

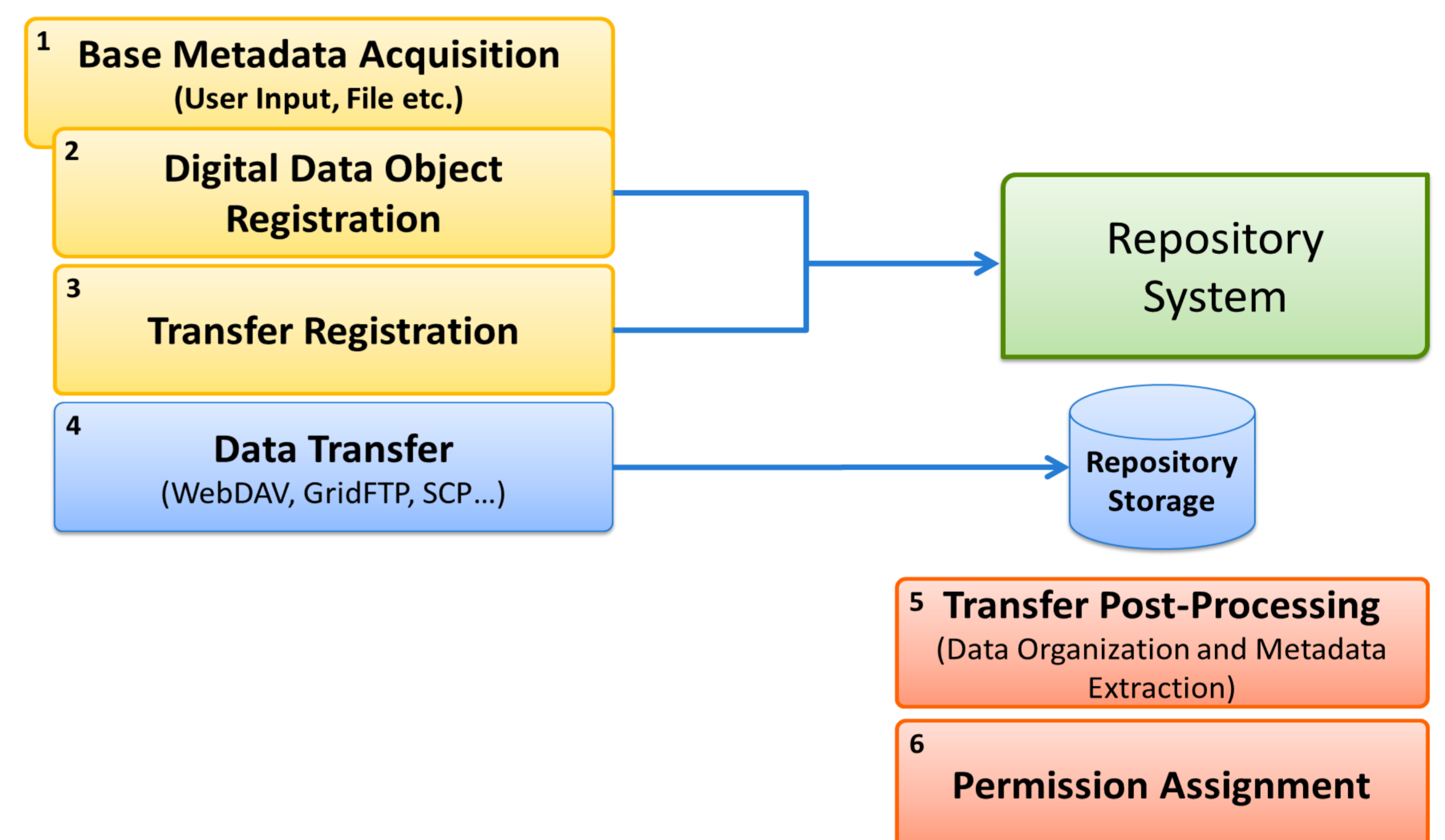
Scientific Repositories for Experimental Data

Thomas Jejkal, Volker Hartmann, Ajinkya Prabhune, Marius Appel, Francesca Rindone, Swati Chandna, Danah Tonne, Alexander Vondrous, Rainer Stotzka

In many scientific domains the necessity of sustainable data storage over a long period of time like decades and beyond is coming more and more into focus. Offered solutions for this challenging task are often limited by traditional file-based approaches. Hence, changing the view to object-based scientific data represents the paradigm shift exceeding these limits resulting in scientific repositories. The presented repository architecture allows to easily build-up scientific repositories and to implement RDA results and recommendations.

