LEXIS 2

Platform

Martin GOLASOWSKI, IT4Innovations

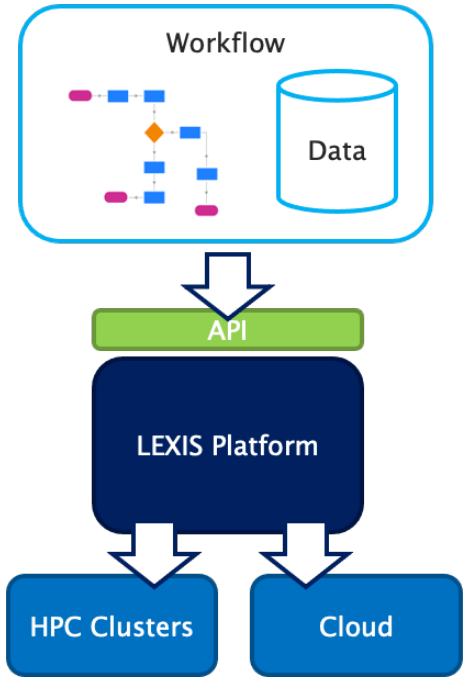
HPC-as-a-Service and distributed data management

Key points

More info:
https://docs.lexis.tech

LEXIS 2 Platform

- **Complex computing workflows across HPC and Cloud**
  - Workflow orchestration on federated HPC using ***Apache Airflow*** *and* ***HEAppE***
  - Secure access to HPC resources through REST API
- **Distributed Data Management**
  - Parallel data transfers by **iRODS**, metadata index based on **Elastic Search**
  - **FAIR** features enabled by EUDAT (PID generation, replication, indexing)
- **Web portal and interfaces for workflow set-up / execution**
  - Unified access to all services via OpenID-based LEXIS AAI
  - Zero trust security model implemented
  - Connected to **B2ACCESS** and **MyAccessID**
- **Easy HPC/Cloud access including SMEs/Industry**
  - **HPC-as-a-Service** approach – REST API for all features of the platform
  - Accounting multiple HPC centres - *CYCLOPS and HEAppE*



Open Source

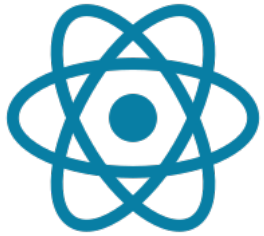
VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

