



Piano Nazionale di Ripresa e Resilienza

WP3 third meeting:

OFED & pipelines update & next roadmap

Federica Bazzocchi
RIT-Area Science Park
Trieste, 28.05.2024



AGENDA

○ WP3 state of art:

□ OFED:

- Breif recap
- Services in production
- Services in test/development

□ FAIR-by-design pipelines:

- Some numbers and remarks
- What has been done so far
- What has to be done yet

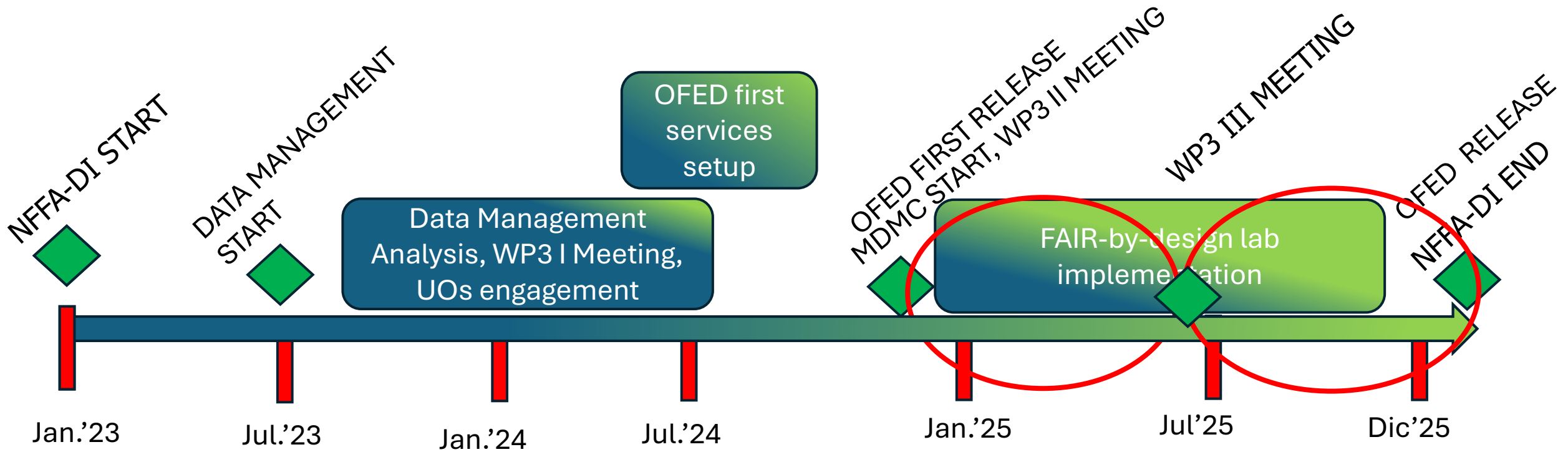
□ Possible Strategies for Next Months:

- Look at your «similar»»
- II MDMC edition
- New tool



Foto di [Brian Penny](#) da [Pixabay](#)

Timeline



The background of the slide is a blue-tinted collage of various electronic components and tools. On the right side, a large, detailed image of a microscope is visible. The rest of the background is filled with smaller, semi-transparent images of electronic parts like capacitors, resistors, and cables.

