Addendum to Project 5 description

DeiC National Integration Portal

Background

This document is an addendum to the proposal for the implementation of the DeiC Project 5 for the construction of a “National Integration Portal”. In agreement with the DeiC board and management, a series of workshops has been held in December 2020 with the relevant stakeholders, including the DeiC national HPC providers. The revised work plan presented here is the main deliverable from the workshops.

Table of content

Vision and expected function of the portal

The vision of the project is to create a user-friendly portal for all researchers in Denmark, which allows easy access to all national HPC and data resources. The solution is based on the UCloud platform; this offers a very solid and flexible foundation and it will allow us to match all the requirements in the DeiC call and provide all the needed functionality.

The National Integration Portal will expand the capabilities of now present in UCloud to:

- be able to connect to all national HPC centers for a number of common tasks such as, e.g. access to files, batch job management, list the hardware or software available at the HPC centers, initiate data transfers among HPC centers;
- collect the reporting statistics and monitor utilization of the national HPC services and projects allocated on them;
- and provide an intuitive user experience and GUI for all the added functionalities.

The National Integration Portal will integrate with the pre-exascale EuroHPC facility LUMI and his portal named “Puhuri”.

In the future, when a detailed plan for DeiC national data management services will be available, more data management functionalities could be added to the National Integration Portal, such as:

- be able to connect to/integrate with national data storage resources and national archives;
- initiate data movement to/from said data resources;
- allow queries of metadata in said data repositories to search for datasets;
- allow publication of datasets in international (e.g. Zenodo) or national repositories and archives.

The National Integration Portal could also provide a way to present the DeiC Danish national services via the EOSC (European Open Science Cloud). We note that UCloud itself is already among the services that will be integrated as part of the EOSC-Nordic project. Services in EOSC have different levels of maturity and have very heterogeneous interfaces. A common interoperability framework and, in the future, API for the EOSC services is being defined.

Project Plan

We plan to start the work in Spring 2021 with a co-design phase of the integration portal and its API. This **co-design phase will be done together with all DeiC HPC centers** starting from January 2021, and it will last several months (see project roadmap below). The early involvement of all DeiC HPC centers is an essential part of the project. The initial co-design phase will ensure that the common design is a good fit to the DeiC national HPC environment. One of the deliverables of the co-design phase is a definition of the work needed at each DeiC National HPC center for the integration with the National Integration Portal.

The project contains three main work packages:

1) the development of the web portal and GUI for the users;
2) the design and implementation of the integration module for the DeiC National HPC centers;
3) and the integration of data services for HPC.

The first two work packages will start immediately in January 2021 on parallel tracks, while the third one is expected to start later in 2022. **The integration with DeiC National HPC systems will be carried out in two phases**: the first phase will involve DeiC Type 1, Type 2 Sophia and Type3; the second phase will include all the remaining DeiC HPC installations, i.e. Type 2 Computerome2, GenomeDK and Type 4. The first phase will involve the consortium partners behind DeiC Project 5, with the aim to develop a first version of the integration with the National Portal, to be later tested, revised and deployed for further testing at the remaining DeiC HPC installations.

Work Packages 1, 2 and 3 will be developed following the methodology presented in the original application. Due to the agile methodology employed, there will be **a continuous feedback loop between the design, development, testing and deployment of the software infrastructure in this project**. This means that there will be a continuous feedback loop between closely related tasks in WP1, WP2 and WP3. In particular the design of the APIs and the reference implementation provided as part of this project, will continue to be refined for the entire duration of the project. The design is such that new features and changes can be made without compromising the operations of the National Integration Portal.

The project is developed as **an open-source project** available on Github.This means that all the development and discussions around specific issues are open and everyone can participate and contribute to the project. This will also guarantee the **transparency in the project**. This is the first time a national infrastructure project in Denmark is being developed as an open-source project.

Roadmap, Milestones, partial deliverables and deadlines

We break the work packages into several tasks, presented in the roadmap Table 1. The roadmap table shows the timeline for the different tasks and the project milestones corresponding to completing the tasks. We list in Table 2 the project milestones, with the deliverables and their deadlines.

We also plan regular status reports to the DeiC board and management every 6 months. These partial deliverables and their deadlines are also reported in Table 2. We plan 3 status updates: each of them presenting the status of the relevant milestones in progress at that time.

