OSCARS

Open Science Clusters' Action for Research & Society

PaN-Finder

Funded by the European Union

PaNOSC OSCARS can provide expertise and guidance for the PaNOSC outcomes (cf. https://leaps-wg3.desy.de/outcomes.html) :

- **AAI** - user authentication and authorisation i.e. login identity
- **Data Catalogues** – SciCat and Icat and other catalogues
- **Data Portals** – Domain specific portals + generic portals
- **Metadata** – NeXus, Sample and ontologies
- **Notebooks** – Jupyter notebooks
- **PaN Search** – searching for PaN data via the data.esrf.fr and other portals
- **Software** – developing, versioning, packaging and deploying software
- **Training** – training material platform
- **VISA VRE** – virtual research environment

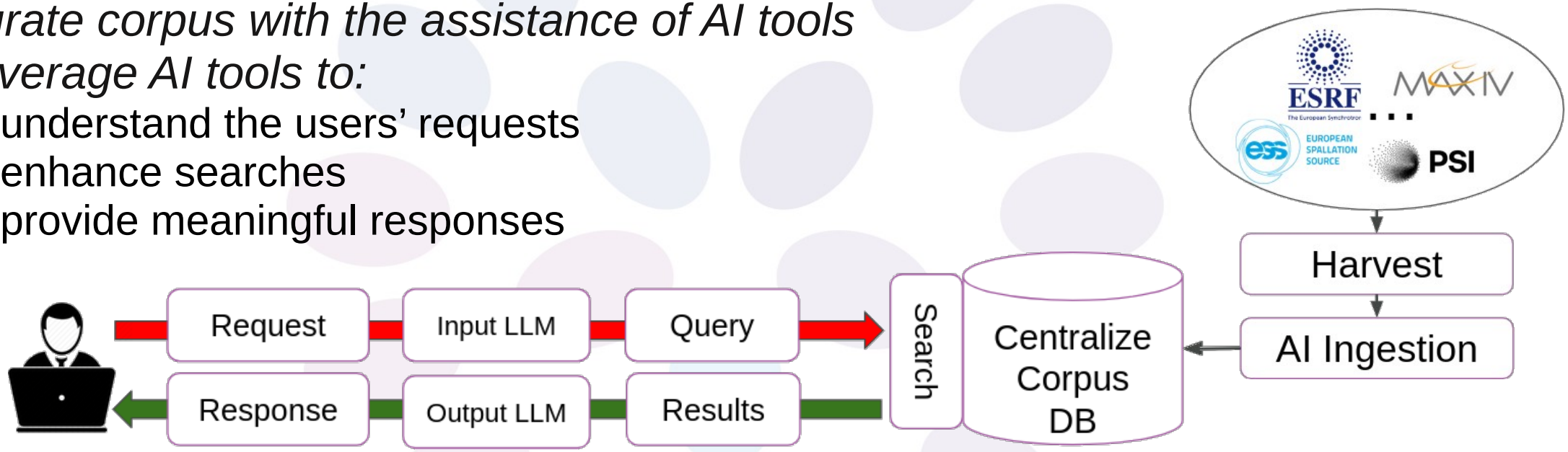
PaNOSC OSCARS Partners have identified the following overlap between the PaNOSC OSCARS projects and the PaNOSC outcomes:

	AI-SCOPE	AMBCAT	CDIF-4-XAS	CODEMETASOFT	Findable	HEFTIE	HiMAGNETOS	MatScatNet	Mc-REDD	mTess-X	OSPARK	PaN-Finder	SHARE	VISA
AAI	Y	Y			Y	Y	Y	Y		Y			Y	Y
Data Catalogues	Y	Y					Y	Y				Y	Y	
Data Portals	Y	Y						Y					Y	
Metadata	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	
Notebooks	Y					Y	Y	Y					Y	Y
PaN Search	Y	Y			Y		Y	Y	Y			Y	Y	
Software	Y			Y		Y	Y		Y	Y			Y	Y
Training	Y	Y		Y		Y	Y	Y	Y	Y	Y		Y	Y
VISA VRE	Y					Y		Y					Y	Y

What is PaN-Finder?

Develop an AI assisted prompt based search system

- *Harvest publicly available data and related metadata*
- *Curate corpus with the assistance of AI tools*
- *Leverage AI tools to:*
  - understand the users' requests
  - enhance searches
  - provide meaningful responses



Desired support from the PaNOSC cluster for your project.

Status of the PaN-Finder project?

PaN-Finder (POC)

Look for datasets where the publisher is MAX IV

Look for datasets where the publisher is ESS. | Look for datasets where the publisher is MAX IV and the beam's shape is ellipse and slit vertical is 0.05. | Document where the DOI is 10.15151/ESRF-ES-1317814821.

Find datasets from the Munich Crystallography BAG experiment conducted at the ID23-1 instrument between March 12 and July 27, 2018, where the resolution is less than 2.1 and the number of images is 2.

