Towards Open Science in the Photon and Neutron Community

Oscar Matilla

ALBA Synchrotron

19-09-2023

The Photon and Neutron Community

http://dx.doi.org/10.1016/j.elspec.2013.12.007

The European Photon and Neutron Community

Photon – LEAPS

https://leaps-initiative.eu

EGI

eli delivery consortium

CERIC
Central European Research Infrastructure Consortium

Neutron - LENS

https://lens-initiative.org

The European Photon Community

LEAPS key figures:

19 facilities - **16** institutions - **10** countries

> 300 operating End Stations

➢ **1.000.000** h beamtime /year
Excellence-driven access free of charge

> 5.000 publications/year

> 15 spin off companies

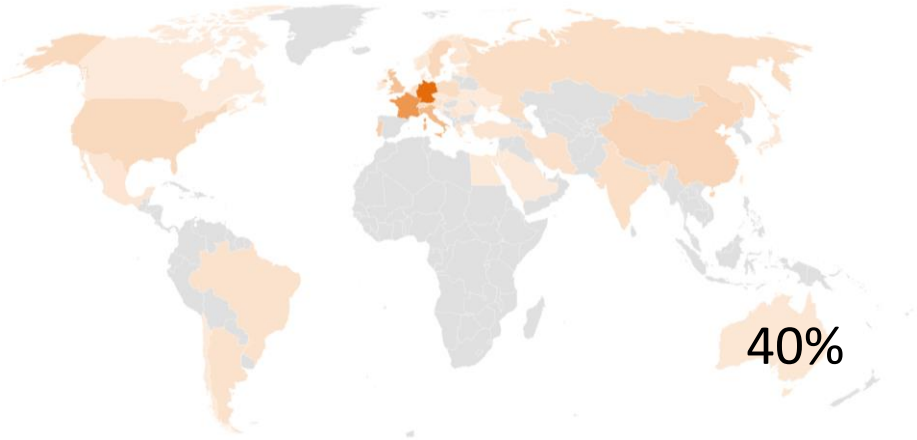
> 35.000 users from all EU & beyond
researchers from all research area

ALBA Synchrotron

National Research Infrastructure with 50% national + 50% regional funding: Ministerio de Ciencia e Innovación and Department de Recerca i Universitats.

SR Parameter	Value
Energy	3 GeV
Circumference	269 m
Emittance	4.4 n mrad
Current	250 mA
Rf frequency	500 MHz
# cavities	6
Long straights	4 (8 m)
Short straights	12 (4 m)

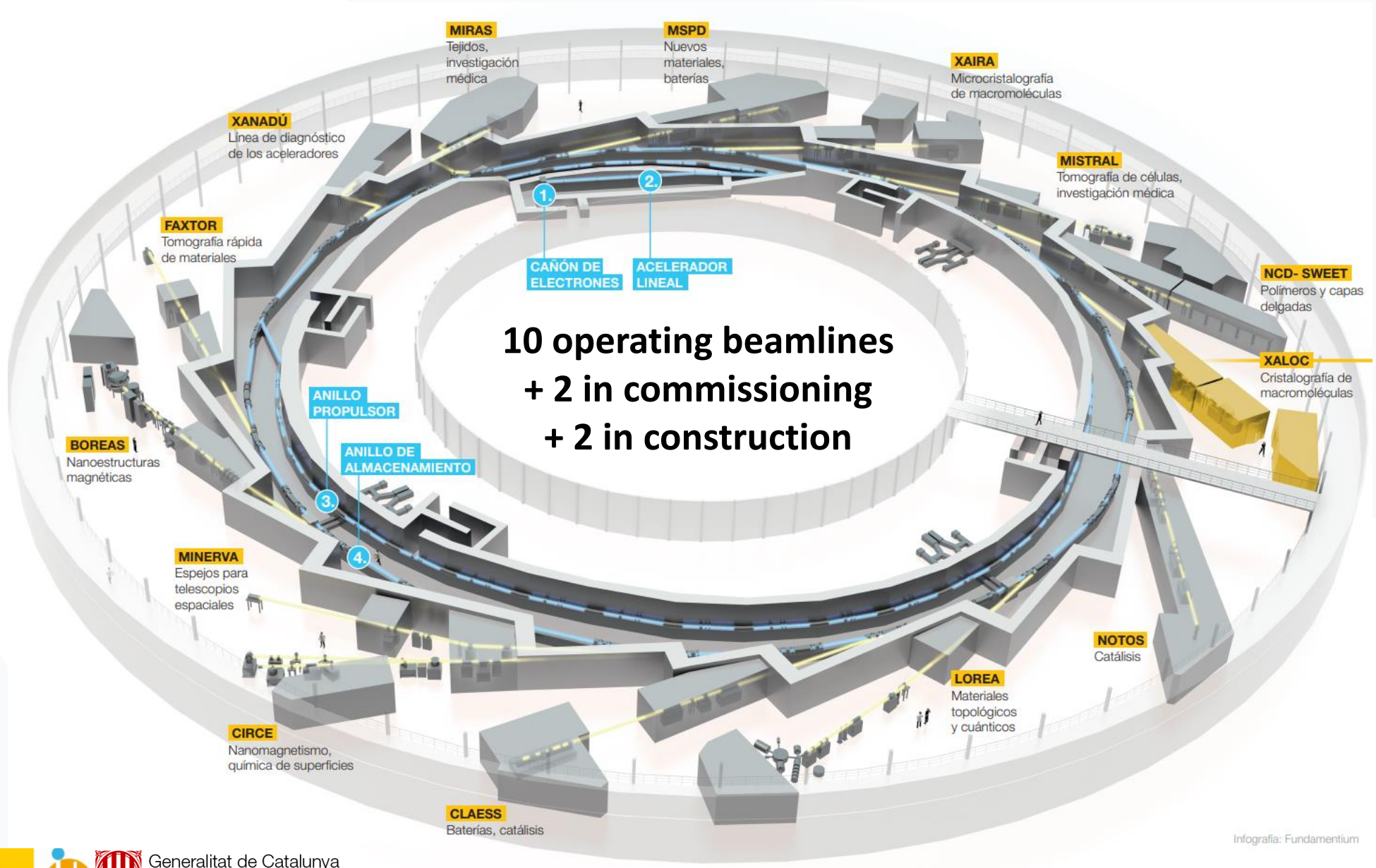
60%



40%

Generalitat de Catalunya
Departament de Recerca i Universitats

ALBA Synchrotron

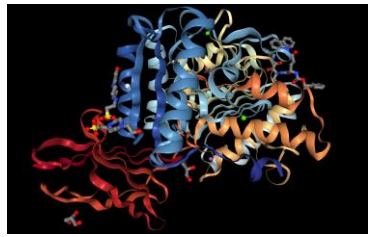


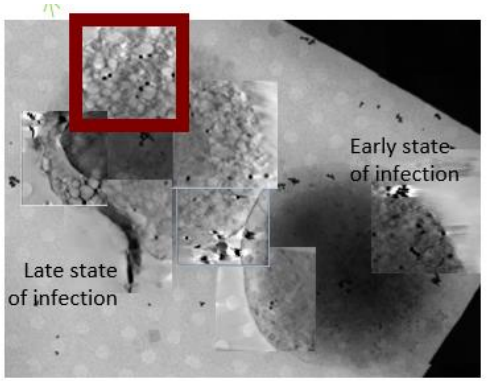
ALBA Synchrotron Research Infrastructure

Three Scientific Sections, three main research lines

Life Science

From the protein, to cell, to tissues

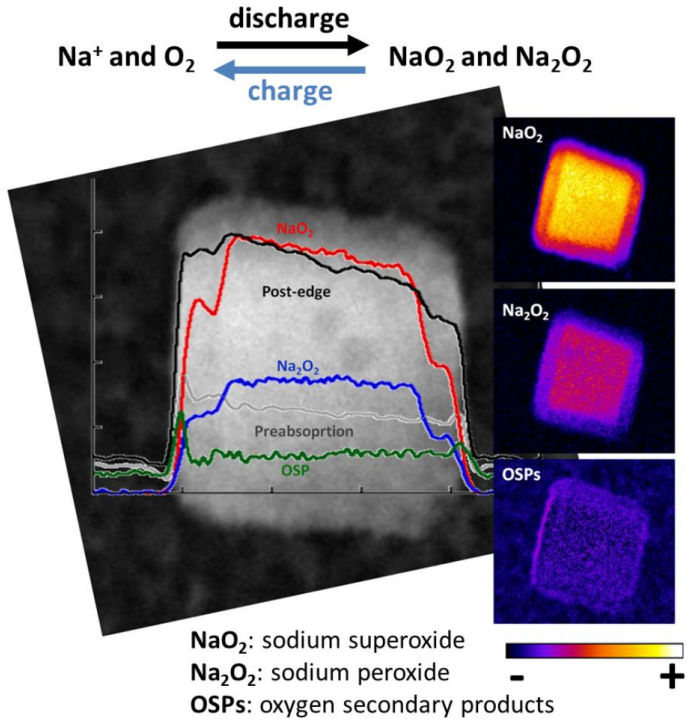




Cell infected by covid-19

Chemistry and Material Science

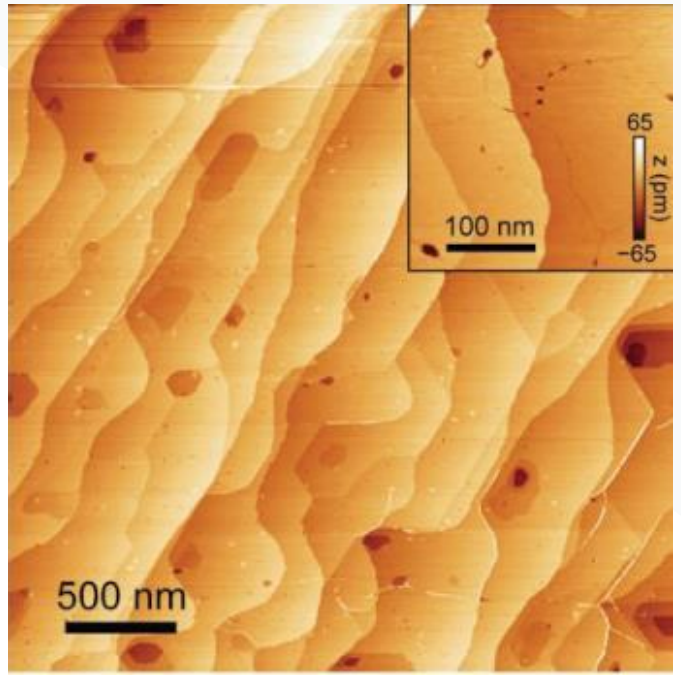
Energy material, catalysts, environment



Battery developments

Electronic and Magnetic Structure of Matter

Advanced materials



Nanomaterials for data storage

ALBA Synchrotron

TABLE 2-1: **MAIN TECHNIQUES OFFERED BY ALBA BEAMLINES**

PORT AND NAME	SCIENTIFIC SECTION	MAIN TECHNIQUES AND STATUS
BL01 - MIRAS	Life Science	*Infrared Spectroscopy & Microscopy – in operation*
BL06 - XAIRA	Life Science	*Macromolecular Microcrystallography – in commissioning; in operation in 2024*
BL09 - MISTRAL	Life Science	*Soft X-ray Microscopy - in operation*
BL13 - XALOC	Life Science	*Macromolecular Crystallography - in operation*
BL31 - FAXTOR	Life Science	*Fast X-ray Tomography and Radioscopy Beamline - being installed; starting operation in 2024*
BL20 - LOREA	Electronic & Magnetic Structure of Matter	*Angle Resolved Photoemission Spectroscopy – in operation*
BL24 - CIRCE	Electronic & Magnetic Structure of Matter	*Photoemission Spectroscopy and Near Ambient Pressure Photoemission – in operation*
BL29 - BOREAS	Electronic & Magnetic Structure of Matter	*Resonant Absorption and Scattering – in operation*
BL04 - MSPD	Chemistry & Material Science	*Materials Science and Powder Diffraction – in operation*
BL11 - NCD-SWEET	Chemistry & Material Science	*Non-Crystalline Diffraction – Small/wide Angle X-ray Scattering – in operation*
BL15 - 3SBAR	Chemistry & Material Science	*Surface Spectroscopy and Structure at 1 bar – in construction; starting operation in 2026*
BL16 - NOTOS	Chemistry & Material Science	*Absorption, Diffraction, Instrumentation innovation and development – in operation*
BL22 - CLÆSS	Chemistry & Material Science	*Core Level Absorption & Emission Spectroscopies – in operation*
BL25 - MINERVA	Instrumentation & optics	*Metrology and instrumentation – in operation in 2023*