Milestones and Deliverables

Milestones and deliverables in WP1, correspond to new “sections” in the portal web application. These sections are intended to offer easy access to the new functionalities on the National Integration Portal. A description of the deliverables of each milestone follows.

M1 New GUI for project management
Design of UX and UI for the described project management functionality (see below in WP1 description). Implementation and testing of the design in the client web application.

M2 New GUI for resource management
Design of UX and UI for the described resource management functionality (see below in WP1 description). Implementation and testing of the design in the client web application.

M3 New GUI for HPC centers reporting and overview
Design of UX and UI for: 1) overview of resources available at DeiC HPC centers; 2) current utilization of DeiC HPC centers. The functionality is described below in WP1. Implementation and testing of the design in the client web application.

M4 Redesigned application section
Design of UX and UI for new “apps store” section. The new section should provide a simple to use overview of applications present at DeiC National HPC centers. The functionality is described below in WP1. Implementation and testing of the design in the client web application.

M5 Co-design of architecture and first version of API
Design of the integration architecture and the necessary APIs between the Integration Module and the core services. Reference design documents are produced as well as an OpenAPI description of the initial APIs. The deliverable is co-designed with all DeiC National HPC centers.

M6 First version of UCloud integration module
First reference implementation of the Integration Module to the install at the HPC centers. Implementation follows design of M5. This first implementation will be the result of a testing phase at the HPC centers which are part of the Project 5 consortium (others can join on a voluntary basis).

M7 Second version of UCloud integration module
Second reference implementation of the Integration Module to the install at the HPC centers. After the experience with the first implementation in M6, a revised design of the architecture and APIs will be produced, as well as a reference implementation in this M7.

M8 Second version of UCloud integration module
Second reference implementation of the Integration Module to the install at the HPC centers. After the

experience with the first implementation in M6, a revised design of the architecture and APIs will be produced, as well as a reference implementation in this M7.

M9 Integration with Puhuri/LUMI
Implementation of the core functionality to interface with the Puhuri portal. Design and implementation of the necessary UI to present data from Puhuri.

M10 Integration of data transfer system among HPC centers
Implementation of the core functionality to interface with the Puhuri portal. Design and implementation of the necessary UI to present data from Puhuri. Interface to Puhuri will be done via the Puhuri portal API.

Risk management

We list below some of the main risks related to this project and the possible mitigation actions that will be taken into consideration.

#	Description of Risk	WP	Mitigation measures
1	Lack of participation in the co-design of integration module	2	Distribute information early. Public discussions and documents available on github. Encourage discussions on critical aspects of the design.
2	Difficulty in the integration of specific HPC sites or specific systems in use.	2	Collect information on the local HPC environment early. Test integrations as soon as possible. Ensure relevant feedback to the design of the integration module.
3	Lack of resources for the integration at DeiC HPC centers.	2	Discuss extent of integration at each DeiC HPC site and resources required for the integration. Discuss with DeiC management resources for the integration in due time for the negotiations of national HPC contracts.
4	Puhuri API not suitable or difficult to integrate (e.g. problems with user mapping, access model, etc)	2	Early coordination with the Puhuri core team and Danish participants. Follow development of Puhuri API for integration with national portals.
5	Delays in the HPC data transfer system or in general with the future DeiC national data management services.	3	Follow development of DeiC national data management services. Revise the plan for WP3 to use alternative data transfer mechanisms available at HPC centers.

Table 1 Project roadmap presenting the milestones and partial deliverable.

