LEXIS 2

Platform

Martin Golasowski, IT4Innovations, Czechia

Mohamad Hayek, LRZ, Germany

The LEXIS Platform V2 and using iRODS for large scale data management

How to supercompute...

LEXIS 2 Platform

- Pick a Supercomputer
- Fill out a request for allocation
- Get it approved
- Set up an account
- Set up a SSH key
- Know how to use Linux terminal
- Login in to the Supercomputer
- Learn how to launch jobs in SLURM
- Learn about modules
- Learn about storages
- … finally compute something

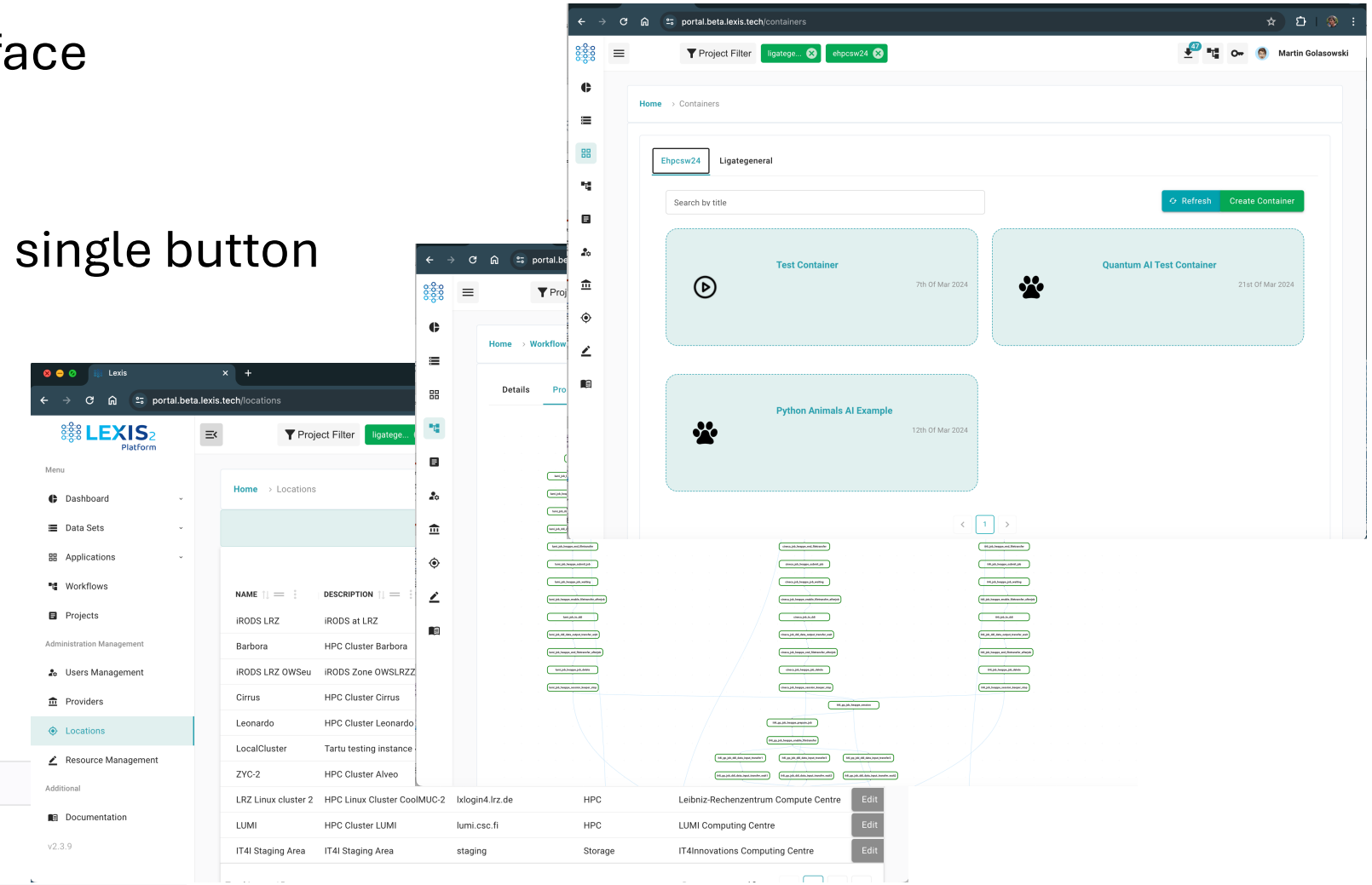
We can do better than that !