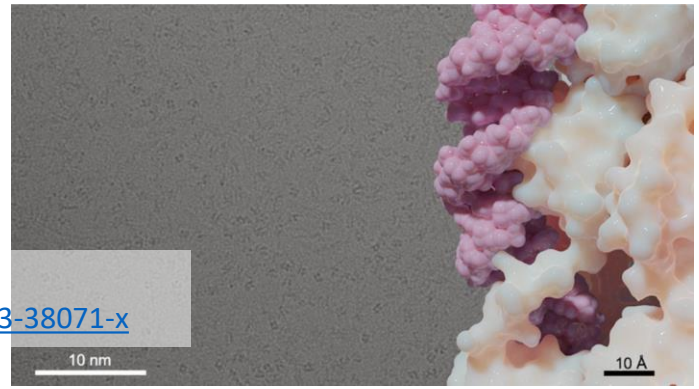
Joint Electron Microscope Center at ALBA (JEMCA)

50% Funded through Catalan ERDF and 50% co-funded by different partners

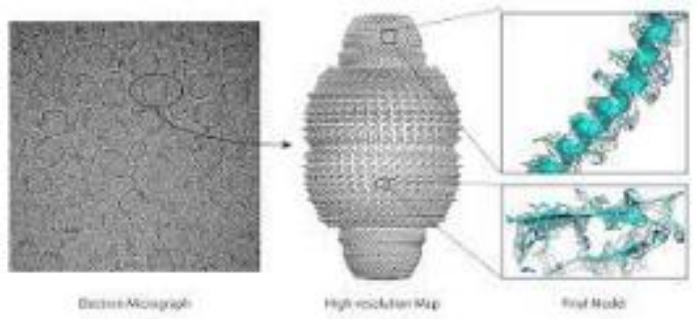
Life science

Cryo-EM

- 200 kV Glacios cryo-TEM
- In operation . Receiving users (overbooking >2)



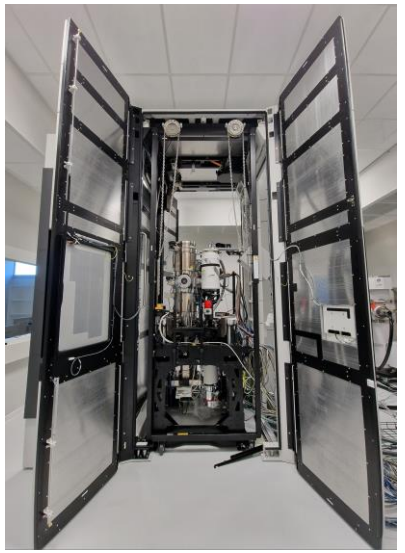
Nature Commun
. DOI: 0.1038/s41467-023-38071-x



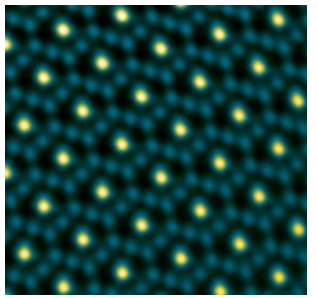
Material Science

TEM

- 60-300 keV Spectra (S)TEM + New EM to be installed in 1 year (InCAEM project)
- First external users in following weeks

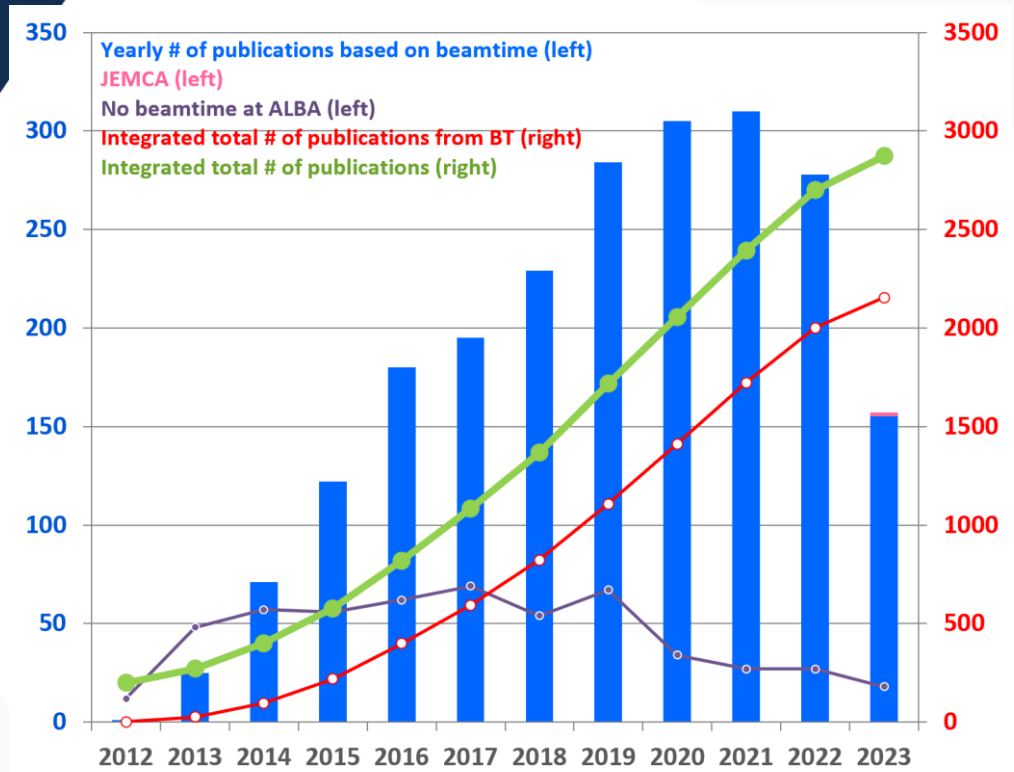


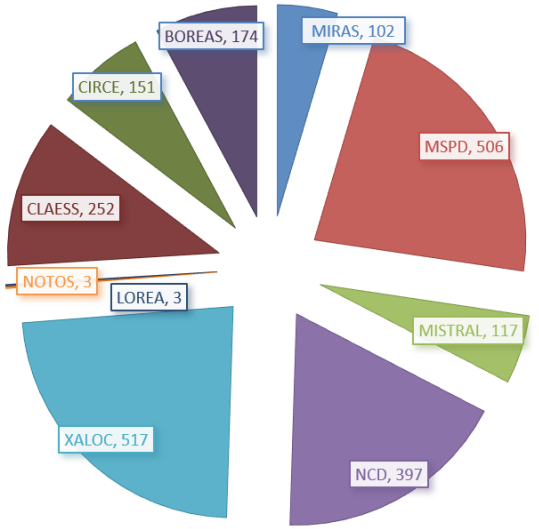
Atomic resolution aberration corrected HAADF STEM images of one of the catalyst nanoparticles and a zoom out of the Co_2FeO_4 cubic spinel structure

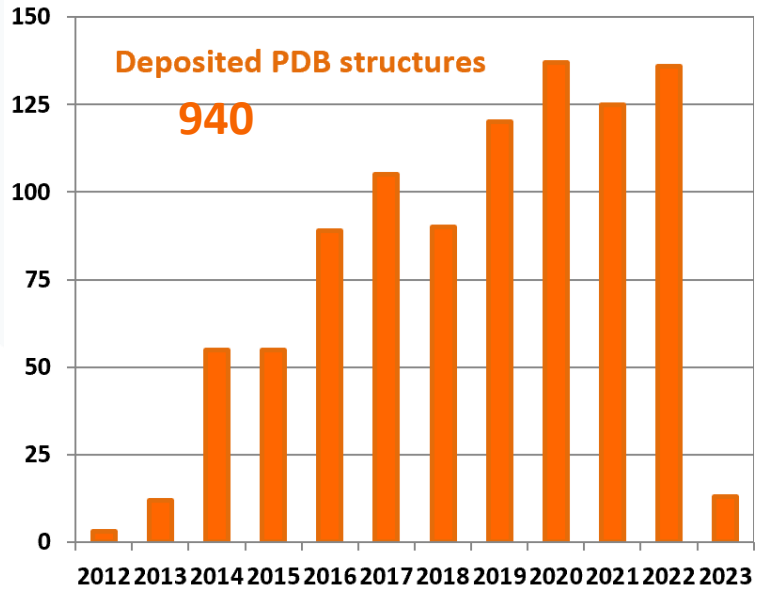


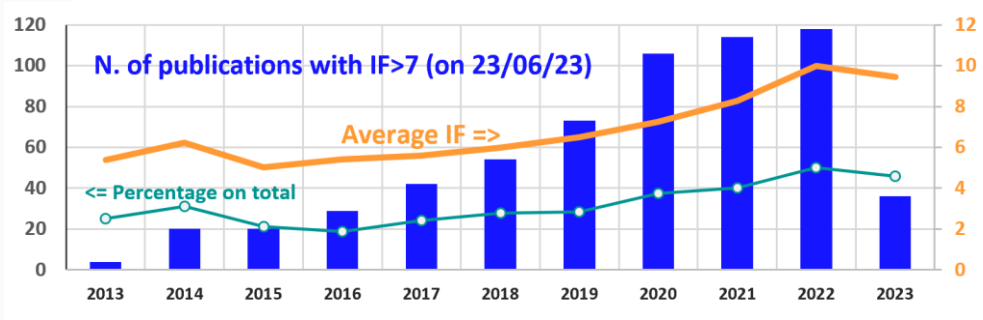
Gobierno de España · Ministerio de Economía, Industria y Competitividad
Generalitat de Catalunya · Departament de Recerca i Universitats
Unió Europea · Fons Europeu de Desenvolupament Regional

ALBA scientific productivity









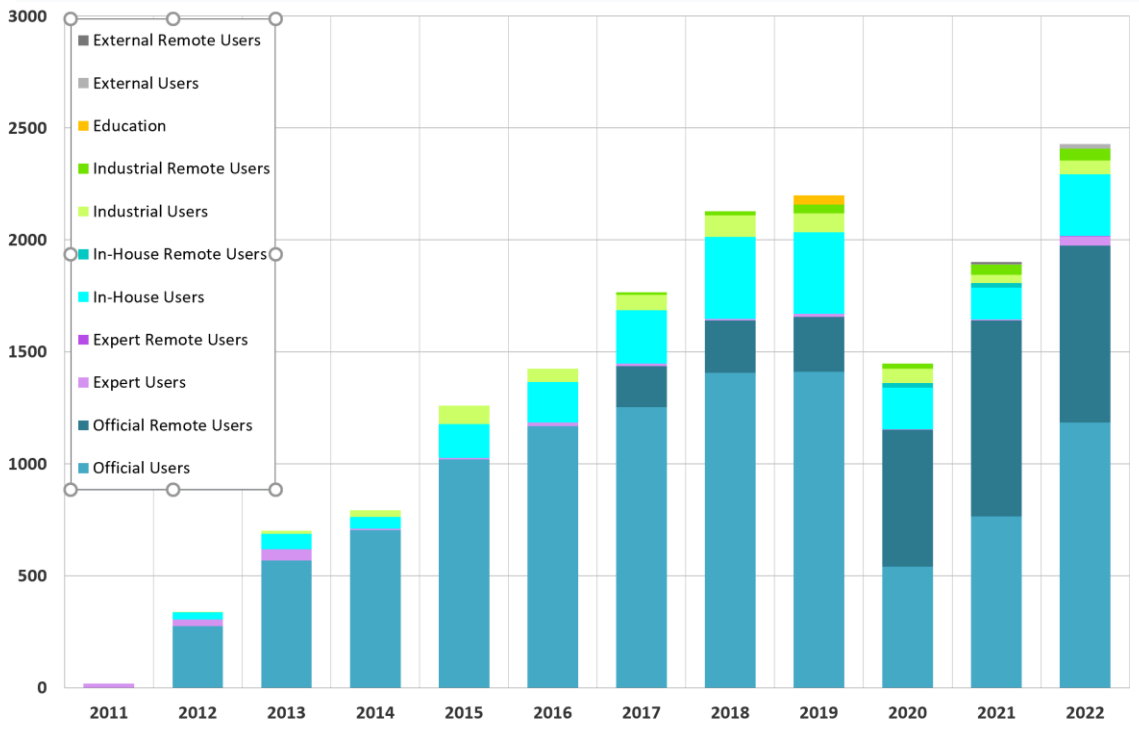
Publications per BL

2020/2021: average # of publications per op. BL > 38
2021: <IF> = 8.3; 41% of publications with IF>7
2022: <IF> = 10; 50% of publications with IF>7

