

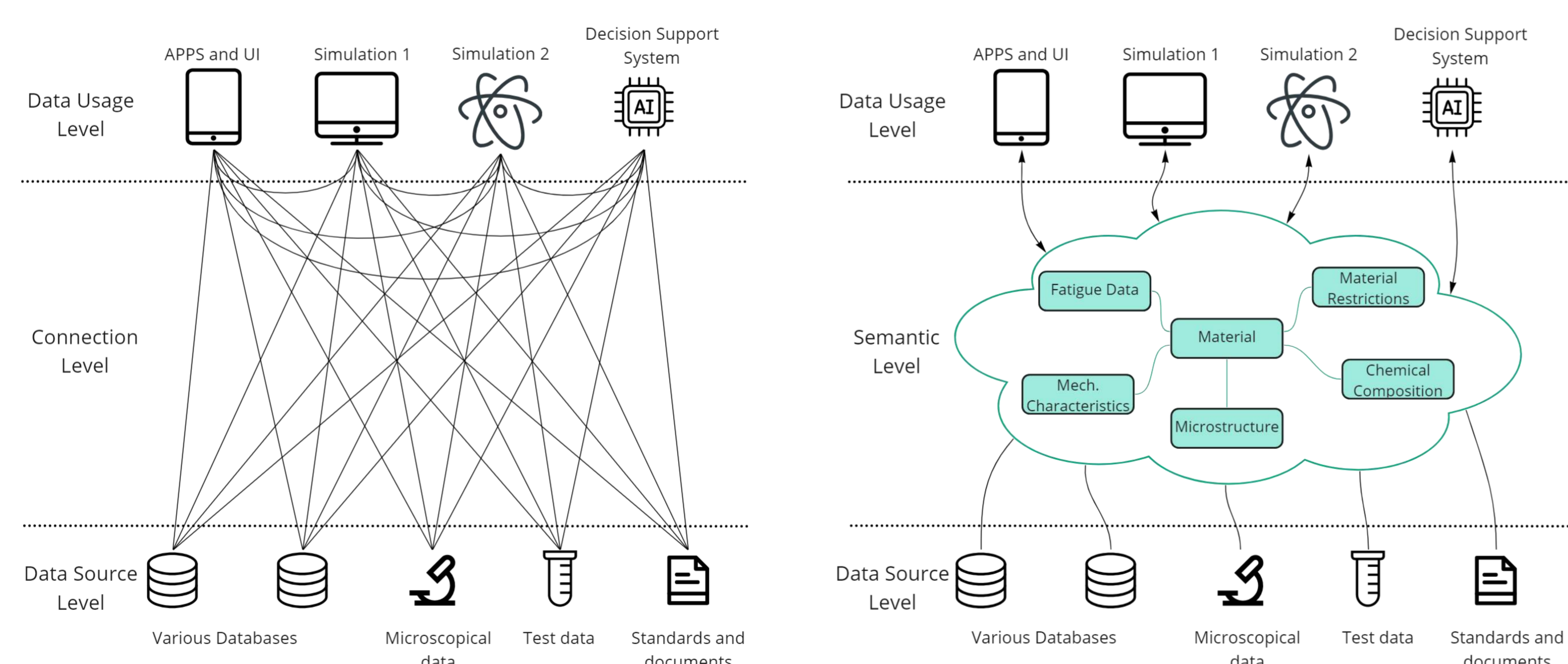
# An intuitive and expandable dataspace management system (DSMS) for materials science and manufacturing processes

Y. Nahshon<sup>1</sup>, T. Huschle<sup>1</sup>, D. Helm<sup>1</sup>

There is a constant need to easily and efficiently share and retrieve data related to characterization of materials, simulations and processes. The Dataspace Management System (DSMS) aims to provide a suitable framework for this challenge in the form of a software solution that combines a dynamic toolbox for typical subtasks, such as data integration and analysis, with an intuitive user interface and user management.