- Let’s have a nice web interface
- Allocations at one place
- Across multiple clusters
- Launch applications with a single button
- Get logs
- Manage data in iRODS
- Use common web login
- … and many more

Visit for more:

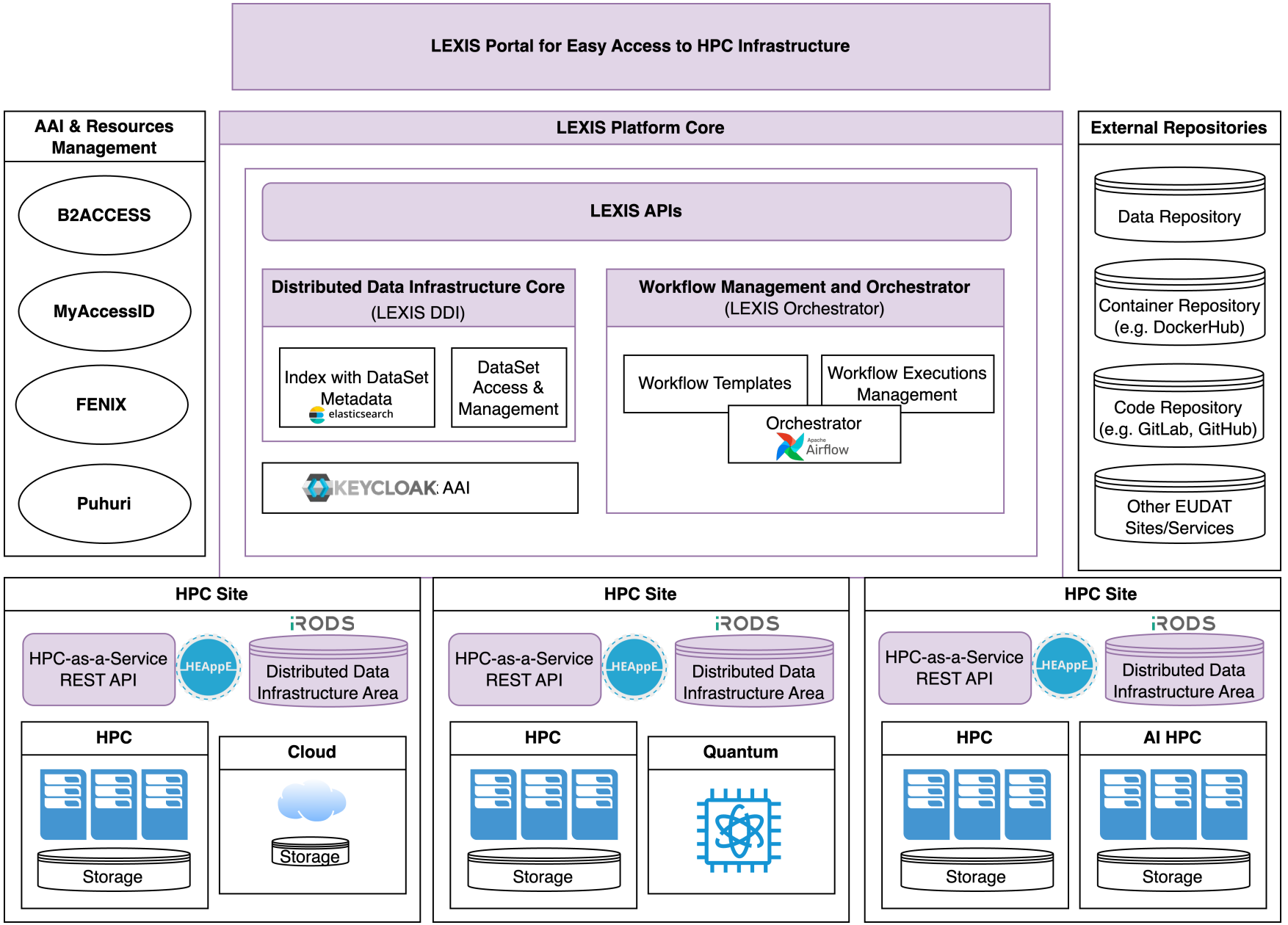
docs.lexis.tech

opencode.it4i.eu/lexis-platform



VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Architecture

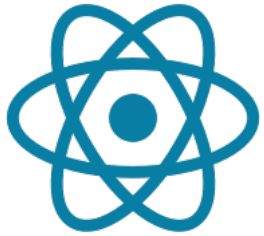


- Federation of **European computing centres**
- **Hiding** of technical and **operational differences** across organizations
- **HPC & Cloud** service providers, Data providers
- Unified & distributed **data management**
- Orchestration
- **Federated Authentication & Authorization** Infrastructure (AAI)

