WP 5.2 – Open Science and (FAIR) Data

EURO-LABS Second Annual Meeting
10/10/2023

A. Lemasson (GANIL), M. Jouvin (IJC Lab),
C. Hornung (GSI), A. Matta (LPC-Caen)

WP5.2 : Open Science and Data

- Task 1 : Data Management Plan (A. Lemasson -- GANIL)

- Task 2 : openNP Research Data catalog (A. Matta -- LPC Caen)

- Task 3 : New Services for improved access to data (and shared services)

  - 3.1 : Authentification and Authorization Infrastructure (M. Jouvin -- IJC Lab)

  - 3.2 : Prototype of data access plateform (C. Hornung -- GSI)

WP5.2 : Open Science and Data

- **Task 1 : Data Management Plan (A. Lemasson -- GANIL)**

- Task 2 : openNP Research Data catalog (A. Matta -- LPC Caen)

- Task 3 : New Services for improved access to data (and shared services)

  - 3.1 : Authentification and Authorization Infrastructure (M. Jouvin -- IJC Lab)

 - 3.2 : Prototype of data access plateform (C. Hornung -- GSI)

Data Management Plan

Deliverable D5.7
Released on zenodo on Feb 27th 2023 :
https://doi.org/10.5281/zenodo.8160229

1) Definition of data produced within the WP of the EURO-LABS project and their management

2) Encourage FAIR (Findable, Accessible, Interoperable Reusable) data pinciple by depositing EURO-LABS productions (reports, slides, codes, …) on zenodo in the EURO-LABS community page to obtain persistent identifier : https://zenodo.org/communities/euro-labs

3) Data from Research Infrstructures do not fall under the EURO-LABS DMP

DOI 10.5281/zenodo.8160229

Data Management Plan
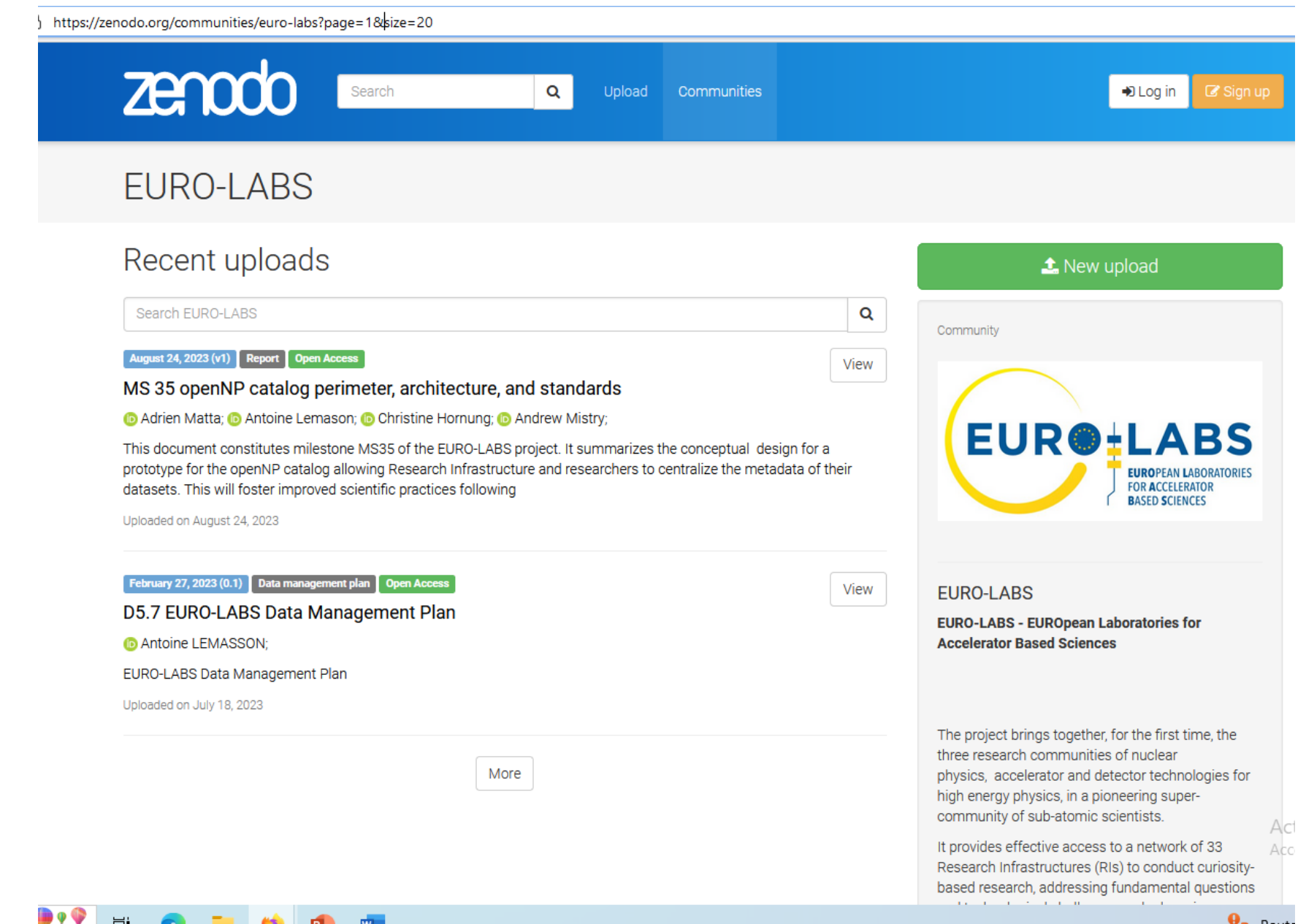
2.2 DATA FROM RESEARCH INFRASTRUCTURES

