

Introduction

The Leibniz Research Alliance "Bioactives and Biotechnology" has formulated the development of "A common platform for data mining, data analysis tools, information and data management and exchange" as one of its goals. With allocated resources of approx. three man years, this is a very ambitious goal. We present progress which has been achieved so far.

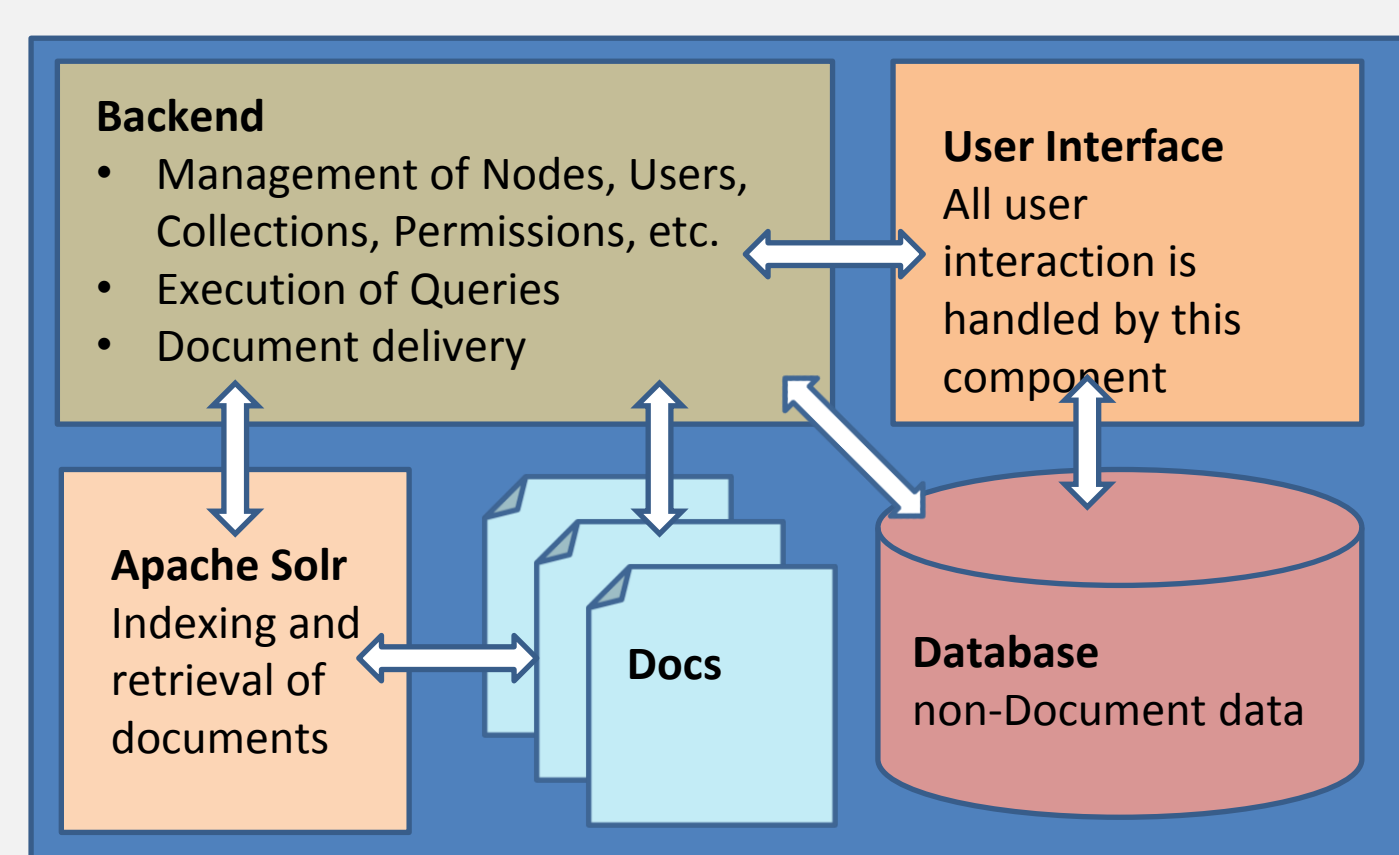
Project Description

The "Leibniz Bioactives Cloud" is conceived as decentralized network of nodes, which will be located at the individual research alliance member institutions. Nodes will be delivered as pre-configured virtual machines or containers to minimize on-site administration expenditure. Each node will keep the data and indexes of its institution, thereby giving the individual institutions more control of their data and avoiding some liability and funding issues, which come with a centralized architecture. However we are aware of some challenges regarding result scoring, storage demand, and responsiveness.