Platform updates

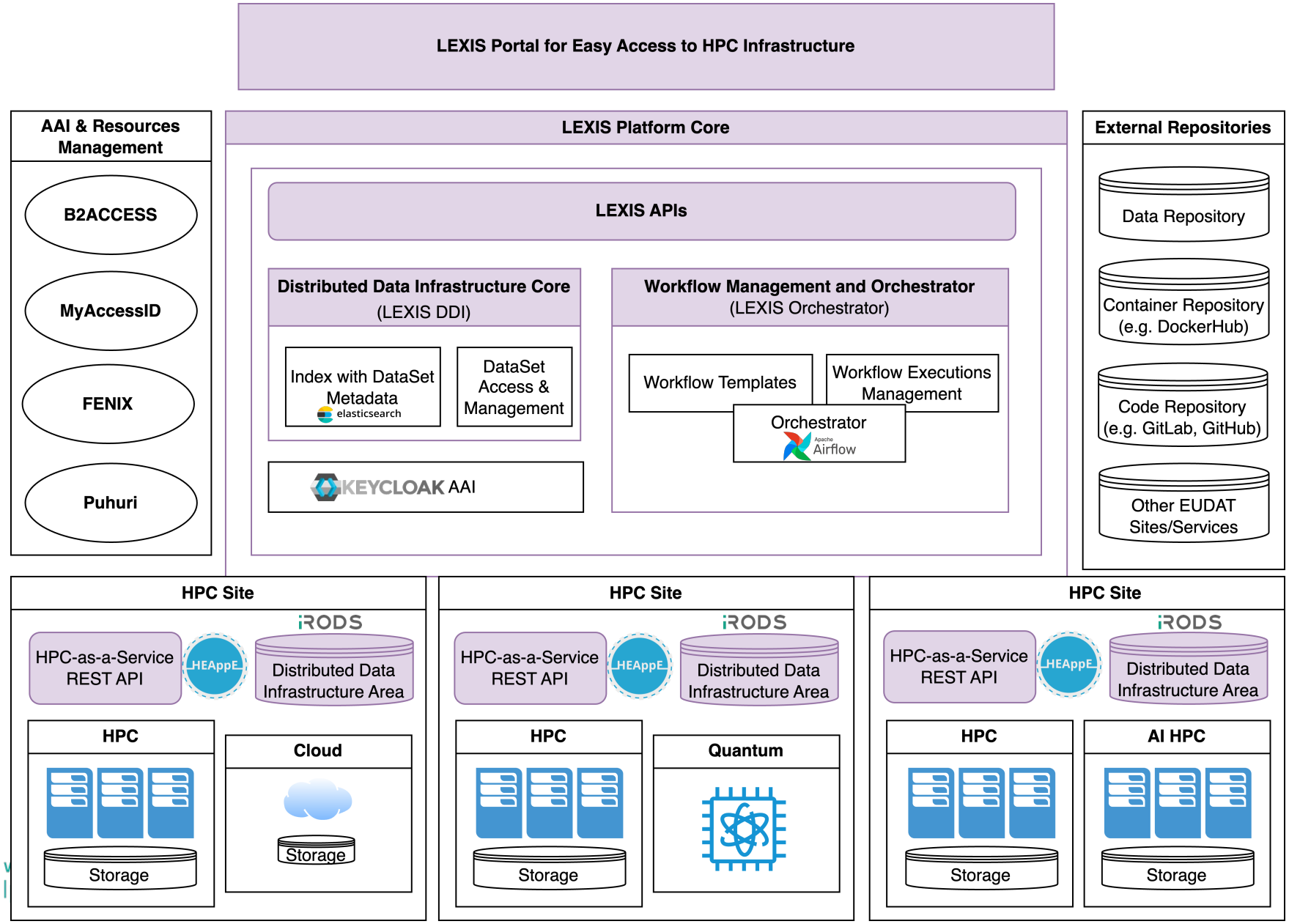
LEXIS 2 Platform

- New orchestrator based on Apache Airflow
- Frontend – version 2 written from ground up
- Better backend DB scheme
- Improved RBAC – fine grained permissions



- Connected locations
  - **IT4I** (data, HPC, Cloud)
  - **LUMI** (data, HPC, Cloud)
  - **IBM ZYC2** (HPC behind VPN)
  - **CINECA** (data & HPC - final step of the integration)
  - **CSC** – Puhti (ongoing)
  - **EPCC** – Cirrus
  - **LRZ** – CoolMUC
  - **DLR -** Terrabyte (ongoing)

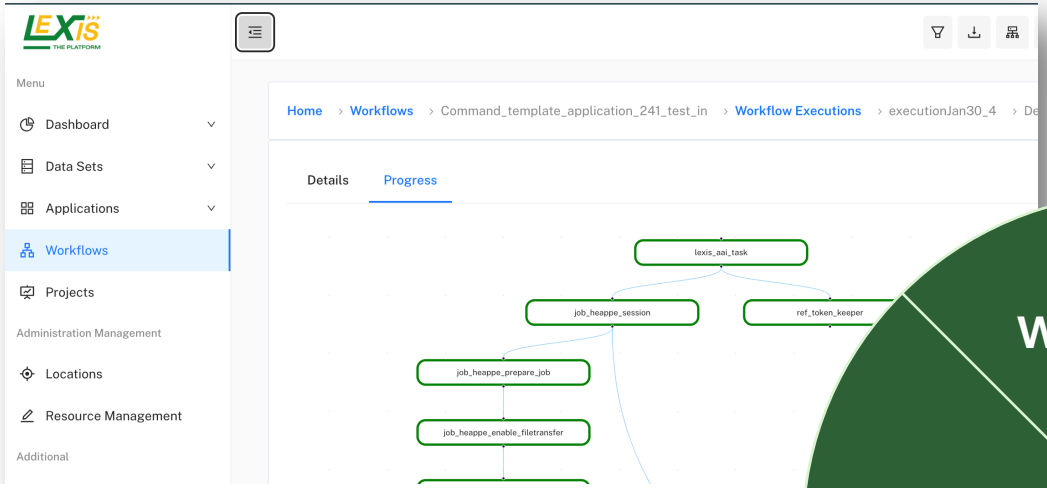
Platform architecture



- Federation of **European computing centres**
- **Hiding** of technical and **operational differences** across organizations
- **HPC & Cloud** service providers, Data providers
- Unified & distributed **data management**
- Orchestration
- **Federated Authentication & Authorization** Infrastructure (AAI)