Transnational Access service program within the EURO-LABS project does not produce directly   data, only through the user teams accessing the participating RIs. The beneficiaries of EURO-LABS **will assure that the teams timely publish any scientific results in conference proceedings and scientific journals as appropriate, making them available in open-access archives.** All work done and presented will acknowledge the project. For a broader impact, the EURO-LABS website **will contain and   display all scientific events and publications, including those reaching out to society and industry.**

Datasets collected at research infrastructures with the TA and VA support of Work packages 2, 3 and 4 **will be managed based on the policy of the facilities**. Research Infrastructure receiving the support from EURO -LABS are expected to provide a Data Management Plan for the data produced in their facility and follow the FAIR principle.

Facilities could request guidance from the WP 5.2 to establish their data management plan

=> promoting FAIR data principles

WP5.2 : Open Science and Data

- Task 1 : Data Management Plan (A. Lemasson -- GANIL)

- Task 2 : openNP Research Data catalog (A. Matta -- LPC Caen)

- Task 3 : New Services for improved access to data (and shared services)

  - 3.1 : Authentification and Authorization Infrastructure (M. Jouvin -- IJC Lab)

  - 3.2 : Prototype of data access plateform (C. Hornung -- GSI)
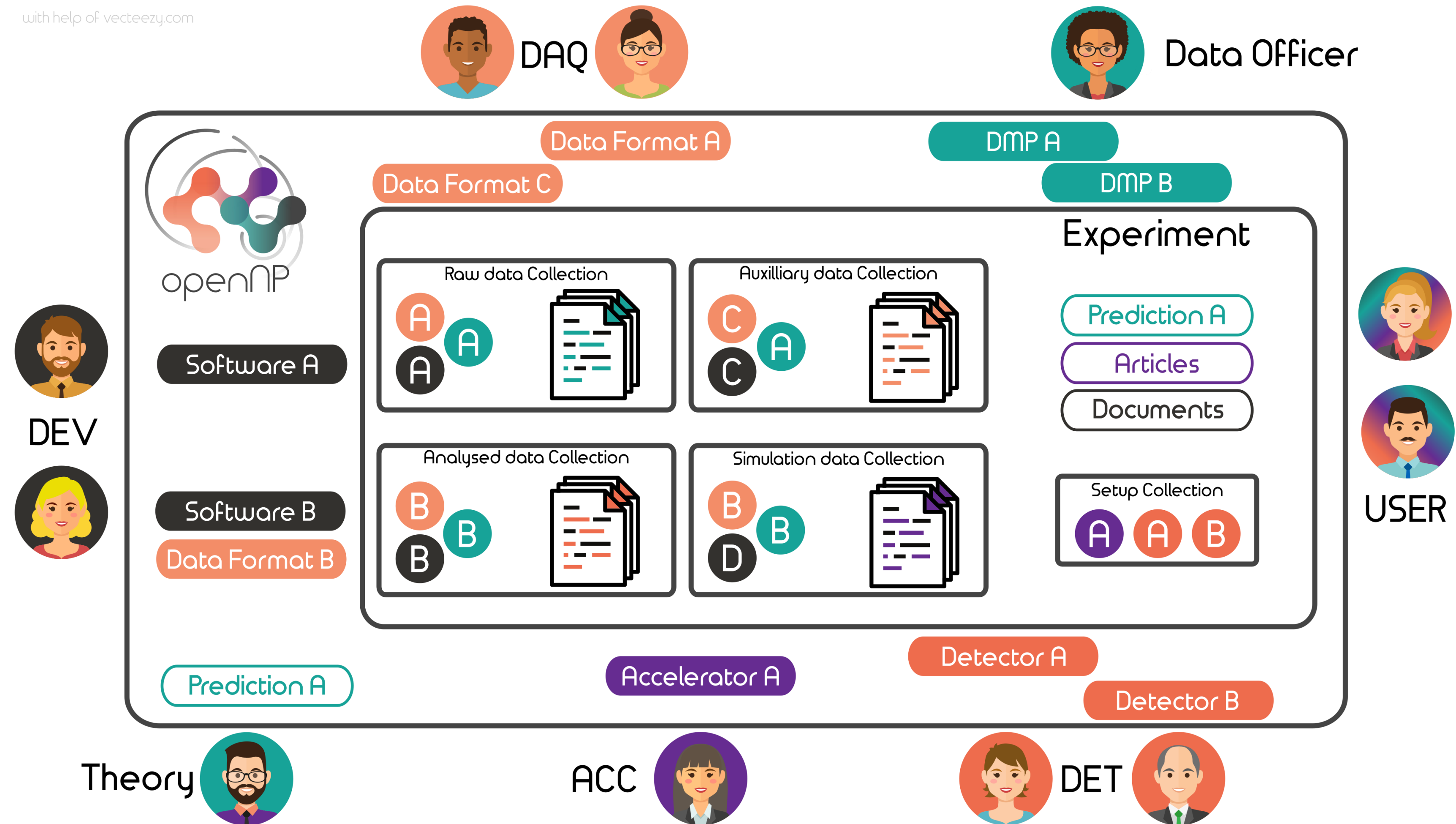
openNP is a catalog of

- experimental datasets from RI
- auxillary Data (elog, …)
- DMP
- DAQ softwares (data format, readout, …)
- Analysis Softwares
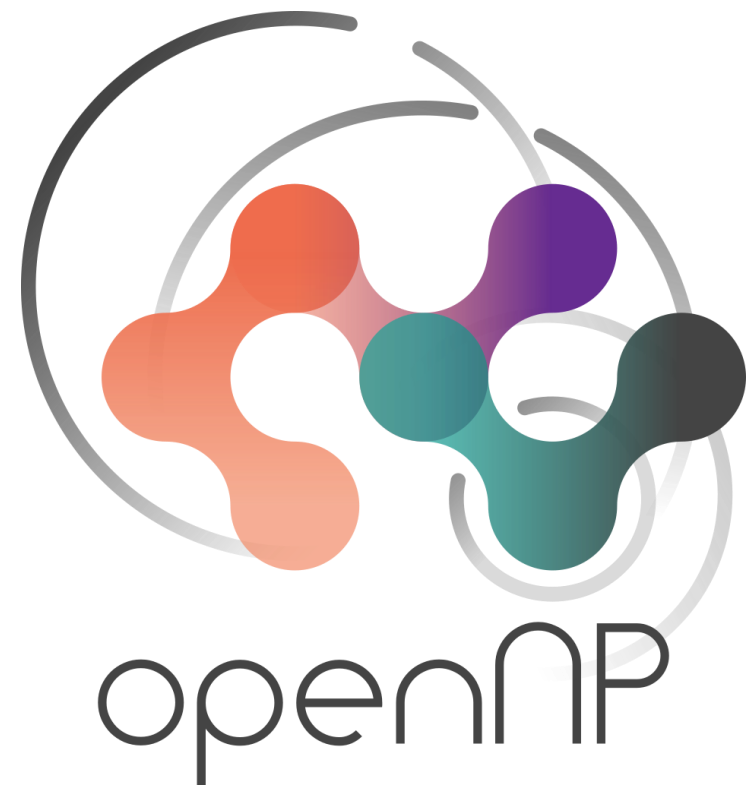- Experimental setup descriptions
- Analysed data