Platform Update

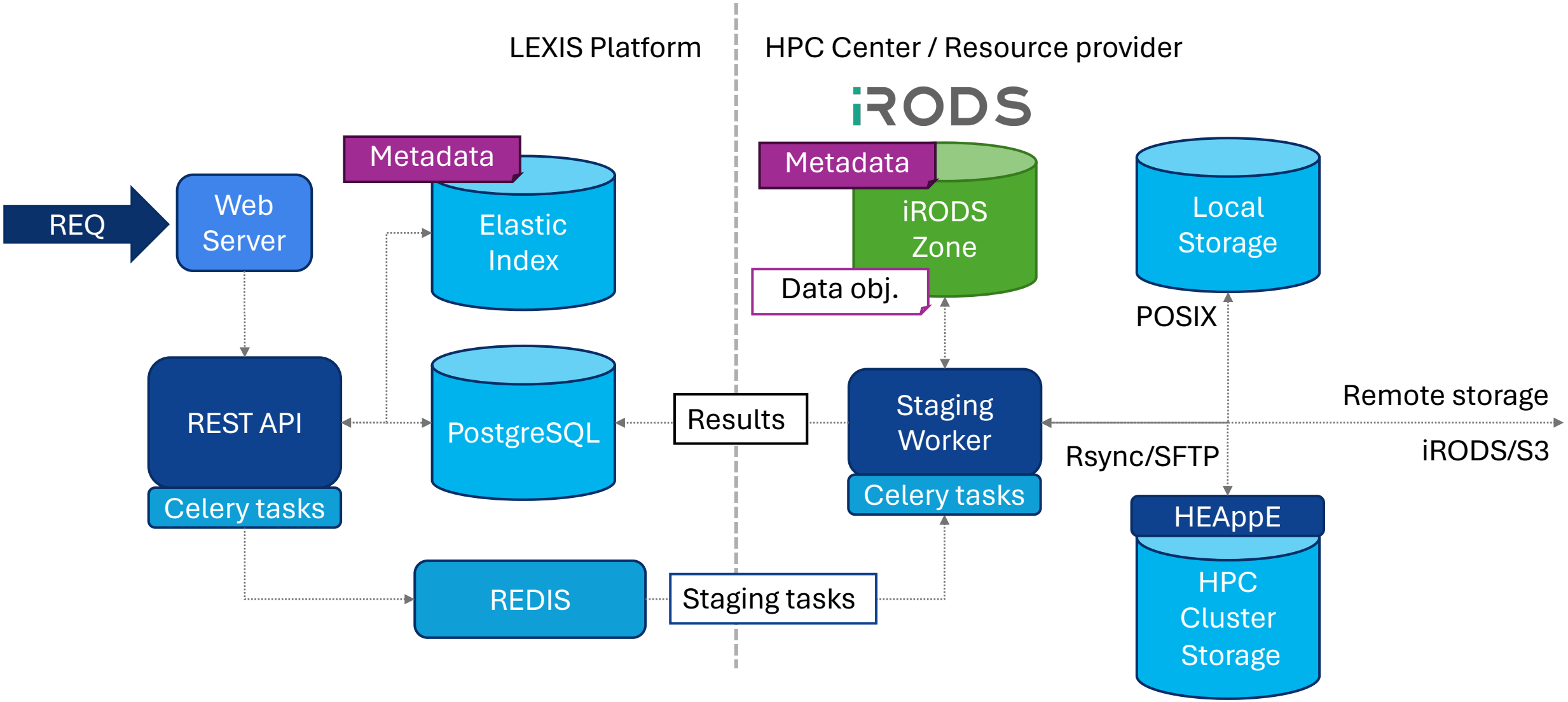
LEXIS 2 Platform

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



- Connected locations
  - **IT4I** (data, HPC, Cloud)
  - **LUMI** (data, HPC, Cloud)
  - **IBM ZYC2** (HPC behind VPN)
  - **CINECA** (data, HPC ongoing)
  - **CSC** – Puhti
  - **EPCC** – Cirrus
  - **LRZ** – CoolMUC
  - **DLR -** Terrabyte (ongoing)
  - **VEGA** (ongoing)
  - **VSC** (ongoing)
  - **SURF** – Snellius (ongoing)

LEXIS Distributed Data Interface - DDI



How do users access data?

