



Co-funded by the Horizon 2020
Framework Programme of the European Union
Grant Agreement Number 825532

Large-scale EXecution for Industry & Society



  www.lexis-project.eu

FEDERATED RESEARCH DATA MANAGEMENT IN LEXIS

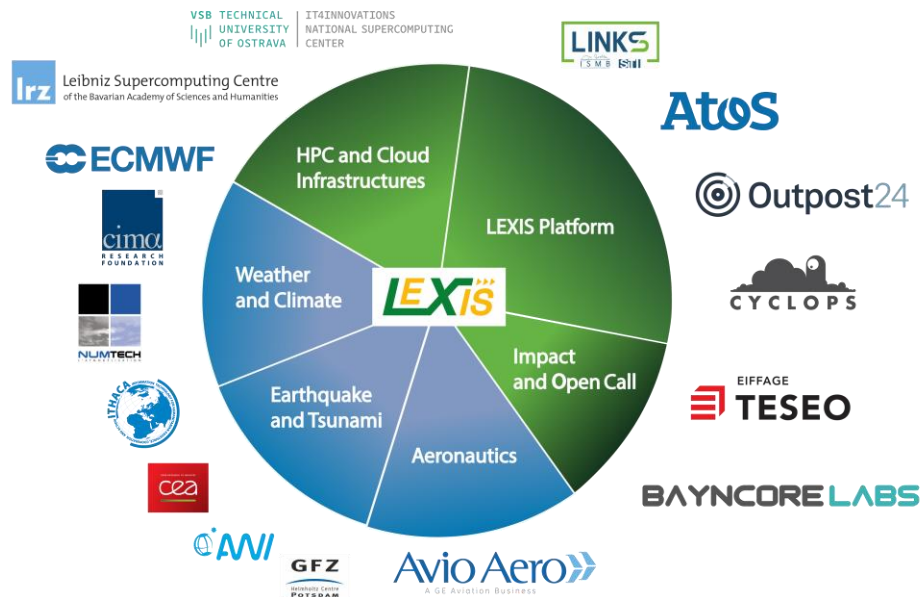
EGI 2020 Conference

Nov 2nd, 2020

Data Management Solutions-Part 1

JOHANNES MUNKE, MOHAMAD HAYEK
(LRZ)

WITH LEXIS DATA SYSTEM TEAM AND
CONSORTIUM



This infrastructure is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825532.



CONTEXT: QUICK OVERVIEW OF LEXIS



Co-funded by the Horizon 2020
Framework Programme of the European Union
Grant Agreement Number 825532

Large-scale EXecution for Industry & Society

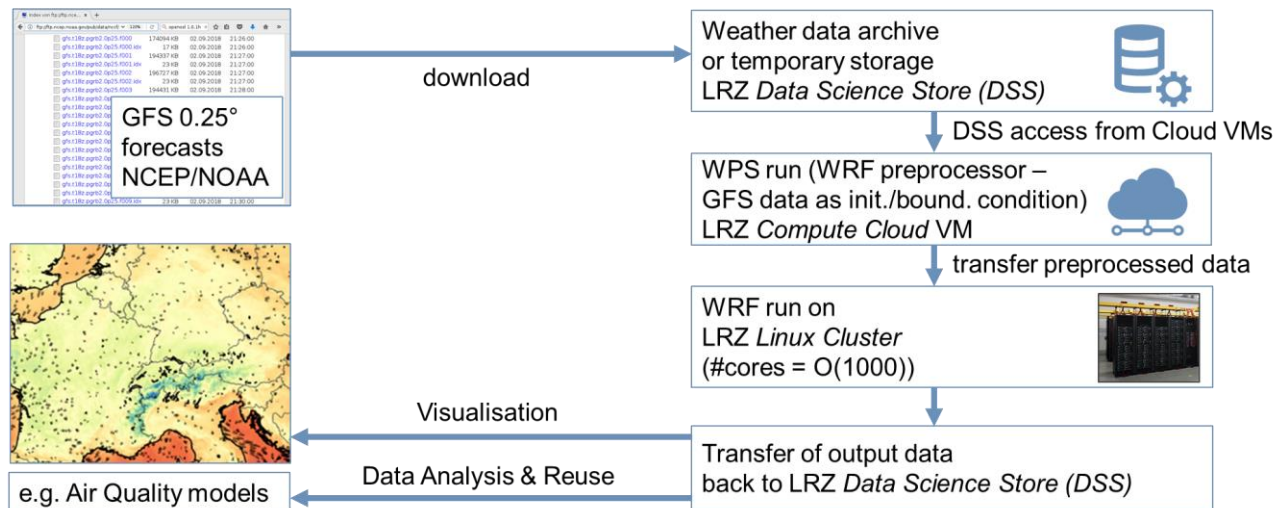


  www.lexis-project.eu

Topic:	HPC and Big Data enabled Large-scale Test-beds and Applications
Topic identifier:	ICT-11-2018-2019
Type	LEXIS project, led by the IT4Innovations (CZ) Supercomputing Centre: (...) is building an advanced engineering platform at the confluence of HPC, Cloud and Big Data, which leverages large-scale geographically-distributed resources from existing HPC infrastructure, employs Big Data analytics solutions and augments them with Cloud services.
Scope	Driven by the requirements of several pilot testcases, the LEXIS platform relies on best-in-class data management solutions (EUDAT) and advanced, distributed orchestration solutions (TOSCA), augmenting them with new, efficient hardware and platform capabilities (e.g. in the form of Data Nodes and federation, usage monitoring and accounting/billing support). Thus realize, LEXIS realises an innovative solution (...)
Project	
Budget	
EC Contribution	
Partners	
Project	

