VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

EUDAT CONFERENCE 2026 / 25–27 NOVEMBER

LEXIS Platform and LUMI AI Factory

Connecting research data with AI computing

LEXIS2 Platform

LUMI AI Factory



Martin Golasowski · IT4Innovations

The gap between finding data and using it

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Discovery is only the first step: researchers also need permission, a usable data copy and computing access.

MANUAL HANDOFFS

Find data → Confirm rights → Transfer data → Configure computing

DATASET-AS-A-SERVICE (DAAS)

Discover → Check access → Prepare a use copy → Run on LUMI

DaaS complements preservation repositories with copies prepared for use near computing.



LEXIS Platform: a common interface to data and computing

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

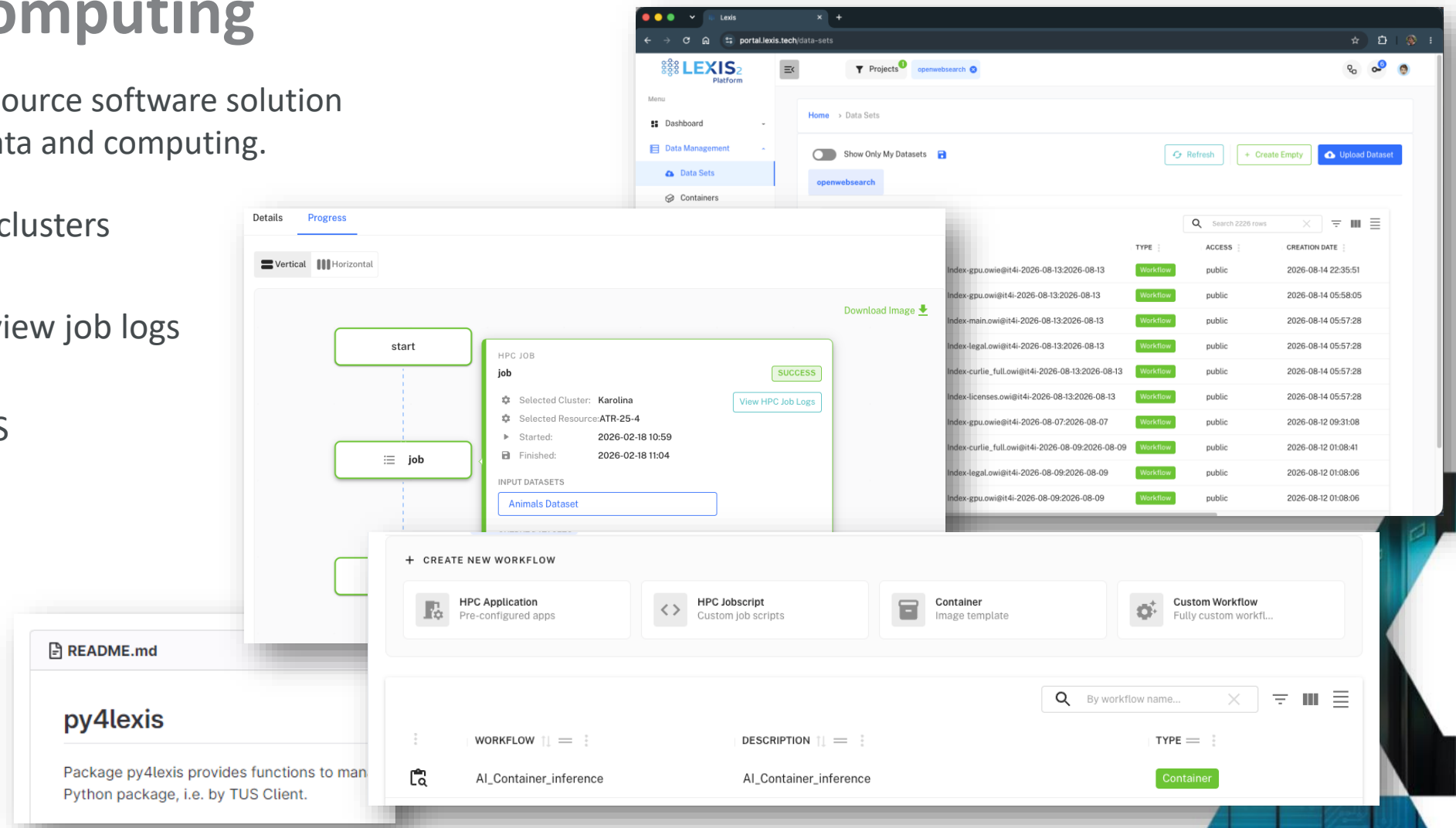
LEXIS Platform is an open source software solution for access to distributed data and computing.