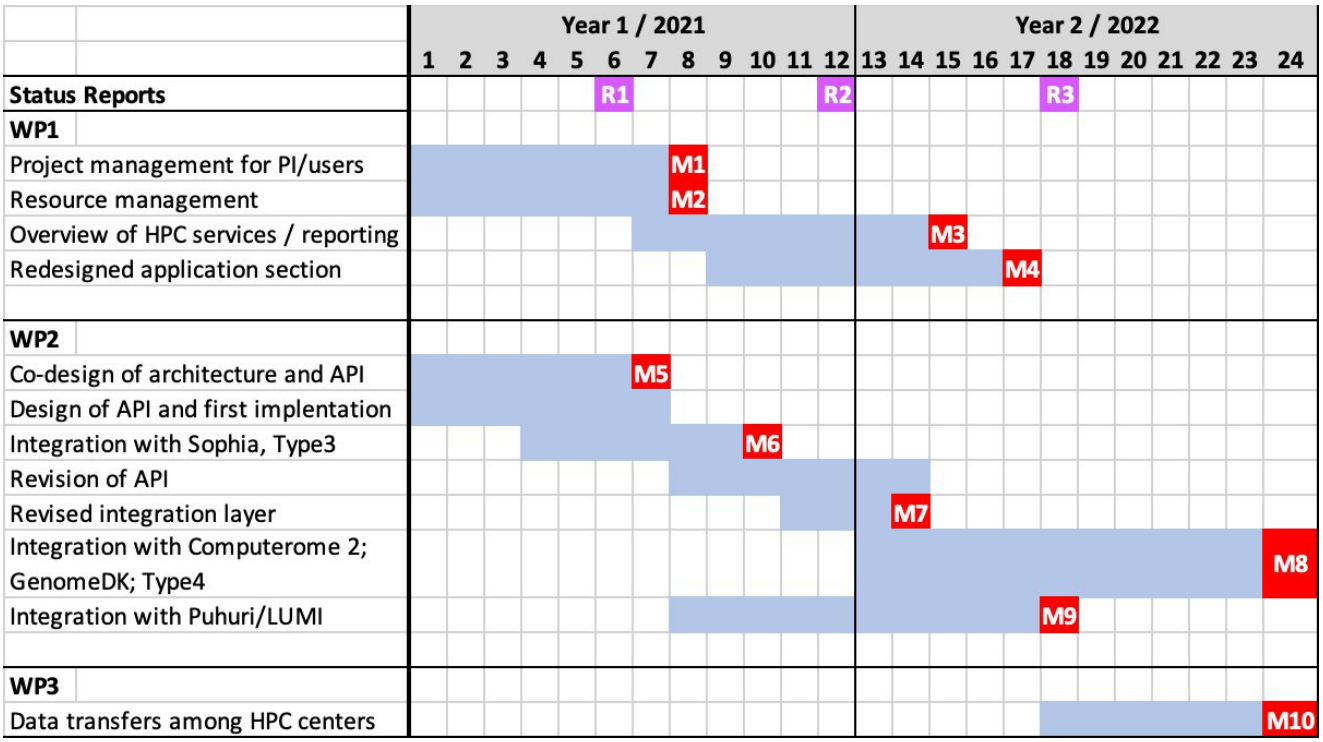


Table 2 List of milestones, deliverables, partial deliverables and their deadlines.

Milestone	Deliverable	Deadline
M1	new GUI for project management	01/09/2021
M2	new GUI for resource management	01/09/2021
M3	new GUI for HPC centers reporting and overview	01/04/2022
M4	new application section	01/06/2022
M5	Co-design of architecture and first version of API	01/08/2021
M6	First version of UCloud integration module	01/11/2021
M7	Second version of UCloud integration module	01/03/2022
M8	Integration of DeiC National HPC systems	31/12/2022
M9	Integration with Puhuri/LUMI	01/07/2022
M10	Integration of data transfer system among HPC centers	31/12/2022
Status Reports	**Description**	**Deadline**
R1	Mid-term report on M1, M2, M5, M6	01/07/2021
R2	Mid-term report on M3, M4 and M7	01/01/2022
R3	Mid-term report on M8, M9 and M10	01/07/2022

Involvement of DeiC National HPC centers

As emerged during the Workshops in 2020, **all DeiC national HPC centers have wished for an early involvement in the project**. Therefore they will be **invited from the start of the project in January 2021**, to participate in the works on WP2 for the task "Co-design of architecture and API". We will also invite the staff in DeiC, which is participating in the NeiC Puhuri project. Design descriptions for the integration with DeiC National HPC system will be presented and discussed with the national providers in dedicated meetings. Such documents will also be available on github for maximum transparency and to give all the opportunity to comment and discuss issues early in the process.

The early involvement of all DeiC HPC centers and the relevant staff in DeiC will give all **the opportunity to co-design the integration module, which will be later installed at all HPC centers**. This will also provide an early feedback loop in the project, between all DeiC National HPC providers, the relevant DeiC staff and the team behind DeiC Project 5, which will provide a good platform for information flow in the project.

As the initial co-design phase is critical for the project, it is expected that all DeiC National HPC centers will participate. This is also the only required involvement in 2021.