ALBA scientific productivity

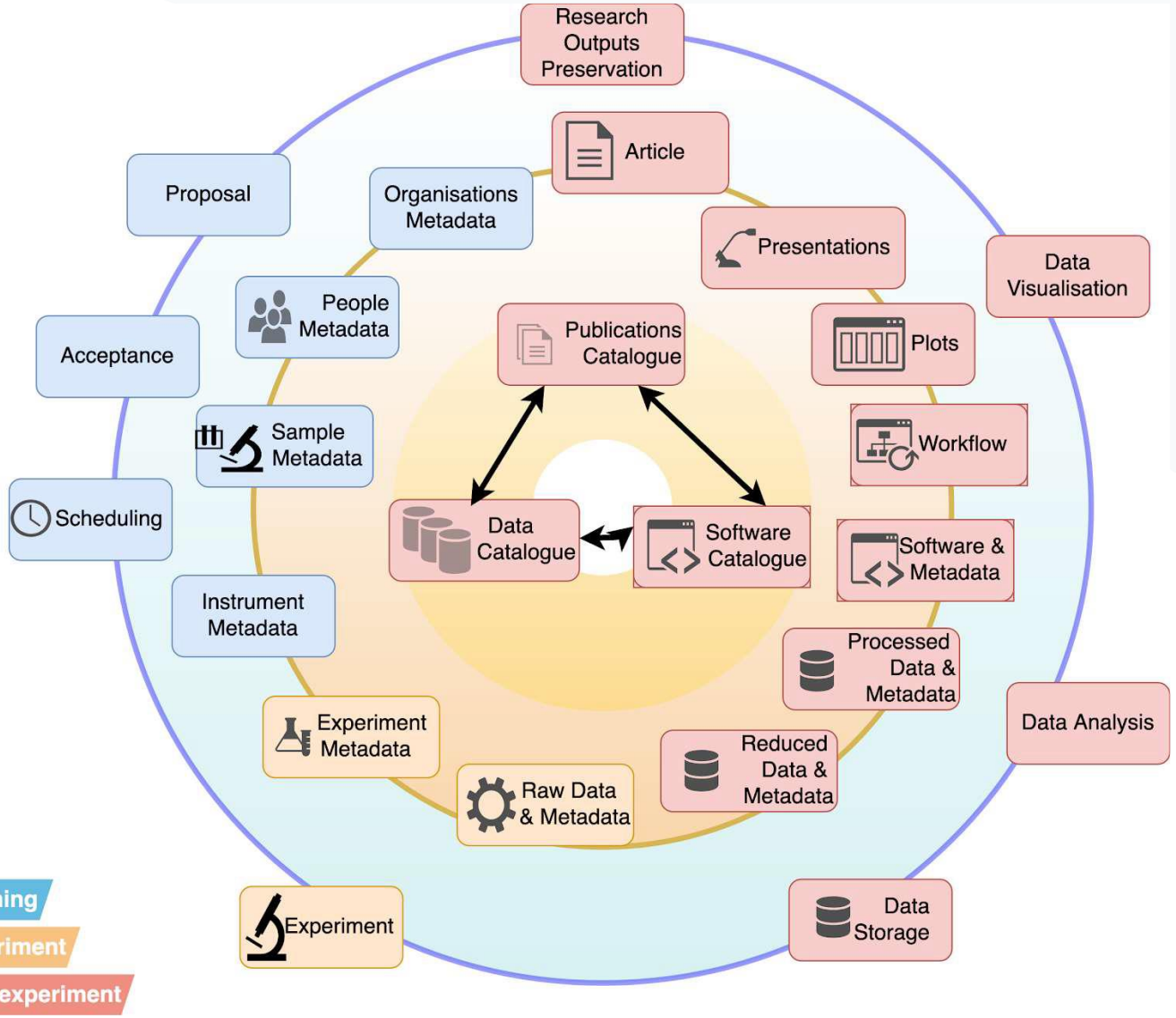
ALBA key figures:

- **+450 yearly experiments**
- **+2800 publications**
- **<IF>2022 = 10**
- +900 Proteins in PDB
- + 2400 yearly user visits
- +500 industrial experiments



Yearly users

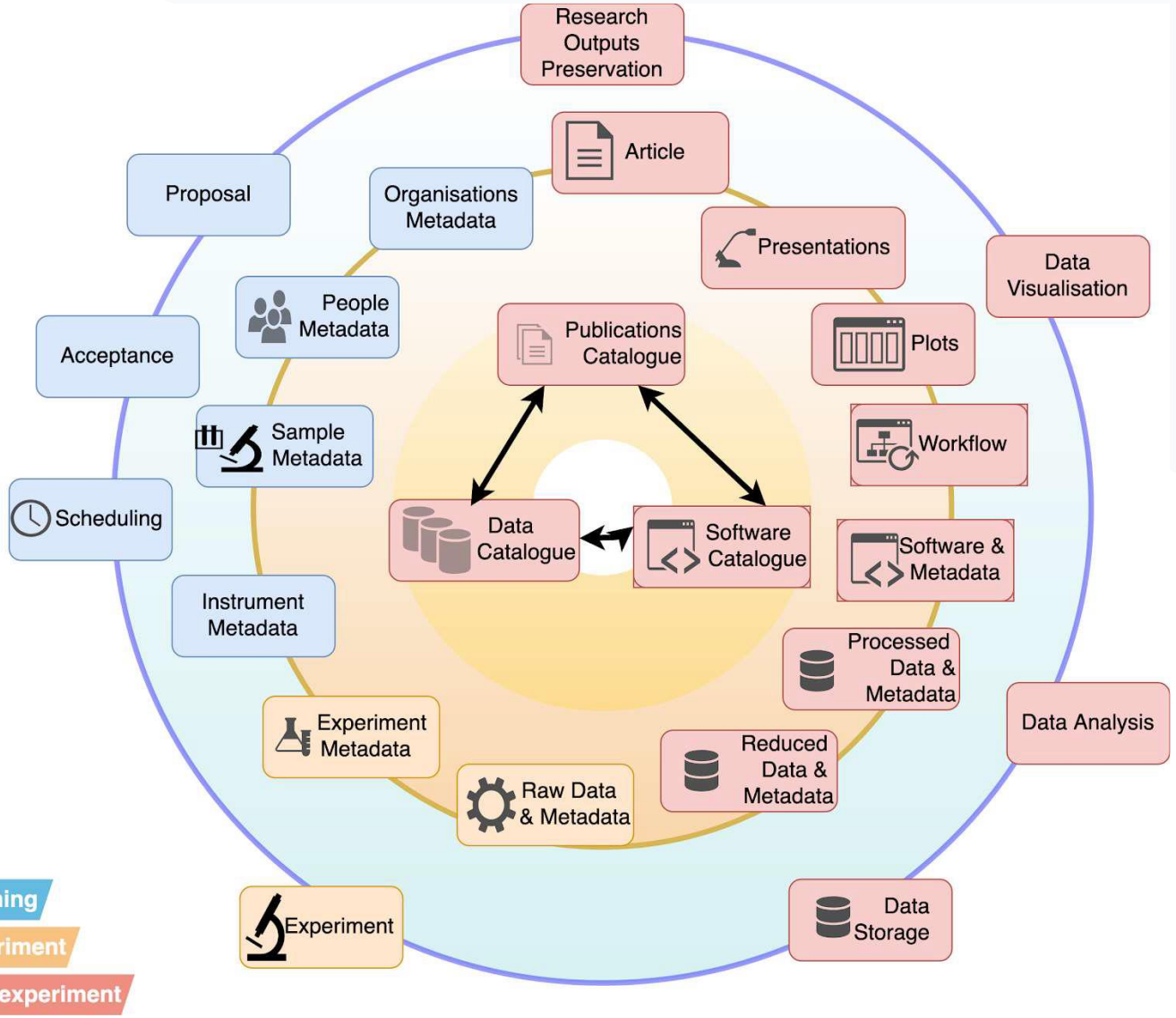
Life Cycle of a PaN experiment



Life Cycle of a PaN experiment

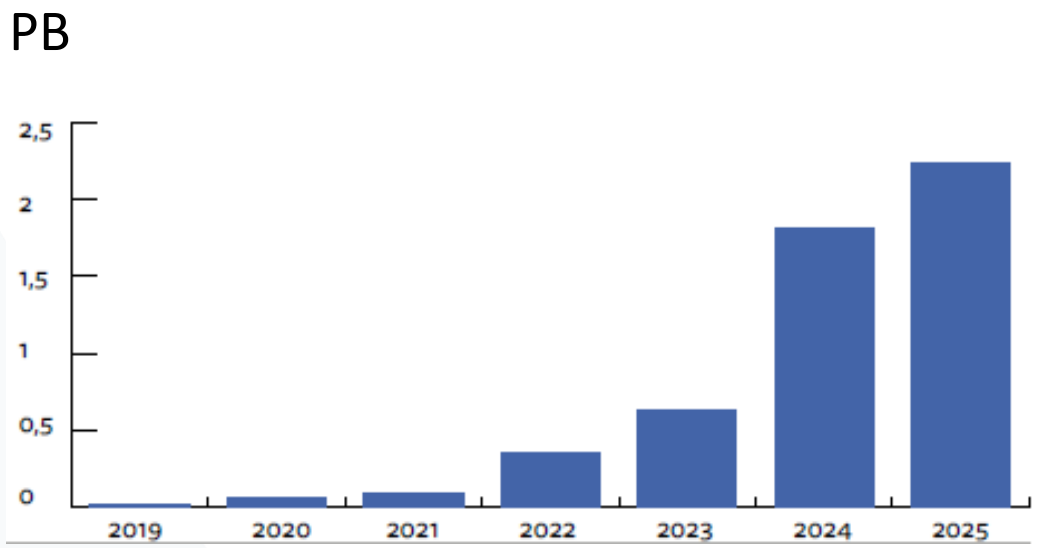
- There are many steps between the user sending a first proposal and achieving publication.
- Ensuring the data FAIRness requires data access and data integrity of every step.
- This role cannot be delegated to our users.
- RIs have to take a step forward and assume not only the custodian role as until now, but also data stewardship.
- Our users (scientists) must keep always the ownership and the IP of the data.