User interfaces

LEXIS 2 Platform



Web portal

Python library SDK

Terminal based interface

Public REST API

README.md

py4lexis

Package py4lexis provides functions to manage datasets within the LEXIS platform. Uploading datasets' files is performed by tuspy Python package, i.e. by TUS Client.

Install

We recommend to use virtual environment, e.g.:

```
python -m venv ./venv
```

ivated virtual environment, type:

```
-m pip install py4lexis
```

```
preprocessing_dataset_path_geographical_data_path": "project/proj934a1763631a81e

flowExecutionName": "Run3",
flowExecutionConsent": True

are_data():
epared = WRF_EXEC_PARAMS
repared["inputParameters"]["preprocessing_start_date"] = datetime.datetime.now().str
#prepared["inputParameters"]["preprocessing_start_date"] = '2022120400'
return prepared

def submit_wf_execution(token):
    res = requests.post(LEXIS_API + WORKFLOW_ID + '/execution',
    json = prepare_data(),
    headers = {'Authorization': 'Bearer ' + token})

    if res.status_code != 201:
        logger.error(res.content)
        return

    return res.json()
```

```
(venv) mgolas@Martins-MacBook-Pro ~/I/L/L/Py4Lexis (develop)> python examples/upload.py
Progress:  |----------| 30.0% Uploaded
```

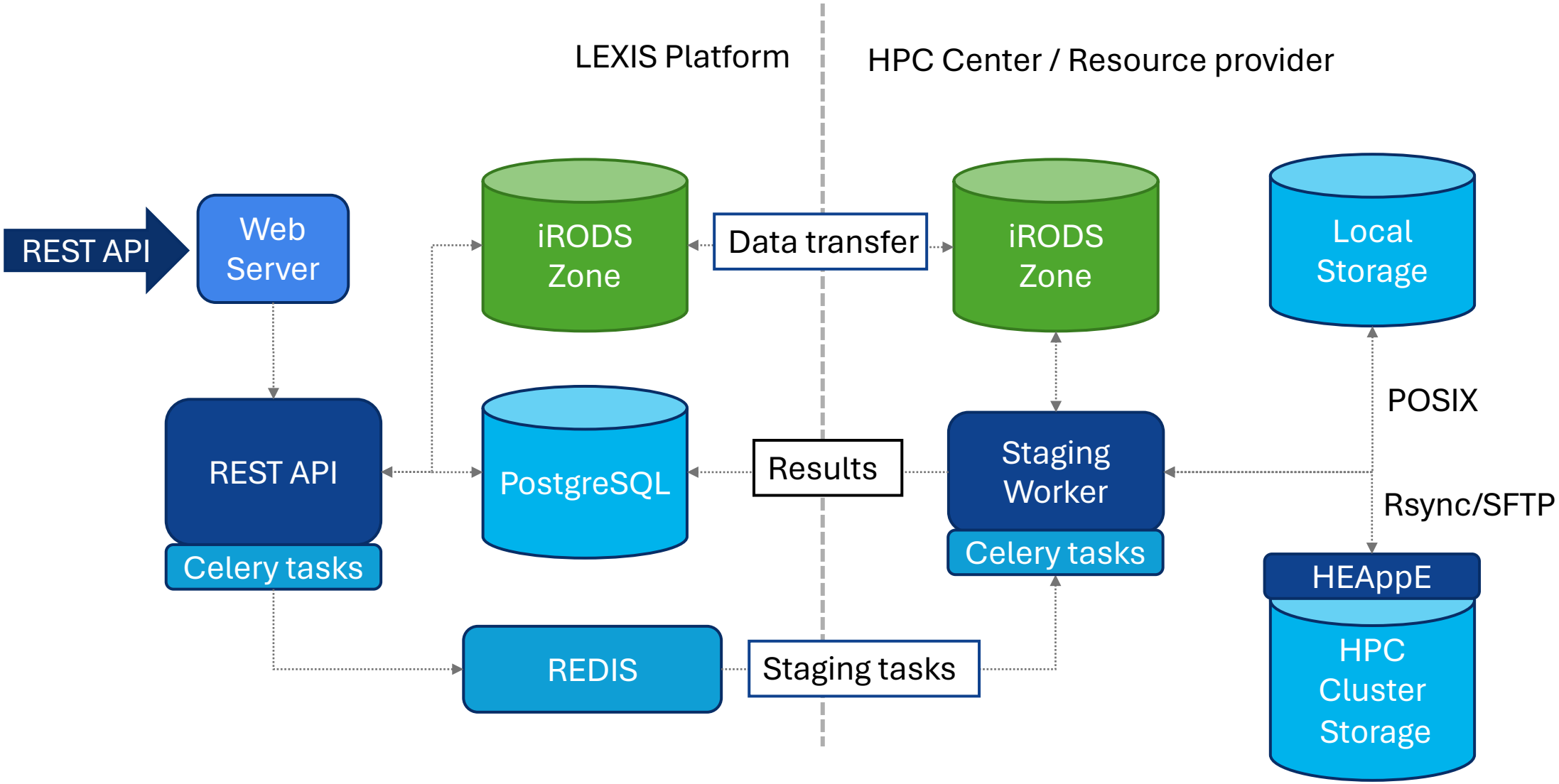
orkflowManagement Actions relating to management of Workflows and Workflow Executions