OFED: Overarching Fair Ecosystem for Data

... more than a datalake

Overarching

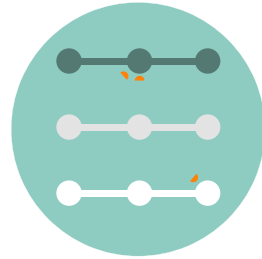
Fair

Ecosystem

Data



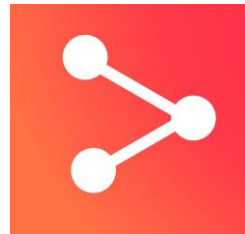
access



plan



**store &
collect**



share



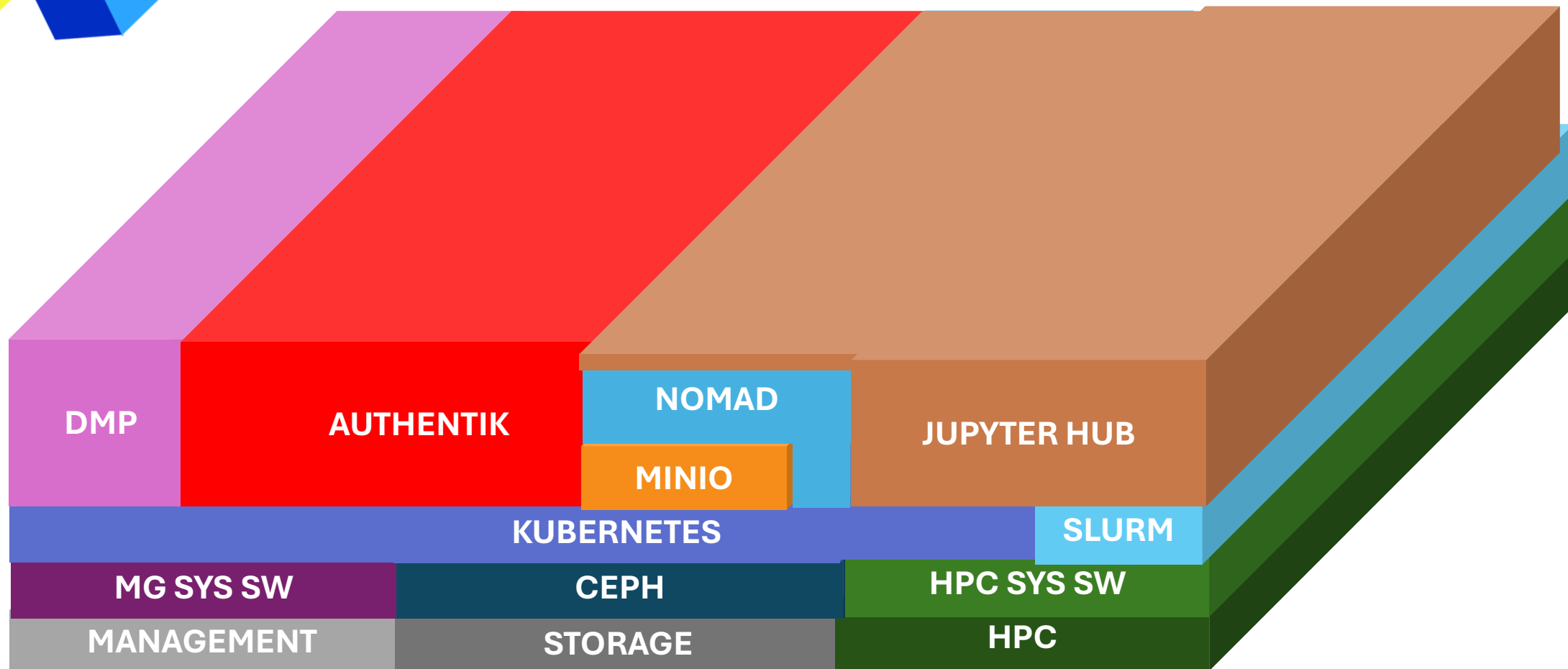
analyze



- Tools and interoperable services
- Find and access metadata engine
- Innovative software through virtual access



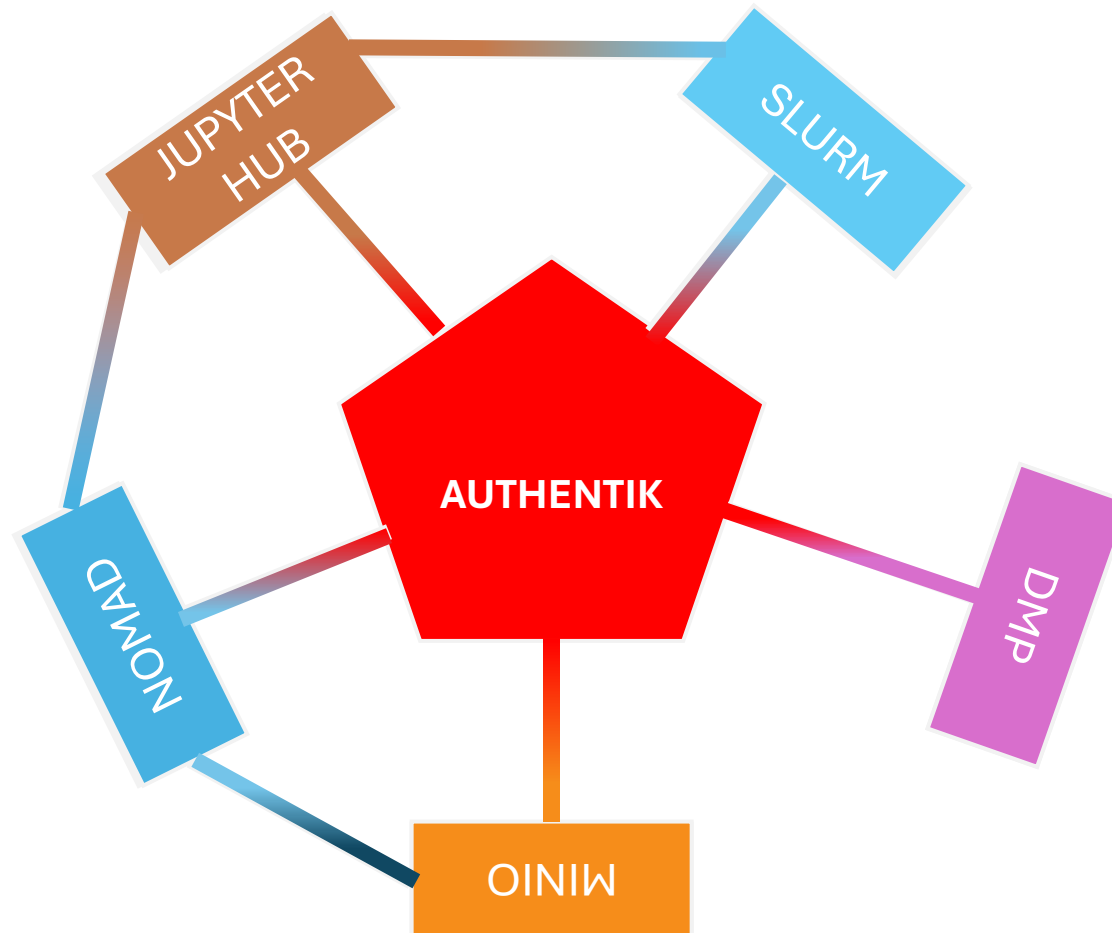
OFED Specific SW Services





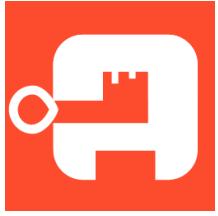
OFED Specific SW Services

Authentik is a CORE SW of OFED Structure because AUTHENTIATION is required by each service



