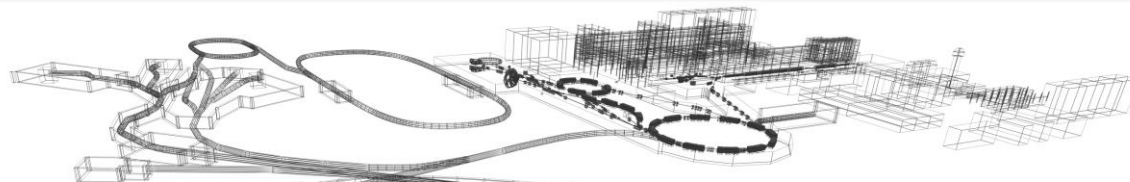




NAPMIX

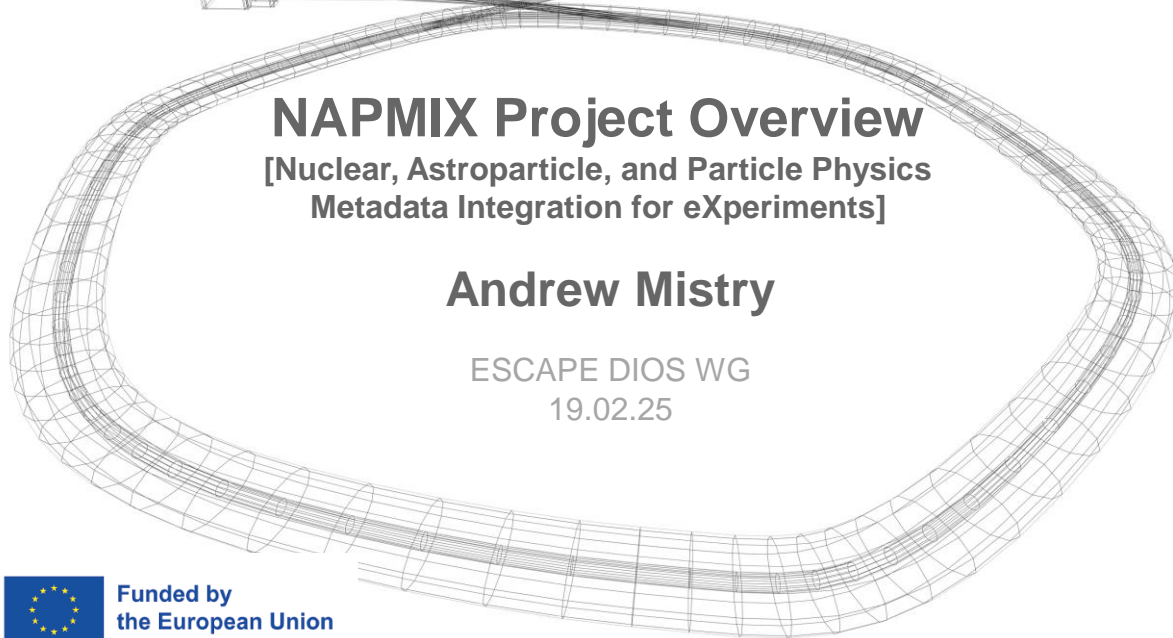


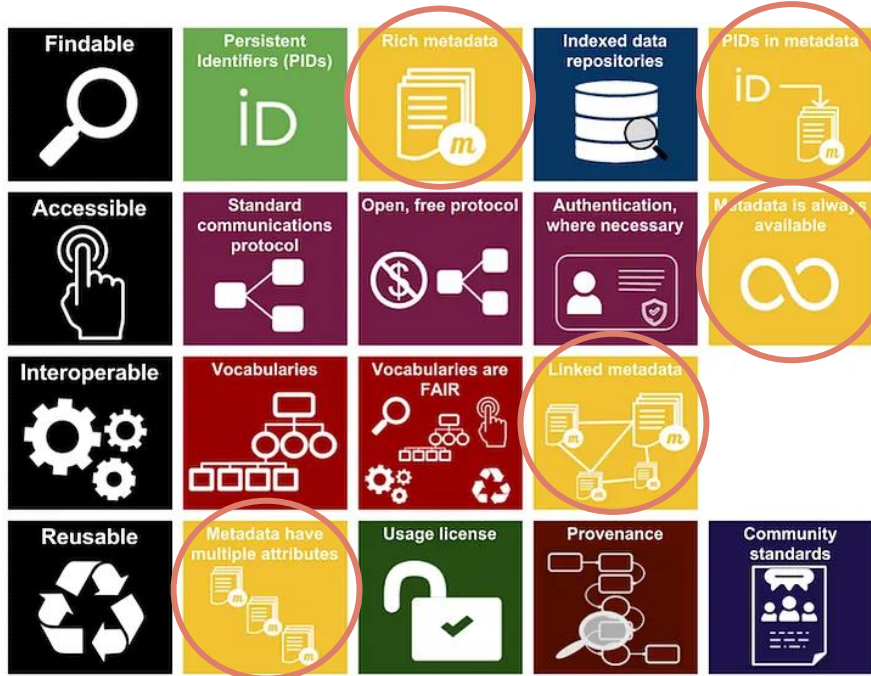
# NAPMIX Project Overview

[Nuclear, Astroparticle, and Particle Physics  
Metadata Integration for eXperiments]

Andrew Mistry

ESCAPE DIOS WG  
19.02.25





Icons made by Freepik are licensed by CC BY 3.0  
FAIR resources graphic is licensed under a CC 4.0 International License

Metadata is essential in Research Data Management and Open Science to enable FAIR data (and code)

- Allows datasets to be searched for and found
- Enables interoperability between datasets
- Facilitates data reprocessing
- Promotes efficient use of resources

Metadata essential in Research Data Management and Open Science to enable FAIR data (and code)  
More info: [Guide to HMC Better Metadata Booklet](#)

2022: *We needed a metadata scheme ‘for GSI’:*

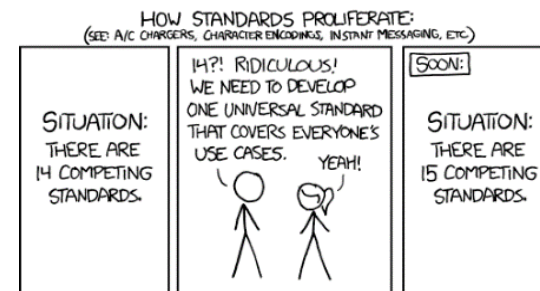
- Ambitious, at the same time not ambitious enough?
- GSI data is broad and complex: *areas of plasma physics, material science, bio physics, hadron physics, atomic physics, nuclear physics, theory etc etc.*
- Where to address?: *Community consultation*

## Starting point: No common schema (yet) between nuclear physics experiments

- Overlaps with other ‘particle accelerator’ fields
- Many labs in Europe to obtain input from
- Is a field that needs to tackle RDM/Open Science, behind the curve