Life Cycle of a PaN experiment

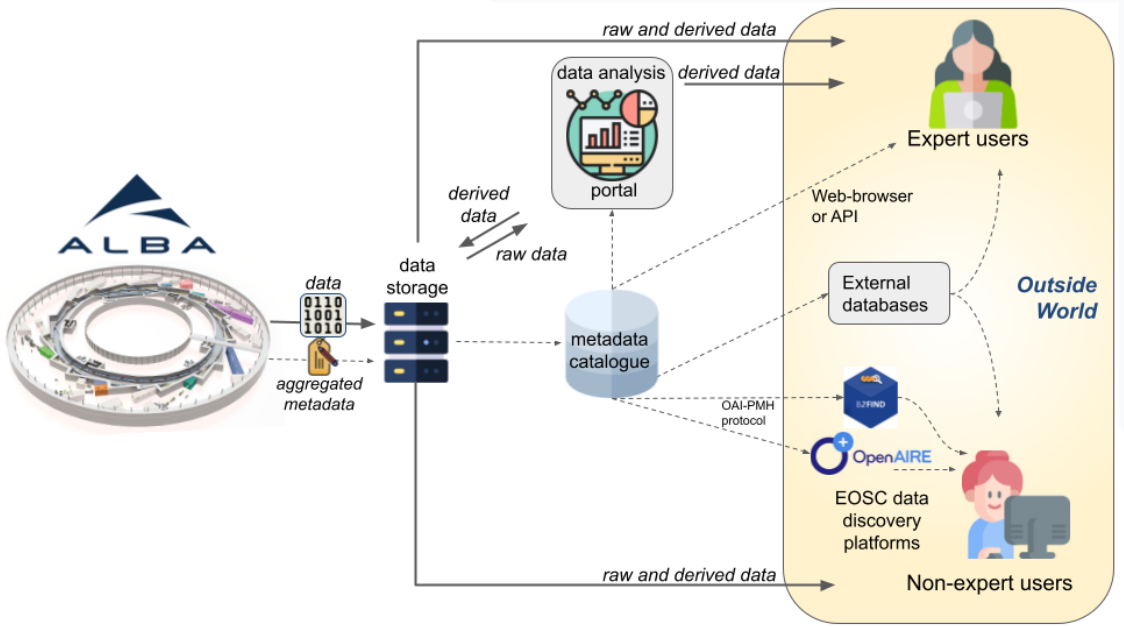


Life Cycle of a PaN experiment

- … But things are becoming more challenging.
- The volume of data generated by new data-intensive beamlines is too large for users to handle at their premises.



ALBA annual raw data production



Users + Facilities feel the weight of data

Users

- Huge data volumes
  Tera → Petabytes
- Sample metadata
- Raw data quality
- Data processing
- Data exporting
- FAIR data

Facilities

- Huge data volumes
  Peta → Exabytes
- Data acquisition
- Metadata collection
- Data curation
- Data archiving
- FAIR data

LEAPS Data Strategy

Three pillars:

1. **Increasing efficiency of experiments**
2. **Open Science**
3. **Sustainability of solutions**

https://doi.org/10.1140/epjp/s13360-023-04189-6

ExPANDS & PaNOSC outcomes

Photon – LEAPS

https://leaps-initiative.eu

egi

eli delivery consortium

CERIC
Central European Research Infrastructure Consortium

Neutron - LENS

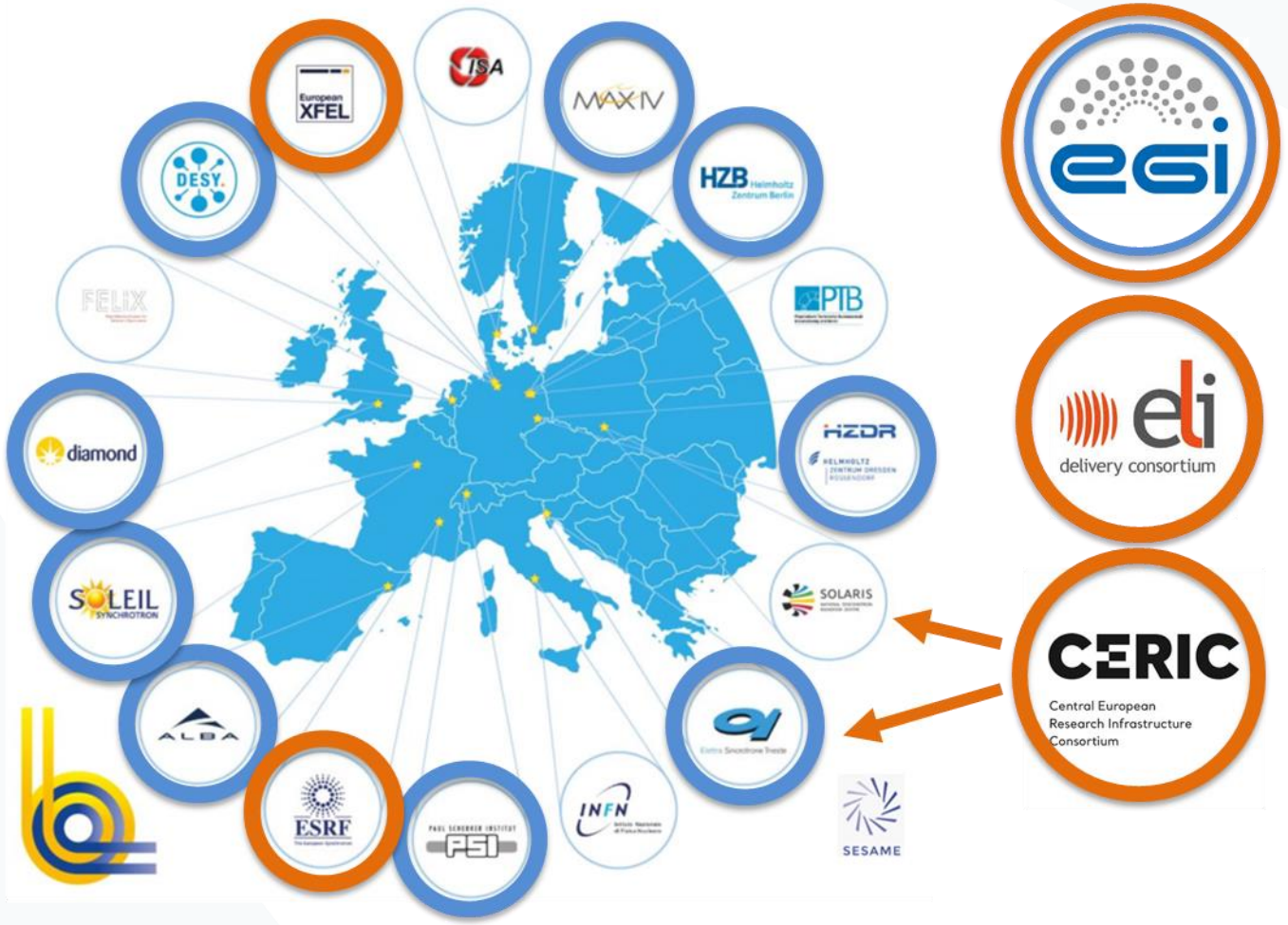
https://lens-initiative.org

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

ExPANDS & PaNOSC outcomes

Photon – LEAPS

https://leaps-initiative.eu



Neutron - LENS

https://lens-initiative.org

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaNOSC + ExPaNDS - 10 Main Achievements

1. **FAIR data policy** and **Data Management Plan (DMP)**
   - We will use Active DMP:
     - An automatic DMP will be generated for every proposal
     - 50 of the 82 questions are automatically filled in.
     - Users can use the DMP for satisfy funders requirements.
     - DMP is the way to ensure that users can manage their data.
2. Standardised metadata (**Nexus/HDF5**)
3. Guidelines for **FAIR data** self-evaluation

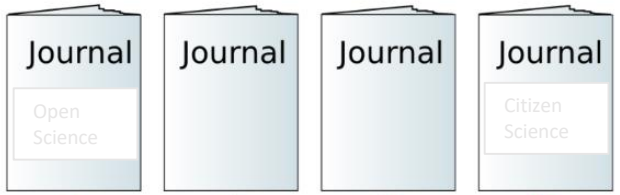
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

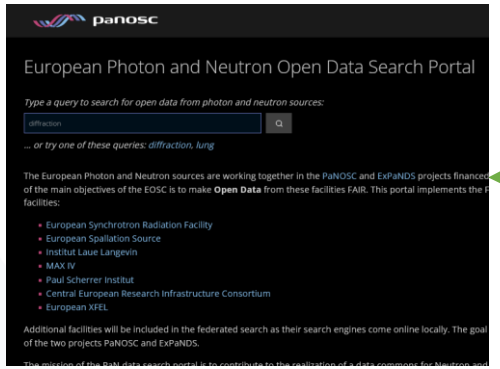
PaNOSC + ExPaNDS - 10 Main Achievements

1. **FAIR data policy** and **Data Management Plan (DMP)**
   - We will use Active DMP:
     - An automatic DMP will be generated for every proposal
     - 50 of the 82 questions are automatically filled in.
     - Users can use the DMP for satisfy funders requirements.
     - DMP is the way to ensure that users can manage their data.
2. Standardised metadata (**Nexus/HDF5**)
3. Guidelines for **FAIR data** self-evaluation
4. **Federated search API** for PaN data catalogues
5. **Open Data portal** for searching + downloading data

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

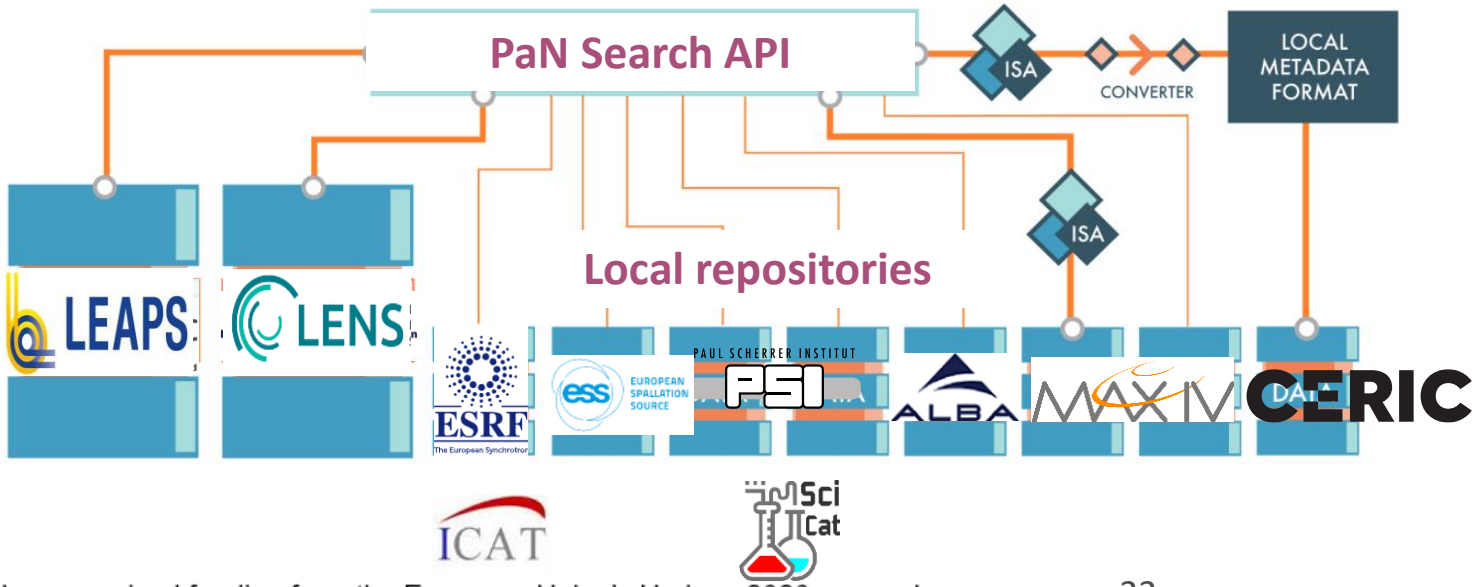
Open Data from federated PaN Repositories





