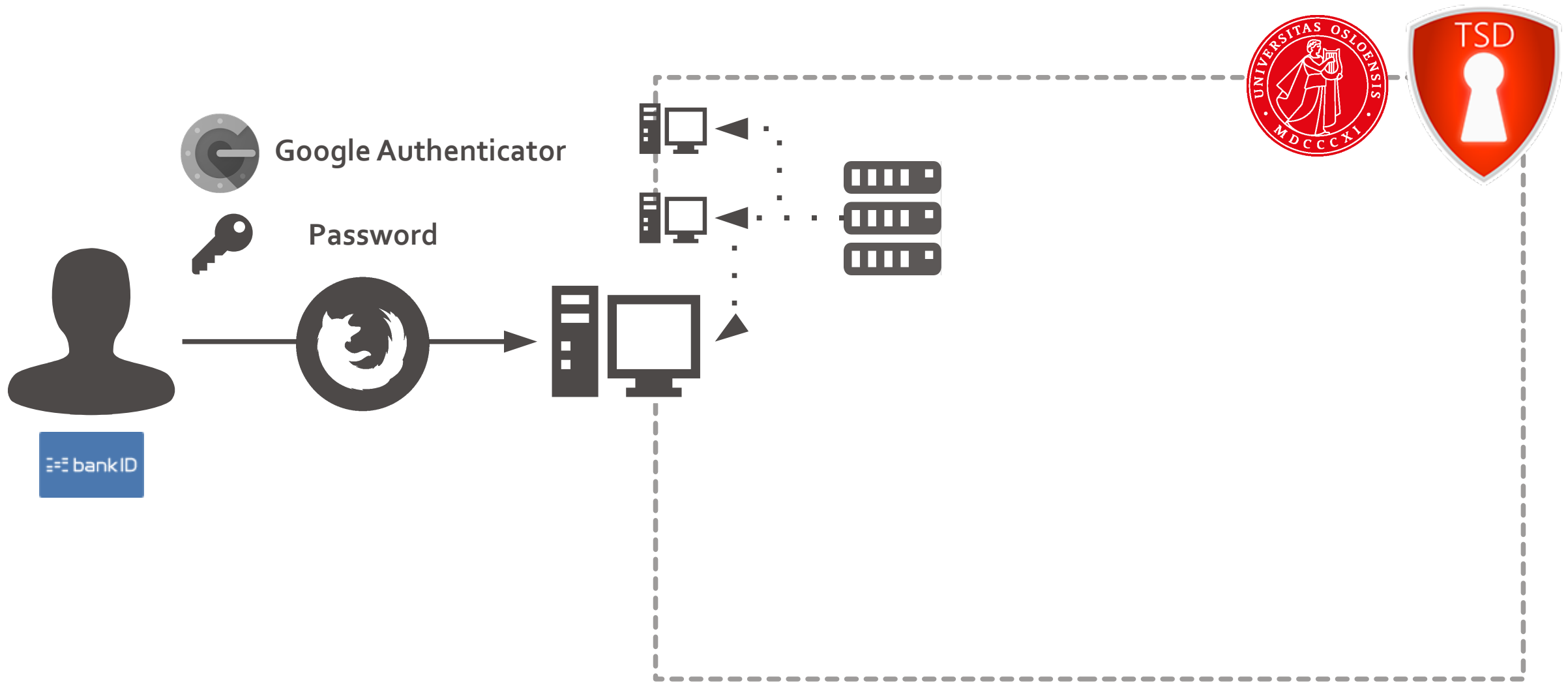
2-factor authentication

Attached HPC resources

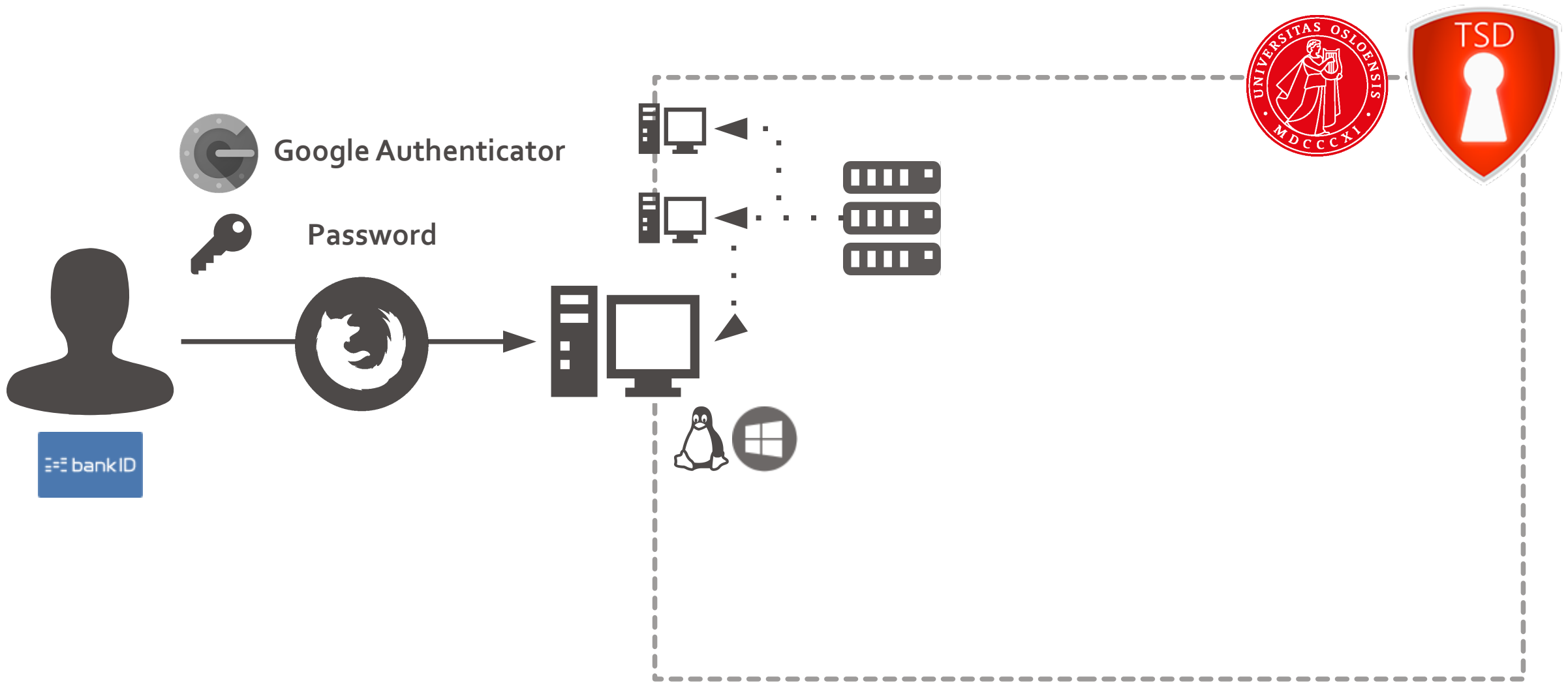
TSD – National sensitive data service