**Caution: Don’t reinvent the wheel!**

**Look for synergies with NeXUS, HELPMI, OpenAIRE et al.**

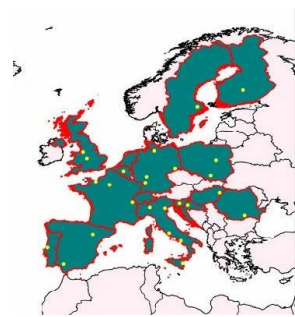


By Randall Munroe: <https://xkcd.com/927>

## Goals:

- Providing efficient access to the available resources to a large fraction of **EURO**pean **L**aboratories for **A**ccelerator **B**ased **S**ciences (**EURO-LABS**)
- Bringing together the three communities engaged in Nuclear Physics, Accelerator and Detector technology for High Energy Physics
- Allowing a synergic implementation of best practices for data management and activities relating to targeted service improvements at these RIs
- Creating synergies and collaborations between the RIs of the Nuclear and High Energy communities
- Enhancing Europe's potential

▪ <https://web.infn.it/EURO-LABS/>



*This project has received funding from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No 101057511.*

# Synergies: EURO-LABS WP 5.2

## Data management and Open Science



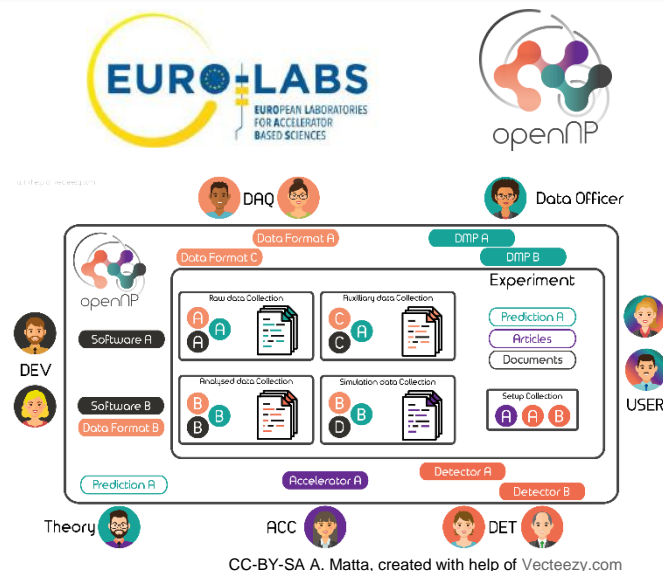
- **EURO-LABS** : Providing **efficient access to the available resources** to a large fraction of **EURO-LABS**
- *Participants*: CSIC, GANIL (Leading partners), INFN, CNRS, IJCLab, GSI

### Goals

- Developing services to enhance F.A.I.R data principles
- Integration of Nuclear Physics community to existing infrastructures/services of EOSC environment - using present experience from ESCAPE/HEP physics community

### Developments

- **Open Science School @ GSI**: November 2024: Training and lectures for PhD Students/Postdocs delivered by experts
- **openNP catalogue** of experimental datasets, auxiliary data, DAQ software, Analysis software...
- **Metadata** for Nuclear Physics Experiments
- **Authentication and Authorization** Infrastructure
- **Http connection** to lustre -> Open Data (large volume)



This project has received funding from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No 101057511.

# Synergies: PUNCH4NFDI (Germany)

## Task Area 2 “Data management” – IT department

- Automatization and optimization of big data management workflows

## Task Area 4 “Data portal” – Research Division

### Mapping and collating metadata

- Build and operate the science data portal

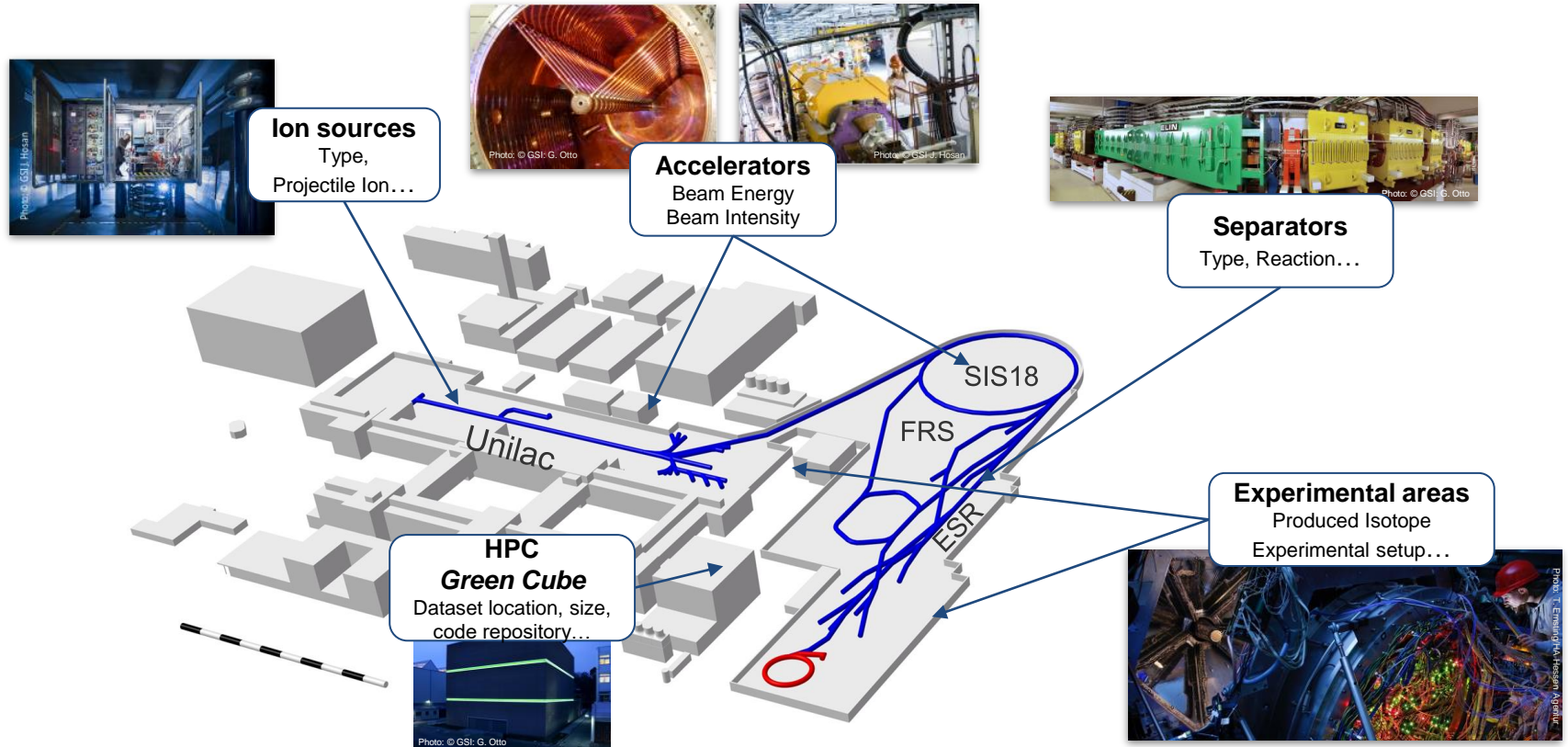
## Task Area 6 “Synergies and services” – IT department