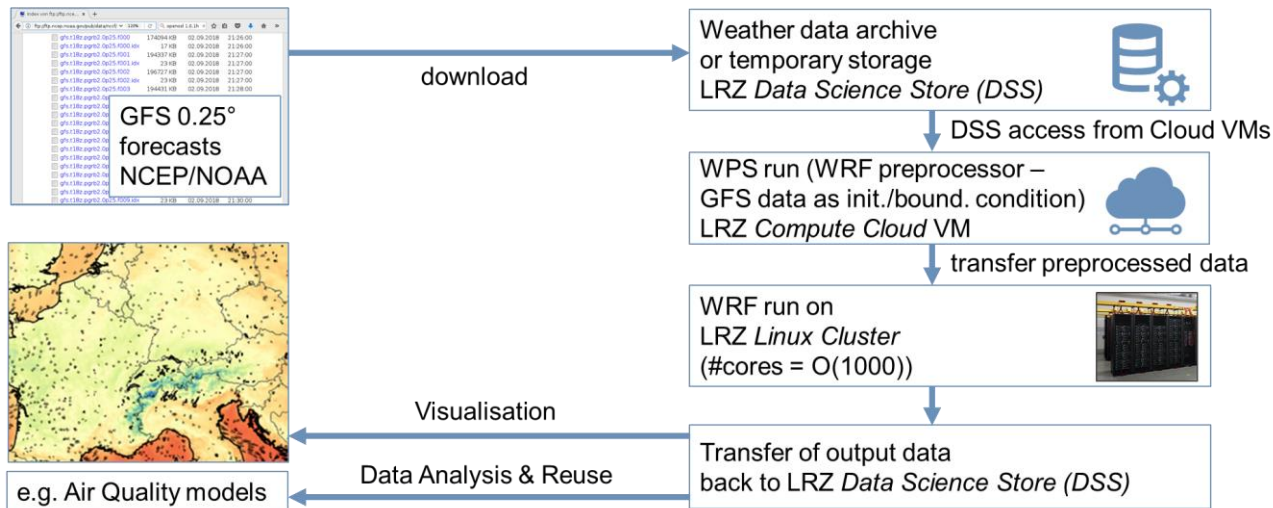
LEXIS: THE PROBLEM ADDRESSED

- Typical High-Performance-Computing (HPC) / Big Data workflow: Weather prediction at LRZ



LEXIS: THE PROBLEM ADDRESSED

- Typical High-Performance-Computing (HPC) / Big Data workflow: Weather prediction at LRZ



- **Complicated!**
 - multiple systems & data transfers to be set up
 - closed system, no flexible usage of European resources
- **Not often leveraged by Industry, SMEs, unexperienced scientists**