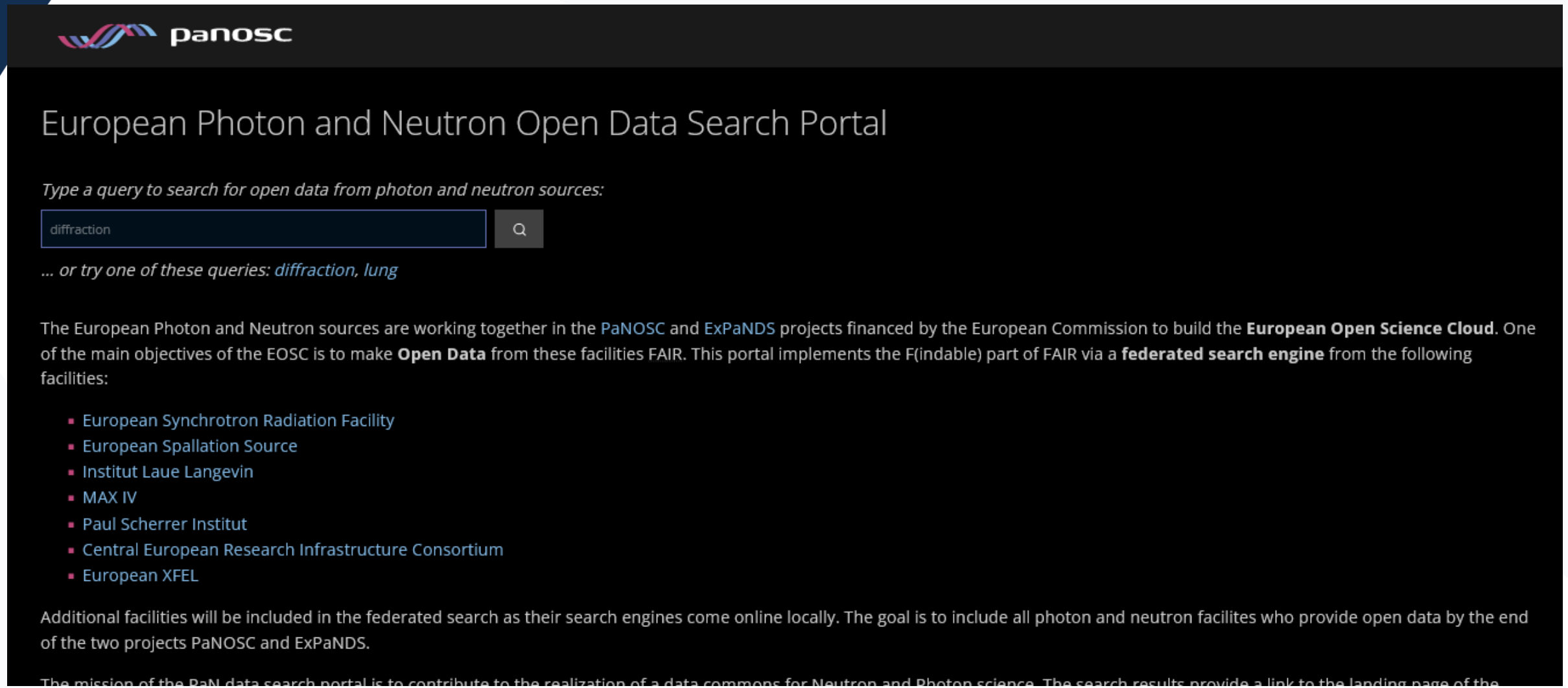
PaN Open Data Repository

PaN Open Data Portal
https://data.panosc.eu



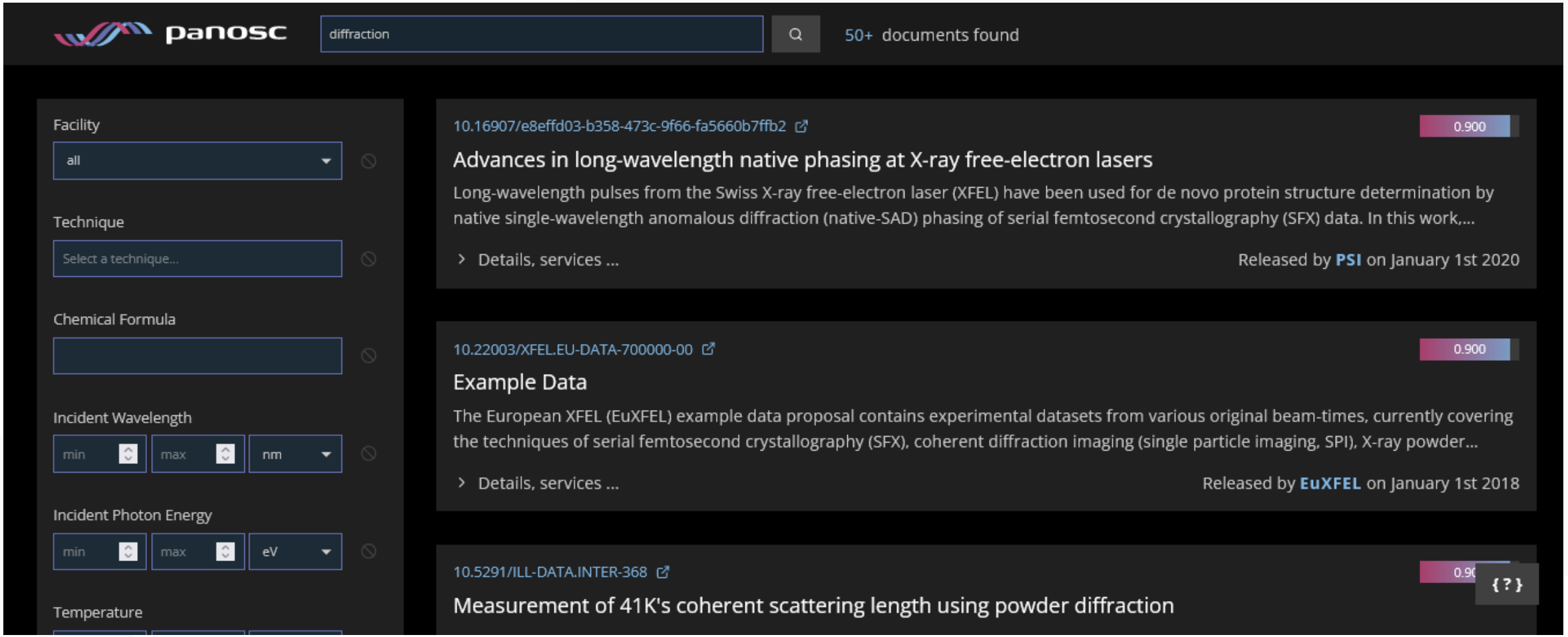
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaN Data Portal - https://data.panosc.eu



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaN Data Portal - https://data.panosc.eu



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaNOSC + ExPaNDS - 10 Main Achievements

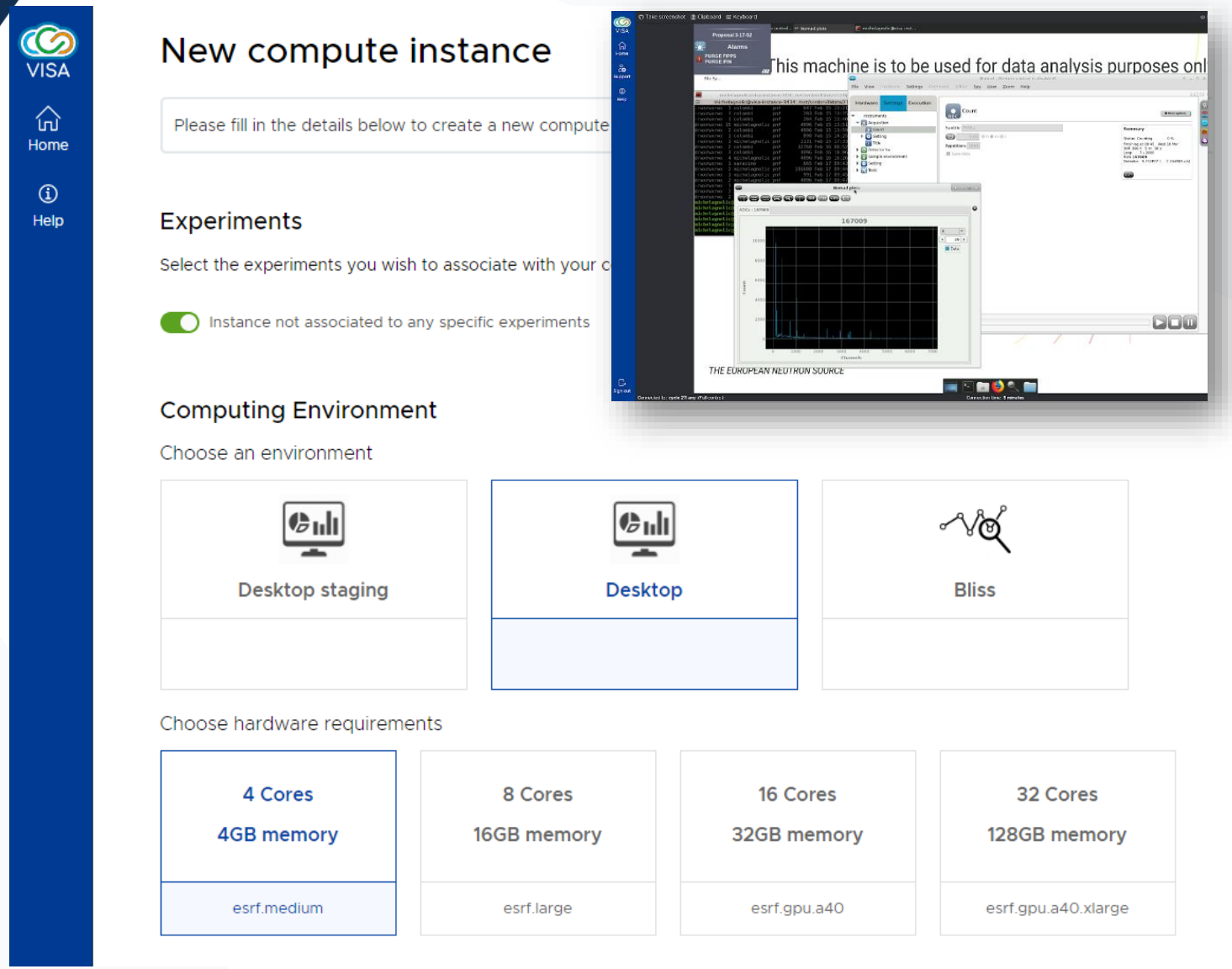
6. Community **AAI UmbrellaId**
   - **https://umbrellaid.org**