- Authorization and authentication infrastructure

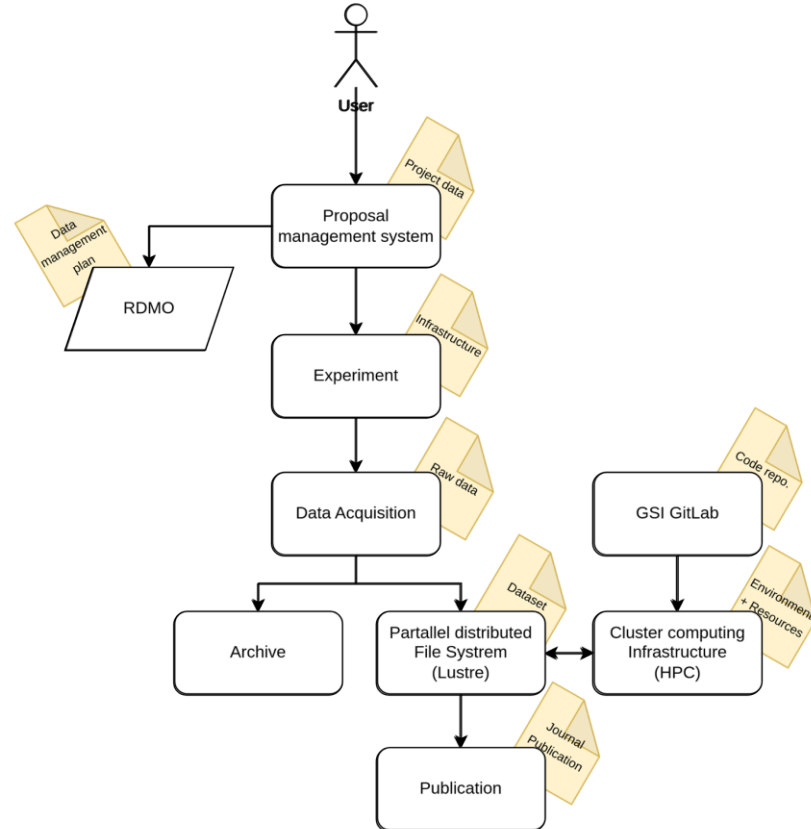
# Particles, Universe, NuClei and Hadrons for the NFDI



The prime goal of PUNCH4NFDI is the setup of a federated and "FAIR" science data platform, offering the infrastructures and interfaces necessary for the access to and use of data and computing resources of the involved communities and beyond. This will help to master the challenges of big data, open data, and FAIR workflows.



- Workflow is representative of experimental procedures at GSI and similar facilities.
- Each stage generates specific metadata crucial for data management.
- Our metadata stratification and categorization is grounded in this workflow.



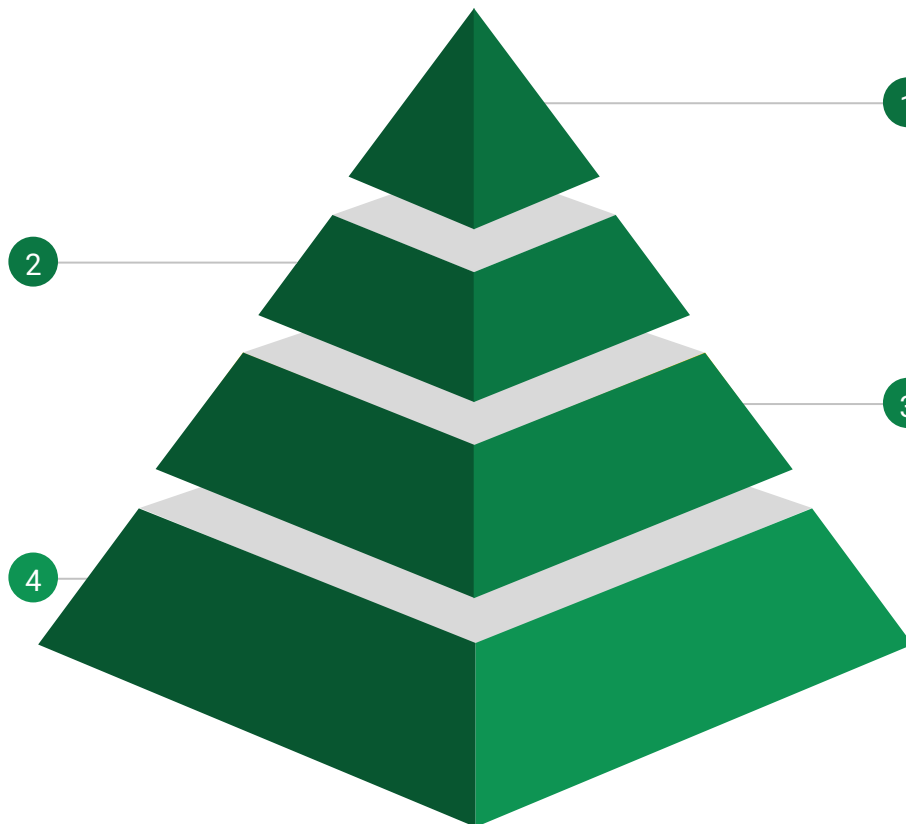


## Discipline specific metadata

Projectile data, target data, infrastructure, detector type, datasets, software repositories

## Dataset specific metadata

Data headers, structure....



## Project specific metadata

Project name, publication date, collaborations, researchers, institutions, journal references

## Experiment specific metadata

Reaction data, gamma-ray energies, time of flight parameters, instrument records

# Pilot: Design and Implementation

- Contact with EuroLabs Partners
- Development
- Support of

CHEP Conference Proceedings (In Publication)

Enhancing Data Management in Nuclear Physics through a F.A.I.R.-Compliant Metadata Schema

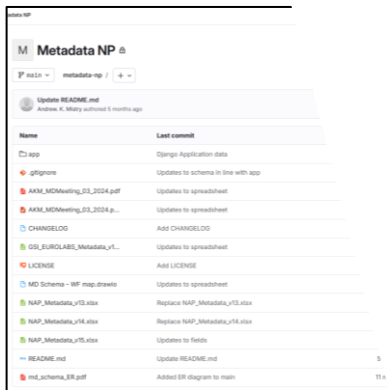
*I. Knežević<sup>1,\*</sup> and A.K. Mistry<sup>1,2,\*</sup>*

<sup>1</sup>GSI Helmholtzzentrum für Schwerionenforschung GmbH, Planckstraße 1, 64291 Darmstadt, Germany  
<sup>2</sup>Facility for Antiproton and Ion Research in Europe (FAIR GmbH), Planckstraße 1, 64291 Darmstadt, Germany

**Abstract.** Within the field of experimental nuclear physics, steps are underway to further advance good Research Data Management and ensuring compliance with these principles. Here, we describe the landscape of Research Data Management in the nuclear physics domain, with a focus on supporting future actions with a F.A.I.R.-compliant metadata schema prototype for the community. The schema's nodal, multi-layered structure supports metadata enrichment across domains, and a user-friendly generator facilitates schema customisation and export in various formats. We detail the development process, key features, and potential applications. An outlook is given toward broader expansion of the schema to cross-domain physics fields.



ATOMKI, LATR-IST



Git Project



Basic Spreadsheet



Django Backend



JSON/XML Schema

Software: <https://doi.org/10.5281/zenodo.14770678>

# Nuclear, Astroparticle, and Particle Physics Metadata Integration for eXperiments (2024 - 2026)

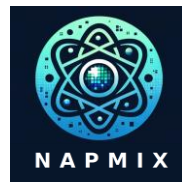


<https://www.oscars-project.eu/projects/napmix-nuclear-astro-and-particle-metadata-integration-experiments>

The NAPMIX project emerged to address a significant gap in nuclear, astroparticle and particle physics: the lack of a unified metadata schema necessary for achieving FAIR datasets. This challenge extends to the fields of particle and astro-particle physics, highlighting the need for a collaborative European effort to **create a common metadata schema with user-friendly infrastructure**. By integrating expertise across these domains, NAPMIX aims to enhance data management practices.