Besides the co-design of the integration module, all the DeiC national HPC center providers will be asked to contribute to:

1) the concrete implementation of any custom plug-ins for the UCloud integration module, which is needed for integration at their site;
2) the deployment and configuration of the UCloud integration module at their site.

It is expected that work required at each DeiC national HPC center will be defined during the co-design phase in Spring 2021.

Coordination with Puhuri and LUMI

Denmark is participating in the NeiC project "Puhuri", for access, resource allocation and tracking for the LUMI supercomputer. Puhuri is being implemented as a web-based portal, which will interact and rely on the national portals such as the present DeiC Project 5. A number of DeiC staff members are participating in the Puhuri project, led by Eske Christiansen member of the Puhuri "Steering Group". Prof. C. Pica has also been invited to be part of the Puhuri "Reference Group", which receives regular updates from and provides feedback to the Puhuri project.

We will invite the DeiC staff participating in the development of the Puhuri portal to both the co-design of the integration module and the later task "integration with LUMI/Puhuri".

The Puhuri portal, by design integrates with the national portals via an API, which is currently being designed (based on the one of Waldur mastermind). We therefore do not foresee any problems for this integration.

Integration with Puhuri is planned to start August 2021, for the first phase of availability of LUMI, but it could be anticipated or delayed, depending on the outcome of the Puhuri project.

It should be noted that the Puhuri project will only offer a limited functionality compared to Project 5. The Puhuri project is limited to project and resource management, and will not provide any "deeper" integration with LUMI. Coordination with LUMI, in particular procedures for resource allocation and management, will be guaranteed by the direct connection with the LUMI "Operational Management Board", which Prof. Claudio Pica is member of.

Description of Work Packages

The work for the DeiC Project 5 will be divided into three work packages over a period of 24 months in total.

Work Package 1 - Web Portal / GUI

We will use the UCloud web UI as the foundation for Work Package 1, and we will add the necessary components to expose the new functionality to the users. **Priority is given to the user experience and to provide an intuitive and clean interface.**

In this work package we will extend the web portal to include:

- an overview of available computing resources, such as the national HPC centers;
- a list specific software present at each HPC center in the apps store section;
- project management functionalities for:
    - monitoring of usage of resources
    - project members and roles
    - access control to project files
    - creation of subprojects
- Resource management functionalities for granting agencies (DeiC, Danish Universities, etc):
    - workflow for allocation of projects
    - monitoring of usage of resources
    - reporting for HPC centers to DeiC

The web portal will thus also be extended to allow DeiC and funding agencies to show the overall utilization of national HPC facilities and job submission statistics.

The new functionalities will be added on top of the current UCloud GUI, which already includes the following functionalities:

- authentication via WAYF and two-factor authentication;
- file browser;
- dashboard for resource utilization;
- submission/cancellation of job to computing backends;
- present a list of available software (apps store).

Work Package 2 - Integration with HPC centers

UCloud provides a high-level abstraction to manage storage and compute resources. In this project, we will design and implement a new “UCloud integration module” and expand the core UCloud functionality to support the national HPC centers as providers of compute and storage resources.
We aim to integrate access to national resources seamlessly within the platform. The integration will be flexible and allow HPC service providers to opt-in or opt-out of the functionalities provided by the National Integration Portal. The minimal functionality required at each DeiC national HPC center will be agreed upon during the co-design phase in Spring 2021.

Interaction with the core UCloud platform will be provided through an API. The core UCloud design abstracts the computing resources in a set of microservices and delegates the work to providers. This means that UCloud is acting as an orchestrator of all types of resources.

We will expand the capabilities of the platform to allow seamless integration with providers of computing resources, such as the national HPC centers.

National HPC providers and, later on, national data/storage providers will be able to register their services/resources with the UCloud platform. This will provide information about which features each provider supports. For example, a traditional HPC provider might declare that it supports storage and batch computing.

We plan to provide an API to interface to external resource providers which will cover:

- user mapping between the national portal/UCloud and resource providers;
- storage/files at HPC centers (disk usage, quotas, file lists, and read/write operations);
- compute/jobs (submit jobs, cancel jobs, list and interact with jobs);
- project management (e.g. project/account usage, quotas, list users to a slurm account, add/remove users).

