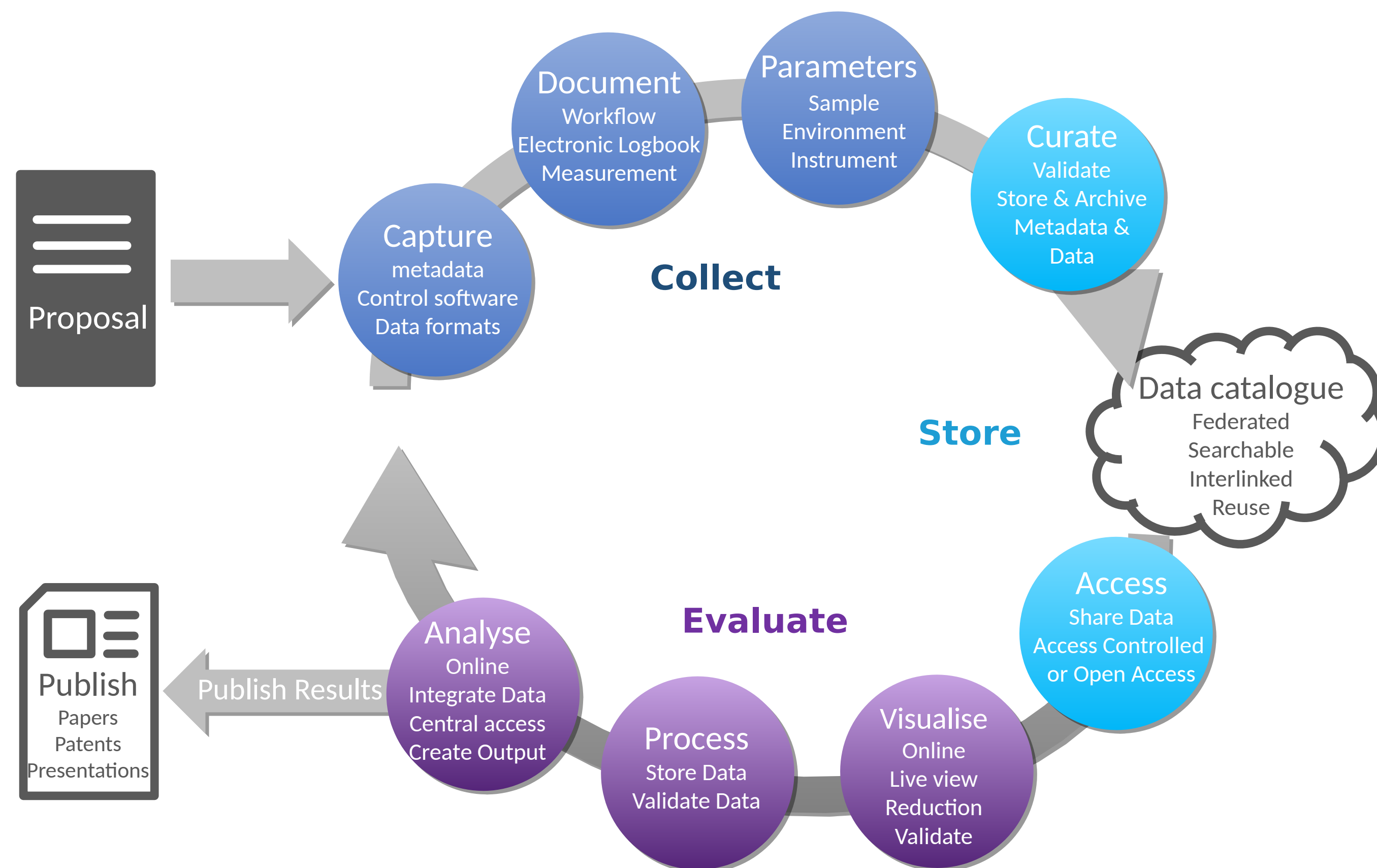


DAPHNE4NFDI

Moritz Hannemann

Jülich Centre for Neutron Science (JCNS) at MLZ, Forschungszentrum Jülich GmbH, Garching, Germany