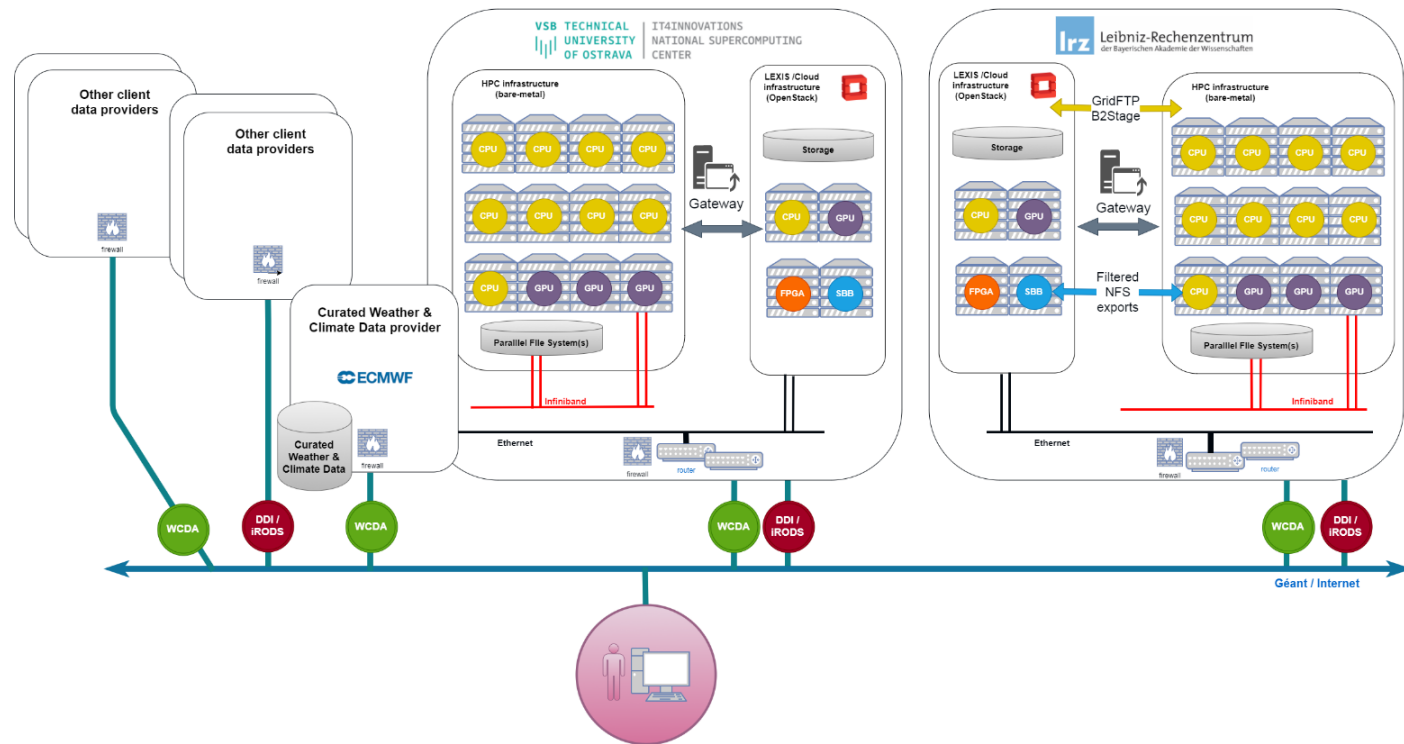
LEXIS H2020 PROJECT OBJECTIVES

European Computing & Big Data Platform for Industry, Society, Science

- **Access to the HPC/Cloud/Big Data resources for SMEs/Industry** (incl. Billing model)
 - **Web Portal** (easy entry point, unified AAI)
 - **Orchestration** for easy, automatic and optimized workflows
 - **Cloud and HPC resources at different European Computing Sites**
 - convenient **user interface** to **define workflows**
 - boosted by **acceleration techniques** (Compute: GPU, FPGA; Data: Burst Buffers)
- Details in the talk of **Alberto Scionti et al.**, 13:15 today, session „**Workflow Management Solutions**“
- **Integration with European Infrastructure**
 - **Demand for a strong Data Management backend** – LEXIS Work Package 3
 - **unified data access** from LEXIS Cloud and HPC resources across Europe
 - **cross-site “Big Data” and metadata management**
 - Data Discovery and FAIR (“Findable, Accessible, Interoperable, Reusable”) Data
- **This is the scope of this talk**

LEXIS PLATFORM

LEXIS Federated data Infrastructure



LEXIS PILOT PROJECTS

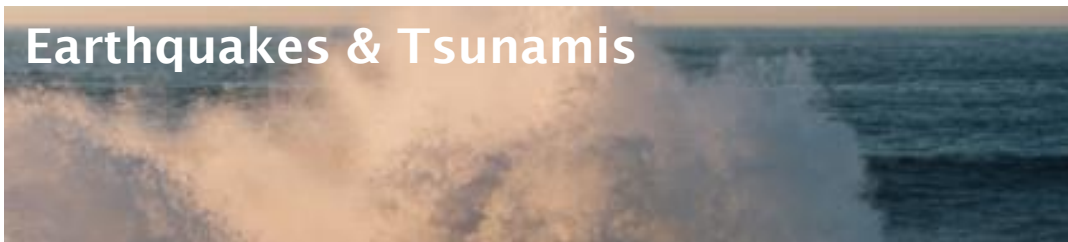
General information - <https://lexis-project.eu/web/>

Aeronautics



Computation Fluid Dynamics (CFD),
Rotating parts (gearboxes),
3D Visualization

Earthquakes & Tsunamis



Earthquakes & Tsunami prediction
models, geographic and urban
databases, emergency organization,
urgent computing

Weather & Climate

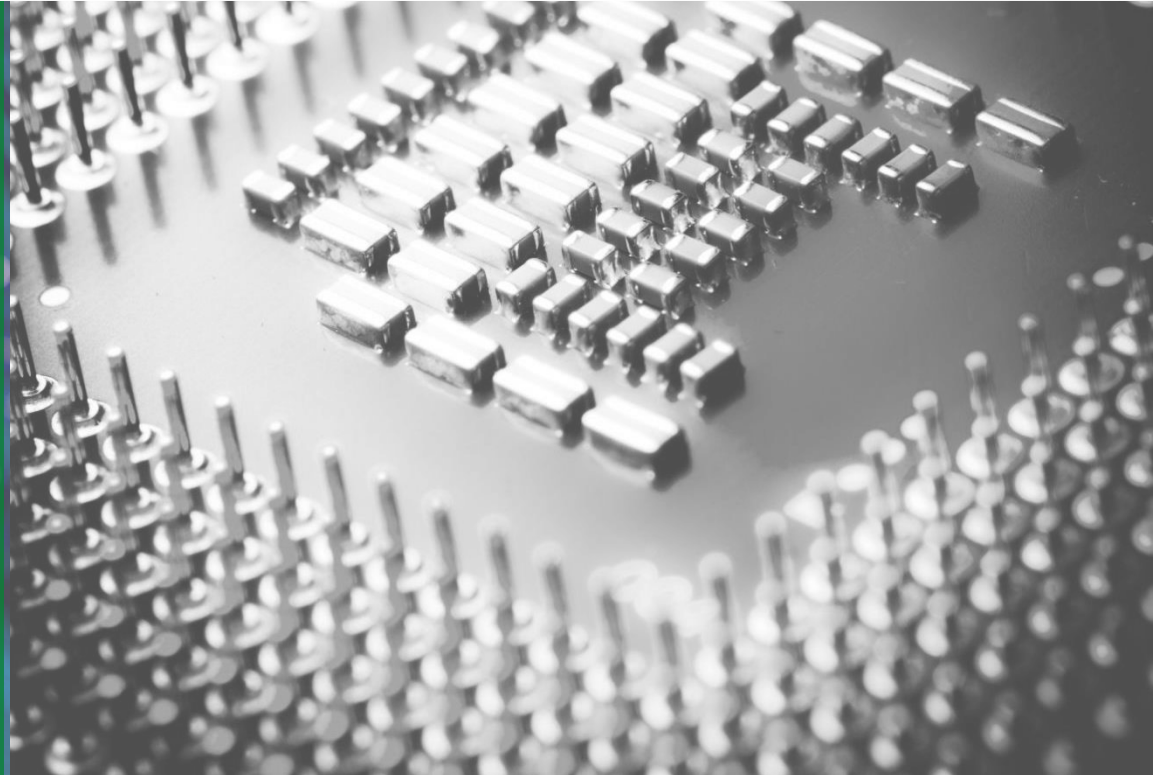


Weather & Climate models (WRF)
and various post-processors for
flood, wildfire & agriculture
applications