7. **JupyterLab notebooks** and **Nexus/HDF5** files visualisation

8. **Remote data analysis** with VISA + data analysis pipelines

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

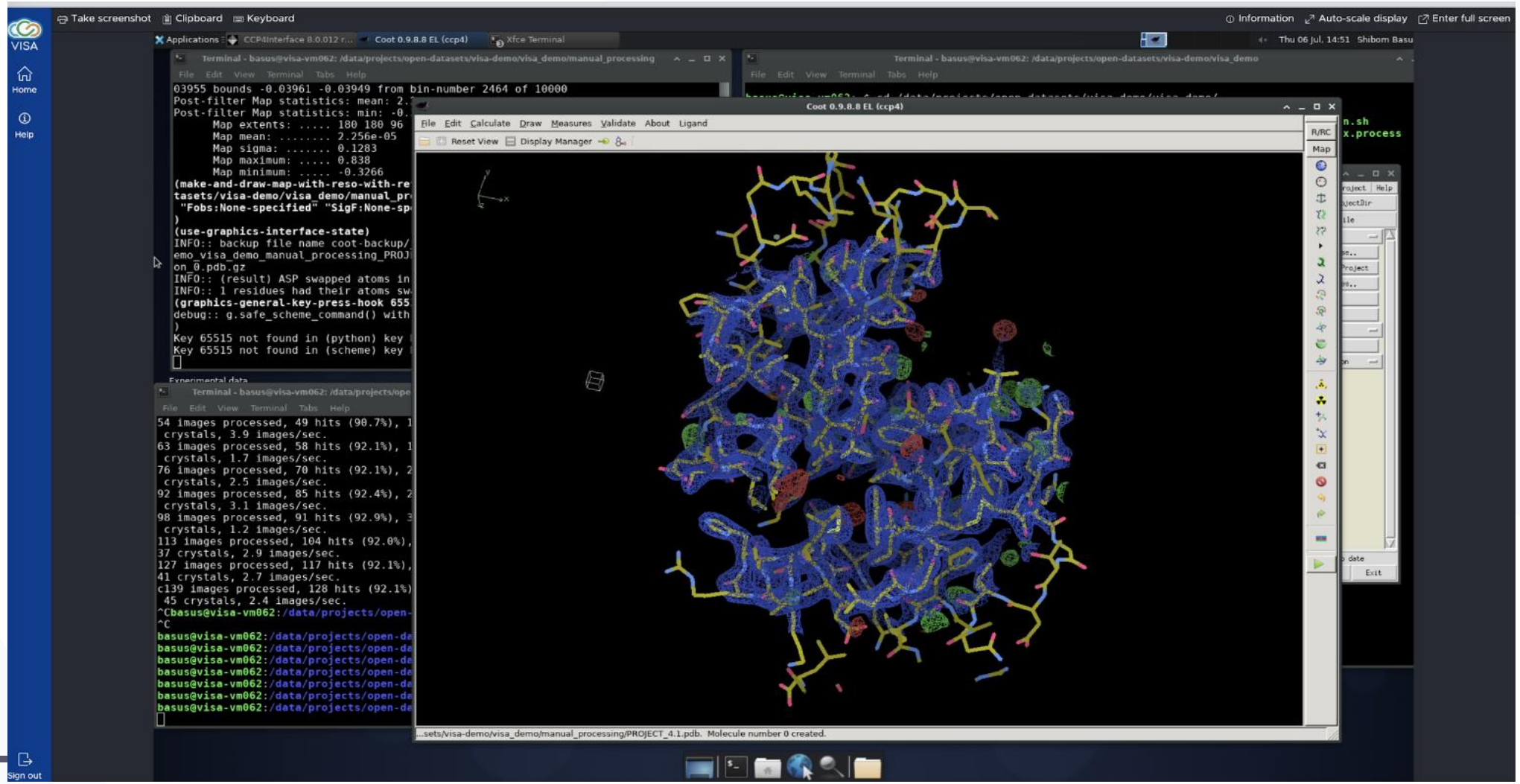
VISA - Remote Data Processing/analyses platform



- Infrastructure for remote data processing / analysis
- Users dedicated VM
- Access to data
- Access to Provisioning of scientific SW using CVMFS and Containers
- Access to the GPUs, HPC cluster
- Infrastructure based on OpenStack
- Development led by ILL in the scope of the PaNOSC project

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Example of DaaS using VISA



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

PaNOSC + ExPaNDS - 10 Main Achievements

6. Community **AAI UmbrellaId**
   - **https://umbrellaid.org**

7. **JupyterLab notebooks** and **Nexus/HDF5** files visualisation

8. **Remote data analysis** with VISA + data analysis pipelines

9. **Simulation** software for simulating experimental data (ViNYL)

10. **PaN-learning** platform (pan-learning.org)

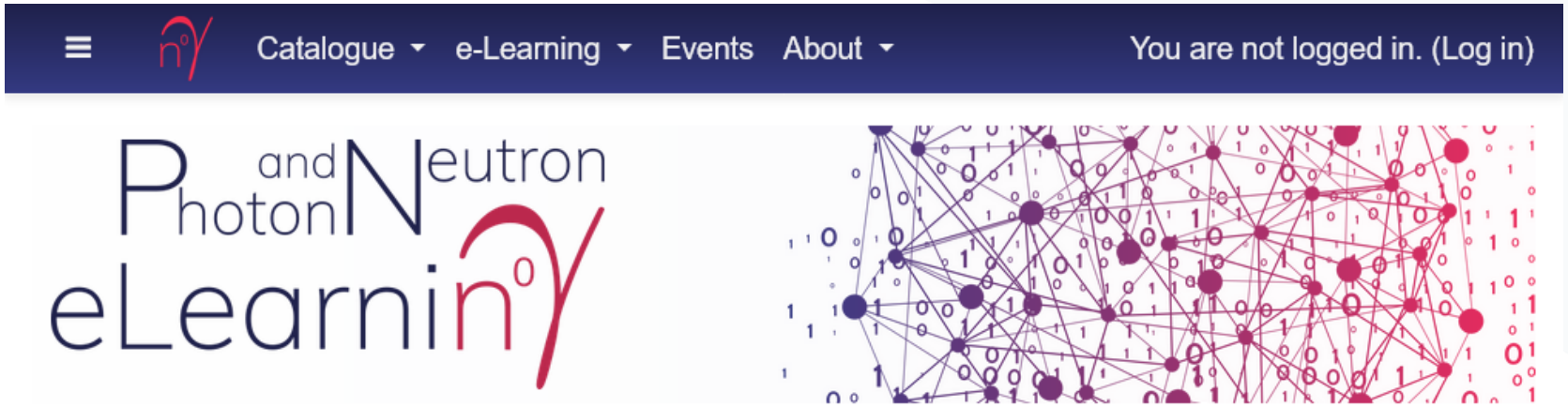
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Open Training - courses on Pan-learning.org

ALBA

Achieving100% Open Educational Resources:

1. *Publish training material on pan-training.org*
2. *Develop learning material on pan-learning.org*



https://e-learning.pan-training.eu/moodle/

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

panosc
photon and neutron open science cloud

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Workshop on leveraging Open Data for ML

https://indico.synchrotron-soleil.fr/event/67/

During the last decade, most European Photon and Neutron (PaN) facilities have adopted **open data policies**, making data available for the benefit of the entire scientific community. At the same time, **machine learning** (ML) is seen as an essential tool to address the exponential growth of data volumes from PaN facilities.