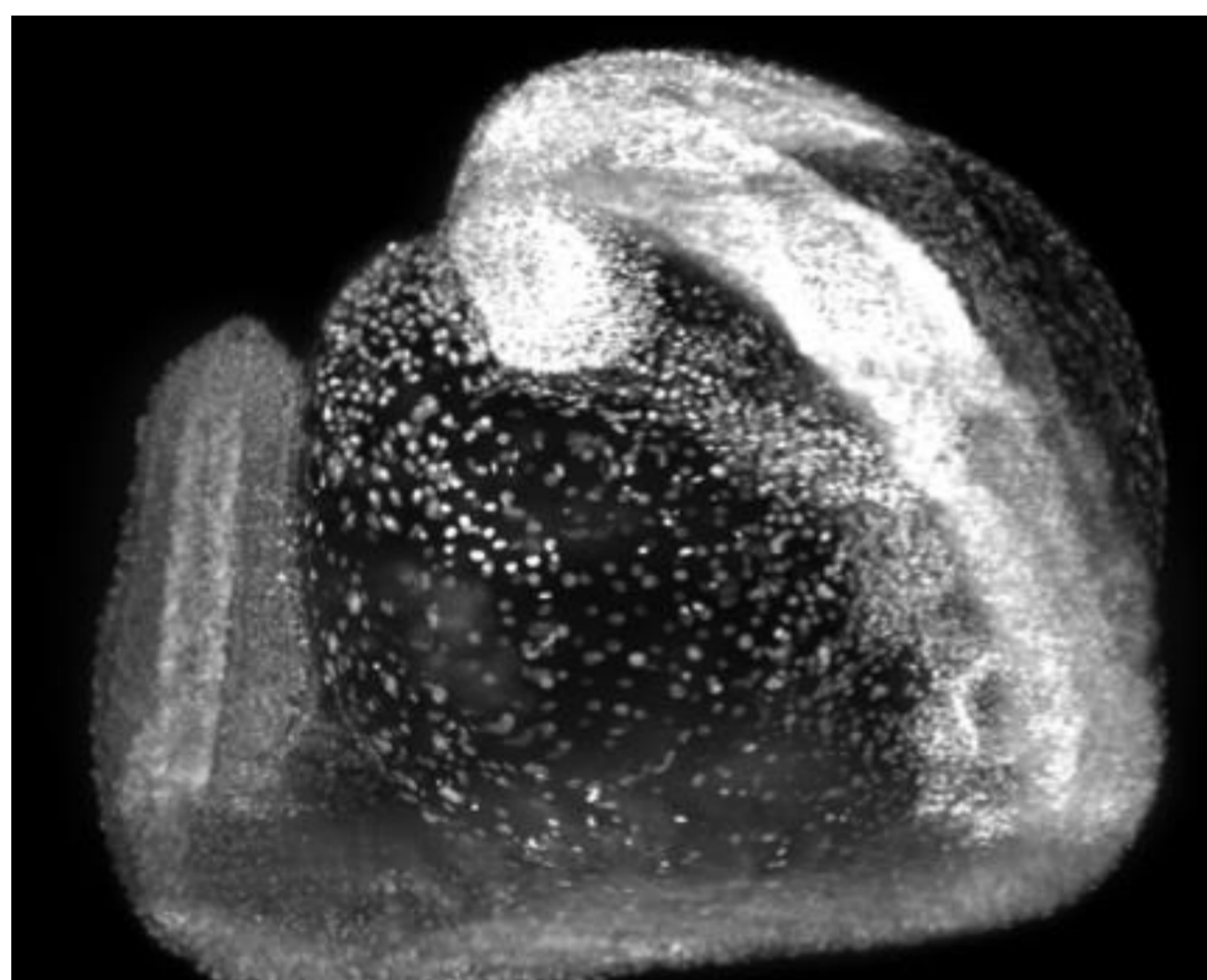
Requirements

- Digital Object-based approach linking data and metadata together
- Flexible data ingest for smooth integration into scientific workflows
- Support for extremely high data rates
- Interfaces to data analysis to allow integrated processing
- Data citation to be able to reference data
- Access policies to enable sharing and publishing data
- Bit and content preservation to enable trust
- Curation to mitigate digital obsolescence

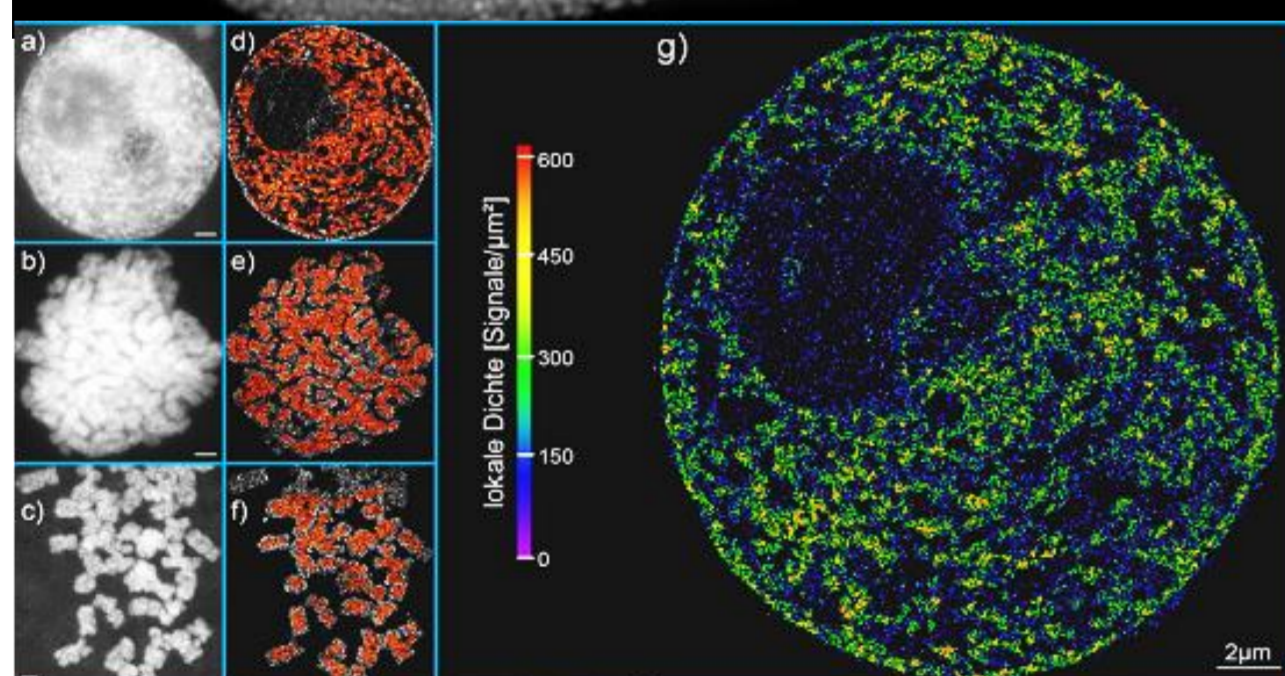


Data ingest workflow including digital object registration (1,2,3), data transfer (4) and post-ingest operations (5,6).

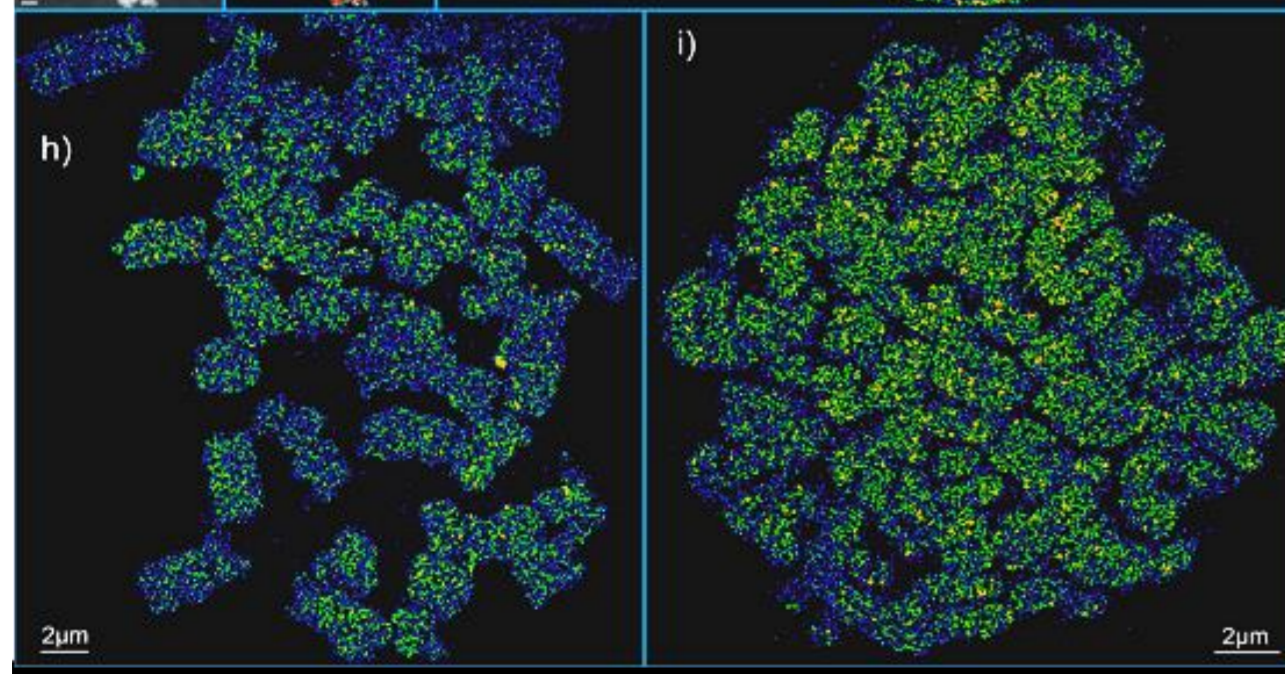
Implemented Use Cases in Data Life Cycle Labs



Extremely high ingest data rates of more than 400 MB/s for High content microscopy.



Setup of a reference data archive for extreme large datasets including data analysis for Nanoscopy.



Data analysis for volume rendering of fast synchrotron X-ray microtomography data.

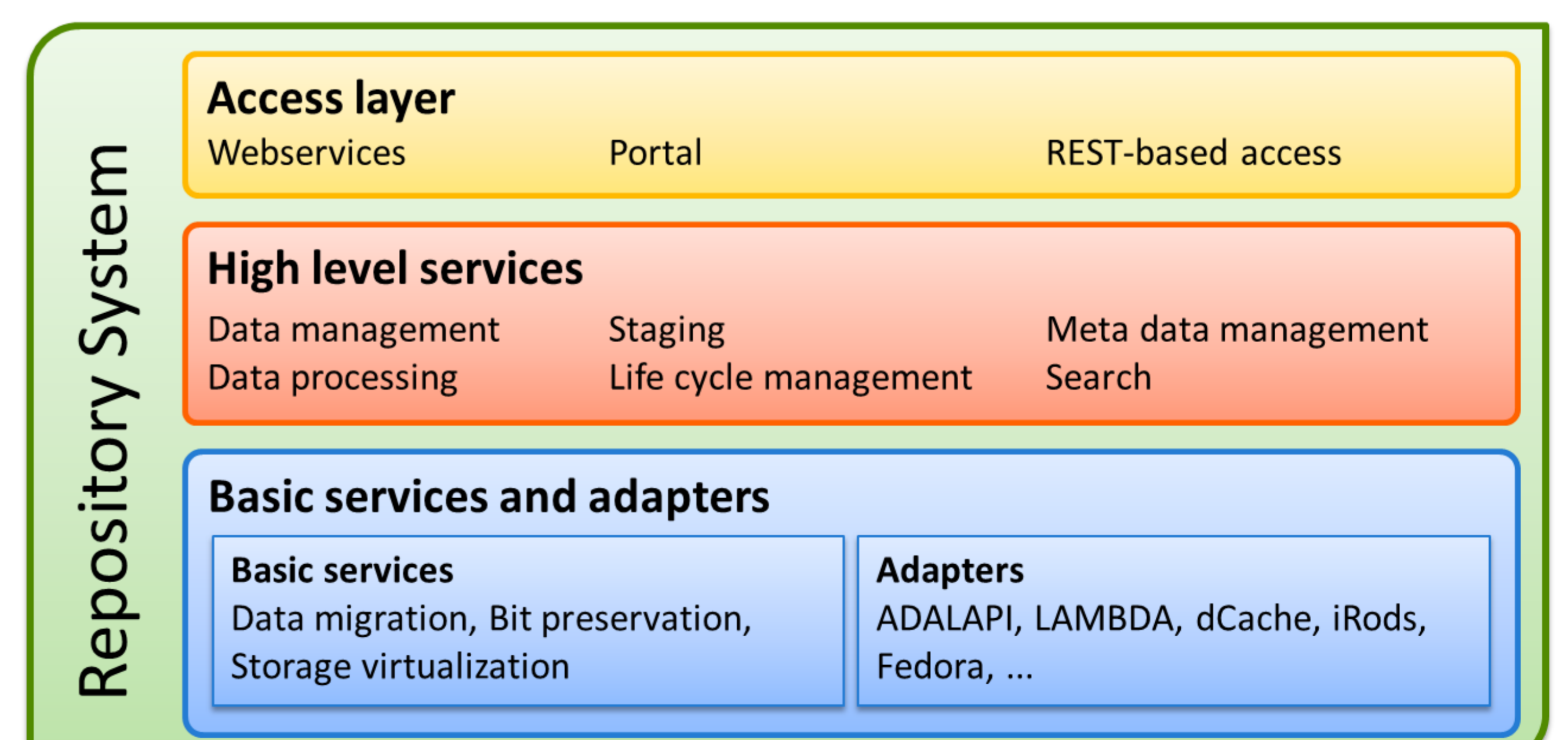


Scans of medieval manuscripts (left) and a high-resolution digital elevation model for archaeology (right).



Repository System Architecture

- Human & machine readable interfaces on Access Layer
- High level services provides generic building blocks applicable for many communities
- Well-defined interfaces on High level services layer should change very rarely
- Community-specific services can be based on High level services
- Basic services (e.g. resource services) accessed via adapters may change frequently
- Support for data migration easily possible without affecting users of higher layers



Overview of the repository architecture.