LEXIS DATA SYSTEM

(WORK PACKAGE 3)



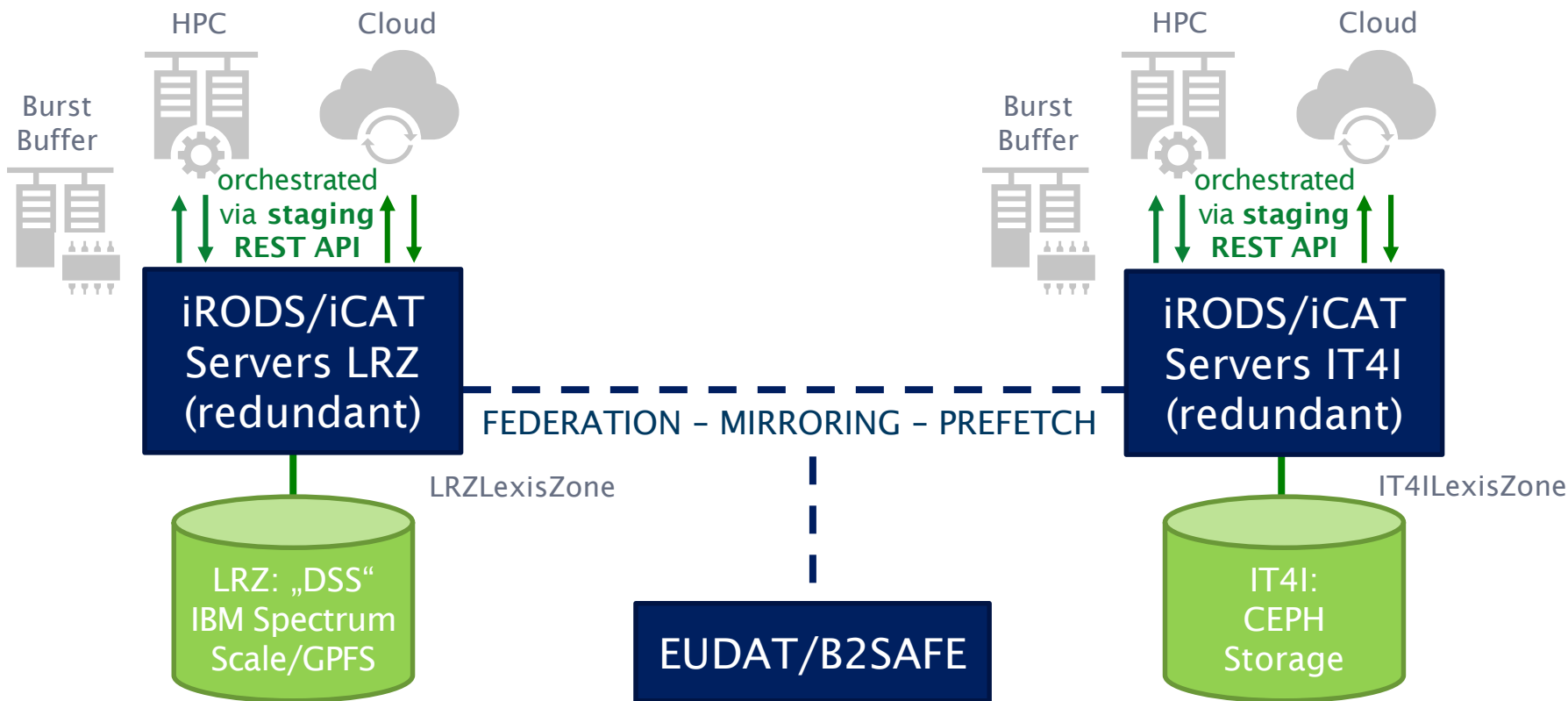
LEXIS DDI - DISTRIBUTED DATA INFRASTRUCTURE

Key Points

- Idea: use an **European system for LEXIS data**
→ **EUDAT/iRODS** (Integrated Rule-Oriented Data System middleware)
- **Unified view on data** in LEXIS – federated **iRODS** zones (IT4I, LRZ, ...)
 - “**filesystem-like**”; top-level directories e.g. `/IT4ILexisZone`
 - **transparent access to all files via all iRODS servers**
- Physical file storage policies implemented as **iRODS rules**, e.g.
 - **cross-site mirroring**
 - **low-level storage tiering** (in each computing/data centre)
- “**Non-invasive**” **data-curation** approach
 - **DataCite-like** metadata stored in iRODS, **EUDAT-B2HANDLE PIDs**
 - directory/access-rights structure fixed on project (top-)level
- Uses **LEXIS cross-provider AAI**

LEXIS DDI

Federation leveraging iRODS & EUDAT B2SAFE (and B2HANDLE, B2STAGE)



DATA PRESERVATION IN LEXIS: B2SAFE

B2SAFE is the data preservation module in EUDAT

- The B2SAFE service runs on top of LEXIS iRODS zone to:
 - provide data safety and resiliency in case of an issue at one of the federated centres, by implementing mirroring policies;
 - optimize the workflows execution by allowing the Orchestrator to get the data from the zone close to the computing resources; and to
 - trigger PID registration.