LEXIS2 Platform

Web browsers / API

Pros

- Intuitive
- Network friendly

Cons

- Slower
- Requires moving the data “out of iRODS” or buffered streaming

iRODS clients

Pros

- Much faster
- No middlewares (except OpenID)

Cons

- Harder to set-up
- Requires passing around iCAT addresses
- Collection must be created through API

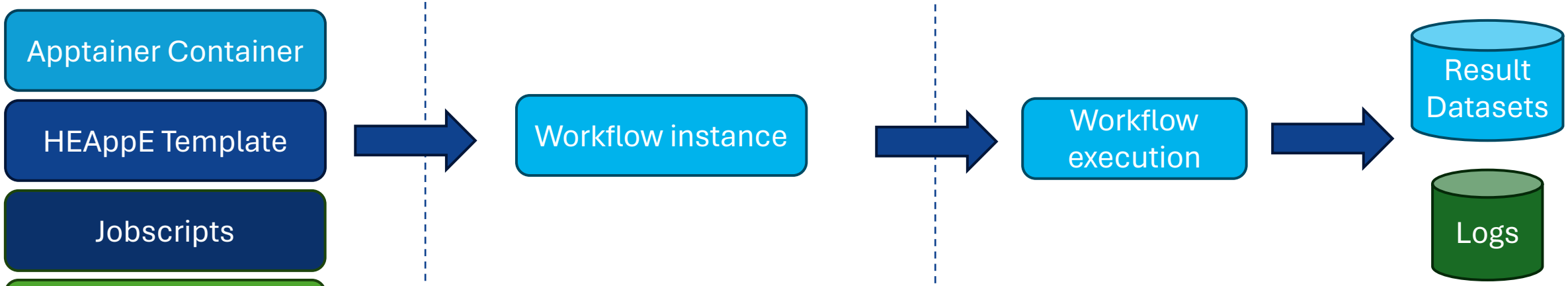
Workflow execution model

LEXIS2 Platform

Templates

Instance

Execution



- 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

Selected use-cases

LEXIS2 Platform

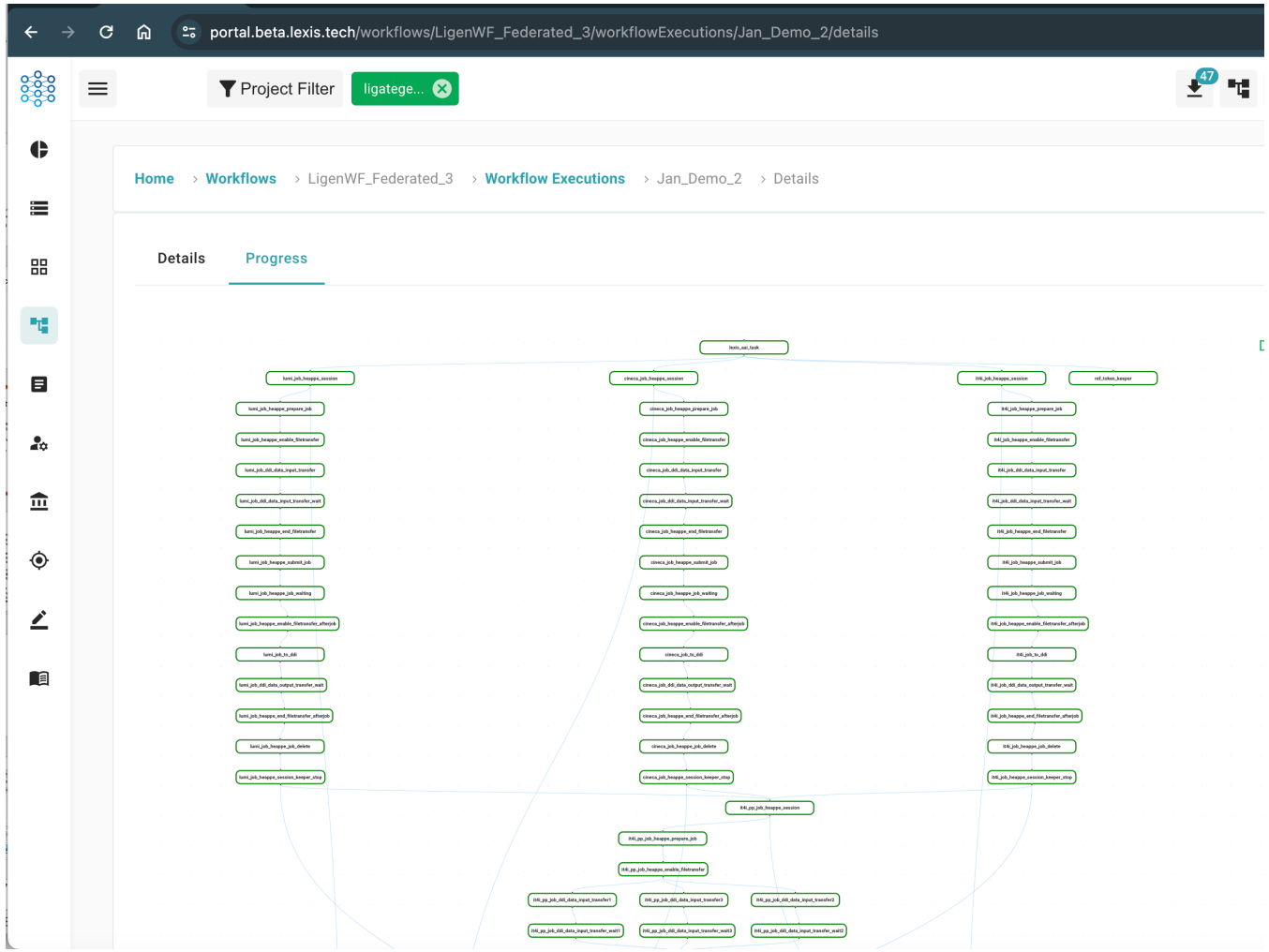
- LIGATE Project
  - Application for molecular docking simulation – private IP by DOMPÉ
  - LEXIS provides access to workflows with this application running on HPC
  - *Without direct access* to the binary or source code