## Challenges in data management

- **Data heterogeneity.** Data originating from different sources is disparate with respect to its vocabularies, which are often ambiguous and conflictual.
- **Myriad of schemas.** Experimental data, generated from different working groups, is often collected using files of the formats xlsx, csv, and even PDF, where each employs a unique schema.
- **Disconnectivity of data.** Lack of explicit connection between data across domains (e.g., materials and manufacturing).
- **Stakeholders with different technological backgrounds.** Data is of interest to various stakeholders, who may possess a wide range of professional backgrounds among themselves. Highly complex systems are desirable for handling the varied aspects of their data, but simultaneously pose a high barrier for data evaluation, exploration, operation, and interpretation.
- **Availability of tools for handling data.** Programs and scripts that were developed for conducting scientific investigations and data analysis are often not readily available and require to be first set up, typically by an expert, in order to use them.

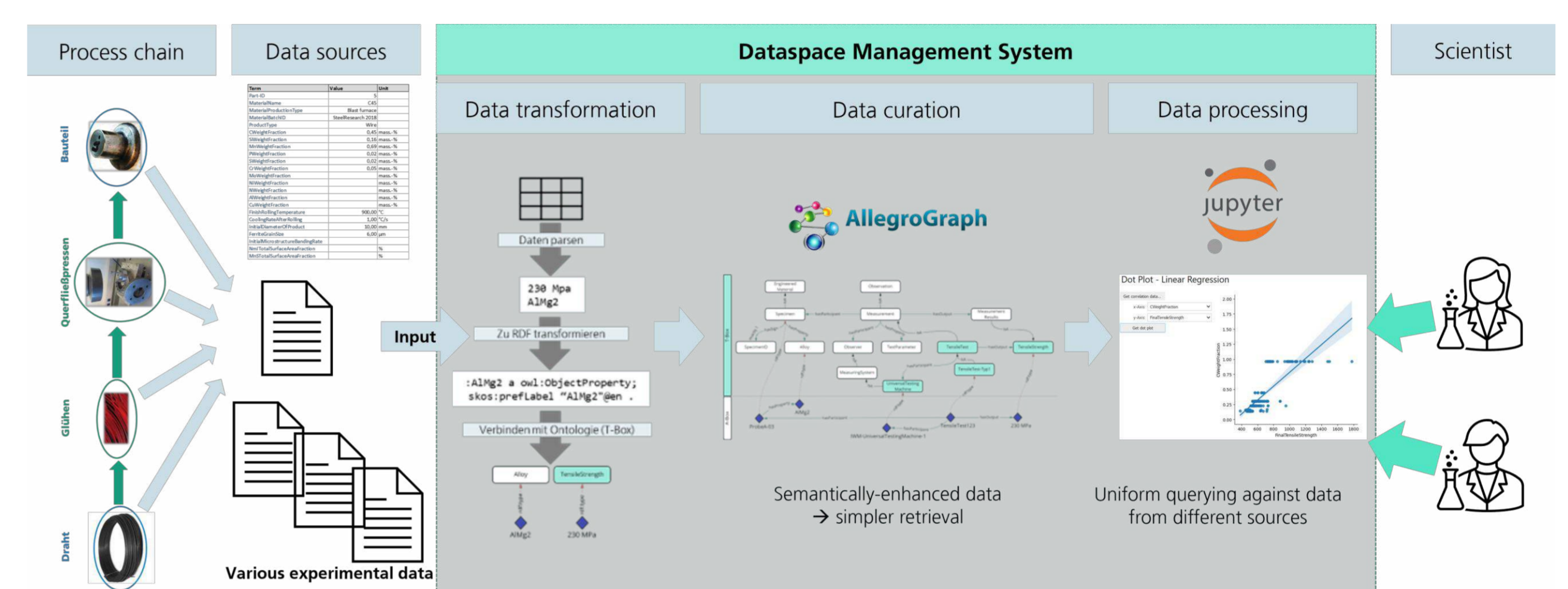


The challenge of using heterogeneous data sources and tools

The benefit of a semantically-enhanced data representation based on uniform vocabulary

## The benefits of dataspaces and how they differ from databases

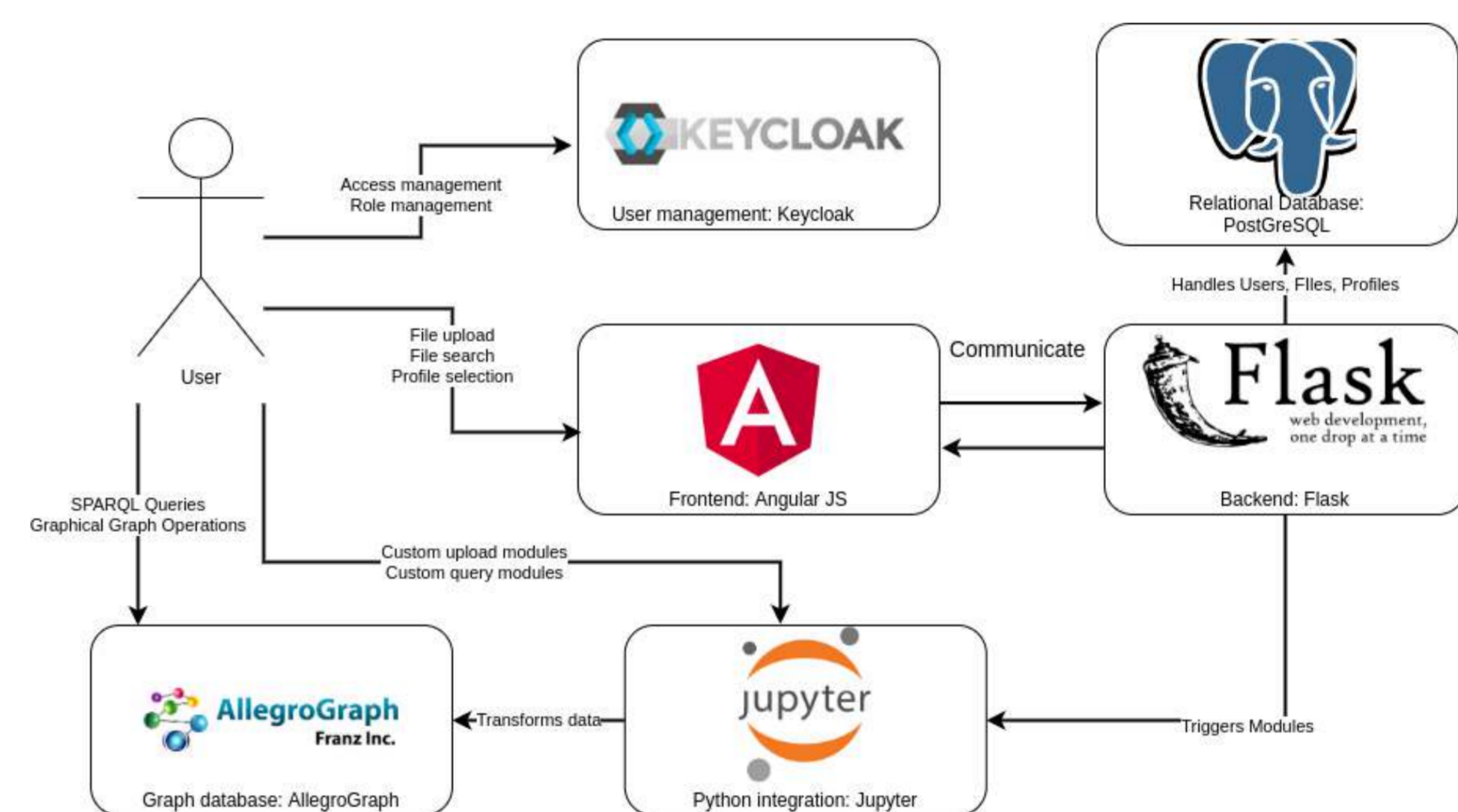
A database is a structured collection of data that is organized in a specific way, typically using a schema or predefined structure. Databases are optimized for processing and querying structured data, such as relational data in tables. On the other hand, a dataspace is a more flexible and dynamic way of managing data, where the data is not constrained by a rigid schema or structure. Instead, a dataspace allows for the integration of a wide range of heterogeneous data sources and types, including structured, semi-structured, and unstructured data. This makes it ideal for managing complex data ecosystems that are constantly evolving and require agility and flexibility in data management.