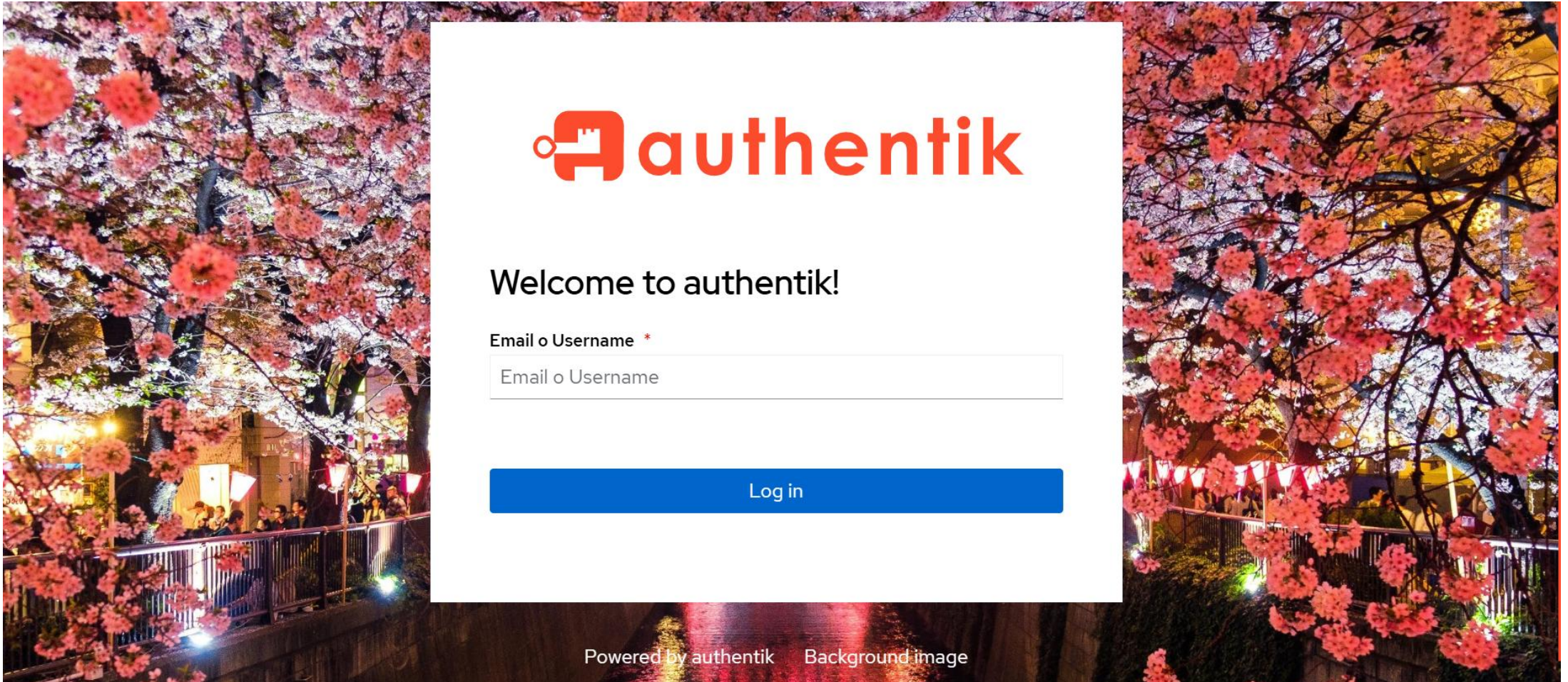
Data Policy

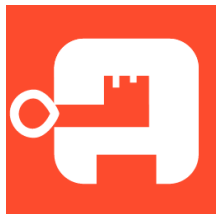
- **Introduction of the FAIR Data Manager** (not a fixed-term revenue)
- **Defined standards** (NeXus or file compatible with OFED)
- **Defined fixed acronym for the UO**
- **Defined buckets naming convention both for users and research-in-house** (*nffa-di_[proposal_id] / nffa-di_[UO]_[project_id]*)
- **Defined file naming convention** (*nffa-di_[proposal_id]_[UO]_[UO_internal_id] / nffa-di_[UO]_[project_id]_key*)



Authentik

access





Authentik

access



 **authentik**

 0



fbazzocchi



My applications

Search...

E

easyDMP

M

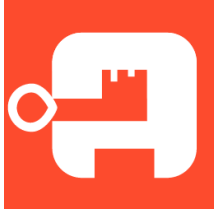
MinIO

N

netbox-area

O

Ollama API



Authentik

access



- **At the moment it requires a VPN**
- **write an email to support@areasciencepark.it to activate a VPN** (with in cc Gianfranco.Gallizia@areasciencepark.it)
- **write an email to Gianfranco.gallizia@areasciencepark.it or federica.bazzocchi@areasciencepark.it to create an account** (soon the registration page will be active)



MinIO

store & collect



High-Performance Object Store

MinIO is a cloud-native object store built to run on any infrastructure – public, private or edge clouds. Primary use cases include data lakes, databases, AI/ML, SaaS applications and fast backup & recovery. MinIO is dual licensed under GNU AGPL v3 and commercial license. To learn more, visit www.min.io.