Access allocations across clusters

Launch applications and view job logs

Manage datasets in iRODS

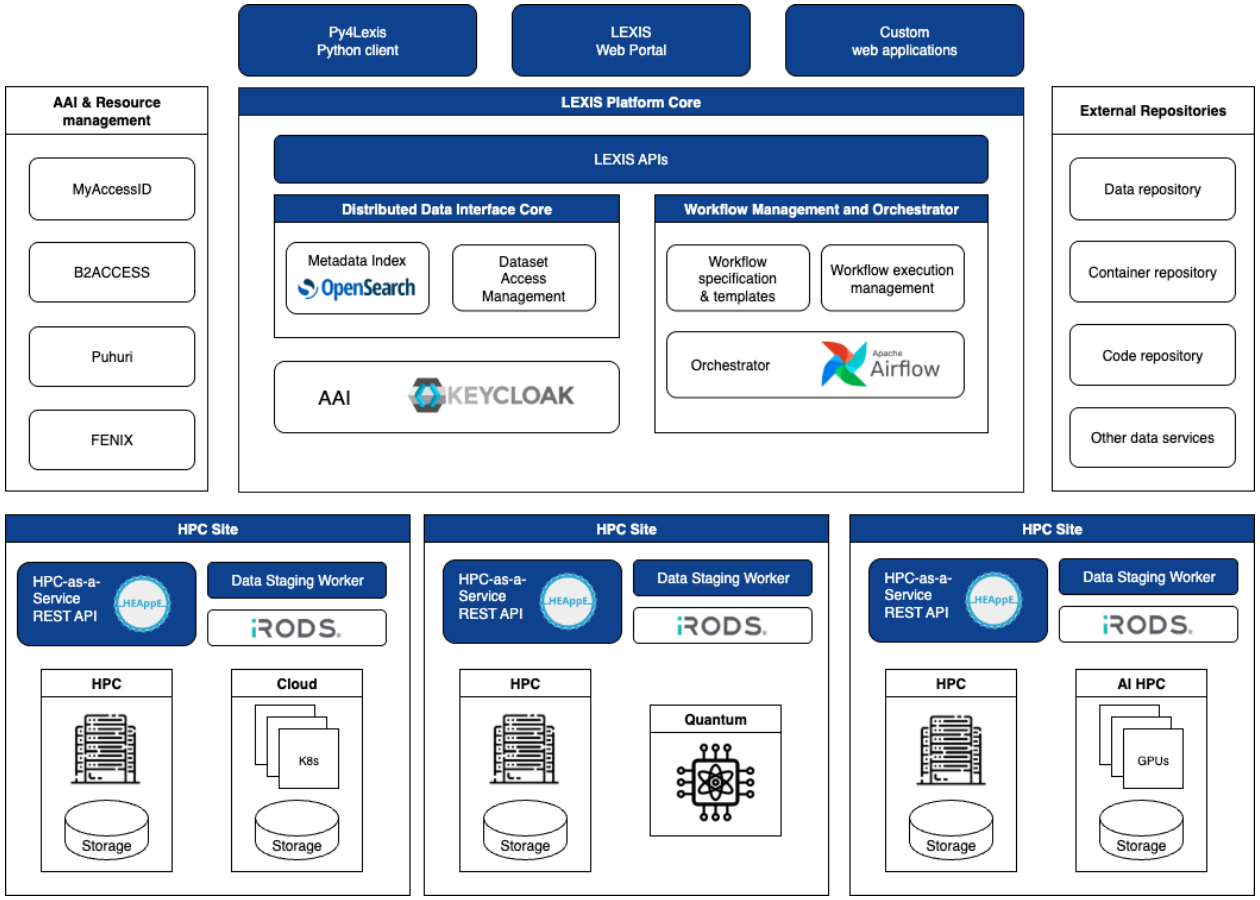
Use a common web login



www.lexis.tech

Federation across computing and data providers

LEXIS Platform coordinates distributed workflows and data through a common interface.



European centres provide HPC, cloud and data services

LEXIS Platform handles differences across providers

Workflows and data span sites

Federated identity supports login and access control

Part of the EuroHPC Federation Platform

Data management in LEXIS Platform

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Searchable descriptions connect to managed files at participating sites.