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

Federated LiGEN Execution



Executed on 3 locations

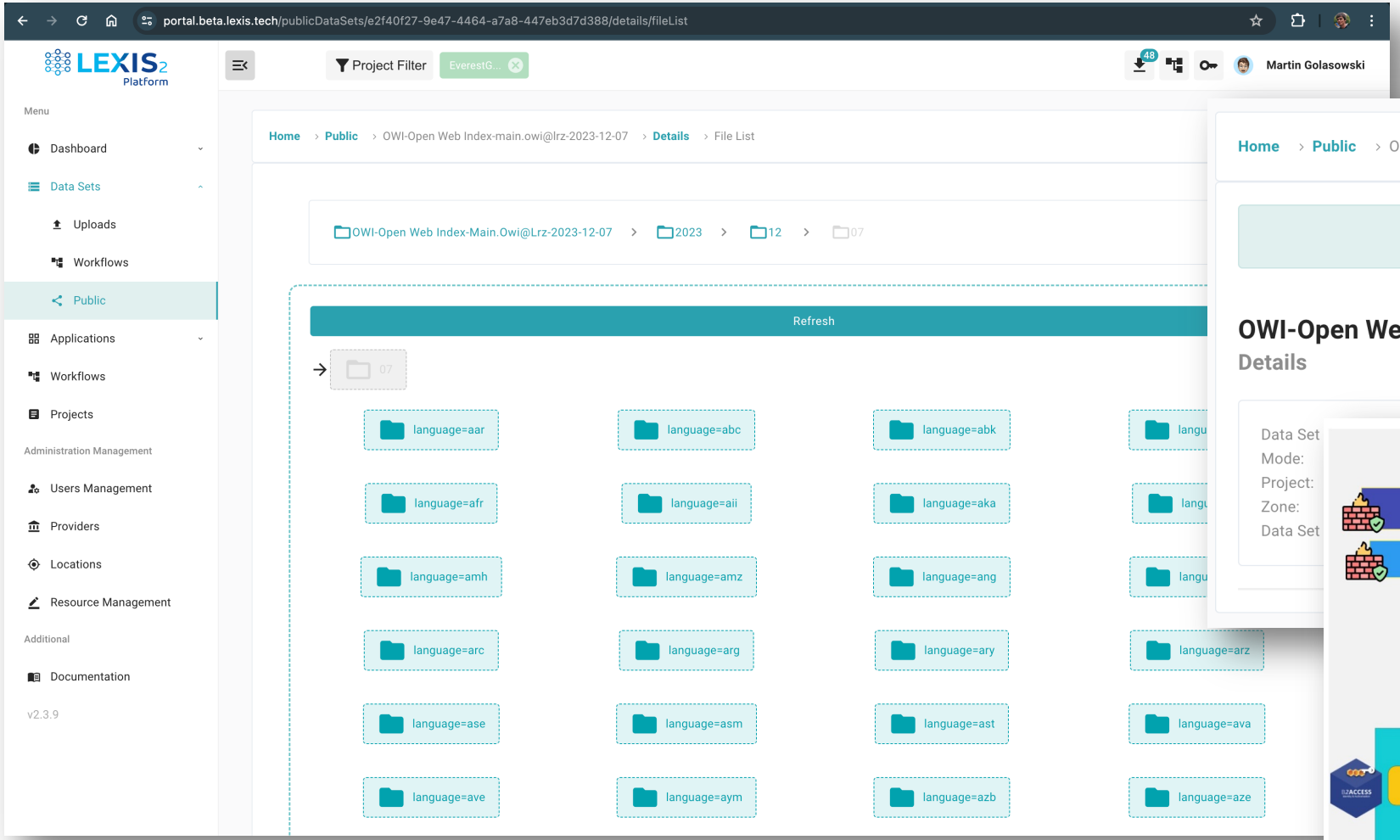