The authors acknowledge the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751



Funded by:



# NAPMIX Project Outline

- A front-end generator for user input, producing outputs in both machine and human-readable formats.
- A multi-layered 'nodal' structure to incorporate use cases based on real experimental datasets.
- Connections to existing ontologies such as DataCite, OpenPMD and NeXus.
- Interfaces to current infrastructures as a pathway to an Open Science Ecosystem.
- Testing and deployment of the schema with multiple use cases to enhance data FAIRness.
- Training for researchers on the importance and application of metadata.



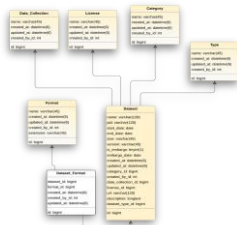
➤ 24 Months (01/12/24 - 30/11/26)

Concept

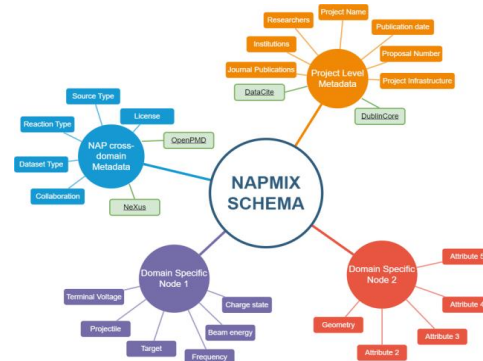
Design

Implementation

Schema



```
graph TD
    subgraph NAPMIX_SCHEMA [NAPMIX SCHEMA]
        direction TB
        Node1((Domain Specific Node 1))
        Node2((Domain Specific Node 2))
        Node3((Project Level Metadata))
        Node4((NAP cross-domain Metadata))
        
        Node1 --- Node2
        Node1 --- Node3
        Node1 --- Node4
        Node2 --- Node3
        Node2 --- Node4
        Node3 --- Node4
    end
```



# NAPMIX Project Goals



**Development of a community in experimental nuclear, astro-particle, and particle physics to generate a common metadata schema and user interface -> Consortium input, options to onboard new partners**

**NAPMIX Database Reference Structure**

Form and spreadsheet pages shown below are for illustrative purposes only.

Consortium 1  
by  
Ministry of most use, none possible  
optional not mandatory one

Optional name from user

The column **Source** for each cell, selecting more than one researcher will generate the **fully Author Name, Author Email, Multiple Sites** (see e.g. Introduction in the example form)

Source	Consortium	Relationship	Example Field First	Name	Field Type	Input Type	App Name	Department	DIRTY	NAPMIX Proposal ID
Project	1	Project Name	Measurement of $^{82}\text{Ca} + ^{10}\text{B}$	Name of Project	String	Text Box	Project Name	Projects		projectName
Project	1	Publication Date	690203	Date of Data Publication	Date	Calendar	ProjectPublication_date	Projects		projectPublicationDate
Project	0/1	Project Number/Project ID	022-01-2022	ORCID iD dependent number obtained assigned at the proposal submission (0-22-0000)	String	Text Box	Projectproposal_number	Projects		projectProposalNumber
Project	0+	POF Type	Control Matter in the Laboratory	Identifies Program/Control Funding	String	Key Value	ProjectPOF	ProjectControlFunding		projectPOF
Project	0+	Journal references	Journal Ref Name: POC	Links to publications derived from data	String	Key Value	Projectjournal_references	JournalReferences		projectJournalReferences
Project	0+	Software references	Software Ref Name: POC Input URL	Links to software used with data	String	Key Value	Projectsoftware_references	SoftwareReferences		projectSoftwareReferences
Project	0+	Field/Institute of Data Generation	GG	2 Data generation	String	Key Value	Projectfield	Institute		projectFieldInst
Project	0+	Website/URL	84FC.COMPASS	Information using in data generation	String	Key Value	ProjectwebsiteURLs	Institute		projectWebsiteURLs
Project	1+	Reactions	48 Ca + $^{10}\text{B}$	Links to Reactions	String	Key Value	Projectreactions	Reactions		projectReactions
Project	0+	Collaboration link to project	84FC Omega Spectroscopy		String	Key Value	Projectcollaborations	Collaboration		projectCollaborations
Project		Created by		Project record creator author	Text	Auto	Projectcreated_by			projectCreatedBy
Project		Created at		Project record creation timestamp	Date	Auto	Projectcreated_at			projectCreatedAt
Project		Updated at		Project record update timestamp	Date	Auto	Projectupdated_at			projectUpdatedAt
Department	1+	Department/Division Name	NUSTAR SHE Physics	Define Department	String	Text Box	Department_name	Institute		departmentName
Department	0/1	Department/Division Short Name	SHE Physics	Abbr. Department name	String	Text Box	Department_short_name	Institute		departmentShortName
Department		Department/Division created by			String	Auto	Department_created_by			departmentCreatedBy
Department		Department/Division created at			Date	Auto	Department_created_at			departmentCreatedAt
Department		Department/Division updated at			Date	Auto	Department_updated_at			departmentUpdatedAt
Institute	1+	Institute Name	GSI Helmholtzzentrum für Schwerionenforschung	Define Institution	String	Text Box	institute_name			instituteName



**Ensure the scheme is enriched with multiple use case 'nodes' based on real experimental datasets**

*-> Use case collation and enhance database for cross domain capabilities*

### NAPMIX Use Case Collection sheet

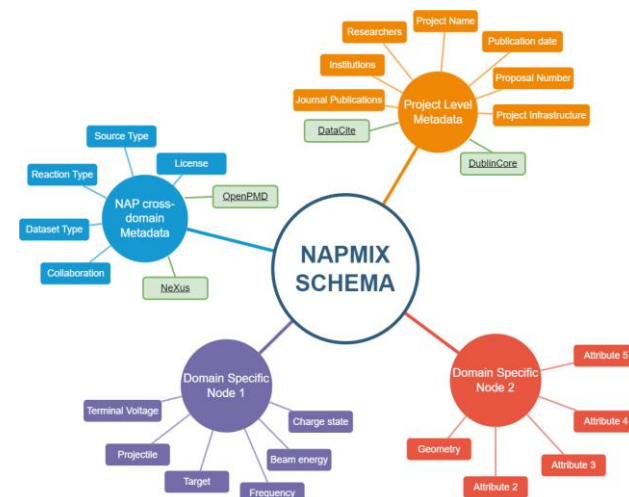
Items are separated into pages when input into excel for ease of use

Please add any additional fields required to the base of the table, or make a new table if needed

The end user has the option to add additional attributes: Elements, Materials, Particles, Units for beam intensity, spread etc.,

The cardinality is hierarchical for each, i.e., selecting more than one researcher will generate the fields Author Name, Author Email etc. Multiple times. (see e.g. Infrastructure in the example here)