Searchable descriptions

OpenSearch index using DataCite 4.5

Local storage and quotas

Files stay at sites under project allocations

Direct data access

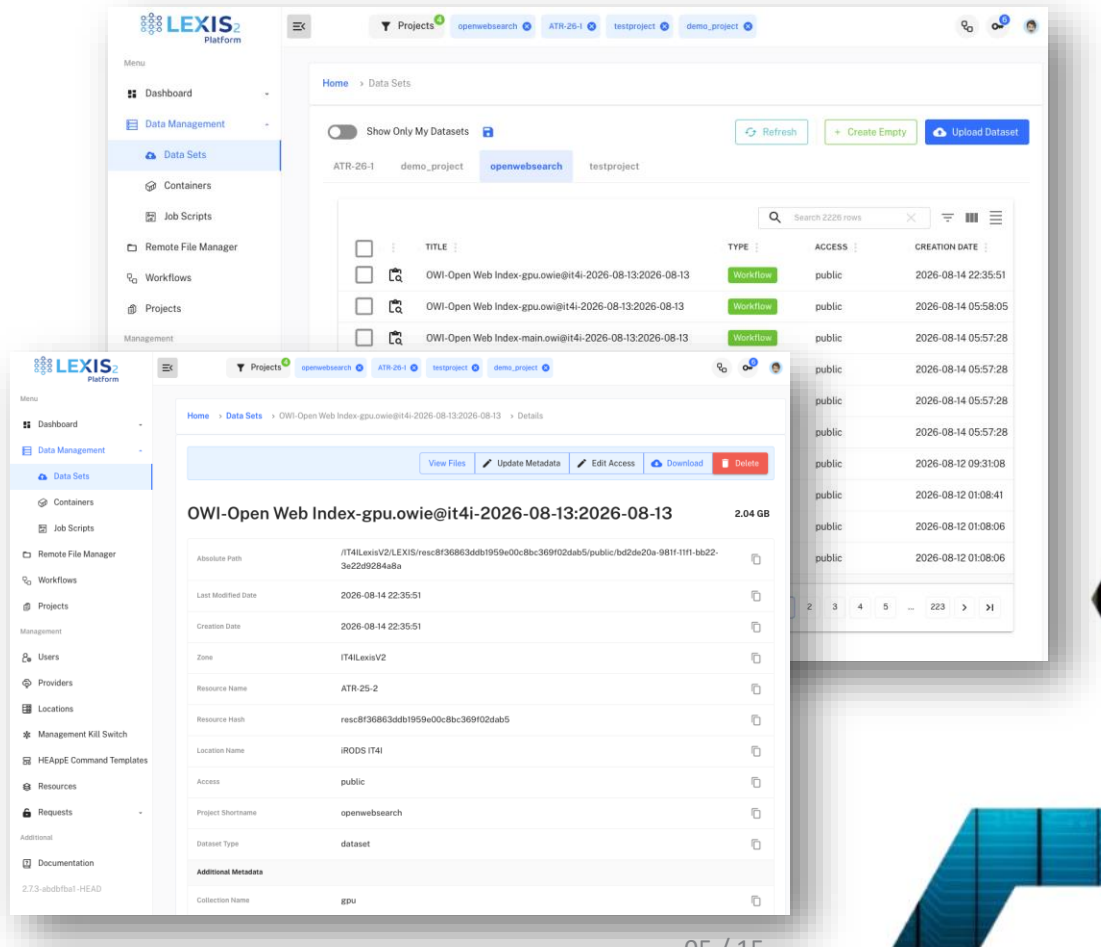
Portal, API or command line over HTTP/2

Data staging

Copies to and from temporary HPC scratch

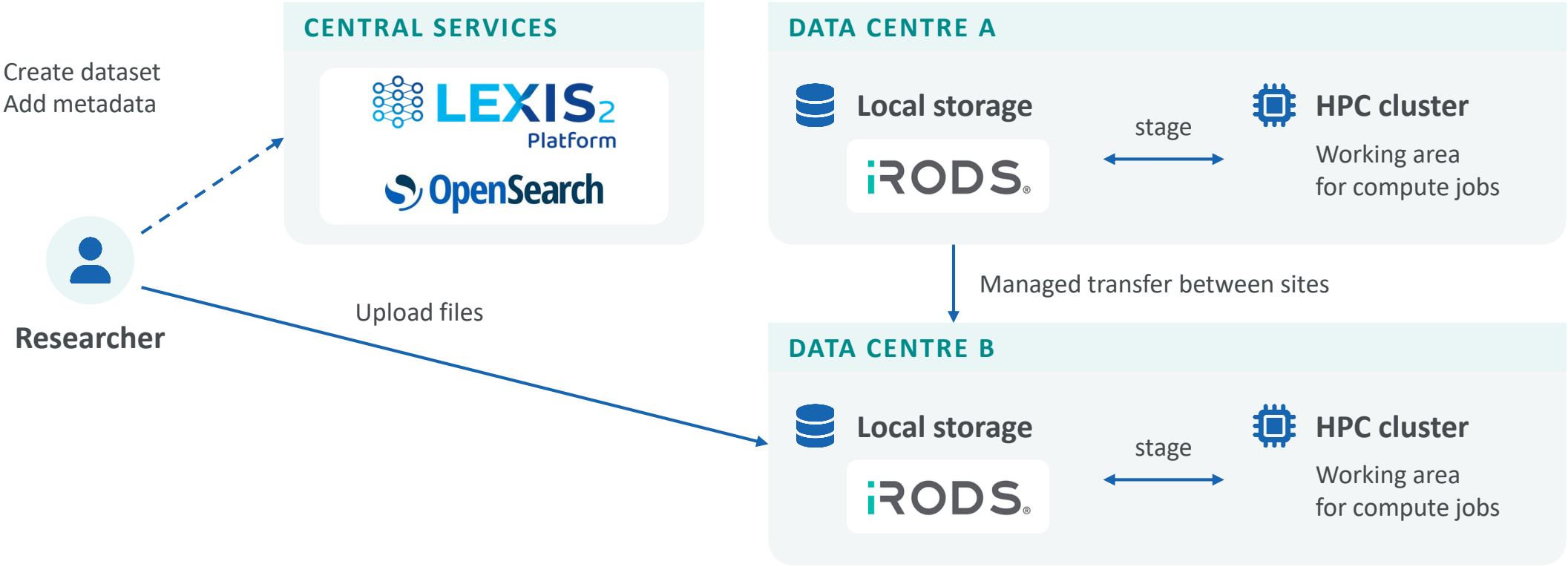
Locality and caching

Fewer repeated transfers between sites



Where the dataset lives and where it runs

LEXIS Platform prepares the data copy needed for a computation.



Staging copies data from managed storage to the working area used by a compute job.



Shared metadata, data under local control

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

LEXIS Platform separates dataset discovery and coordination from the systems that hold the files.

CENTRAL DISCOVERY AND COORDINATION

Metadata service

OpenSearch index

Staging requests

Queued data movement

Workflow service

Workflow execution

SITE A

iRODS storage zone

Worker stages to HPC scratch

SITE B

iRODS storage zone

Worker stages to cloud / HPC

SITE C

iRODS storage zone

Worker stages to AI / HPC

Site workers use the iRODS HTTP API. Providers retain control of their storage.



LEXIS2 Platform

iRODS: the data management layer

Logical collections, metadata, permissions and policies sit above provider storage.

iRODS®

Consistent access across storage systems

Provider storage keeps its local implementation.

LEXIS Platform clients	Portal, scripts and workflows
iRODS HTTP API	Authenticated service access
iRODS storage zone	Collections, catalogue and access rules
Storage resources	Filesystems and object stores