GET	/workflow	Return list of available LEXIS Workflows	
POST	/workflow	Create a new LEXIS Workflow on the system	
GET	/workflow/{workflowId}	Return detailed info on LEXIS Workflow for given Workflow ID	
DELETE	/workflow/{workflowId}	Delete LEXIS Workflow on the system	
GET	/workflow/{workflowId}/execution	List the current available LEXIS Workflow Executions.	listWorkflowExecutions
POST	/workflow/{workflowId}/execution	TODO: Needs implemented with TOSCA 1.3 Capabilitise. Create a new LEXIS Workflow Execution by providing remaining inputs	createWorkflowExecution
GET	/workflow/{workflowId}/execution/{workflowExecutionId}	Returns LEXIS Workflow Execution detail.	getWorkflowExecutionDetail
DELETE	/workflow/{workflowId}/execution/{workflowExecutionId}	Cancel a LEXIS Workflow Execution.	cancelWorkflowExecution

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

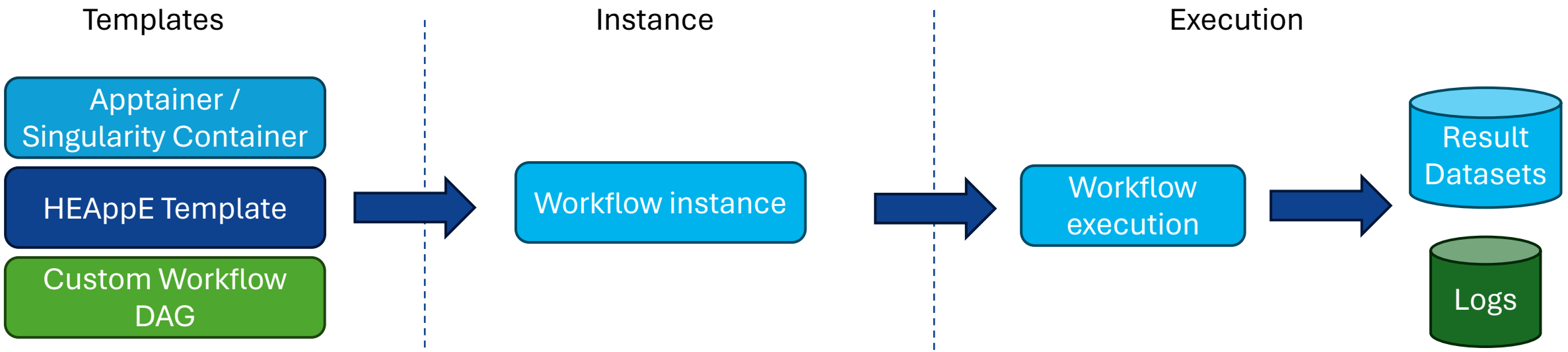
Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