Model	Cardinality	Metadata Attribute	Example GSI SHE Filled Field	Example for Use case	Notes	Field type	UI
<b>Project</b>	<b>1</b>	<b>Project Metadata</b>		<i>Please fill this column for an example dataset</i>			
Project	1	Project Name	Measurement of 48Ca + 181Ta		Name of Project	String	TR
Project	1	Publication Date	5/6/2023		Date of Data Publication	Date	CR
Project	0:1	Proposal Number/Project ID	022-01-2022		Official experiment number obtained assigned at the proposal submission (3-22-00)	String	TR
Project	0+	POF Type	Cosmic Matter in the Laboratory		Heinrich Heine Program-Oriented Funding	String	KI
Project	0+	Journal references	Journal Ref Name: PFD		Links to publications derived from data	String	KI
Project	0+	Software references	Software Ref Name: PFD-Repo URL		Links to Software used with data	String	KI
Project	1+	Facility/Institute of Data Generation	GSI		CI data generation	String	KI
Project	1+	Reactions	48Ca + 181Ta		Links to Reactions	String	KI
Project	0+	Collaboration link to project	SHEP Decay Spectroscopy			String	KI
<b>Researcher</b>	<b>1+</b>	<b>Author Metadata</b>					
Researcher	1	Author ID	1		Auto Fills	int	AR
Researcher	1	Author Name	Andrew Mistry		Author Name	String	TR
Researcher	1	Author Email	a.k.mistry@gsi.de		Author email	String	TR
Researcher	0:1	Author Title	Dr		Author title	String	TR
Researcher	0:1	Author ORCID	https://orcid.org/0000-0002-0951-0475		Author ORCID	String	TR
Researcher	0+	Author Institutes	GSI		Author Institution(s)	String	KI
Dataset/Researcher	1	Author Role (Project)	Data Manager		Author Role in Project	String	TR
Research/Institute	0:1	Author Position (Institute)	Data Steward		Author Position at Institute	String	TR
Research/Institute	0:1	Author Start Date (Institute)	01.03.2023		Author Start Date at Institute	Date	CR
Research/Institute	0:1	Author End Date (Institute)			Author end Date at Institute	Date	CR



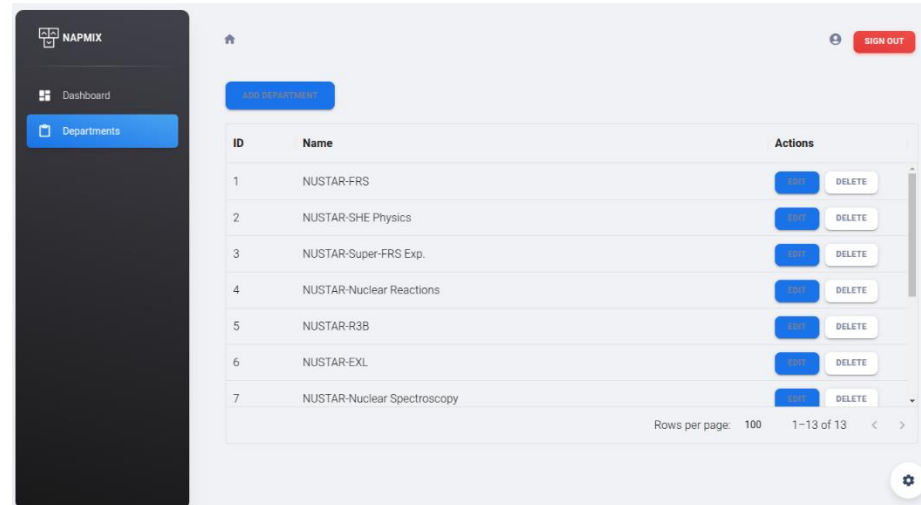
**Develop a frontend interface for end-user input with outputs in machine and human readable formats -> SciCat/REACT/InvenioRDM... Outputs to JSON, XML + XSD Validator**

Investigate available options

Open Source Solutions: e.g. SciCat



REACT



The screenshot shows the REACT application interface. On the left is a dark sidebar with the NAPMIX logo and navigation options for 'Dashboard' and 'Departments'. The main content area features a 'ADD DEPARTMENT' button and a table with the following data:

ID	Name	Actions
1	NUSTAR-FRS	<a href="#">EDIT</a> <a href="#">DELETE</a>
2	NUSTAR-SHE Physics	<a href="#">EDIT</a> <a href="#">DELETE</a>
3	NUSTAR-Super-FRS Exp.	<a href="#">EDIT</a> <a href="#">DELETE</a>
4	NUSTAR-Nuclear Reactions	<a href="#">EDIT</a> <a href="#">DELETE</a>
5	NUSTAR-R3B	<a href="#">EDIT</a> <a href="#">DELETE</a>
6	NUSTAR-EXL	<a href="#">EDIT</a> <a href="#">DELETE</a>
7	NUSTAR-Nuclear Spectroscopy	<a href="#">EDIT</a> <a href="#">DELETE</a>

At the bottom right of the table, it indicates 'Rows per page: 100' and '1-13 of 13'. A settings gear icon is visible in the bottom right corner of the interface.

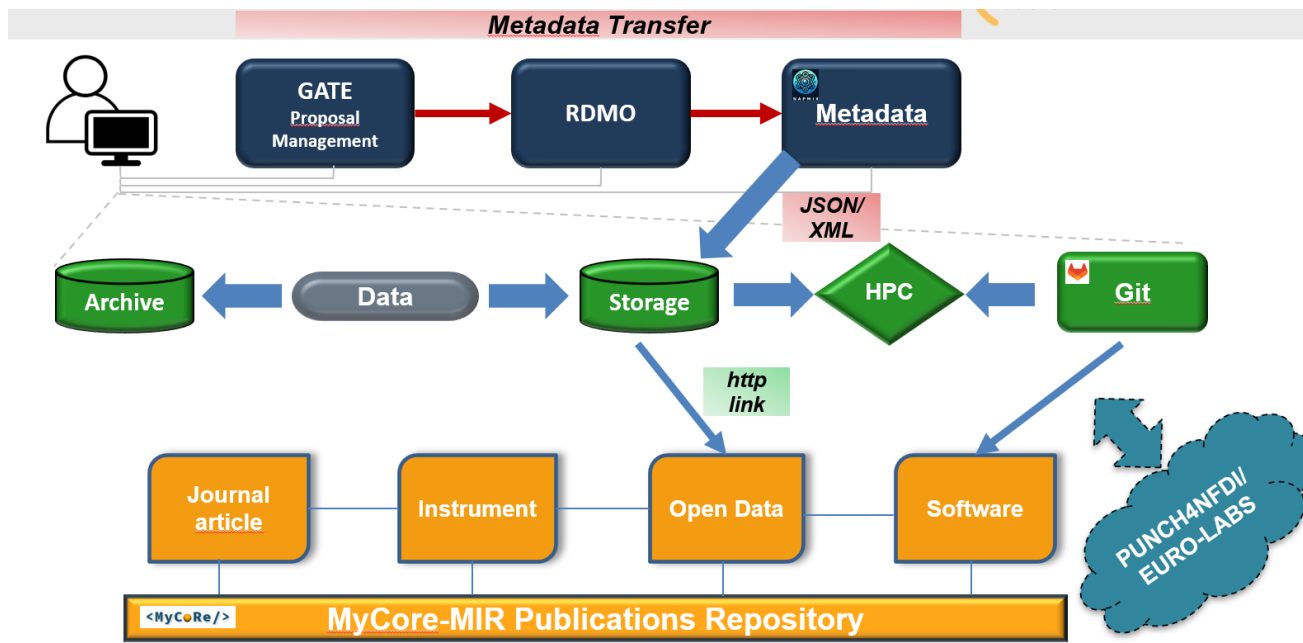
**Establish connections and mappings to existing/ongoing ontologies -> DataCite, NeXuS, OpenPMD, HELPMI...**