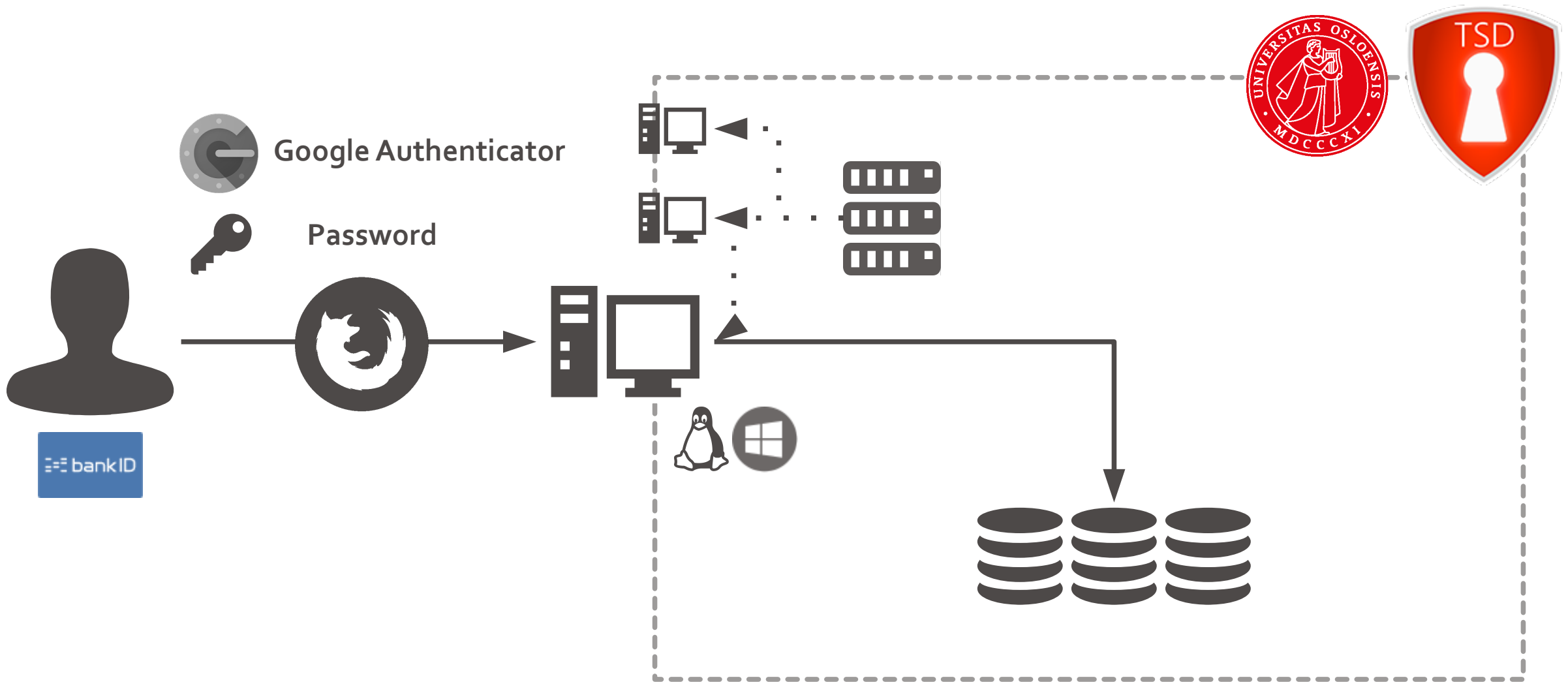
TSD – National sensitive data service



TSD – National sensitive data service

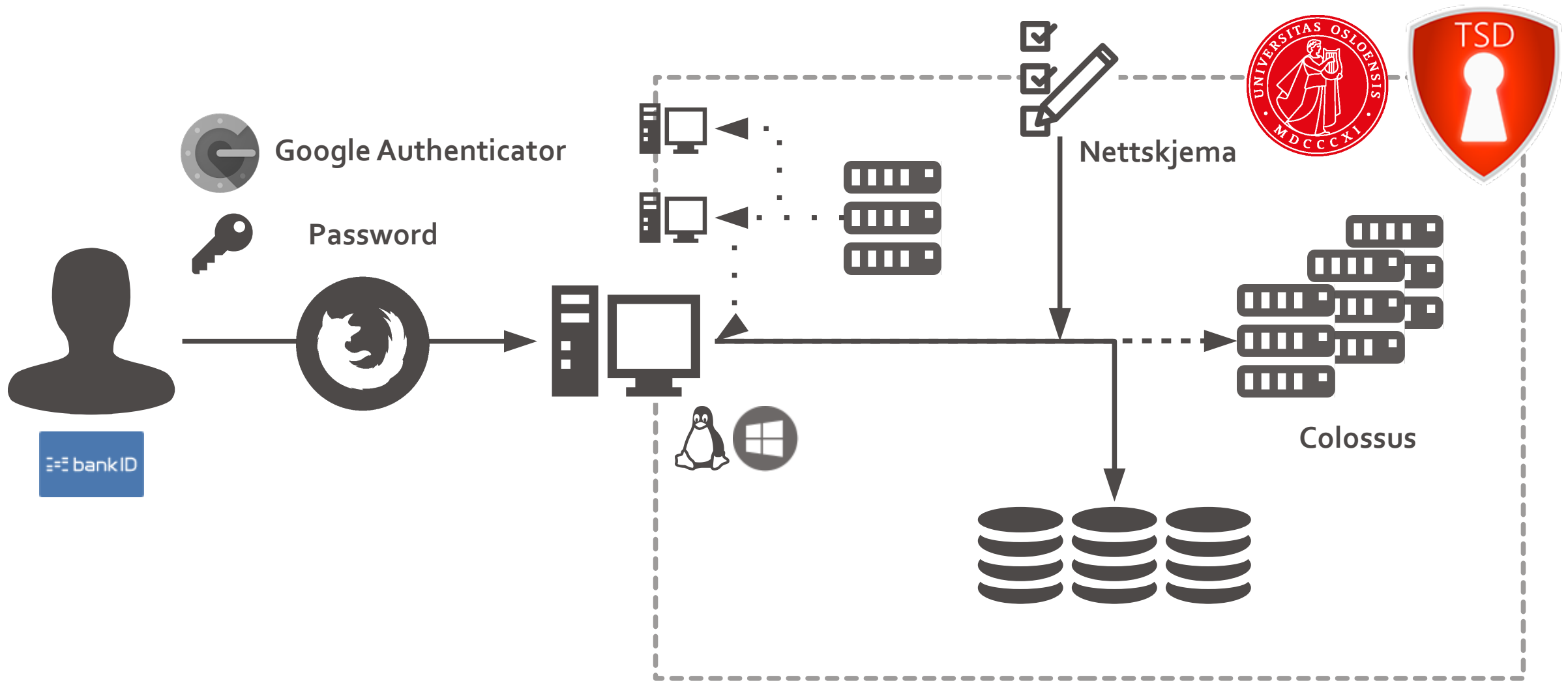


TSD – National sensitive data service



[illegible]

TSD – National sensitive data service



TSD – Getting Access



Subsidized access through Sigma2 calls

User contribution model (non UiO: min 15kNOK/a)

Free storage quotas through NeLS