MINIO
OBJECT STORE
 **LICENSE**

Login with SSO (Authentik)

Other Authentication Methods



[Documentation](#) | [GitHub](#) | [Support](#) | [Download](#)



MinIO

store & collect



MINIO

OBJECT STORE

AGPL LICENSE

User

Object Browser

Access Keys

Documentation

Administrator

Buckets

Policies

Identity

Monitoring

Object Browser



Filter Buckets

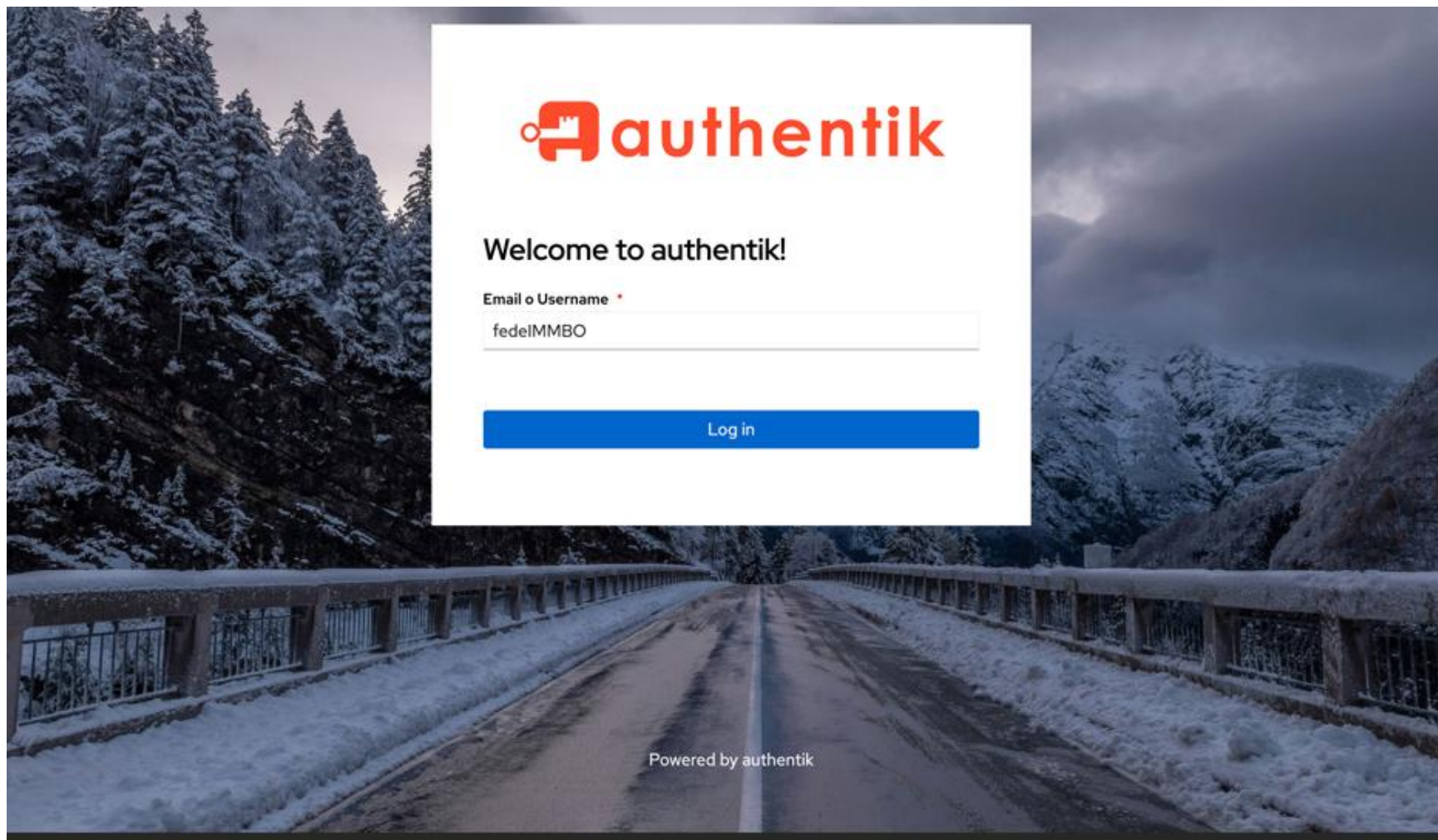
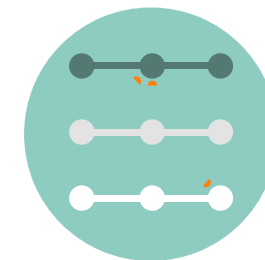


Name	Objects	Size	Access
nffa-di-276	0	0.0 B	R/W
nffa-di-277	0	0.0 B	R/W
nffa-di-279	0	0.0 B	R/W
nffa-di-280	0	0.0 B	R/W
nffa-di-283	0	0.0 B	R/W
nffa-di-287	0	0.0 B	R/W
nffa-di-288	0	0.0 B	R/W
...	-	- - -	- - -



easyDMP

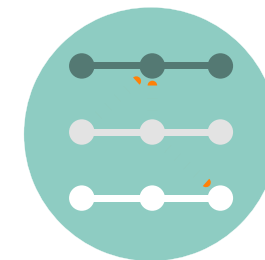
plan





easyDMP

plan



NFFA-DI EASY DMP

Your UO is Consiglio Nazionale delle Ricerche -Istituto per lo studio dei materiali nanostrutturati- Bologna

Experimental technique

- This field is required.

Select your Technique:

Select

Instrument

Select

instrument

Save

NFFA-DI EASY DMP

Your UO is Consiglio Nazionale delle Ricerche -Istituto per lo studio dei materiali nanostrutturati- Bologna

Experimental technique

Select your Technique:

Select

Instrument

Select your Instrument:

Select

I Administrative Information

1. Installation Information

A) Institution Name

Consiglio Nazionale delle Ricerche -Istituto per lo studio dei materiali nanostrutturati- Bologna

B) Institution Site

Bologna

C) Facility Type

Laboratory source

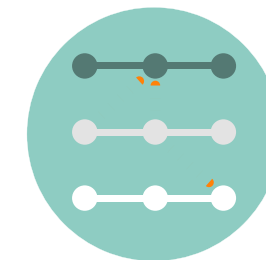
D) Kind of Installation

Growth & Synthesis



easyDMP

plan



Yes. Any scientific output will be uploaded on a trusted open access repository such Zenodo.

D) Will scientific publication or any other scientific document or output (presentations, posters, softwares) be licensed?

Yes. They will be licenced with the most suitable open licence.

V Data and Metadata storage preservation

1. Storage

A) Where will raw data be stored ?

Raw data will be stored both locally and on the infrastructure datalake named OFED.

B) How long will raw data be preserved ?

On OFED raw data will be preserved for 3 years. After this period if raw data have not be uploaded on NOMAD or or

2. Backups

A) Which backup policy will be applied to published data or raw data related to published data?

Raw data related to published datasets as well as raw data published will undergo the same backup policy of the digi

B) Which backup policy will be applied to not published data or raw data not related to published data?

Not published datasets will be saved on a distributed file system equipped with erasure code 8+2 the failure domain is the Datacenter. There is no backup for this data but different types of redundancy will be set up based on the size of the data: for datasets size lower than 20 GB there will be redundancy between the OUs and the Data Lake while for those larger than 20 GB there will be no redundancy unless different policies defined on a case-by-case basis via Lab-DMP.

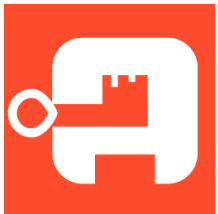
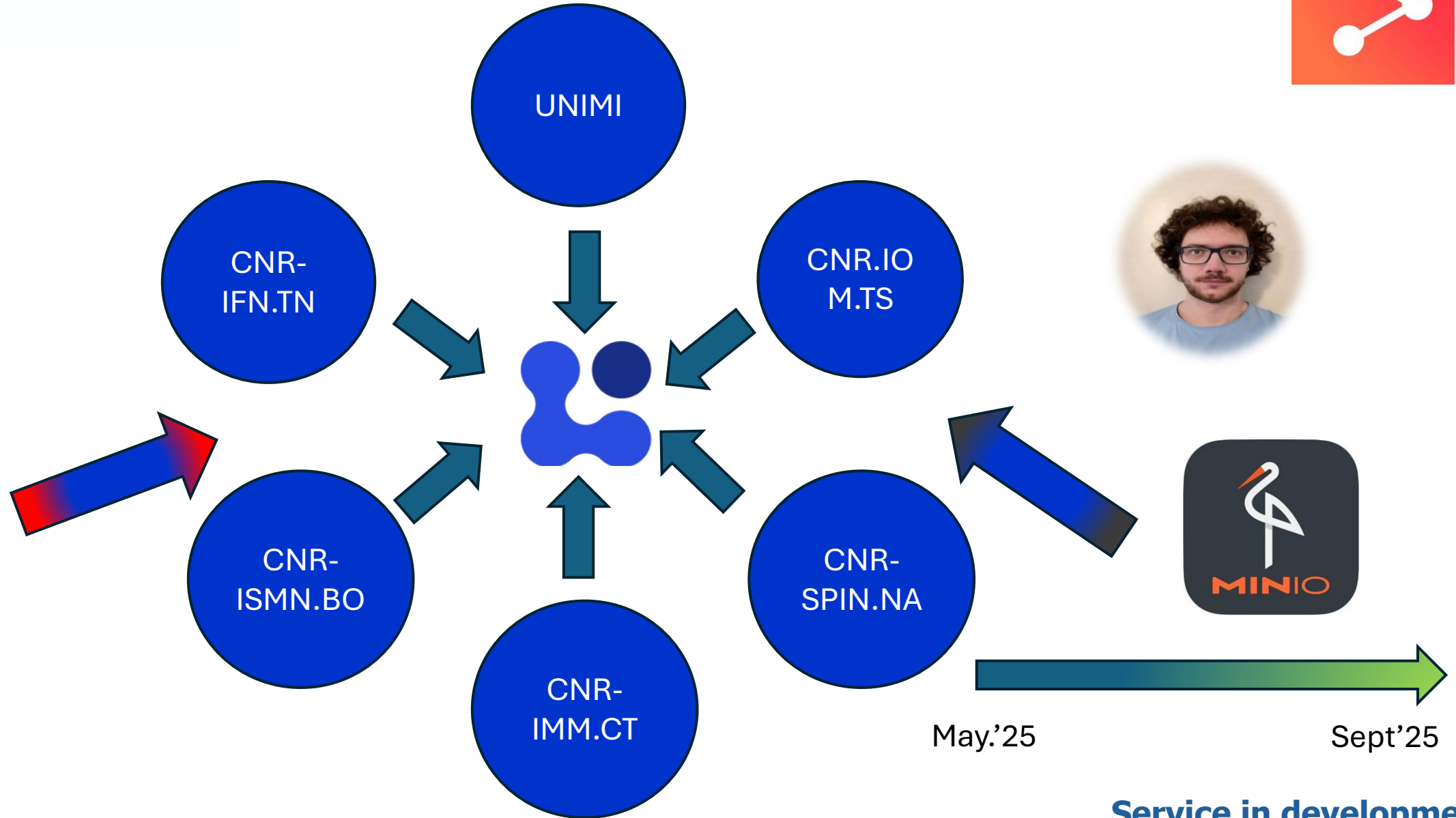
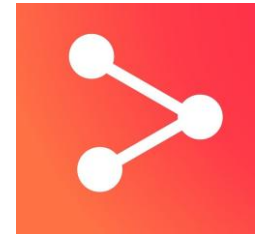
Save

Your DMP has been created!

You may proceed to download it.

[download](#)

Logout





analyze



July'25

Dec'25

FAIR-by-design pipelines

Some numbers regarding NFFA-DI offer

UO	# Techniques	# Instruments
UNIMI-Fisica	15	20
CNR-IOM@TS	14	22
POLIMI-POLIFAB	14	17
CNR-IMM@CT	11	15
CNR-ISMN@BO	8	18
CNR-NANOTEC@LE	8	9
CNR-SPIN@NA	7	8
CNR-IFN@TN	5	6
CNR-ISM@RM	2	2
AREA-RIT	2	2
CNR-IFN@MI	1	5

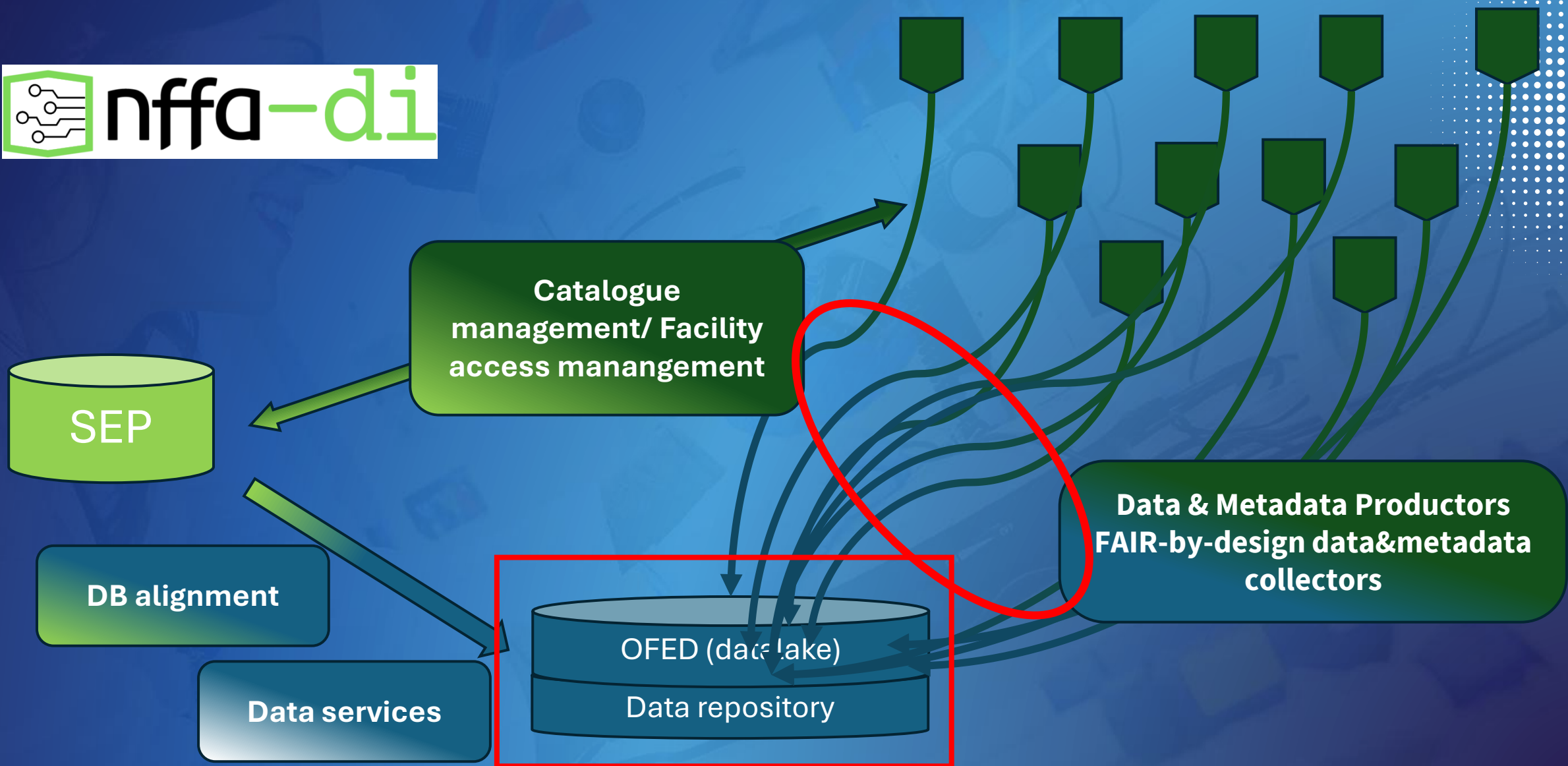
124!!