One dataset reference connects metadata and files

A LEXIS Platform identifier links a searchable description to files in provider-local iRODS storage.

DATASET IDENTIFIER — **UUID / logical dataset reference**

Searchable metadata

OpenSearch — DataCite 4.5

TITLE	Open Web Index – main
OWNER	OpenWebSearch.eu
ACCESS	Public / terms apply

Provider-local files

- **/openwebsearch/owi/**
  - metadata_000.parquet
  - index.ciff.gz

Illustrative dataset fields

Workflows use the dataset ID.
LEXIS Platform resolves locations and stages the required files.

LUMI AI Factory

Computing, data services and expertise help researchers use AI infrastructure.



COMPUTING

LUMI supercomputer and AI computing services

DATA

Dataset-as-a-Service prepares curated data close to computing

SUPPORT

Technical guidance, training and user support

What Dataset-as-a-Service provides

Researchers get clear descriptions, governed access and a usable copy, while providers retain control.

SERVICE NEED	VALUE FOR RESEARCHERS	SERVICE COMPONENT
Documented data	Understand relevance and limitations	Provider review and documentation
Findable metadata	Search by topic, terms and year	Fairdata Etsin and Metax
Governed access	Clear licences and approval conditions	REMS and support workflows
Data near computing	Reduce repeated large transfers	A use copy in LUMI-O
Managed movement	Stage data between systems	LEXIS Platform transfer and coordination

Initial DaaS service: some steps still require expert support



How the DaaS services work together

Each service has a distinct responsibility in making data usable on LUMI.