Most of these APIs will be optional. A provider can declare which APIs are supported in a metadata document. This will allow providers of resources to interface gradually with the national portal. For example a provider can initially only support a subset of functionality to allow the users/grant providers to see the project information and usage at its facility, but not to change this information or interact with the jobs.

As part of this project, we will provide a reference implementation for all of these APIs both for the national portal/UCloud and the HPC centers. The reference implementations will be provided through the "UCloud integration module". The integration module will provide a plugin-based architecture. Default plugins, part of **the implementation provided** by the UCloud team, **will contain functionality for common systems, which have been put forward by the DeiC National HPC providers** during the workshops in December 2020. The list of systems from the workshops is included in a later section. Custom plugins allow providers to implement their own scripts in reaction to events in UCloud. This allows providers to fine-tune the UCloud integration module to their concrete setup. A high level overview of the integration design is included in a later section.

During the workshops in December 2020, **several HPC service providers expressed their interest in having user administration being part of the national portal.** IDs would still be maintained in the local user repositories at each site, which are available for WAYF authentication. However roles and permissions concerning the use of the national HPC infrastructure could be maintained as part of the portal functionality. This would provide an alternative mechanism to the local identity management procedures, however optional for each HPC service provider to use. Given the interest in a central/national solution for user management, a discussion on the desired functionality and the possibility to include it within the scope of DeiC Project 5 will be carried out during the co-design phase of the National Integration Portal, starting from January 2021.

Work Package 3 - Integration of data services

At a later stage the platform can be expanded to include APIs to interface with national data storage resources and archives. UCloud already provides the user with a unified environment for data and computing. By interfacing the future national storage facilities with the platform, the users can be presented a common view for both computing and storage resources. By providing an integration layer on top of the existing data repositories and future national data facilities, a powerful national data management system can be put into force.

The envisioned integration could provide APIs for operations such as:

- interface with national authentication and authorization system;
- access datasets located in national data storage resources and national archives;
- initiate data movement to/from said data resources;
- allow queries of metadata in said data repositories to search for datasets;
- allow publication of datasets in international (e.g. Zenodo) or national repositories and archives.

Such functionality can be integrated into this project and presented to the users in the provided GUI.

For national HPC centers, we plan to integrate functionality to

- initiate data movement to/from DeiC HPC centers;

based on the prototype ScienceDMZ project between SDU and AAU, currently in progress in 2020.

The plan for national data management services is not available yet but should become available at the end of 2021. As the national data management services become available in 2022, the National Integration Portal could include more functionality as envisioned above.