DAPHNE4NFDI - What we want to accomplish



Automatic sharing of data

- Data management is a hassle
- Manually copying data to other servers is tedious and can lead to errors or to forget the data after some time
- Solution: An automatic data transfer system between the participating parties (institutes & universities)

Machine Learning

- More and more demand
- FAIR principles help to prepare for an infrastructure that supports ML
- Automatically searching data by some criteria, fetching and reading this data will be standardised, thus ready for everyone and any (ML) approach

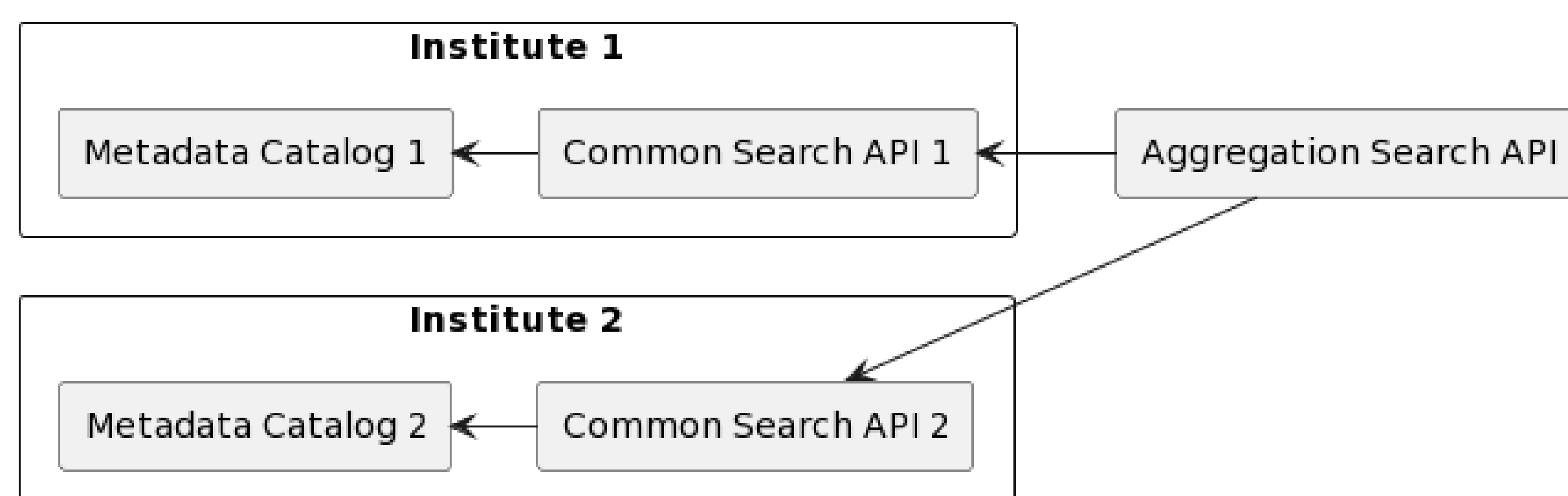
DAPHNE4NFDI - How we plan to do this

Describe data with metadata

Data describing data is called metadata. For example, it describes properties of data, like when it was recorded, under what pressure, by whom and so on.

Common search

- The data is stored in a decentralized way
- A search query first goes to a central gateway, which then calls all institute APIs
- Every institutes evaluates the query and returns its data



Ontology

- Common names for common concepts, eg. **t** for temperature
- A collection of those agreements is called an ontology
- These help that everyone knows exactly what to search for in their data and to return everything in the same manner
- This is not limited to variable names, other examples would be techniques or any other concepts