Semantic data integration of experimental data in DSMS

## The approach of DSMS

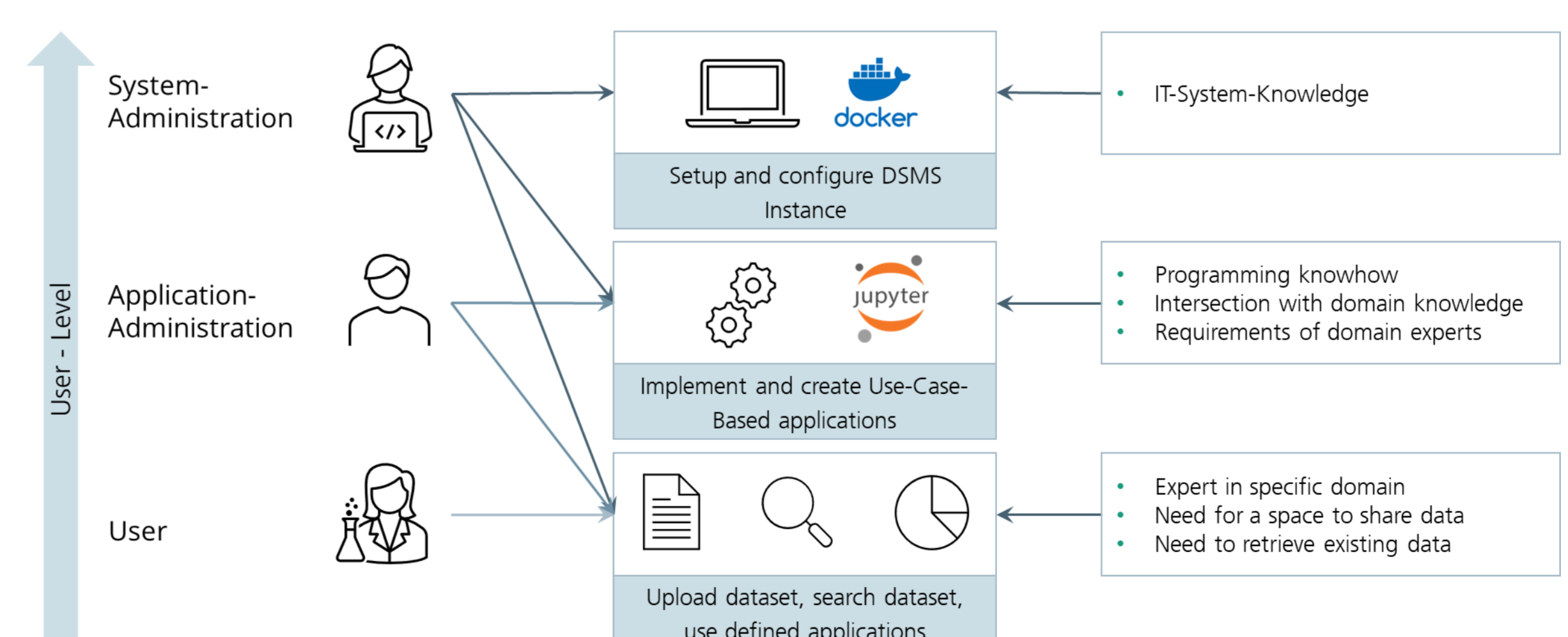
DSMS provides a friendly graphical user interface (GUI) that combines useful features such as user and role management, ability to organize user data and processing it via Jupyter-based and user-written plugins.



The technological stack of DSMS

## By users for users

To take advantage of users who possess programming skills, DSMS offers a plugin mechanism that is based on JupyterLab. The advantage of this approach is multifold. First, the dependency on the platform developers is being reduced by enabling the delegation of writing scripts to users with the appropriate programming knowledge. Second, those tools are integrated in the system in such a way that other users, who may have lesser programming skills, are free from concerns when it comes to setting up the tools and connecting them with the data – this is being carried out by the platform. By doing so, such tools are immediately available for all users with the corresponding access rights.



Delegation of tasks according to user roles