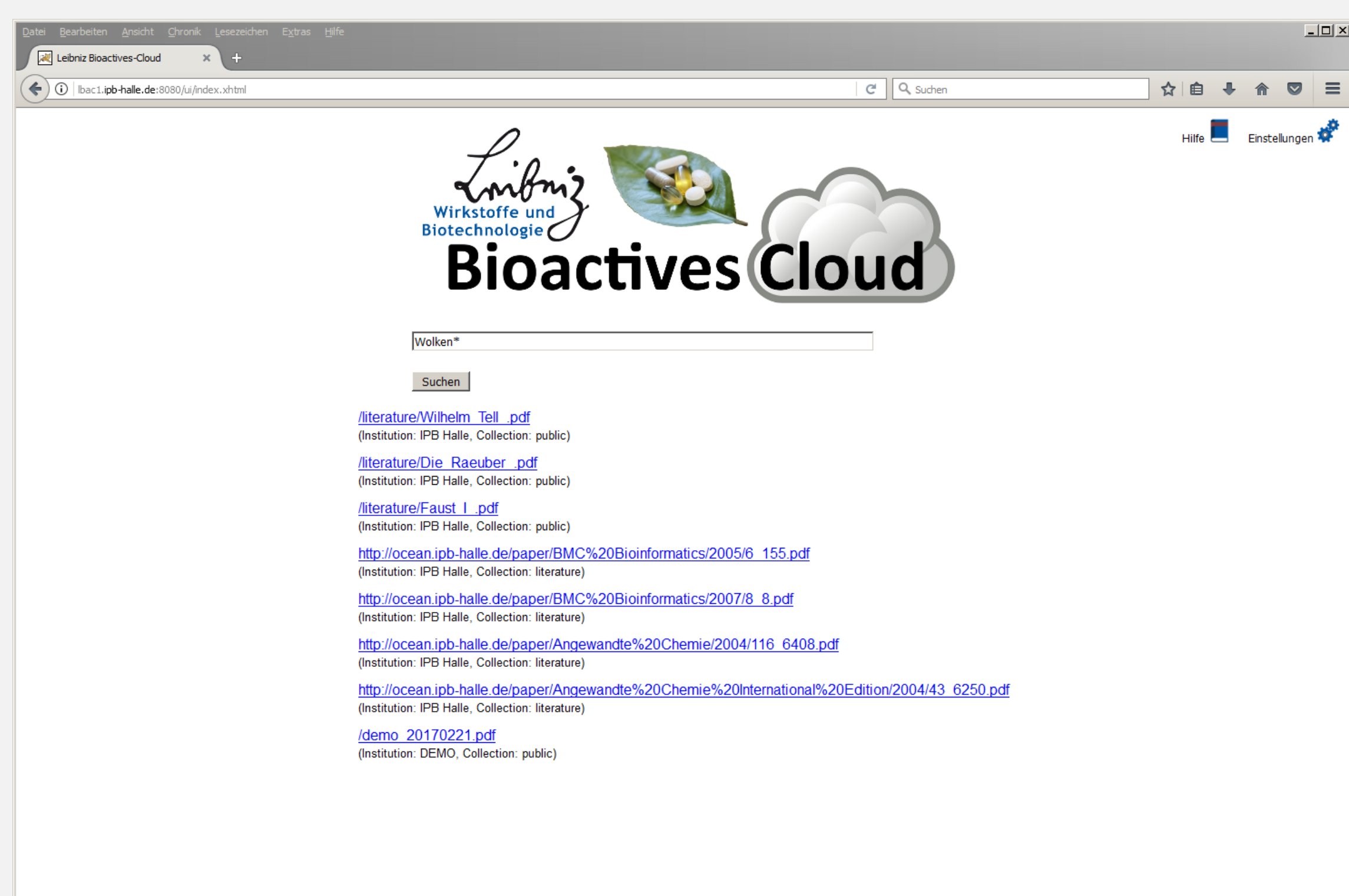
Development will be mainly JavaEE based and relies on well established Open Source software technologies. Related initiatives^[1-6] will be closely inspected to learn established approaches and reuse them where applicable.