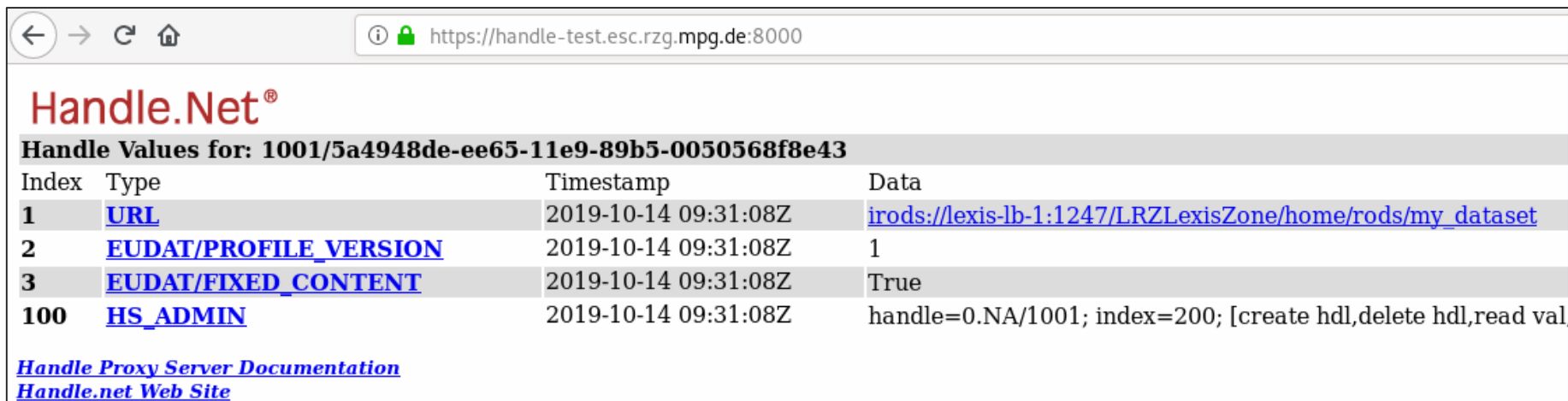


EUDAT Collaborative Data Infrastructure, "B2SAFE - EUDAT," EUDAT Ltd, 2020. [Online]. Available: <https://www.eudat.eu/services/b2safe>.

PERSISTENT IDENTIFIERS IN LEXIS (PIDs): B2HANDLE

B2HANDLE is the PID module in EUDAT

- The LEXIS DDI makes use of B2HANDLE PIDs to guarantee data access via long-lasting unique identifiers, in order to to
 - facilitate publication of data; and
 - facilitate data retrieval in the long term.



The screenshot shows a web browser window with the address bar displaying `https://handle-test.esc.rzg.mpg.de:8000`. The page title is "Handle.Net®". Below the title, it says "Handle Values for: 1001/5a4948de-ee65-11e9-89b5-0050568f8e43". The main content is a table with four columns: Index, Type, Timestamp, and Data.

Index	Type	Timestamp	Data
1	URL	2019-10-14 09:31:08Z	irods://lexis-lb-1:1247/LRZLexisZone/home/rods/my_dataset
2	EUDAT/PROFILE_VERSION	2019-10-14 09:31:08Z	1
3	EUDAT/FIXED_CONTENT	2019-10-14 09:31:08Z	True
100	HS_ADMIN	2019-10-14 09:31:08Z	handle=0.NA/1001; index=200; [create hdl,delete hdl,read val

Below the table, there are two links: [Handle Proxy Server Documentation](#) and [Handle.net Web Site](#).

THE LEXIS DDI PROMOTES FAIR DATA MANAGEMENT

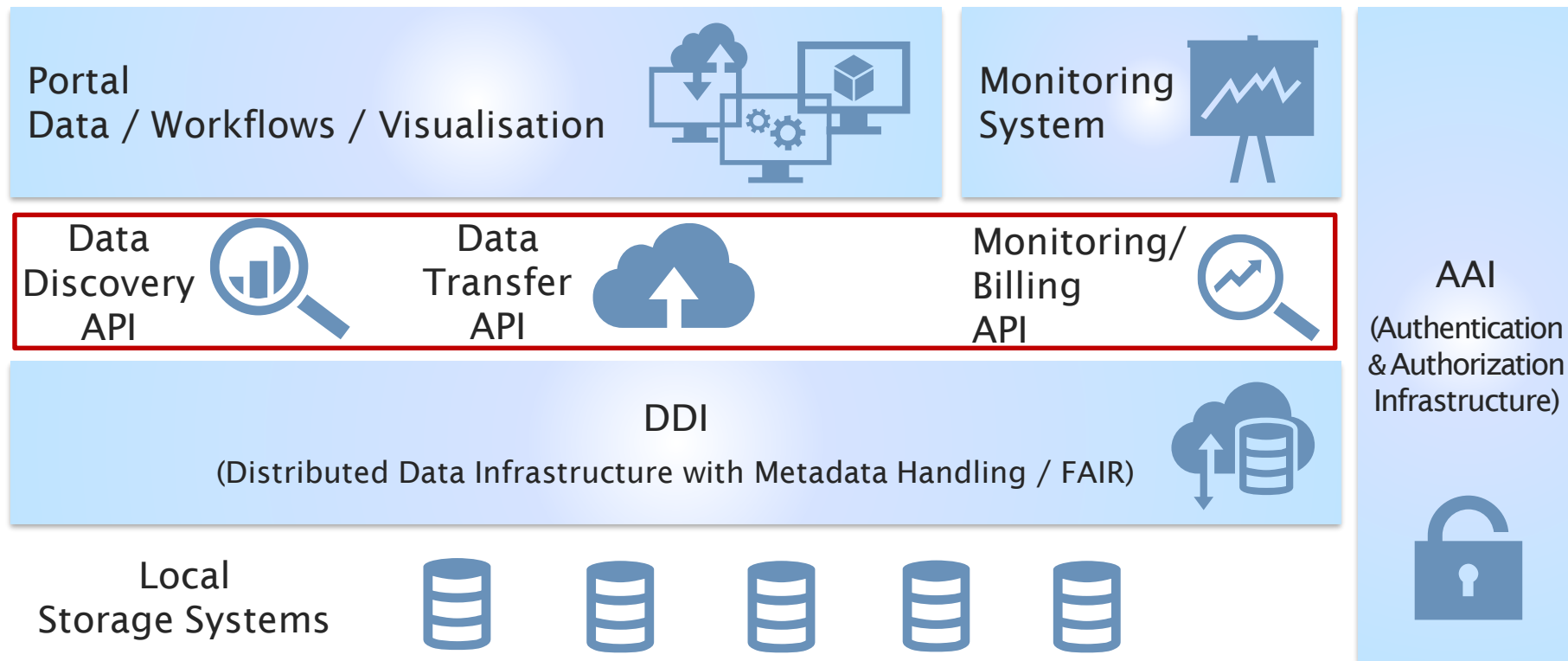
Findable, Accessible, Interoperable, Reusable Research Data