Data user

FAIRDATA
Etsin catalogue
Metax metadata

REMS
Access requests and approvals

LEXIS Platform
Data transfer and workflows

LUMI-O
Object storage near computing

LUMI
Training, inference and analysis

METADATA AND PERMISSIONS GUIDE ACCESS

LEXIS Platform moves the required data copy.



Use case: OpenWebSearch.eu

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Open WebSearch .eu

A distributed workflow produces the Open Web Index for search, analysis and AI research.

LEXIS Platform coordinates processing and data movement between sites

CRAWL
Collect web material
URLs / WARC

→

PROCESS
Extract text and metadata
Enrichment

→

INDEX
Topic and language partitions
Parquet + CIFF

→

STORE
Publish datasets through iRODS
Data centres

→

REUSE
Search, analysis and AI research
Daily releases

FEDERATED iRODS STORAGE

Site A Site B Site C



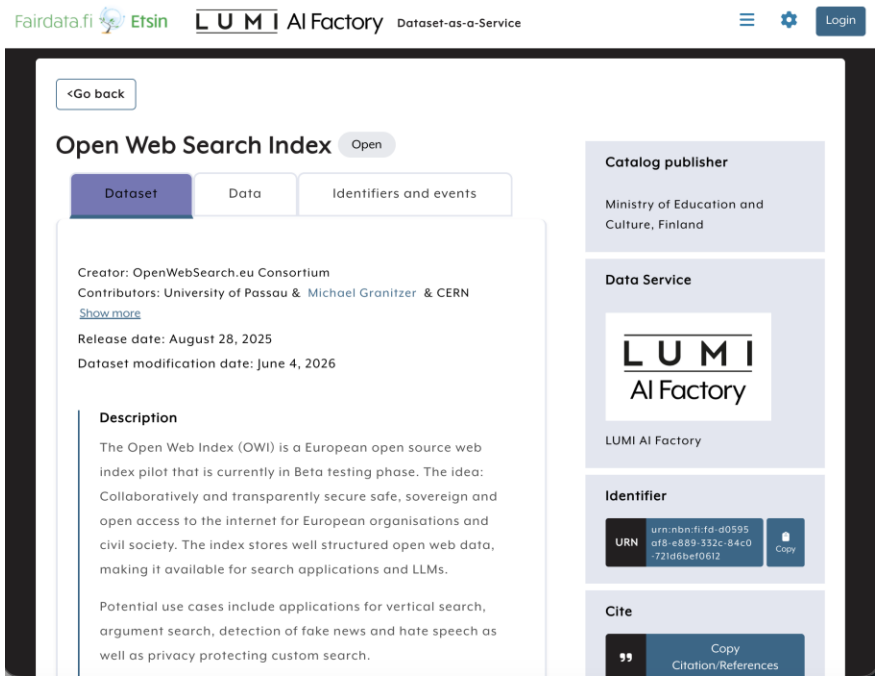
LEXIS2 Platform

The Open Web Index in LEXIS Platform and LUMI DaaS

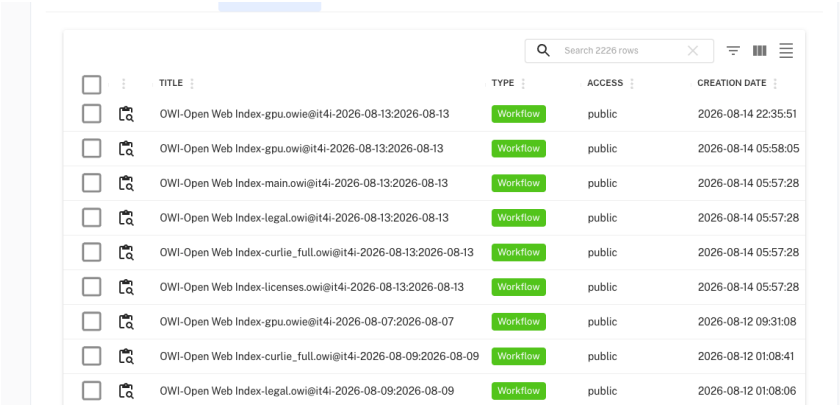
VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Catalogue discovery and operational data management support different parts of reuse.

Illustrative source captures



Discovery in LUMI AI Factory DaaS



Data management in LEXIS Platform

- **FINDABLE** — Searchable titles and metadata
- **ACCESS CONTEXT** — Recorded project and access context
- **STORAGE LOCATION** — Identifiable iRODS zone and resource

LEXIS2 Platform

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

EUDAT CONFERENCE 2026

Thank you

Questions welcome · Enjoy Prague!



Prague · Pierre-Selim Huard · CC BY 4.0