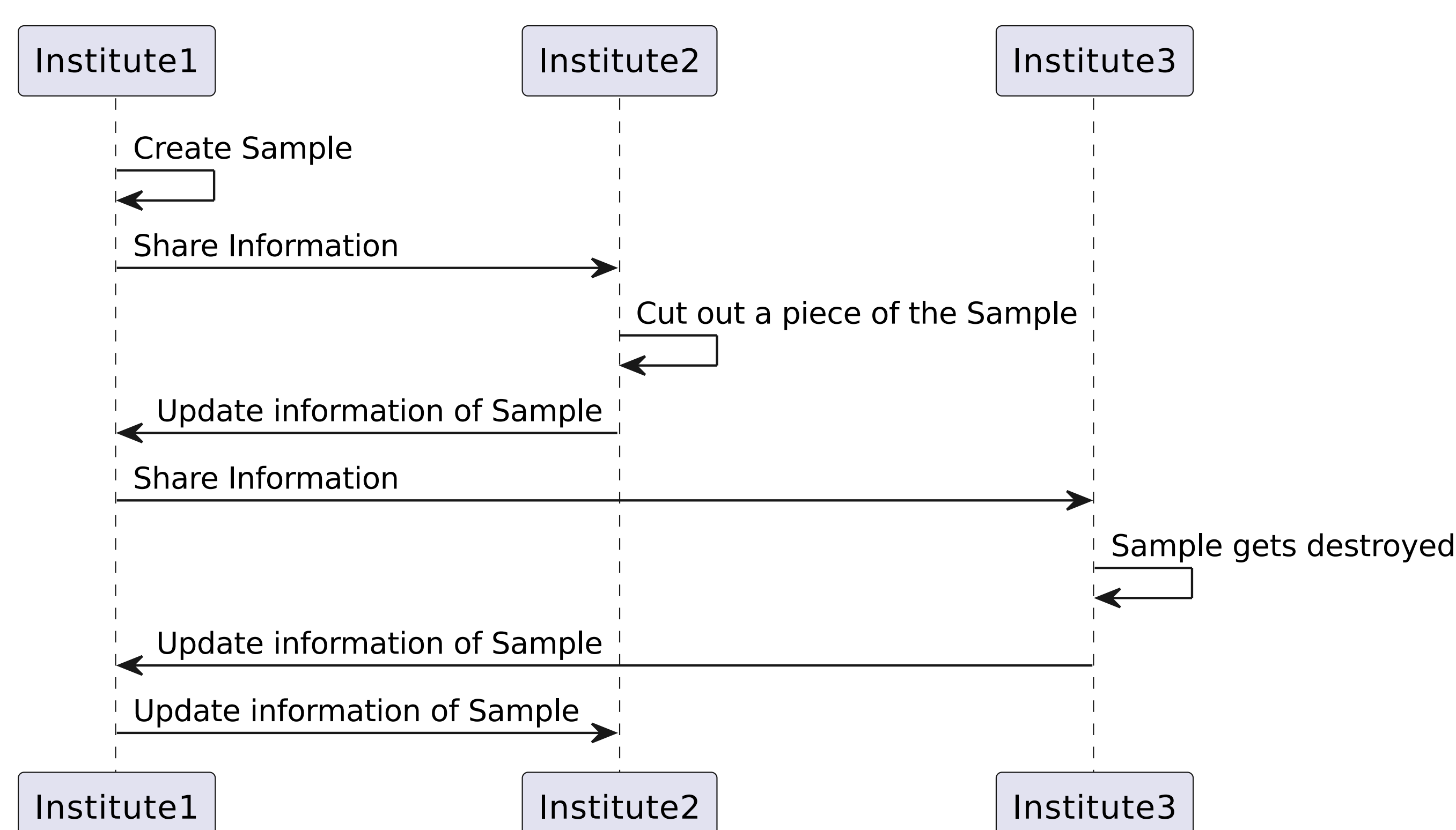
Metadata Catalogs

- Databases that store metadata in certain schemes
- Provides functionality to search data by metadata
- Real data is only referenced in the metadata catalog



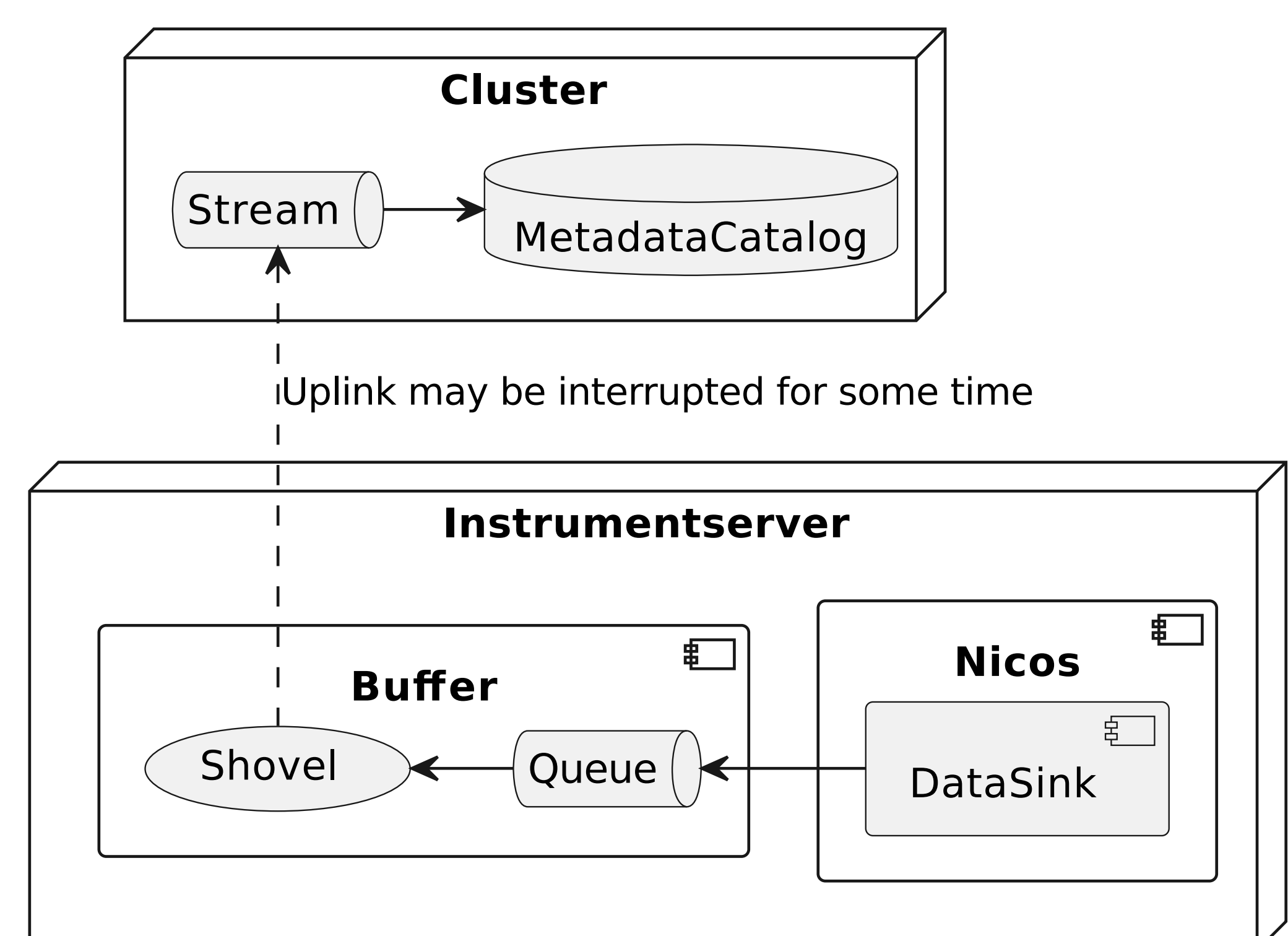
Problems & Visions

Sample federation problem



- Sample data should be createable & accessible everywhere
- Sample data should be editable everywhere
- Different datasets can mean the same sample but in different states → Sample data needs different versions
- How to synchronise sample data edits

Vision @ MLZ



- One instrument server for every instrument
- Decoupling daq from metadata ingest