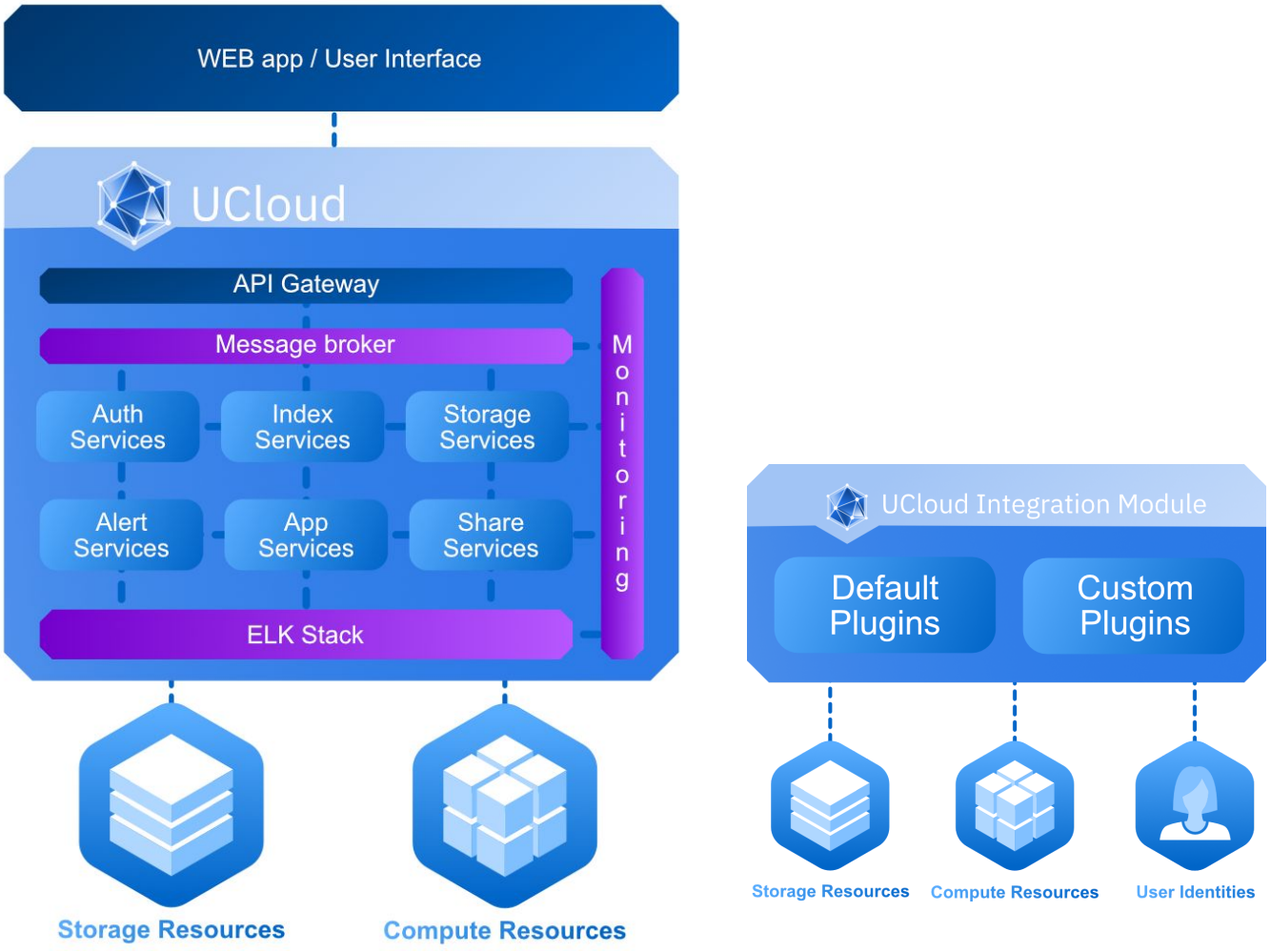
This can be done using the same design developed in this project and by expanding the UCloud integration module functionality to support new APIs.

The national integration portal will act as an orchestrator, controlling/interfacing to lower-level components for, e.g., national authorization to access datasets or parallel data transfer between locations, etc. These lower-level components already exist in many if not all cases and can be reused by the national storage resources and archives (for example, among HPC centers one could use, e.g., gridftp as a standard tool for fast data transfer). When national archives already implement APIs for the envisioned functionalities, these will be used.

High level view of the integration design

The high level design of the planned integration was presented and discussed at the workshops in December 2020. While the individual components will be co-designed starting from January 2021, the high level design presented here was agreed upon during the workshops.

The DeiC National Integration Portal is based on UCloud. UCloud has a modular “core” which acts as a high-level orchestration platform for compute and storage resources. The UCloud core is based on a modern microservice architecture, in which different services manage specific resources (e.g., compute or storage) or perform specific tasks (e.g. authentication or logging). The users access the system via a modern web application, which presents an intuitive GUI. On the opposite side, the service providers connect resources for HPC compute and storage to the national portal via the “UCloud Integration module”.



List of systems in use at DeiC National HPC providers

During the workshops in December 2020, we have collected a list of common systems in use at the DeiC National HPC providers. This list of systems will be used for the development of the “default plugins” for the integration module.

DeiC HPC	Provider	user management	storage systems	job management	project management	Responsible person
Type 1	AAU	AD	CEPH	OpenStack	OpenStack	R. D. Jensen
	SDU	UCloud	CephFS/GPFS	UCloud	UCloud	C. Pica
Type 2	Genome DK	NIS/local unix users	BeeGFS	slurm	slurm + custom scripts	D. A. Søndergaard
	Computerome 2	local users / LDAP / 2FA	OneFS	Moab	Moab accounting manager	E. Hochheim
	Sophia	AD	BeeGFS/CEPH	slurm	slurm + custom scripts	M. Rasmussen
Type 3	SDU	IPA	GPFS	slurm	slurm + custom scripts	C. Pica
Type 4	KU	TBD	TBD	TBD	TBD	H. H. Happe
Notes:						
Other relevant storage:		S3 object storage