LUMI

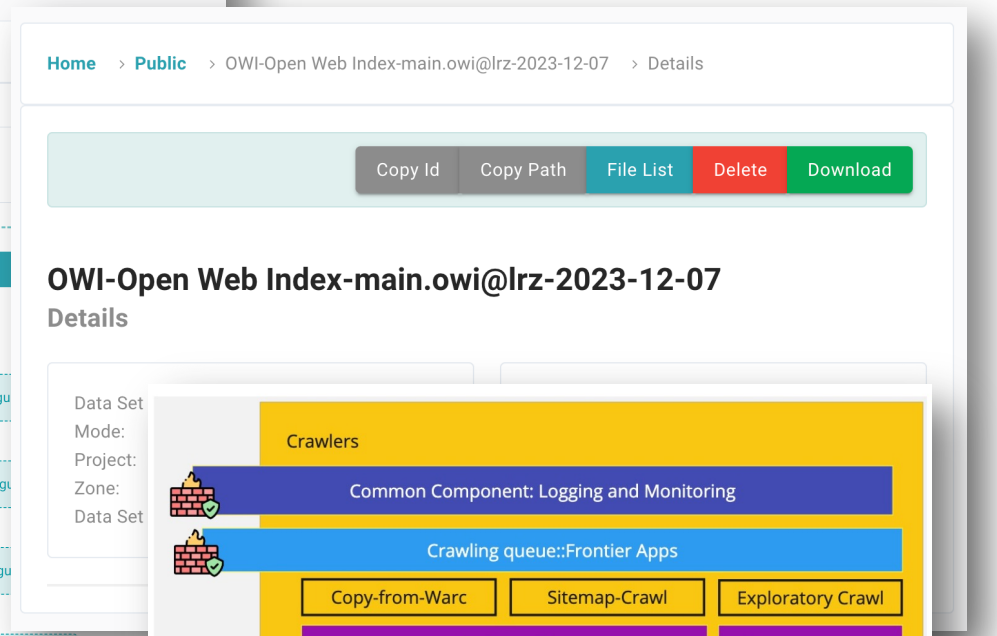
Leonardo (CINECA)

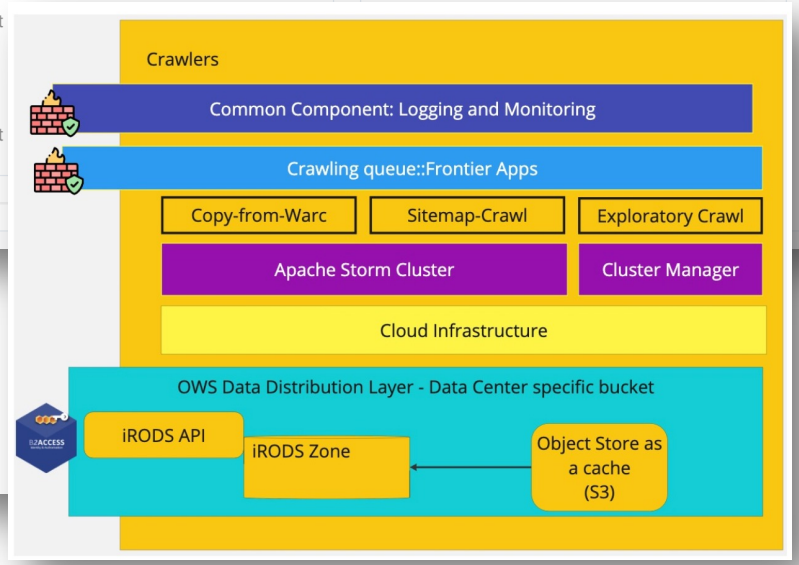
Karolina (IT4I)