I suppose you know ALL the LABS need their FAIR-BY-DESIGN pipelines towards OFED as described in the project

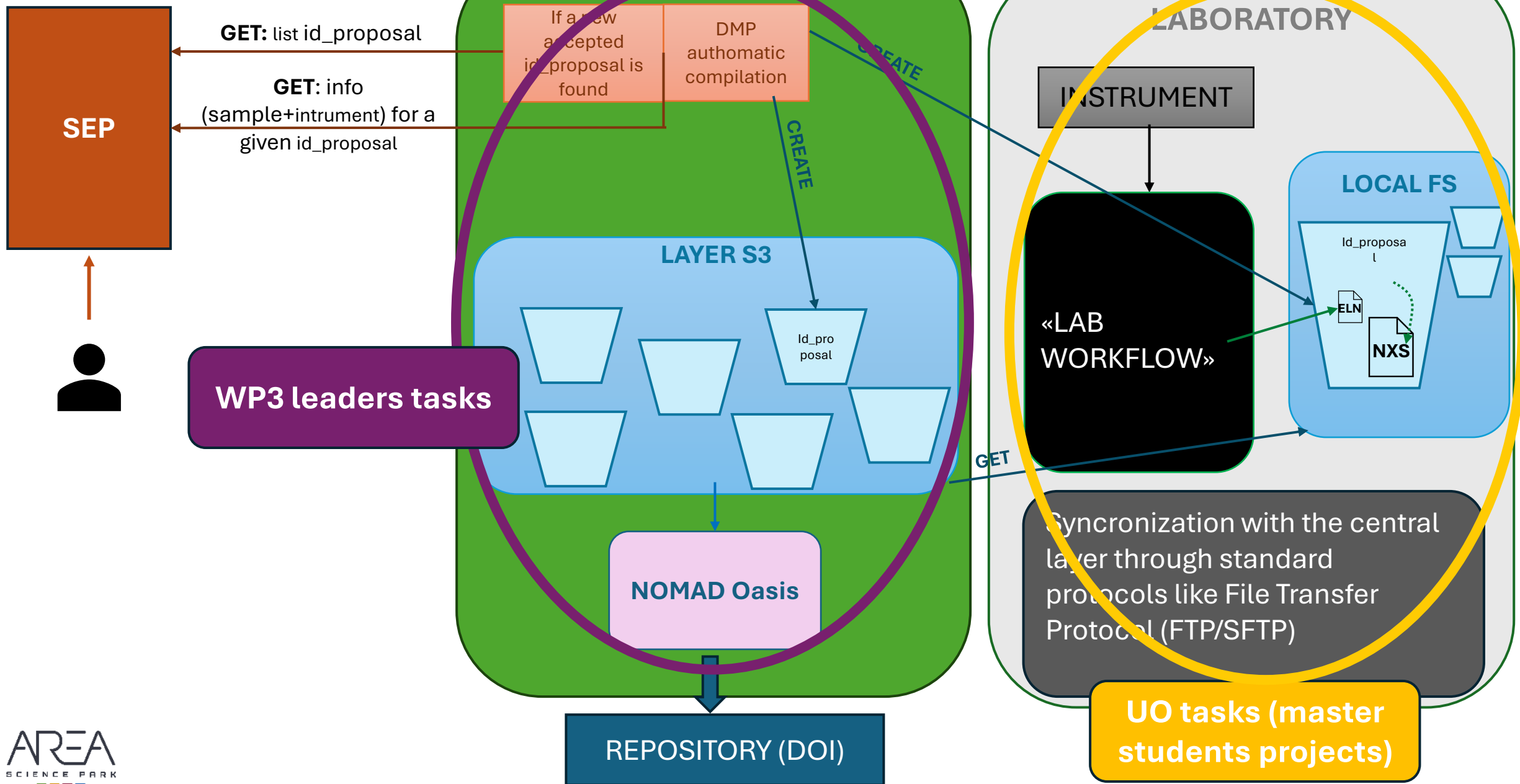
Most popular techniques

Technique	# UO
Scanning Electron Microscopy	5
Chemical Vapour Deposition	4
Photo-Electron Spectroscopy	4
Focused Ion Beam	3
Pulsed Laser Deposition	3
PVD sputtering	3
Scanning Tunneling Microscopy	3
Ultrafast-Spectroscopy	3
X-Ray Diffraction	3
Angle Resolved Photoelectron Spectroscopy	2
Atomic Force Microscopy	2
Atomic Layer Deposition	2
Dicing-saw	2
Electron Beam Lithography	2
FIB-Scanning Electron Microscopy	2
Magneto-Optic Kerr Effect	2
Molecular Beam Epitaxy	2
Optical-Lithography	2
Reactive Ion Etching	2
Transmission Electron Microscopy	2

To remember ... elements of the Research Digital Infrastructure



How we left in october



WHAT WE HAVE SEEN IN THE LAST TWO DAYS

Maybe one of the largest
number of FAIR-by-design
(prototype) pipelines in
experimental labs belonging
to the same project ...

I think the largest so far in ITALY



- ❑ **Collaboration among:**
 - CNR-IFN.TN&CNR-ISMN.BO
 - CNR-IFN.MI&CNR-ISM.RM
- ❑ **Dialogue with FAIRMAT (UNIMI, CNR-IMM.CT, CNR-IOM.TS)**
- ❑ **Reuse of the tool developed (CNR-NANOTEC.LE vs AREA, CNR-SPIN.NA vs CNR-IFN.TN&CNR-ISMN.BO)**
- ❑ **The EM NFFA-DI community agreed on NXem**



- ❑ **Same technique approached in **different ways** (simulation and synthesis)**
- ❑ **No complete semantic harmonization**
- ❑ **Many **different approaches** to the sample**
- ❑ **If it is difficult finding an agreement between a subset of the community, **how** we may thing to reach **standards**? (this is a wider discussion)**

Now....