Evolution of Synchrotron Radiation Sources

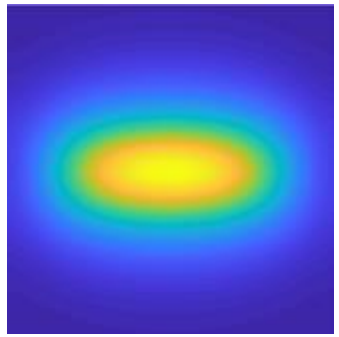
Increase in Brilliance

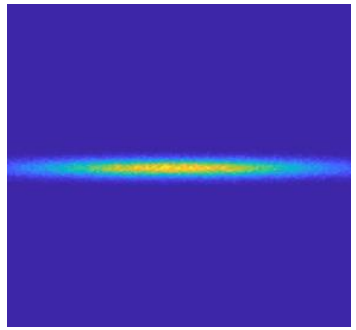
Increase in Resolution Power

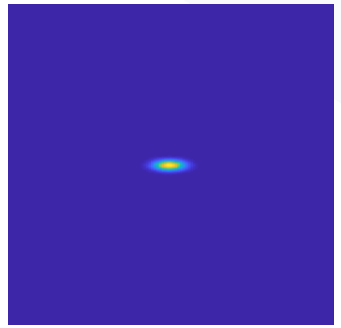
Increase in Coherence

Increase in Data management capacity

Representation of photon beam on sample →







1960 1970 1990 2020

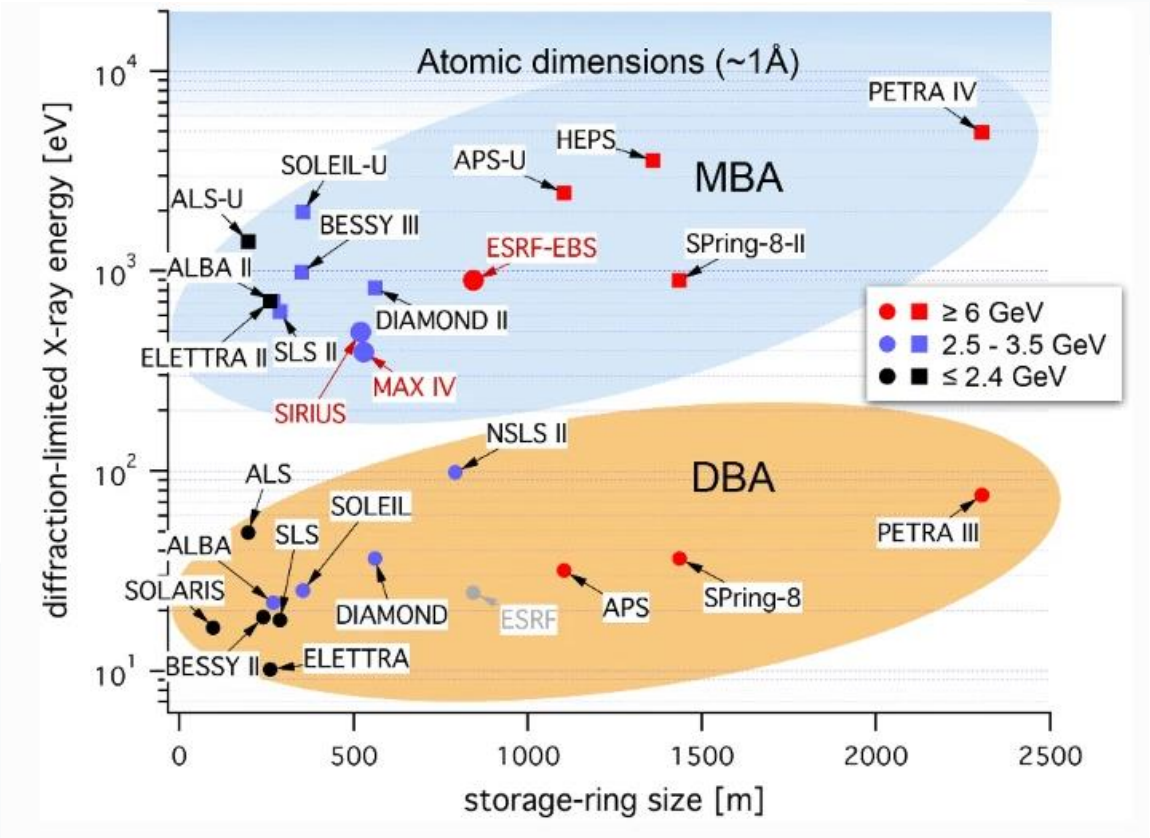
1st Gen

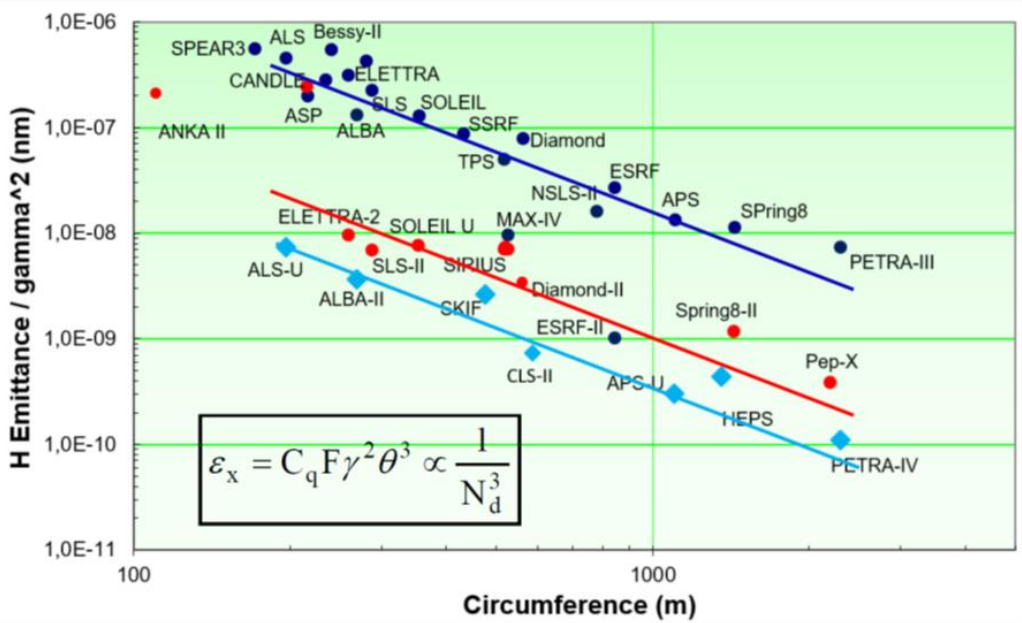
2nd Gen

3rd Gen

4th Gen

Towards 4th Generation light sources

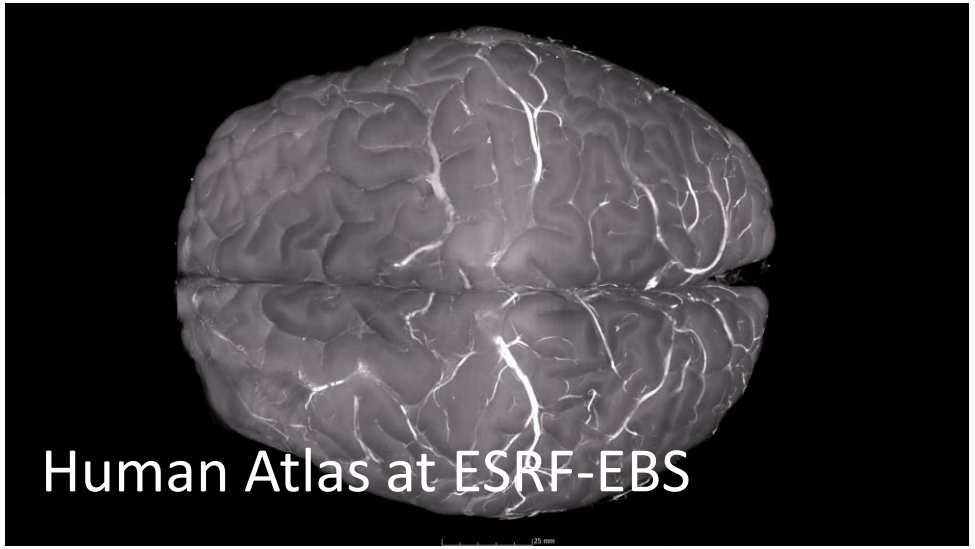


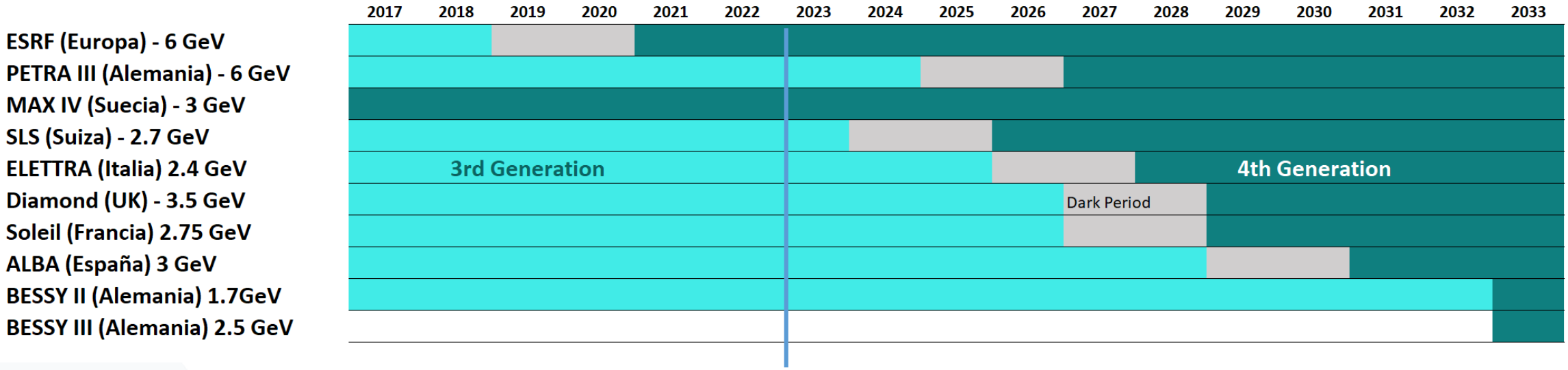


Abela, R., Biscari, C., Daillant, J. *et al.* The European strategy for accelerator-based photon science. *Eur. Phys. J. Plus* **138**, 355 (2023). https://doi.org/10.1140/epjp/s13360-023-03947-w

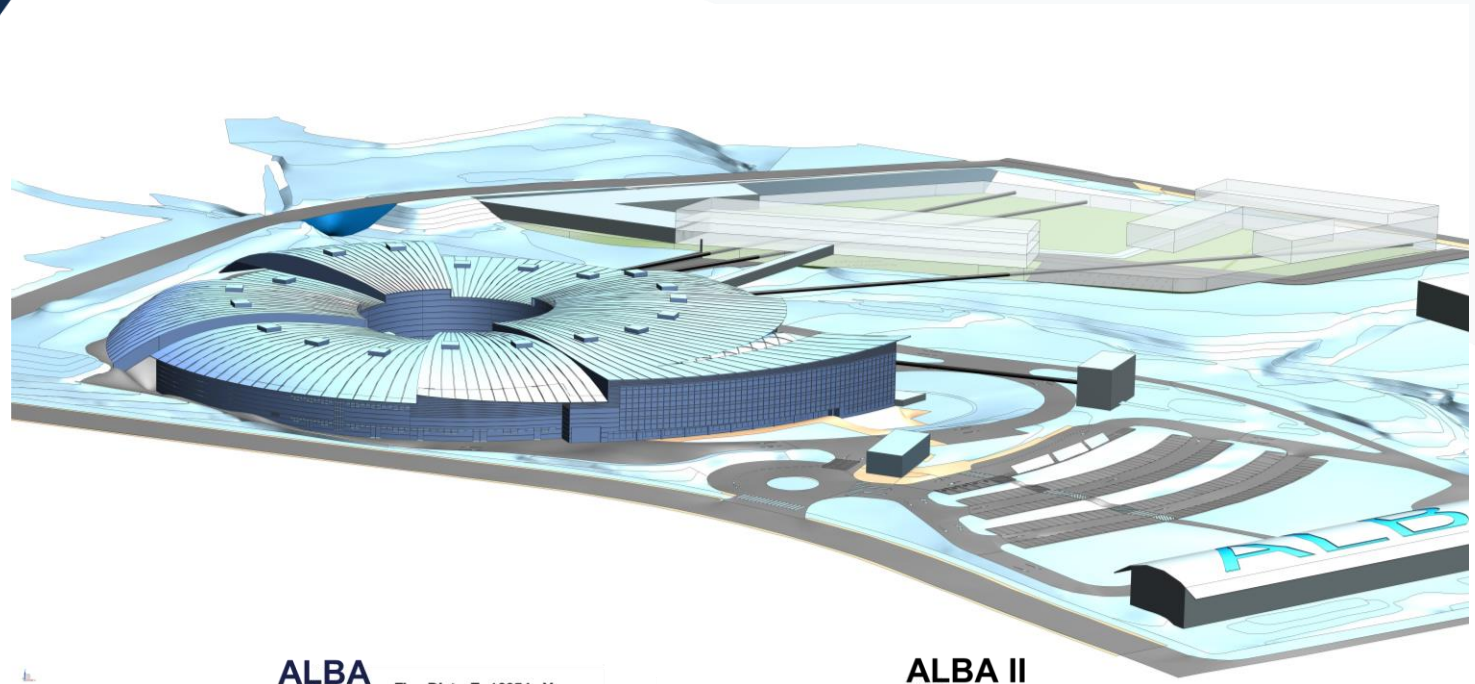
Evolution from 3rd to 4th generation in Europe

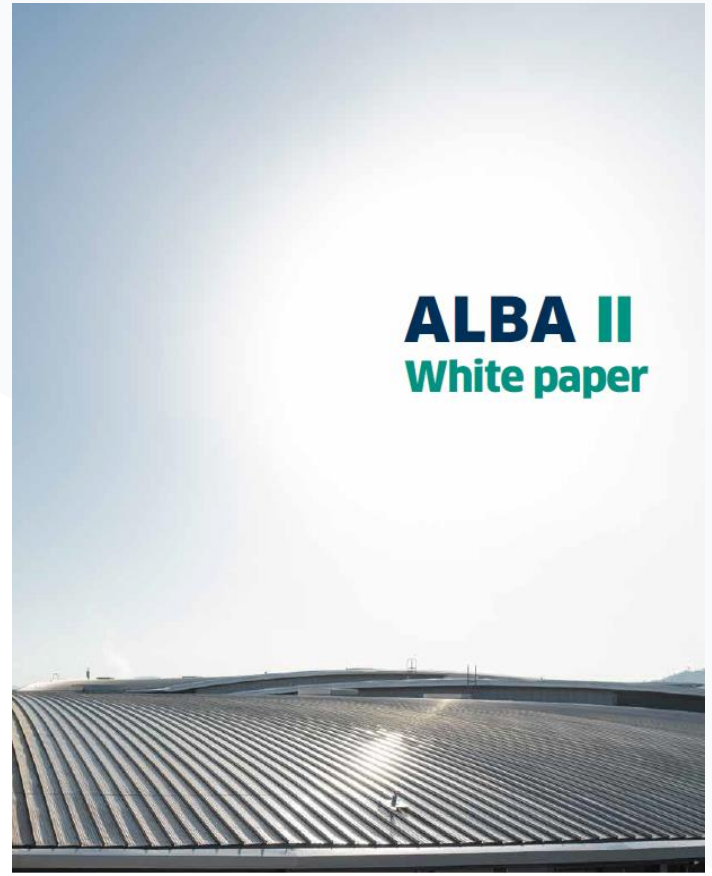
All synchrotrons in Europe (and in America and Asia) are evolving towards the 4th generation