OpenWebSearch.eu - Public web indexes

LEXIS2 Platform







VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Next steps and outline

LEXIS 2 Platform

- Implement more iRODS features in the web GUI
  - Tickets (i.e. link sharing)
  - Full-fledged file browsing
  - Metadata editing
- Improve downloads and uploads
  - Use direct connection to iRODS zones with buffering
- Apache Airflow iRODS provider
- Tracked replication between zones
- Metadata publishing to external services

iRODS

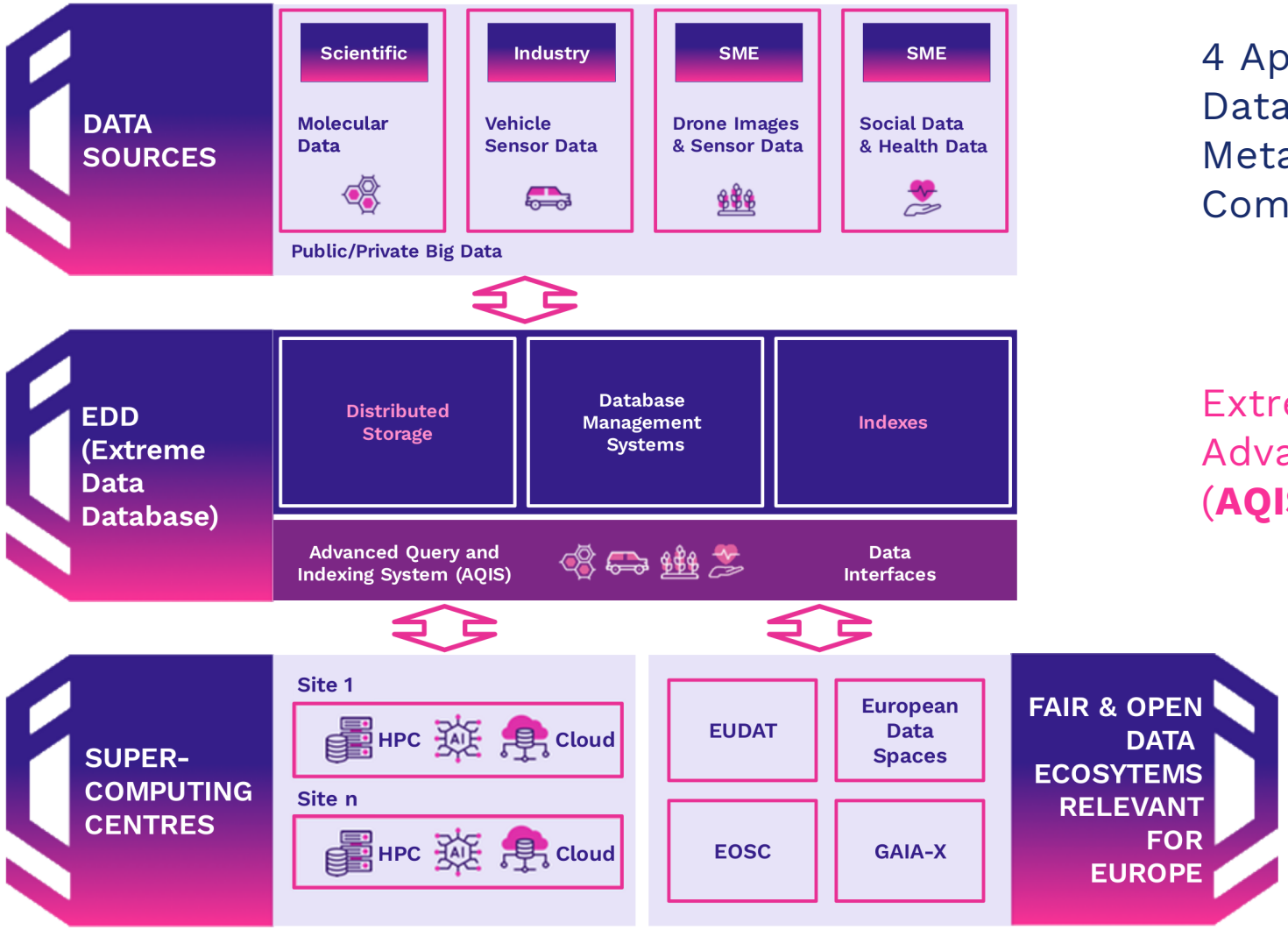
Apache Airflow

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

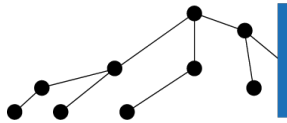
Projects & partners


LIGATE


BIODT
biodiversitydigitaltwin


EXAHMIND


CompBioMed


EVEREST

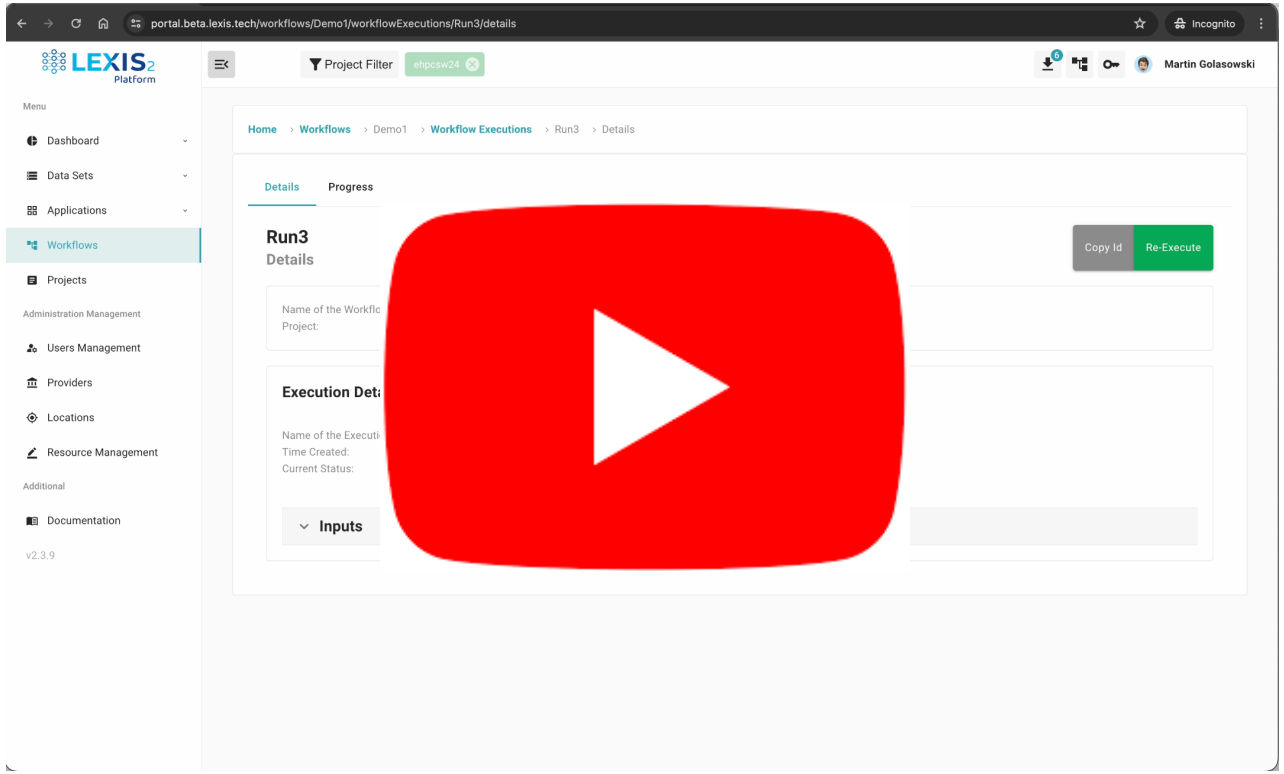

Open WebSearch .eu


ACROSS


e-INFRA CZ

AI workflow demo with Python in Apptainer

LEXIS2 Platform



Check it out!

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER