

 Search[Contact information](#)**About Science-IT News & Events Resources and Usage Contacts**[For Triton users](#) [For Grid users](#) [Scientific Computing in Practice](#) [Aalto Data Repository](#) [Summary for proposals](#)

PHYS Sites &gt; Science-IT at Aalto University School of Science &gt; Aalto Data Repository

## Aalto Data Repository

### Overview

We are creating a central online location for data sharing for all researchers within Aalto University. This will host both data and metadata: the name, description, ownership, source, and information on usage. Other dataset hosting sites exist, so our main target use case of expanding EUDAT scope is intra-Aalto University interaction. Researchers with data analysis skills will be able to find data related to their work, as well as the domain experts responsible for that data. Furthermore, the solutions should be tightly integrated to existing computing resources and Big Data platforms available nationally.

### The scientific & technical challenge

In Aalto University, Big Data and Data Science have been recognized as key areas of ICT and digitalization at all levels of rapidly developing socio-economic societies. These systems generate ever-increasing amounts digital data, which can in unprecedented ways serve as a gold mine for researcher of various disciplines to study as well as enable the private sector players and public sector to develop their services, processes and technologies. Hence, there is need to respond and find solutions to this data deluge, which is also reflected in the Aalto University application for profiling of Finnish Universities.

For the solution, we have identified a set of design requirements that may pose also as a technical challenge as follows:

1. The metadata of datasets is full text searchable.
2. Published datasets are assigned a persistent identifiers.
3. There are no restrictions to the type of uploaded data.
4. The datasets in the system can be made public for all the world to see.
5. The tool imposes no restrictions to the type of research data stored.
6. The metadata templates offered by the system satisfy the needs of different fields of science and also national requirements.
7. The system should be integrated to the already existing user management system.

In the first phase, the system will have some tens of users. If found successful, the solution will be scaled first within University and possibly even beyond to the national level. The possible user base may increase tenfold or even further in these cases. As Aalto users are working with large data sets, the data volumes can already in the pilot phase potentially extend to tens of terabytes.

### Why EUDAT?

We engaged into discussion with Centre for Scientific computing (CSC) that coordinates EUDAT operations in Finland. Based on these discussions, we have assessed especially B2SHARE & B2DROP functionalities based on the requirements. The results look quite promising and from this analysis we can see that the B2SHARE service would probably be the most suitable tool for addressing both our publishing and data management requirements.

As a part of the pilot, a suitable metadata templates can be provided in B2SHARE and customized for Aalto University. Further, as federated authentication via B2ACCESS is supported in B2SHARE, it is should be possible to authenticate to the national identity service within EUDAT services. In addition to that, interfacing to national metadata and storage solutions is of interest to us in the EUDAT pilot.

### The expected outcomes for Aalto University

The resulting data platform enables system and method-level development, resulting in research innovations but it also opens up possibilities for educational and training purposes on Data Science and Big Data. In the initial phase, there is a need to concentrate on data policy and repository practices for contributing to increased research effectiveness and generating wider goals of data sharing and open data in alignment with CSC's planned Big Data platform, but also to start widening and enhancing the skill base for data related research. We see that EUDAT solutions can play a major role in supporting the implementation of the ambitious research data management goals set by Aalto University.

### The legacy for Aalto University's scientific domain

This approach can enhance the system and process level understanding of any area of research, basic and applied, from science to engineering and humanities. It will strengthen multi and cross disciplinary research and step up product development, as well as facilitate novel innovations and services. At the same time, it lowers disciplinary boundaries both in the public and private sectors, thus adding to innovativeness and new breakthroughs in R&D, and covering all areas of society and science from discovery and security to general competitiveness. The data repository will be initially used mainly for research purposes, but yet at the same time, it will serve as a pilot and education platform for methodological development, training, and data repository technology.