Involving

- Researchers,
- Data Officers,
- IT Departments,
- …

Do not intend to store the
large experimental datasets

Open NP Catalog

Milestone MS35 openNP catalog perimeter, architecture, and standards
Released on zenodo Aug 24th 2023: **https://doi.org/10.5281/zenodo.8279798**

1) Definition of the data type and metadata fields
Beams, Accelerator, Experimental Setup, DataSet, Softwares, ... => Subject to evolution
Synergies with ongoing activities on metadata standardization at GSI and PUNCH4NFDI project
(C. Hornung, A. Mistry *et al. GSI*)

2) Planned implementation of the catalog using DataVerse (https://dataverse.org)
- already used in a wide range of reasearch context
- a flexible handling of nested collections (compared to zenodo)
- import/export capabilites with other existing repositories

3) Prototype expected 2024 (hosted at CC-IN2P3)
- Prototype tests and adjustment with the help of RI
- deliverable is final service in M36

data manager position at GANIL to be publicized in the coming week
/!\ recruitment challenge for emerging skills => Need for training in the community

WP5.2 : Open Science and Data

- Task 1 : Data Management Plan (A. Lemasson -- GANIL)

- Task 2 : openNP Research Data catalog (A. Matta -- LPC Caen)

- Task 3 : New Services for improved access to data (and shared services)

  - 3.1 : Authentification and Authorization Infrastructure (M. Jouvin -- IJC Lab)

  - 3.2 : Prototype of data access plateform (C. Hornung -- GSI)

Authentification and Authorization Infrastructure

Task 3.1 (M. Jouvin IJCLab)

Common identification and authentification service

- Identification using existing federations (eduGAIN, …) using single sign on (SSO) or ORCID
- Easy registration of new services (compared to registration directly to the federation)
- Management of users based on groups to handle/differenrtian authorizations levels on the service side
- Benefiting from the experience gained from ESCAPE

=> INDIGO IAM

Authentification and Authorization Infrastructure
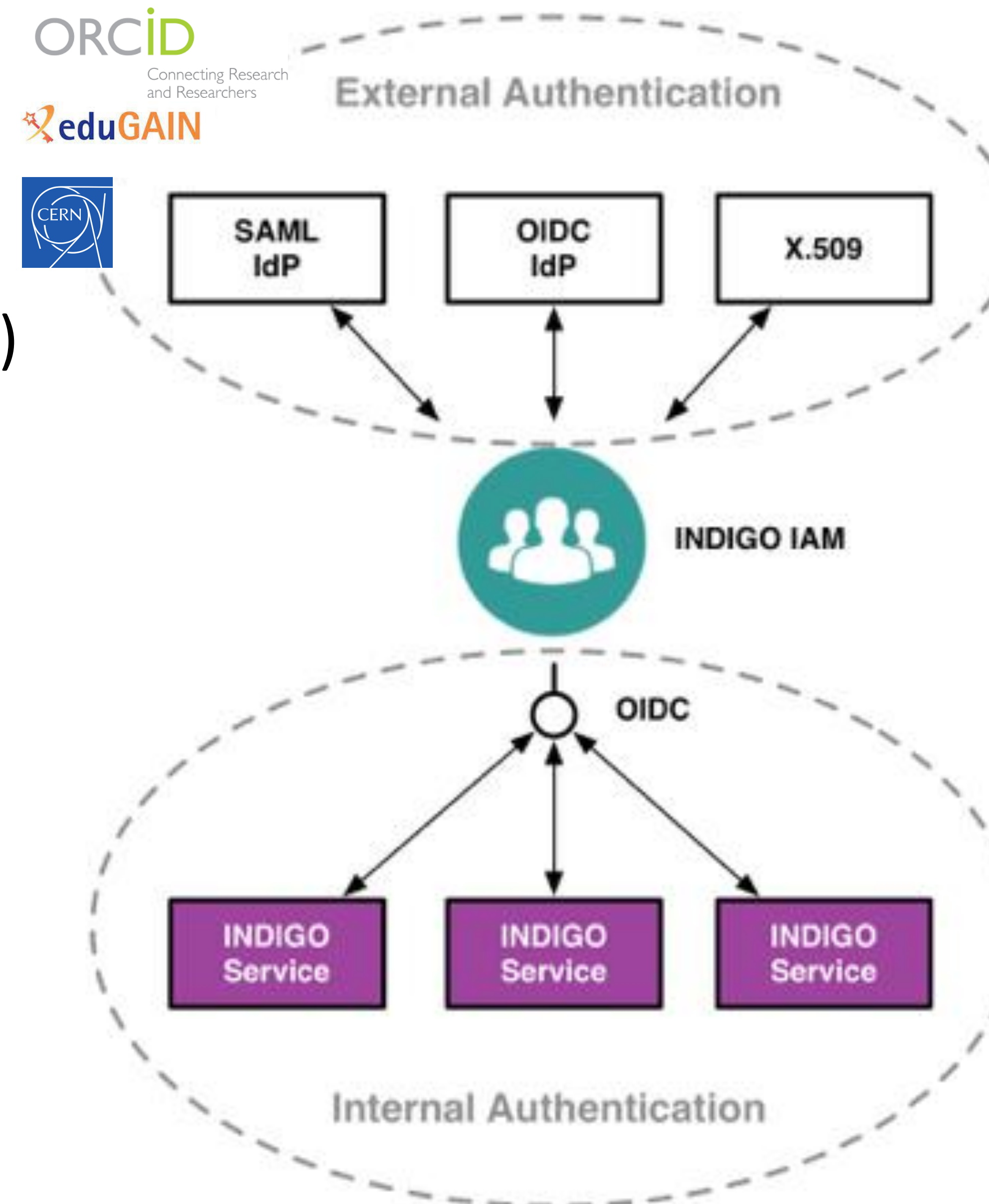
Service was released in June (you can register now)