Model	Cardinality	Metadata	Dependent on	Entry Section	NAPMIX Proposed Standard	DataCite
		<i>General Information</i>				
Project	1	Project Name		Projects	projectName	Title
Project	1	Publication Date		Projects	projectPublicationDate	PublicationYear
Project	0/1	Proposal Number/Project ID		Projects	projectProposalNumber	Identifier? If Unique
Project	0+	POF Topic	ProgramOrientedFunding	Projects	projectPoF	N/A
Project	0+	Journal references	JournalReference	Projects	projectJournalReferences	relatedIdentifier (relationType: IsReferencedBy, relatedIdentifierType: DOI/URL)
Project	0+	Software references	SoftwareReference	Projects	projectSoftwareReferences	relatedIdentifier (relationType: IsReferencedBy, relatedIdentifierType: DOI/URL)
Project	1+	Facility/Institute of Data Generation	Institute	Projects	projectInstitutes	contributor (contributorType: HostingInstitution)
Project	0+	Infrastructure	Infrastructure	Projects	projectInfrastructures	TBD
Project	1+	Reactions	Reaction	Projects	projectReactions	N/A
Project	0+	Collaboration link to project	Collaboration	Projects	projectCollaborations	
Project		Created by			projectCreatedBy	contributor (contributorType: ContactPerson)
Project		Created at			projectCreatedAt	created_at: date (dateType: Created)
Project		Updated at			projectUpdatedAt	updated_at: date (dateType: Updated)





**Implement interfaces to existing infrastructure as a stepping stone towards an Open Science Ecosystem -> API/GraphQL capabilities, Repository Ingestion**



**Test the deployment of the schema and interface with a number of use cases, publishing use-cases with data to enhance the F.A.I.R-ness of the data -> Institutional/Discipline specific repositories + Platforms (e.g., PUNCH RPR) Schema Ingestion, Metrics on uptake**



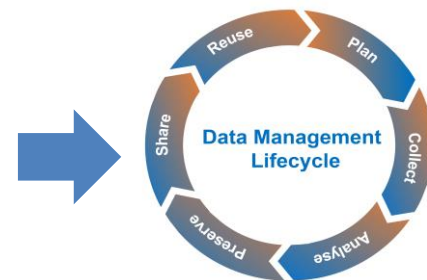
## For example:

- Deployment of app in partner institutions
- Apply to given use cases (including data publication!)

**Teaching and training on the purpose, scope, and use of metadata to researchers -> Drive Uptake among researchers**

## Hands-On Metadata/RDM workshop for researchers (2026)

- Basic training on the purpose of metadata
- Apply metadata to their projects including application



### Expectations



### Reality



### Aim



***Publish the schema and broadening connections within the ESCAPE community and EOSC -> Links and synergies, ensure longevity of outcomes***



**Broaden engagement to other RI's and universities**

**Additional use cases can always be added, the project forked in Git etc**

# Project Schedule (Basic)

Year	2024												2025												2026											
Month	Dec 1	Jan 2	Feb 3	Mar 4	Apr 5	May 6	Jun 7	Jul 8	Aug 9	Sep 10	Oct 11	Nov 12	Dec 13	Jan 14	Feb 15	Mar 16	Apr 17	May 18	Jun 19	Jul 20	Aug 21	Sep 22	Oct 23	Nov 24												
Step																																				
1																																				
2																																				
3																																				
4																																				
5																																				
6																																				
7																																				

## Tasks:

Kick-off Workshop

Database Model Evaluation and Improvement

Data Enrichment Strategy Implementation

API Integration for Metadata Transfer

Metadata Schema Testing and Publication

Comprehensive Training Workshop

EOSC integration and project dissemination

**Deliverable D1:** Written report on the pilot model and publishing the prototype database structure to the Open GitLab project -> CHEP Proceedings

**Deliverable D2:** Written report/Presentation on the choice of front end and implementation of the structure

**Deliverable D3:** Published article or white paper on the NAPMIX schema including use case examples and proof of F.A.I.R. data publication

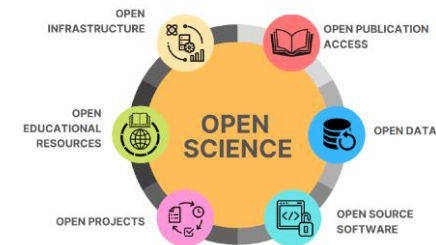
**Deliverable D4:** EOSC presentation event



- Links to DIOS: exchange and interoperability of the scheme into the interface?
- REST APIs (or GraphQL) endpoints -> API specifications
- Community building and schema enhancements, further use cases?
- AAI mechanisms
- Data cataloguing (including OpenNP project from EURO-LABS)



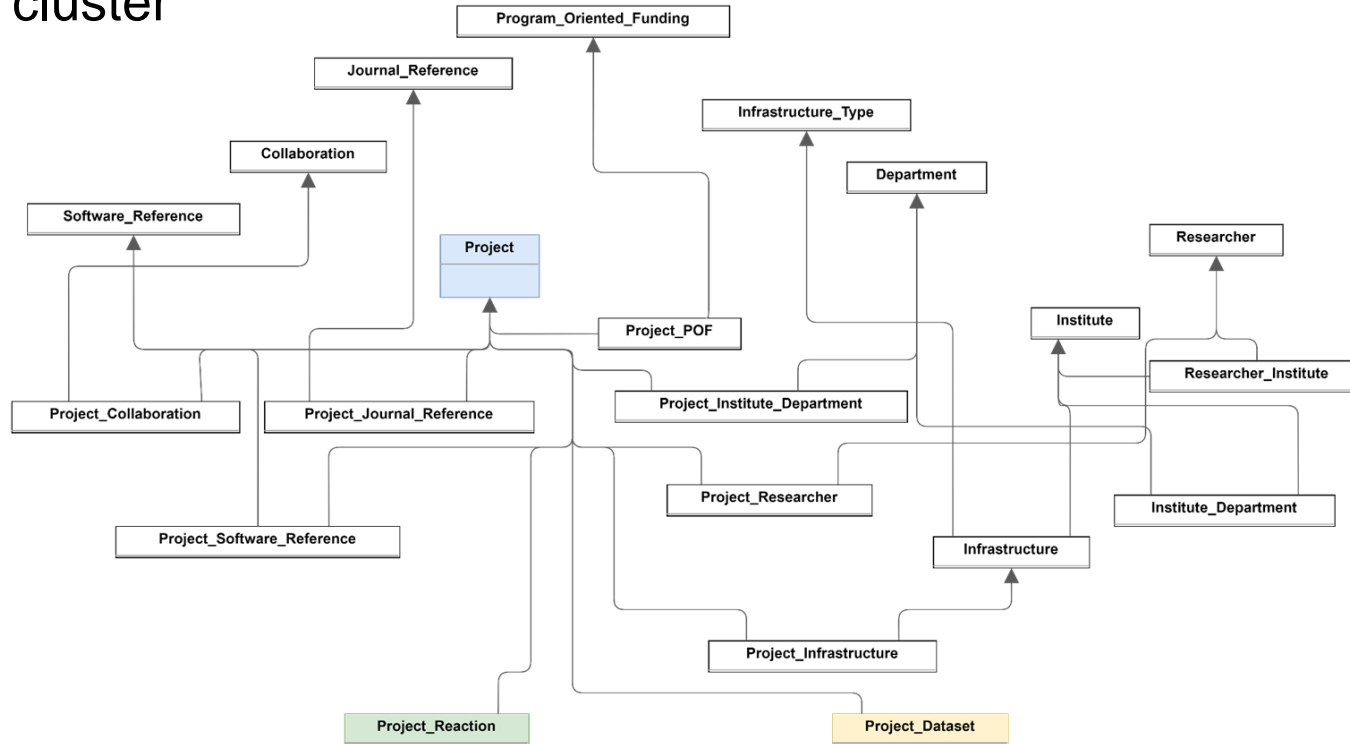
- Next steps:**
- Use Case Development
  - Development of Database
  - Investigations into Suitable Front-End