- Two of the most basic FAIR requirements:
 - metadata
 - a (world-)unique identifier for the data
- Metadata in LEXIS:
 - stored in iRODS Attribute-Value(-Unit) store for each data set
 - schema oriented at the basics from DataCite (schema.datacite.org)

```
@lexis-lb-1:~$ ils
/LRZLexisZone/home/rods/my_dataset:
@lexis-lb-1:~$ iput opensearch.txt
@lexis-lb-1:~$ ils
/LRZLexisZone/home/rods/my_dataset:
  opensearch.txt
@lexis-lb-1:~$ irule -F eudatPidsColl.r
*newPID = 1001/5a4948de-ee65-11e9-89b5-0050568f8e43
@lexis-lb-1:~$ imeta ls -C /LRZLexisZone/home/rods/my_dataset
AVUs defined for collection /LRZLexisZone/home/rods/my_dataset:
attribute: EUDAT/FIXED_CONTENT
value: True
units:
----
attribute: PID
value: 1001/5a4948de-ee65-11e9-89b5-0050568f8e43
units:
```

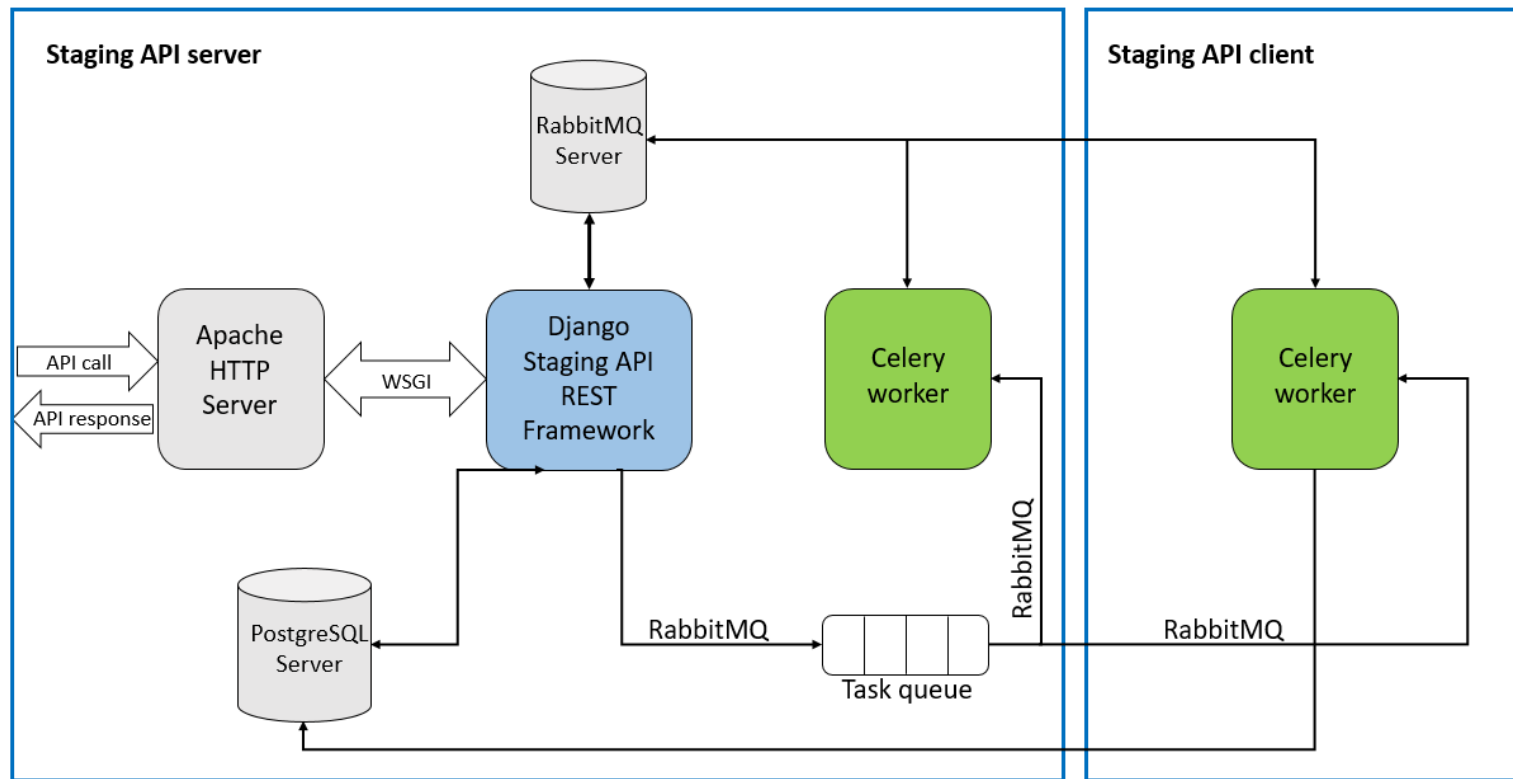

THE DDI IN THE LEXIS ECOSYSTEM

Communication via APIs



LEXIS DATA TRANSFER/STAGING API (*continued*)