LEXIS Staging API - Architecture



Workflow execution model

LEXIS 2 Platform



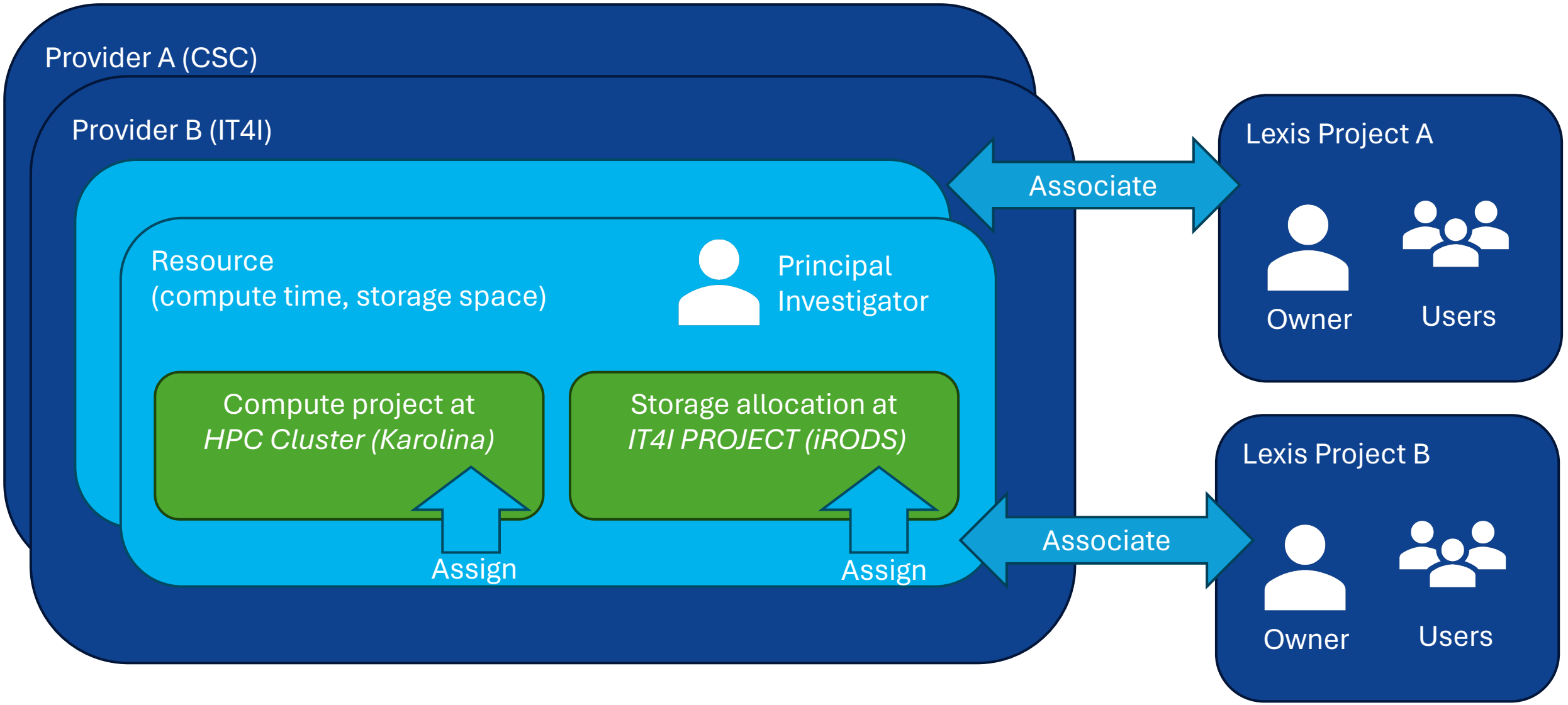
- Custom container
- HPC application deployed (compiled) by specialist
- Custom workflows with K8S, schedules, etc.

- Target allocation / resource
- Data handling
- Default execution-time-parameters

- Input / output datasets
- Application params / environment vars
- Scheduling policy
- Static location

Resource abstraction model

LEXIS 2 Platform



Selected use-cases

LEXIS 2 Platform

- LIGATE Project
  - Application for molecular docking simulation – private IP by DOMPÉ
  - LEXIS provides access to workflows with this application running on HPC
  - *Without direct access* to the binary or source code

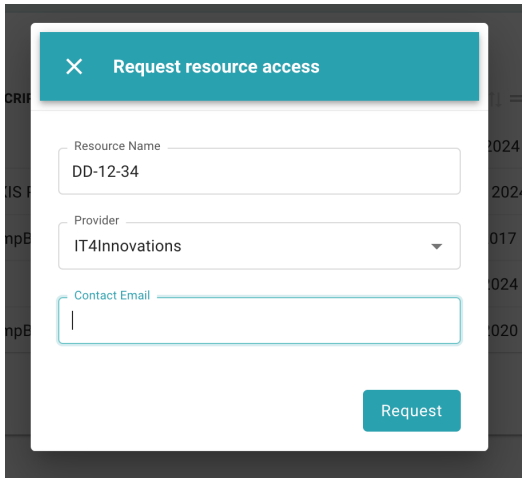
LIGATE

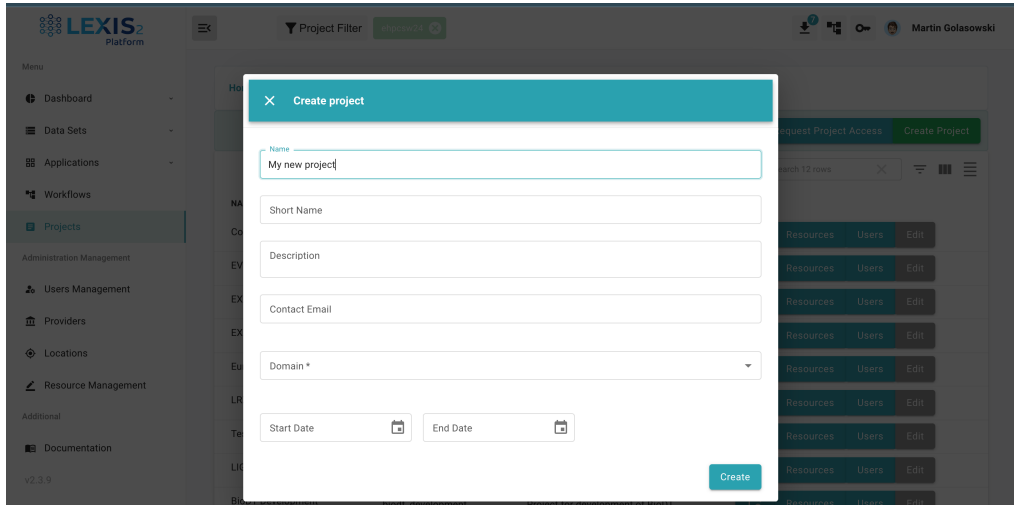
- OpenWebSearch.eu
  - European open web index processed through LEXIS on several HPC locations (LRZ, IT4I, CSC, DLR)
  - Public indices made available trhough the LEXIS Portal

OpenWebSearch.eu

Early adopters access for v2

- Want to use LEXIS Platform?
  - Obtain allocation on IT4I or other connected EuroHPC system
  - Login to the Portal at https://portal.lexis.tech
  - Create your LEXIS project
  - Request connection of your allocation to your LEXIS project
  - Contact us at support@lexis.tech



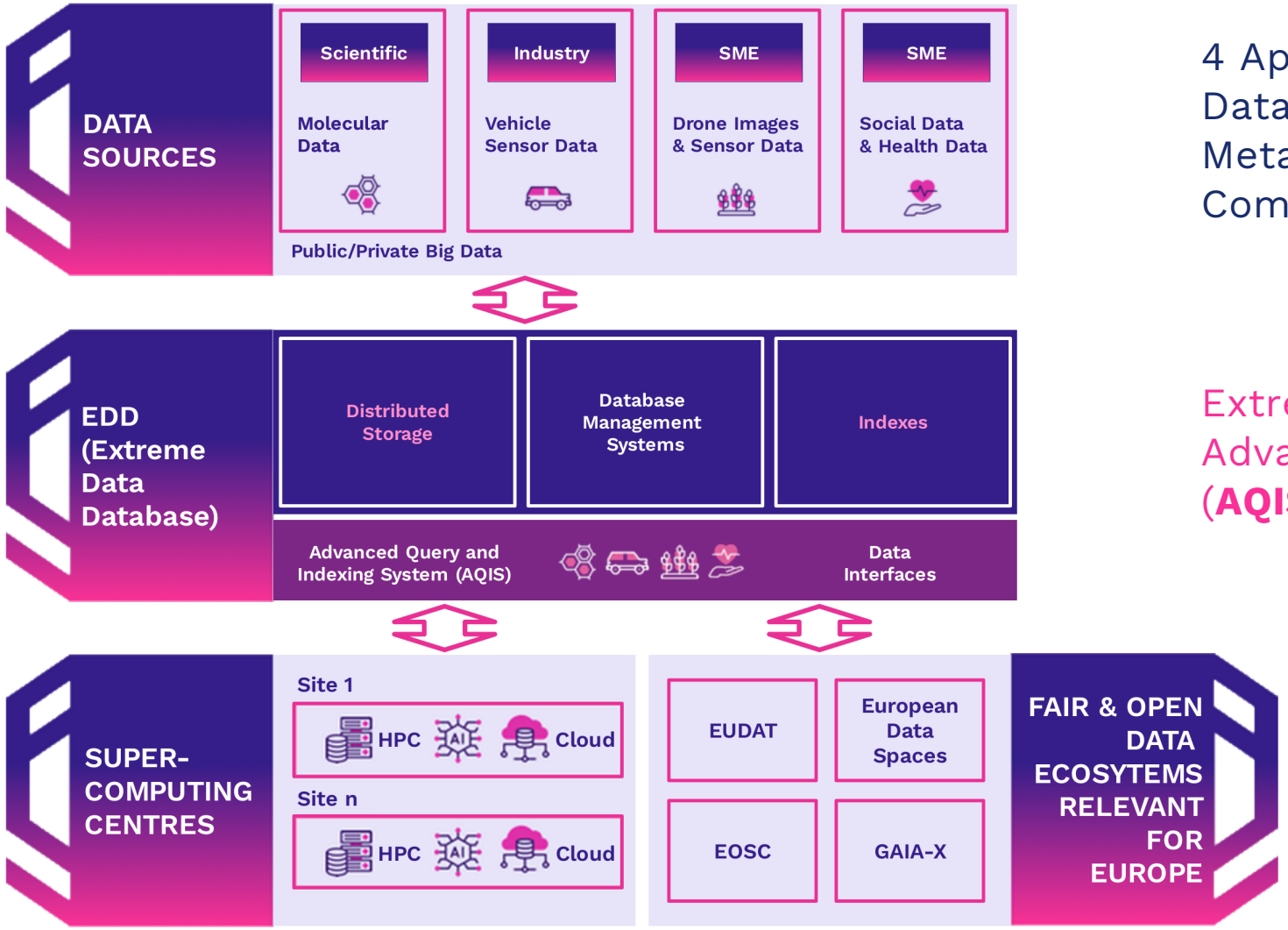


EXAHMIND

EXtreme Analytics For MINing Data spaces

Funded by
the European Union

What we do



4 Application Cases
Data transfer
Metadata ingestion and indexing
Complex computing pipelines

Extreme Data Database (**EDD**)
Advanced Query and Indexing System (**AQIS**)

Staging and workflow automation
FAIR Data access & publication

LEXIS and EXA4MIND Platforms Integration Plan

LEXIS 2 Platform

EXA4MIND

User interfaces
Web GUI
REST APIs

Resource management

Workflow orchestration

Metadata & control APIs

Indexes

HPC Parallel FS

POSIX

S3 Objects

CACHE

SQL

iRODS

AAI & User management

Monitoring

Smart scheduling

Distributed Data Infrastructure

Data staging and transfer for Extreme Data

Supercomputing Centre Resources

Cloud compute

HPC Clusters

Remote storage

LEXIS2 Platform

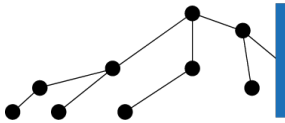
Projects & partners

LIGATE

BIODT
biodiversitydigitaltwin

EXAHMIND

CompBioMed


EVEREST

Open WebSearch .eu

ACROSS

e-INFRA CZ