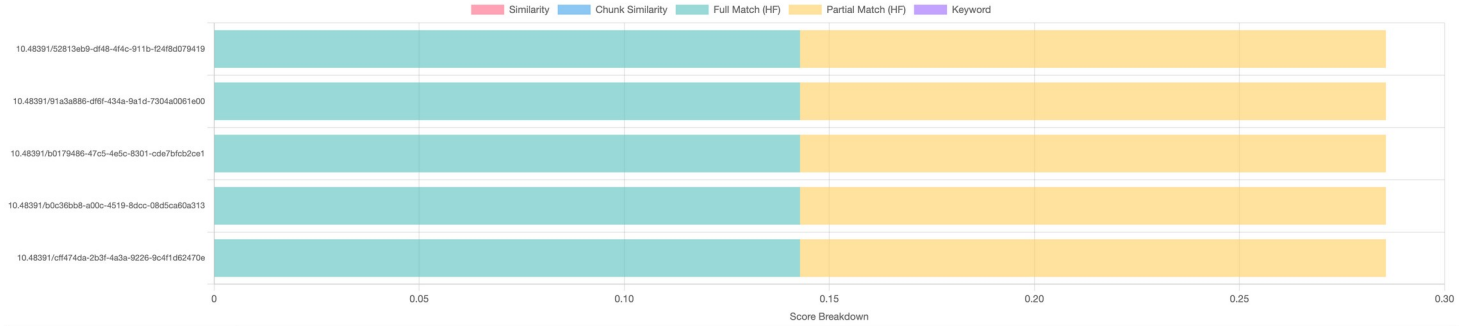
Look for research proposals involving the D50 T tomograph where the sample formula includes Si, O, K, Al, Na and the publication year is 2018. | Look for studies on solid polarizers for cold neutrons where the experimental energy is less than -19 A and the publication year is 2016 or 2017.

Look for research on magnetic diffuse scattering in CuMnO2 where the temperature is between 1.5 K and 300 K and the sample mass is 10,000. | Look for research about Crystal structure where the publication year is 2025.

Similarity K: 6 Chunk K: 6 Full-Match K: 6 Partial-Match K: 6 Keyword K: 10 Smaller value of rrf_k gives more weight to higher ranked items => RRF Score: 1.0 / (rank + k)

Search

DOI	Overall Score	Similarity Score	Chunk Similarity Score	Full Match Score (HF)	Partial Match Score (HF)	Keyword Score
10.48391/52813eb9-df48-4f4c-911b-f24f8d079419	0.2857142857142857	0.000000	0.000000	0.142857 ✔	0.142857	0.000000
10.48391/91a3a886-df6f-434a-9a1d-7304a0061e00	0.2857142857142857	0.000000	0.000000	0.142857 ✔	0.142857	0.000000
10.48391/b0179486-47c5-4e5c-8301-cde7bfcb2ce1	0.2857142857142857	0.000000	0.000000	0.142857 ✔	0.142857	0.000000
10.48391/b0c36bb8-a00c-4519-8dcc-08d5ca60a313	0.2857142857142857	0.000000	0.000000	0.142857 ✔	0.142857	0.000000
10.48391/cff474da-2b3f-4a3a-9226-9c4f1d62470e	0.2857142857142857	0.000000	0.000000	0.142857 ✔	0.142857	0.000000



Debug Key	Debug Value
Extracted Data (LLM)	{ "intention": "", "keywords": [], "filters": { "logic": "AND", "conditions": [{ "name": "publisher", "operator": "=", "value": "MAX IV" }] } } Search with Modified Data
LLM Response (Elapsed Time)	1.11s
Resolved Filter Names	**publisher**: [('publisher', 0.0), ('publicationYear', 0.4629048429187 9874), ('authors', 0.46653155489815534), ('creator', 0.5588360490149091), ('owner', 0.5660204133938727)]

Are the dependencies on the PaNOSC outcomes identified in the previous slide correct?

The dependencies identified are correct.

Data Catalogues

Metadata

PaN Search

Please elaborate, which PANOSC outcomes have you managed to adopt?

1.Data Catalogues

PaN Finder retrieve its data directly from each individual facility data catalogue leveraging the PaNOSC API.

2.Metadata

PaN Finder can utilize the techniques included in PaNET to enhance its search capabilities

3.PaN Search

PaN Finder will be integrated in the current PaNOSC Data Portal

Desired support from the PaNOSC cluster for your project.

Please discuss the PaNOSC outcomes do you need help in:

1. ***Data Catalogues***
   Expand the information retrievable through the PaNOSC API or OAI-PMH protocol, so PaN Finder can operate on more data.
2. ***Metadata***
   Expand PaNET with additional definitions, synonyms and keywords that can be used to expand search functionalities in PaN Finder.
3. ***PaN Search***
   Design a user interface to enhance user experience and collect user feedback

Describe how your project's developments contribute to the needs and challenges of the photon and neutron community.

1. PaNOSC single search portal

The PaN Finder project aims to become the single search portal for all the open data of the PaNOSC community.

2. Intuitive and Flexible search

Find data in an intuitive and flexible way with minimal knowledge of the variety present across facilities.

3. Data/Metadata trends

Discover trends in the data and metadata fostering the creation of data curation best practices and meaningful feedback on each facility's data

How will you encourage users to adopt your solutions? Are there any partnerships, funding plans, or integration efforts that will help sustain the outcomes beyond the project's duration?

Adoption and visibility will be fostered by:

1. integrating the PaN Finder in the current PaNOSC Data Portal.
2. integrating the PaNOSC Data Portal into the PaNOSC EOSC Node.
3. transfer all PaNOSC open data searches from each facility to the common portal


OSCARS
Thank you