The authors acknowledge the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751

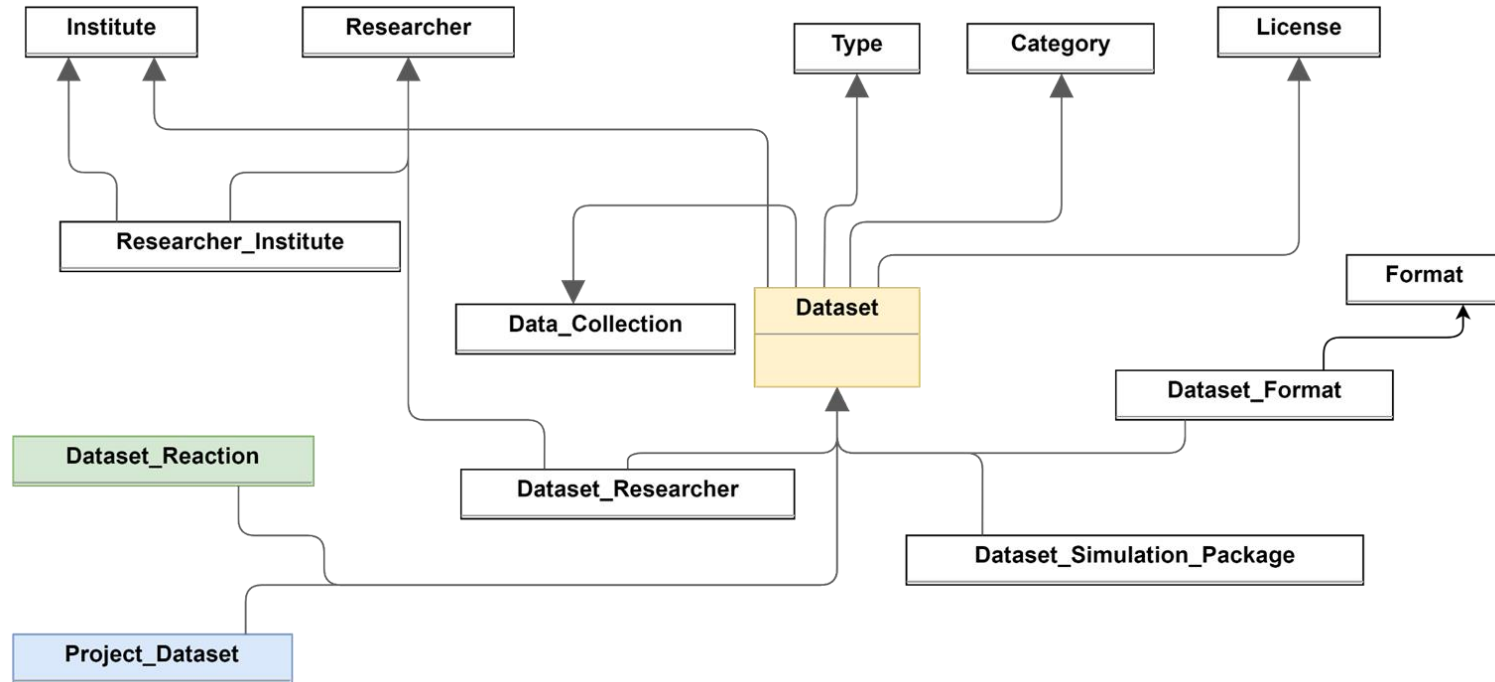


## Project cluster

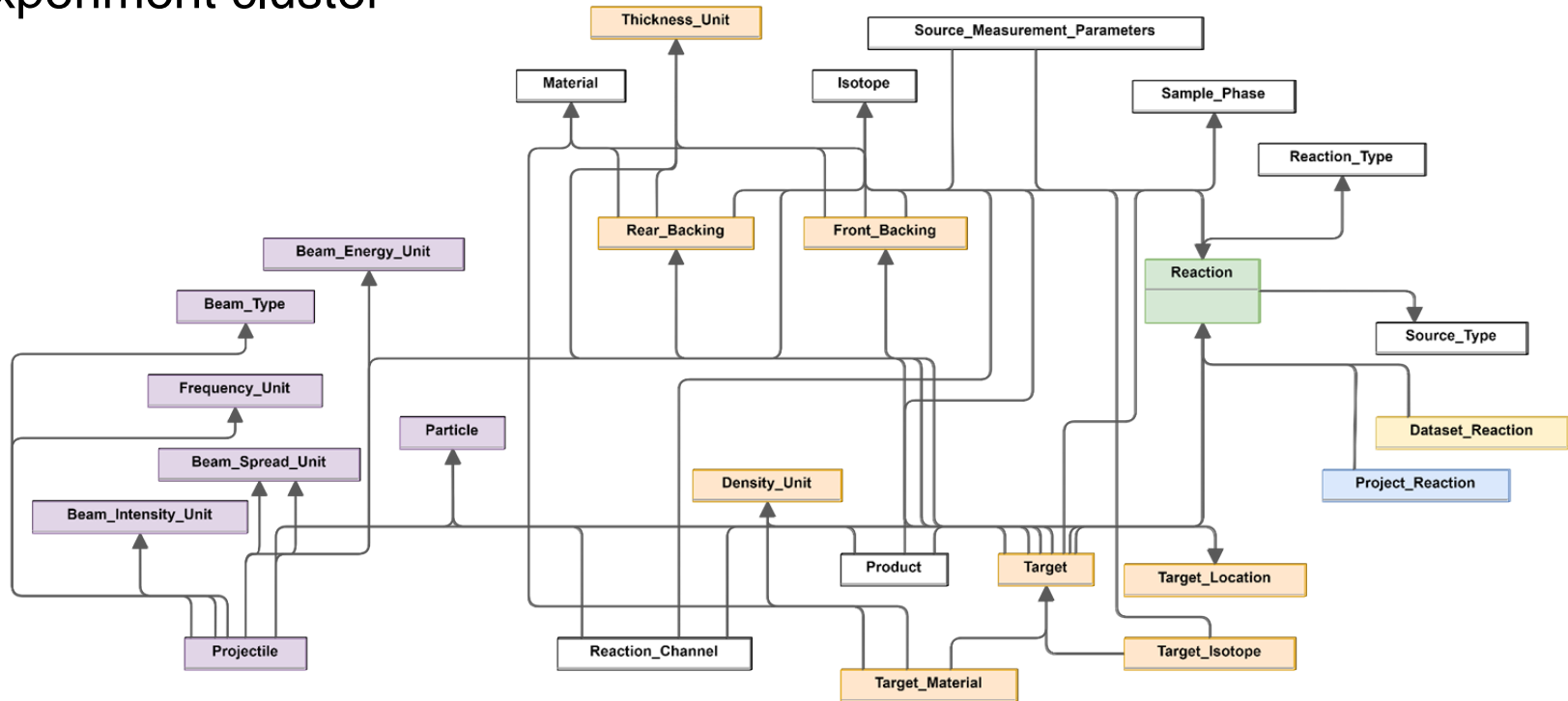


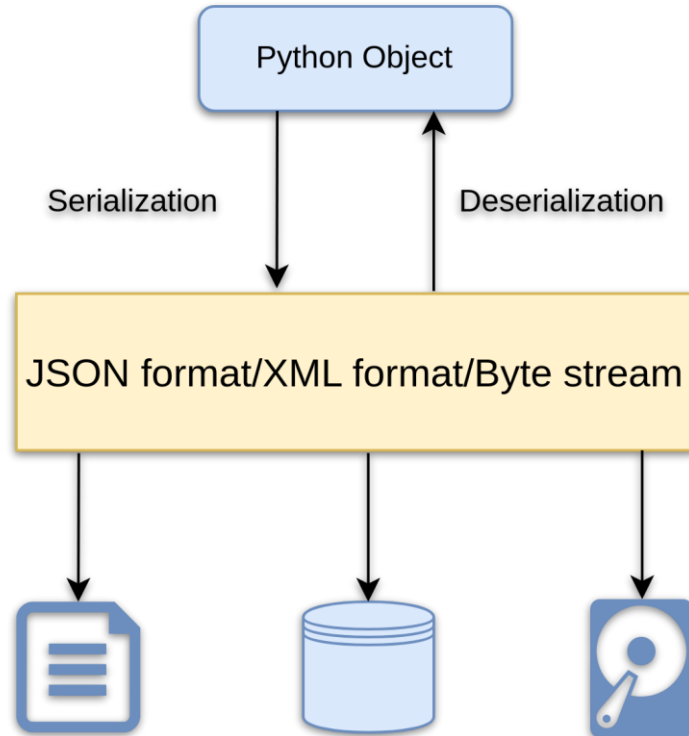


## Dataset cluster



## Experiment cluster





```
def serialize_projectile(projectile, exclude_fields=None, field_name_mapping=None): 1.usgagf 4.iknezevic
    exclude_fields = ['reaction'] if exclude_fields is None else exclude_fields + ['reaction']
    projectile_data = {}

    # Prepare reverse mapping for special fields
    reverse_field_name_mapping = {v: k for k, v in (field_name_mapping or {}).items()}
    special_fields = {
        'isotope': reverse_field_name_mapping.get('isotope', 'isotope'),
        'beam_energy_unit': reverse_field_name_mapping.get('beam_energy_unit', 'beam_energy_unit'),
        'beam_intensity_unit': reverse_field_name_mapping.get('beam_intensity_unit', 'beam_intensity_unit'),
    }

    field_order = ['isotope', 'percentage', 'beam_energy', 'beam_energy_unit', 'beam_operation', 'frequency',
                  'average_beam_intensity', 'beam_intensity_unit', 'charge_state', 'beam_destination', 'energy_spread',
                  'beam_spread_unit', 'beam_spot_size', 'beam_spot_size_unit', 'beam_type']
```

- Serializing each smaller cluster separately,
- Increased schema flexibility,
- Relationships maintained

Start typing to filter...

### AUTHENTICATION AND AUTHORIZATION

Groups + Add

Users + Add

### GSIMETADATA

Datasets + Add

Infrastructures + Add

Institutes + Add

Projects + Add

Reactions + Add

Researchers + Add

### Select Project to view

ADD PROJECT +

Action: [dropdown] Go 1 of 5 selected

		PUBLICATION DATE	INSTITUTES	PROPOSAL NUMBER
<input type="checkbox"/>	NA			
<input type="checkbox"/>	My	July 15, 2024	GSI	-
<input type="checkbox"/>	S10	July 1, 2024	GSI, GSI	S100
<input type="checkbox"/>	Exp	June 28, 2024	GSI	S482
<input type="checkbox"/>	Test of all Metadata	June 13, 2024	GSI, GSI	G111-XYZ
<input checked="" type="checkbox"/>	The 48Ca+181Ta reaction: Cross section studies and investigation of neutron-deficient 86sZs93 isotopes	April 11, 2024	GSI	G2024-123

5 Projects

- A tool that allows users to export comprehensive metadata for a selected project.