Architecture



DATA MANAGEMENT VIA LEXIS PORTAL

Work Package 8, Task 8.1 – “Work in Progress”

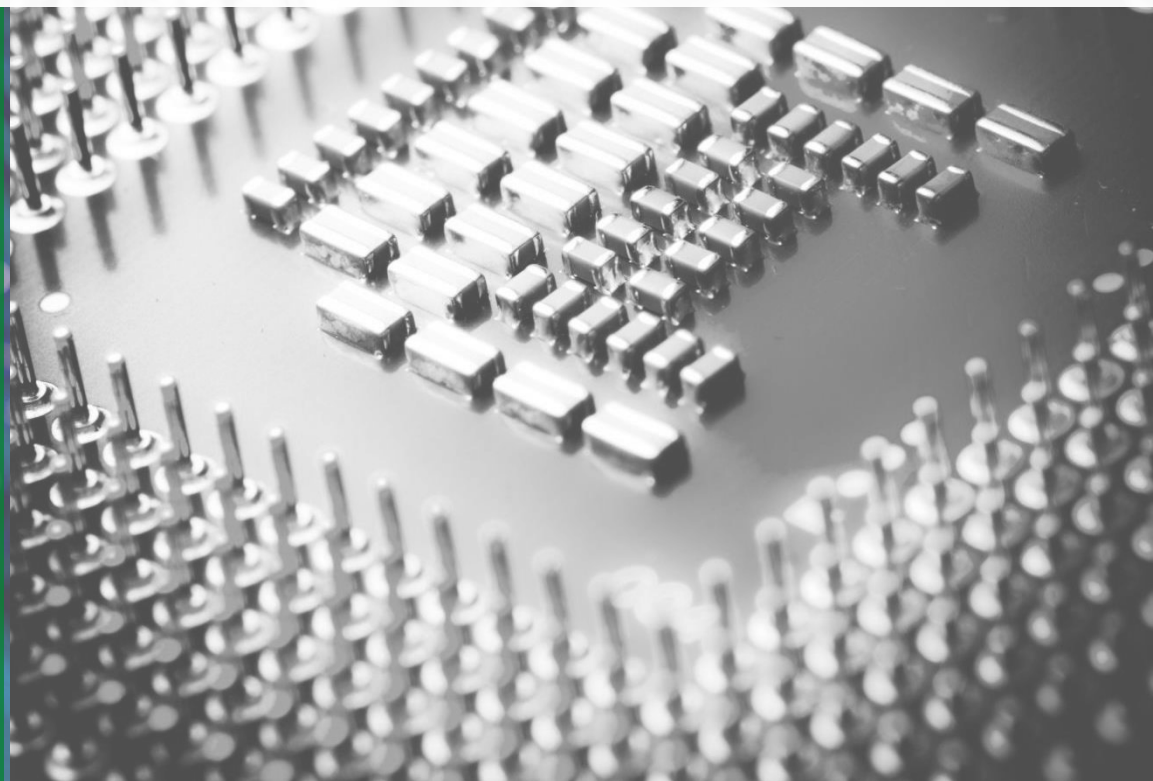
- Login using LEXIS AAI (Keycloak-based)
- User accesses DDI data with rights encoded in token:
LEXIS DDI validates token using strongly customized iRODS-OpenID Plugin.
- Next months:
from working demonstrator to
convenient everyday data management!

The screenshot shows a web form titled "Upload new dataset". It is divided into three main sections: "Dataset", "Metadata", and "Resource Type".

- Dataset section:** Contains instructions for uploading a new dataset. It offers two options: "Unencrypted Zip with dataset (decompressed server side)" and "Single file upload (use this also if you want your (possibly encrypted) zip file to be stored as-is)". Both options have a "Durchsuchen..." button and the text "Keine Datei ausgewählt."
- Metadata section:** Contains several input fields for dataset information:
 - Project:** A text input field with a dropdown arrow on the right.
 - Access:** A text input field with a dropdown arrow on the right.
 - Creator:** A text input field with an "Add" button next to it.
 - Title:** A text input field with the placeholder text "Please enter Title".
 - Publisher:** A text input field with an "Add" button next to it.
 - Publication Year:** A text input field with the placeholder text "Please enter Publication Year".
- Resource Type:** A text input field at the bottom of the form.