Design Goals

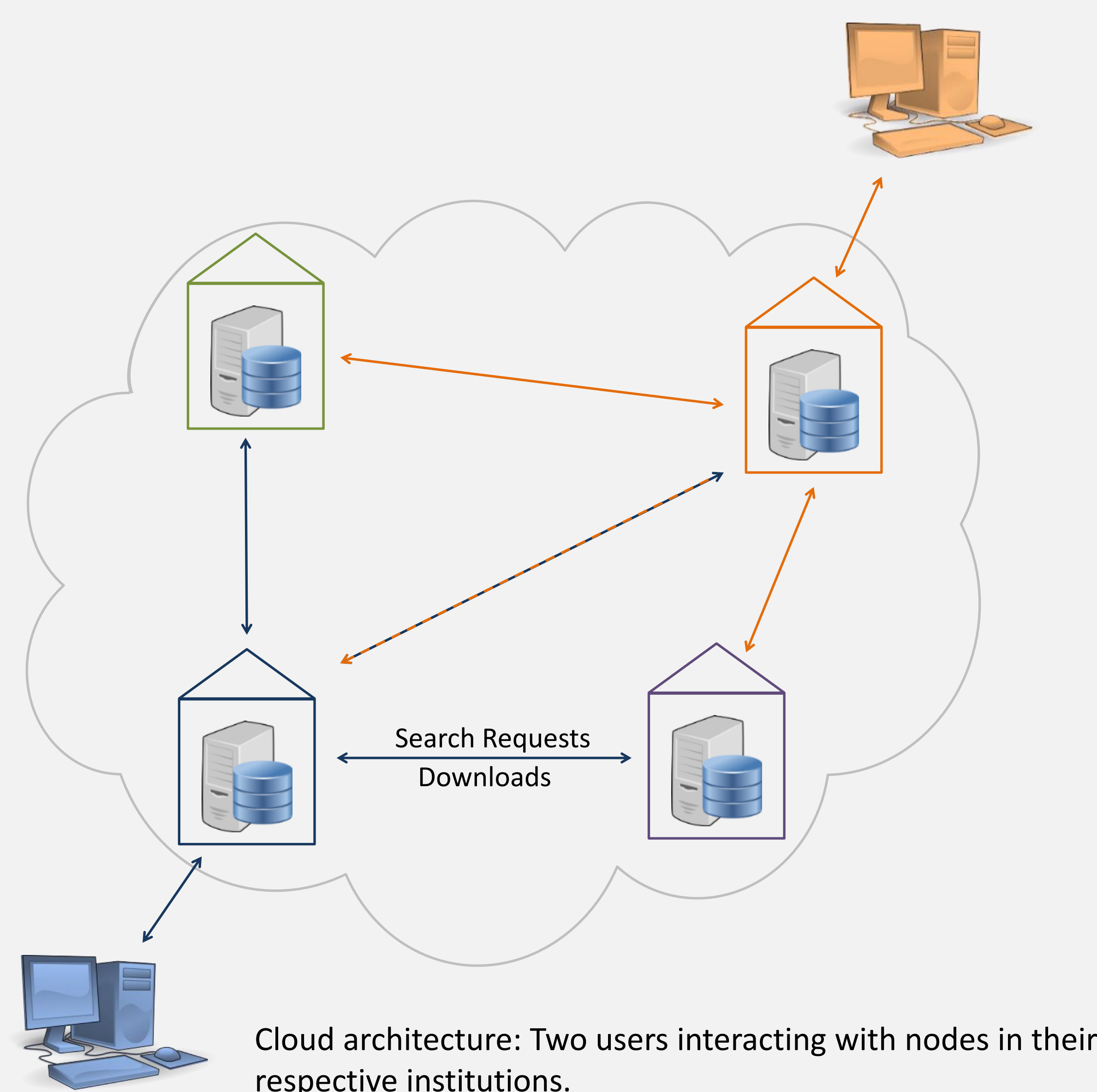
- **Security**
The system must be protected against unauthorized access to data and fraudulent misuse of the infrastructure.
- **Usability**
The system must be intuitive to use. Uploading and publishing as well as querying and download of data must be easy to accomplish.
- **Relevance**
The results and services provided by the Leibniz Bioactives Cloud must be relevant to their users. Special care will be given to chemical structures and assays.



Preliminary node architecture



Screenshot



Cloud architecture: Two users interacting with nodes in their respective institutions.

Prototype

The prototype can currently be characterized as follows:

- Distributed network of nodes, each hosting at least one public collection of documents
- Network drive interface for upload of new documents and automatic indexing by the search engine
- Execution of distributed queries and possibility to download result documents if liberated

Next Steps

- User authentication and authorization
- Non-public document collections
- Encryption of the network traffic
- Case studies on ontologies and chemical structures

Contribution

The development team welcomes ideas, comments and other contributions to this project. However, limited resources will place constraints on what will be achievable. We plan to establish regular web conferences to report on development progress.