- Supports multiple formats and encoding standards.

## XML schema

```
<Reactions>
  <Reaction>
    <name>208Pb + 9Be</name>
    <reaction_type>Fragmentation</reaction_type>
    <source_type>Accelerator</source_type>
    <Projectiles>
      <Projectile>
        <Isotope>Lead (Pb-208)</Isotope>
        <Percentage>100.0</Percentage>
        <Beam_energy>1050.0</Beam_energy>
        <Beam_energy_unit>Megaelectronvolt/nucleon (MeV/u)</Beam_energy_unit>
        <Beam_operation>Pulsed</Beam_operation>
        <Average_beam_intensity>1500000000.0</Average_beam_intensity>
        <Beam_intensity_unit>Particles per spill (pps)</Beam_intensity_unit>
        <Beam_destination>HFS</Beam_destination>
        <Beam_type>Stable</Beam_type>
      </Projectile>
    </Projectiles>
  </Reaction>
</Reactions>
```

## JSON schema

```
"reactions": [
  {
    "name": "208Pb + 9Be",
    "reaction_type": "Fragmentation",
    "source_type": "Accelerator",
    "projectiles": [
      {
        "isotope": "Lead (Pb-208)",
        "percentage": 100.0,
        "beamEnergy": 1050.0,
        "beamEnergyUnit": "Megaelectronvolt/nucleon (MeV/u)",
        "frequency_type": "Pulsed",
        "beamIntensity": 1500000000.0,
        "beamIntensityUnit": "Particles per spill (pps)",
        "beam_destination": "HFS",
        "beam_type": "Stable"
      }
    ]
  }
]
```

Original attribute name

```
def export_to_json_datacite(self, request, queryset): 1 usage  @l.knezevic
if queryset.count() == 1:
    project = queryset.first()
    # Define your custom field mappings here
    custom_field_name_mapping = {
        'journal_article_pid': 'articlePID',
        'is_prime_investigator': 'PrimeInvestigator',
        'energy_spread': 'beamEnergySpread',
        'energy_spread_unit': 'beamEnergySpreadUnit',
        'beam_spot_size': 'beamSpotSize',
        'beam_spot_size_unit': 'beamSpotSizeUnit',
        'beam_energy': 'beamEnergy',
        'beam_energy_unit': 'beamEnergyUnit',
        'average_beam_intensity': 'beamIntensity',
```

Desired attribute name

- Attribute names can be customized via code (array mapping),
- Supports interoperability across platforms and institutions,
- Provides flexibility to meet specific metadata standards,
- Plans for making attribute customization more accessible through the UI.