https://iam-eurolabs.ijclab.in2p3.fr

Already used by

- Milano for the NS4EXP virtual access theory plateform
 (see talk from Manuela Rodriguez-Gallardo)
- Grafana service identification at GANIL for EXOGAM@NFS experiment

=> presently 30 users

If such a service may seem useful to you, please contact us

Next steps :

1) build an operation committee with members from the various labs
to manage the service (approval of registration, group management),
define policies
2) Extend to ORCID identity provider (in the coming week)
3) Distribute exemples of implementations (APACHE server, python, …)

=> 18 months position at IJCLab

WP5.2 : Open Science and Data

- Task 1 : Data Management Plan (A. Lemasson -- GANIL)

- Task 2 : openNP Research Data catalog (A. Matta -- LPC Caen)

- Task 3 : New Services for improved access to data (and shared services)

    - 3.1 : Authentification and Authorization Infrastructure (M. Jouvin -- IJC Lab)

    - 3.2 : Prototype of data access plateform (C. Hornung -- GSI)

Data Access Plateform

Task 3.2 (C. Hornung, GSI)
Prototype of data access plateform

Test of different data access plateform prototypes (Data Lakes)
Different data lake prototypes are being developed/tested at the GSI.
- For the CERN data lake, the GSI contributes two rucio storage elements (RSEs). This data lake was build and tested within the WP2 in the ESCAPE project.
- Another data lake that span three participating sites, GSI, Karlsruhe Institute of technology (KIT) and the university of Groningen (RUG) was also build and tested within the same ESCAPE project.
- A further prototype based on XRootD is also under investigation within the activities of the PUNCH4NFDI project.

The results from these prototypes combined with the other related activities like AAI and cashing will lead to a federated infrastructure concept that will be implemented and used for FAIR and could be implemented in other Research Infrastructures.

Deliverable is a prototype of plateform by M36.

Postdoc hiring procedure ongoing => Starting date beginning of 2024

WP5.2 : Open Science and Data

- **Task 1 : Data Management Plan (A. Lemasson -- GANIL)**
  => Delivered 02/2023
  => Encourage deposting EURO-LABS output to zenodo (in addtition to EURO-LABS website)
  => Encourage to establish Data Management Plan for data produced at all RI
  => If needed please contact us for discussion about DMP at your infrastructure

- **Task 2 : openNP Research Data catalog (A. Matta -- LPC Caen)**
  => MS 35 Delivered
  => Deliverable : Catalog to be released by M36

- **Task 3 : New services for improved access to data (and shared services)**
  3.1 Authentification and Authorization Infrastructure (M. Jouvin -- IJC Lab)
    => Service delivered, already used within EURO-LABS task 2.4
    => Policies and Management to be constructed with the different facilities
  3.2 Prototype of data access plateform (C. Hornung -- GSI)
    => Deliverable : Prototype of data plateform to be released by M36

Research Data Management is not a threat but an opportunity for RI and researchers.
A lot of tools already exists, that needs to be made available to the community.
Chosing the right tools and initiate a slight change of habits !