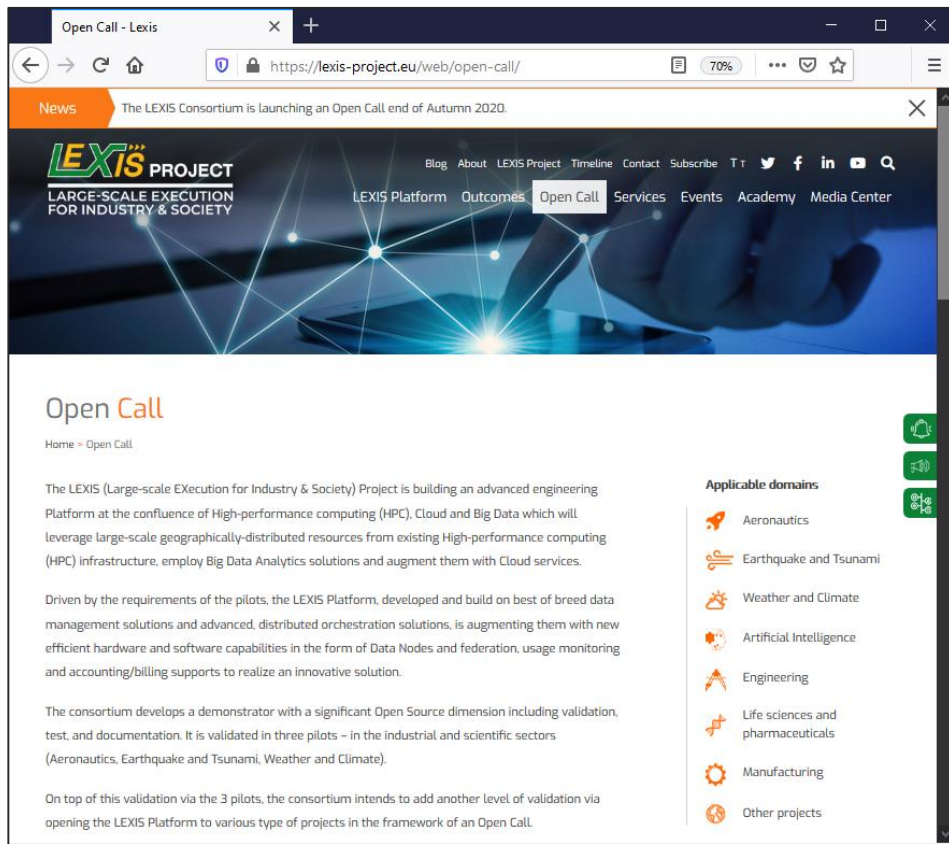
**LEXIS OPEN CALL,
EUROPEAN
COLLABORATION &
OUTLOOK**



LEXIS – OPEN CALL: OBJECTIVES

- Idea: additional validation of platform via various “external” use cases
- Selected Open Call projects have the opportunity to
 - access HPC/Cloud workflow orchestration & data platform with large resources, and
 - get technical support and training (individual "project manager" from LEXIS team).
- These projects also include other EU projects (e.g. CompBioMed2) → collaboration
- **Current stage:** Further applicants are submitting their projects and the Open Call Board will select the best ones
 - larger focus on Industry / Private sector (but not exclusive)
 - collaboration & integration with European & third party infrastructures (Computing/Data)

LEXIS – OPEN CALL ON PROJECT WEBSITE



Open Call - Lexis

News The LEXIS Consortium is launching an Open Call end of Autumn 2020.

LEXIS PROJECT
LARGE-SCALE EXECUTION FOR INDUSTRY & SOCIETY

Blog About LEXIS Project Timeline Contact Subscribe T T [Twitter](#) [Facebook](#) [LinkedIn](#) [YouTube](#) [Search](#)

LEXIS Platform Outcomes **Open Call** Services Events Academy Media Center

Open Call

Home » Open Call

The LEXIS (Large-scale EXecution for Industry & Society) Project is building an advanced engineering Platform at the confluence of High-performance computing (HPC), Cloud and Big Data which will leverage large-scale geographically-distributed resources from existing High-performance computing (HPC) infrastructure, employ Big Data Analytics solutions and augment them with Cloud services.

Driven by the requirements of the pilots, the LEXIS Platform, developed and build on best of breed data management solutions and advanced, distributed orchestration solutions, is augmenting them with new efficient hardware and software capabilities in the form of Data Nodes and federation, usage monitoring and accounting/billing supports to realize an innovative solution.

The consortium develops a demonstrator with a significant Open Source dimension including validation, test, and documentation. It is validated in three pilots – in the industrial and scientific sectors (Aeronautics, Earthquake and Tsunami, Weather and Climate).

On top of this validation via the 3 pilots, the consortium intends to add another level of validation via opening the LEXIS Platform to various type of projects in the framework of an Open Call.

Applicable domains

- Aeronautics
- Earthquake and Tsunami
- Weather and Climate
- Artificial Intelligence
- Engineering
- Life sciences and pharmaceuticals
- Manufacturing
- Other projects

For further information and registration visit:

lexis-project.eu/web/open-call/

CONCLUSIONS AND OUTLOOK

- **LEXIS European Cloud-HPC Workflow Platform** (H2020) works with a **Distributed Data Infrastructure** based on **iRODS/EUDAT-B2SAFE**
- Philosophy: **Easy usability** via Portal, **clean architecture** (REST APIs...)
- **Data Federation** of IT4I (CZ) and LRZ (DE) to be extended to **further participants**
- Platform to be further **immersed with EUDAT and beyond** (ideas: BDVA, EGI federated cloud, EOSC, ...)
- We are aiming at (and **open** for) **collaborations!**
 - with centres joining the LEXIS platform
 - with EU projects and initiatives

... in the frame of the **LEXIS Open Call** and in **general**.
- Focus in 2021: benchmarking, optimisation & sustainability of the platform (operation & funding model, etc.)

CONTACT

Johannes Munke / Mohamad Hayek
LEXIS Task 2.3 / 3.3 leads
munke@lrz.de / hayek@lrz.de

Jan Martinovič
LEXIS Project Coordinator
jan.martinovic@vsb.cz

Olivier Terzo
LEXIS Co-Design Manager
olivier.terzo@linksfoundation.com

Large-scale EXecution
for Industry & Society

LEXIS

Thanks for your attention!

CONSORTIUM

VSb TECHNICAL
UNIVERSITY
OF OSTRAVA

IT4INNOVATIONS
NATIONAL SUPERCOMPUTING
CENTER