Where we are now ...

UO	STATUS (DMP/ PIPELINE)						TOTAL
	YES&DONE	YES&ALMOST DONE	YES&IN PROGRESS	NO&ALMOST DONE	YES&NOT DONE	NO&NOT DONE	
CNR-IOM@TS		21				1	22
UNIMI-Fisica			4		2	14	20
CNR-ISMN@BO			18				18
POLIMI-POLIFAB			1			16	17
CNR-IMM@CT		1	1			13	15
CNR-NANOTEC@LE			1			8	9
CNR-SPIN@NA		1		1	1	5	8
CNR-IFN@TN			6				6
CNR-IFN@MI		3			2		5
CNR-ISM@RM	2						2
AREA-RIT			2				2
TOTAL	2	26	33	1	5	57	124



Possible Strategies

I. MDMC NEW EDITION 2025-2026

Shaping Data, Shaping the Future!

The MDMC is a full-time program, which consists of 1500 hours of training activities, corresponding to 60 ECTS credits:

Program Component	Hours	ECTS Credits
Classroom lectures	250	10
Internship	500	20
Final Project (Thesis)	125	5
Individual Study	625	25
Total	1500	60

- **Total Duration:** 10 months, from September 2025 to June 2026
- **Teaching Language:** English
- **Teaching Mode:** In-person, with some remote sessions during the internship period
- **Location of in-presence lectures:** Trieste – Area Science Park & SISSA
- **Internship:** 6 months of hands-on experience in research laboratories and data-intensive environments

Full tuition fee € 4.500,00
Scholarships available!!!



II. LOOK FOR YOUR SIMILAR!



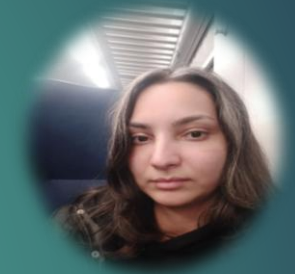
TECHNIQUE	AREA-RIT	CNR-IFN.MI	CNR-IFN.TN	CNR-IMM.CT	CNR-IOM.TS	CNR-ISM.RM	CNR-ISMN.BO	CNR-NANOTEC.LE	CNR-SPIN.NA	POLIFAB	UNIMI	TOTAL
Chemical Vapour Deposition			1	2	2		6					11
Ultrafast-Spectroscopy		5			3	1						9
Photo-electron Spectroscopy					3				1	1	1	6
Reactive Ion Etching			1				5					6
Scanning Electron Microscopy			1	1				1		1	1	5
Molecular Beam Epitaxy					3					1		4
Pulsed Laser Deposition					1				2	1		4
PVD sputtering				1			2		1			4
Scanning Tunneling Microscopy					2				1		1	4
Focused Ion Beam			1	1				1				3
Atomic Force Microscopy										1	2	3
X-Ray Diffraction				1				1	1			3
Transmission Electron Microscopy	1			1								2
Atomic Layer Deposition								1		1		2
Dicing-saw							1			1		2
Electron Beam Lithography			1							1		2
Optical-Lithography							1			1		2
RAMAN											2	2
X-Ray Reflectivity				1				1				2
Angle-resolved photoelectron spectroscopy					1					1		2
FIB - Scanning Electron Microscopy	1						1					2
Magneto-Optic Kerr Effect					1					1		2
Structural and ground-state electronic properties											2	2

**LOOK FOR
YOUR
SIMILAR!**

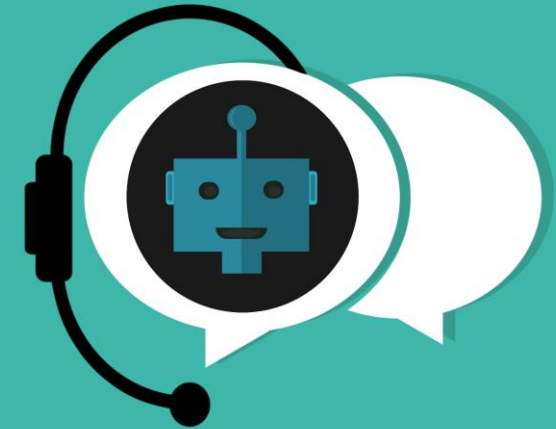
- ☐ **All the codes developed are uploaded on **github**** (https://github.com/Master-Data-Management-and-Curation/Thesis_Code.git)
- ☐ **All the thesis will be at disposal**
- ☐ **UO network have been fruitful during the master so let's learn from the youngest!**

III.

New tool : Data Curator Assistant



- Creating an AI agent that could be deployed as additional OFED service to help FAIRify data
- Bottom up approach from the techniques better implemented so far
- Do not substitute a FAIR-by-design pipeline



The background is a blue-toned collage of various electronic components and tools. On the right side, there is a large, semi-transparent image of a microscope. Scattered throughout the background are various electronic parts, including resistors, capacitors, integrated circuits, and wires. The overall aesthetic is technical and scientific.

Deploy
~~Submit~~ and ... do not relax yet

Final Remarks

- **NFFA-DI** is still far from reach the goal
- **OFED** and the **UO** have now a solid base
- Many **deadlines** are coming (services, deliverables,..) so a deep analysis of the resources at disposal is needed by all of us