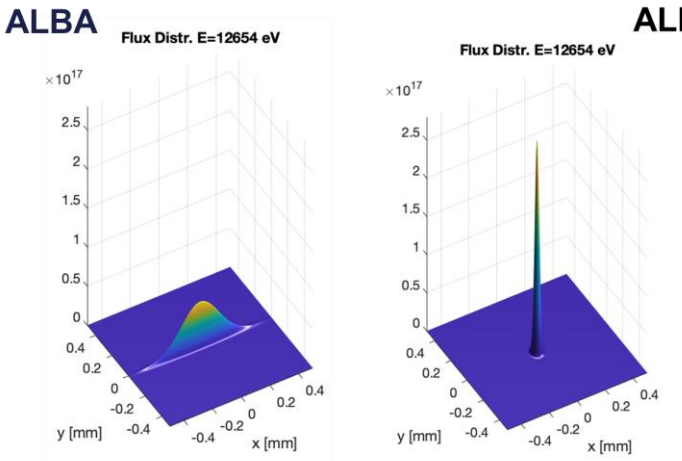


ALBA II



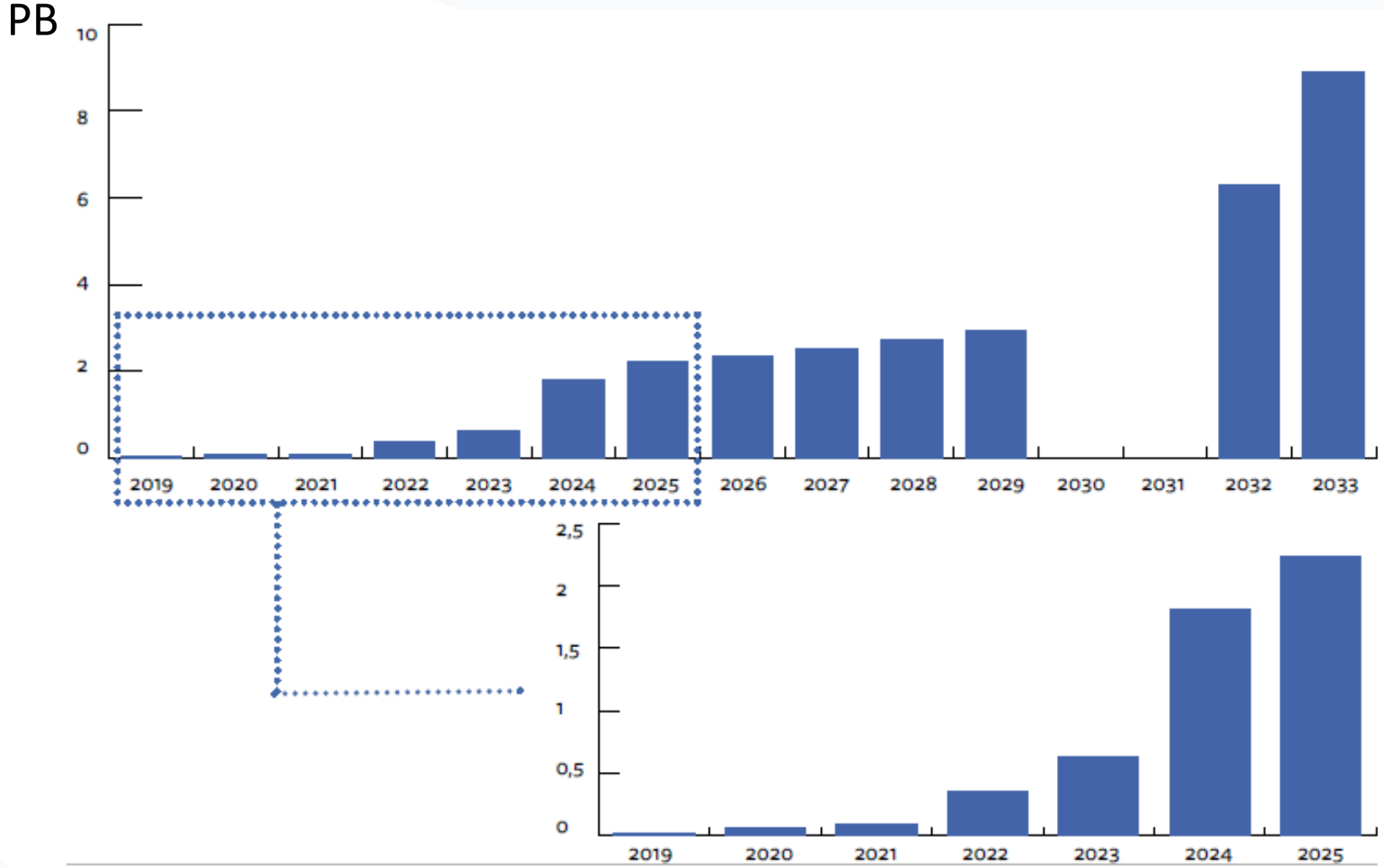


e- beam transverse dimensions: 50 μm (like the diameter of a hair, and like the eye resolution power)



https://www.cells.es/en/science-at-alba/alba-ii-upgrade/alba-ii

Data Produced in ALBA II



ALBA annual raw data production

IT infrastructures @ ALBA

Evolución del Sincrotrón ALBA			
	2023 - ALBA	2029 - ALBA	2032 – ALBA II
Aceleradores	3ª generación	3ª generación ALBA II preparado para la instalación	4ª generación Mayor brillanza, resolución, coherencia
Beamlines	10 operativas 4 en construcción	14 operativas 3 en avanzado estado de construcción	17 operativas (dos extra largas). Potencial de ampliar hasta 26 líneas
Otras capacidades experimentales	JEMCA (2 TEM con colaboradores), laboratorio de baterías	JEMCA e InCAEM (3 TEM + 3 microscopios) + laboratorio de catálisis	Se añaden capacidades con laboratorios en los nuevos terrenos en colaboración con otros institutos
Infraestructuras almacenamiento y análisis de datos	**Almacenamiento: 2 PB Procesado CPU: 500 Procesado GPU: 12.500**	**Almacenamiento: 18 PB Procesado CPU: 13.000 Procesado GPU: 190.000**	**Almacenamiento: 34 PB Procesado CPU: 28.000 Procesado GPU: 400.000**
Usuarios, experimentos y tiempo de luz de sincrotrón anual	2500 usuarios / 300 experimentos / 46.500 h para usuarios	3500 usuarios / 400 experimentos / 65000 h para usuarios	4250 usuarios / 500 experimentos / 79 000 h para usuarios
Publicaciones	300 / año	400 / año	500 / año
Uso industrial	60 experimentos/año	85 experimentos/año	100 experimentos/año
Impulso innovación	[Hablar de Patentes, contratos mayores con componente innovación]	[Hablar de Patentes, contratos mayores con componente innovación]	[Hablar de Patentes, contratos mayores con componente innovación]

- **To provide service to the ALBA II and the JEMCA.**
- **Including acquisition, processing and DaaS services to the user.**
- **EOSC required services are difficult to be seized.**

Infraestructuras almacenamiento y análisis de datos	**Almacenamiento: 2 PB Procesado CPU: 500 Procesado GPU: 12.500**	**Almacenamiento: 18 PB Procesado CPU: 13.000 Procesado GPU: 190.000**	**Almacenamiento: 34 PB Procesado CPU: 28.000 Procesado GPU: 400.000**

Conclusions

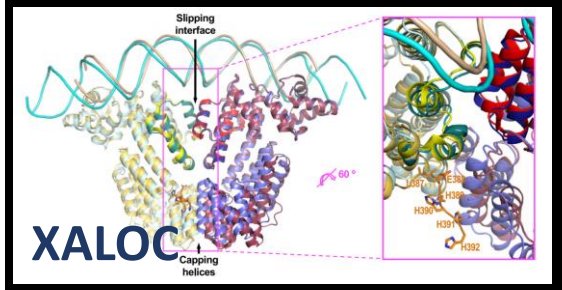
- **The PaN community is taking important steps towards Open Science.**
  - **Common Data Catalogues tools (ICAT, SCICAT) have been developed**
  - **PaN Data Portal is operative (and compatible with Open AIRE and B2FIND)**
  - **All facilities are in the process of linking their acquired data, creating a federated repository.**
  - **The Common DaaS platform (VISA) is functional, but it still has a long development roadmap ahead.**

Conclusions

- **Future challenges:**
  - **The volume of data is exploding (every facility will generate 10s PB/year).**
  - **Such volumes will require new techniques for leveraging Open Data or transferring/processing data from various sources.**
  - **Data Analysis algorithms must be optimized for this new scenario.**
  - **As of 2023, it appears to be a Herculean task to unify Ontologies, Data formats, and data analysis tools for every field.**

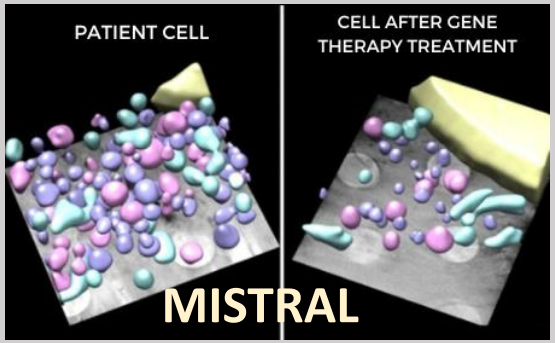
Some highlights from the scientific results published in the last months

MECHANISM OF ACTION OF THE ARBITRIUM COMMUNICATION SYSTEM IN SPBETA PHAGES



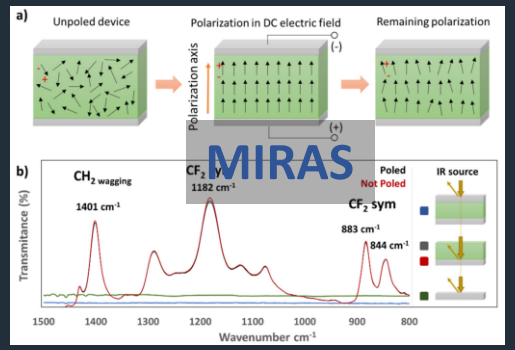
Nature Communications **volume 13**, Article number: 3627 (2022)

GENE THERAPY PROVED AGAINST MUSCULAR DYSTROPHY



Int. J. Mol. Sci. **2022**, 23, 7651doi.org/10.3390/ijms23147651

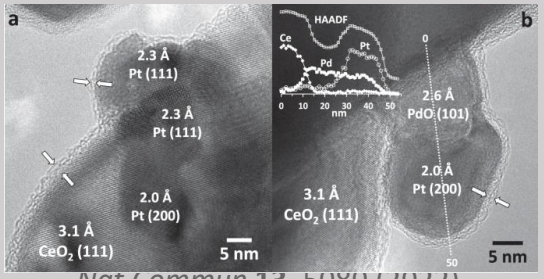
CONTROLLED POLING OF A FULLY 3D PRINTED PIEZOELECTRIC DEVICE



J. Mater. Chem. C, 2022, 10, 11555

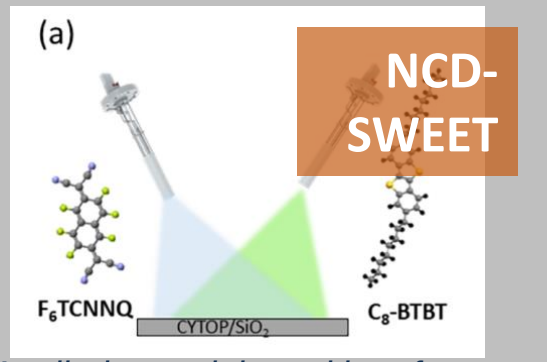
CATALYST TO REMOVE EMISSIONS OF METHANE

CIRCE (+CLAESS & MSPD)



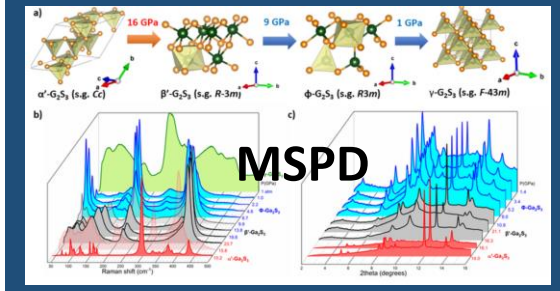
Nat Commun **13**, 5080 (2022). doi.org/10.1038/s41467-022-32765-4

CHARGE-TRANSFER COMPLEXES IN ORGANIC FIELD-EFFECT TRANSISTORS



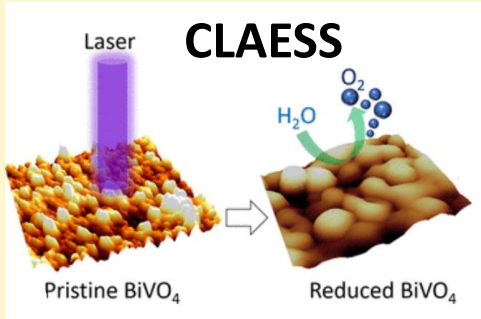
Applied materials and interfaces 10.1021/acsami.2c09168

PRESSURE-INDUCED PARAELECTRIC TO FERROELECTRIC PHASE TRANSITION



Chem. Mater (2022), 34, 13, 6068–6086. https://doi.org/10.1021/acs.chemmater.2c01169

ENHANCED PHOTOELECTROCHEMICAL WATER SPLITTING



ACS Applied Materials & Interfaces, 2022, 14, 29, 33200–33210

LIGHT-INDUCED SPIN SWITCHING DEVICE WITH APPLICATIONS IN SPINTRONICS

Advanced Materials (2022). **doi.org/10.1002/adma.202202551**