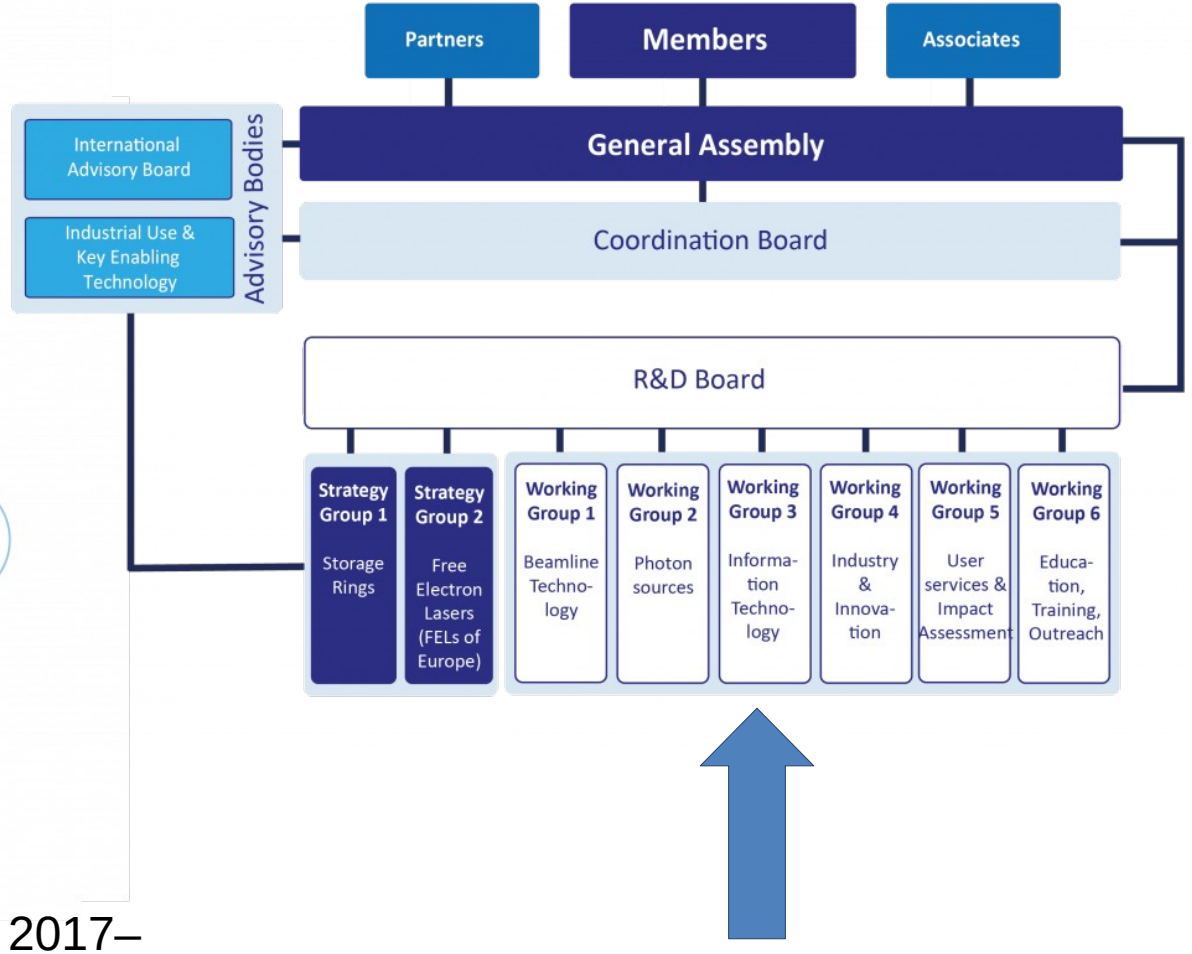
LEAPS League of European Accelerator-based Photon Sources

Update from LEAPS WG3

Paul Millar

WG3 Spokesperson

What is LEAPS? What is WG3?



2017–

SESAME LENS Canadian Light Source Centre canadien de rayonnement synchrotron ESUO EUROPEAN SYNCHROTRON AND FEL USER ORGANISATION

LEAPS League of European Accelerator-based Photon Sources

Projects: ExPaNDS and PaNOSC

Facility	Search	OAI-PMH	FAIR-IF	PaNET	Search-API	VISA
ALBA	Y	P	WIP	U	Y	WIP
DESY		WIP	WIP	WIP	WIP	Y
Diamond				Y		
Elettra	Y	Y	Y		Y	U
ESRF	Y	Y	Y	WIP	Y	Y
ESS	Y	Y			Y	WIP
EuXFEL	Y	Y	Y	WIP	Y	WIP
HZB		Y		Y	WIP	N
HZDR		Y	WIP	U	WIP	U
ILL	Y	Y	Y		WIP	Y
ISIS		Y	Y	U	Y	
MAX IV	Y	Y	WIP		Y	
PSI	Y	Y	P	Y	Y	N
SOLEIL		Y	WIP		Y	WIP

Yes, already adopted (Y)	Will not be adopted (N)
Planned to be adopted (P)	Under evaluation (U)
In the progress of being adopted (WIP)	

Last updated: 2023

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

2019 - 2023

panosc
photon and neutron open science cloud

2018 - 2022

LEAPS League of European Accelerator-based Photon Sources

LEAPS Data Strategy paper

Eur. Phys. J. Plus (2023) 138:617
https://doi.org/10.1140/epjp/s13360-023-04189-6

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

LEAPS data strategy

Andy Götz[1,a], Erwan le Gall[2,b], Uwe Konrad[3], George Kourousias[4,c], Oliver Knodel[3,d], Salman Matalgah[5,e], Oscar Matilla[6], Darren Spruce[7,f], Ana Valceril Orti[8], Majid Ounsy[9], Thomas H. Rod[9], Frank Schluenzen[10]

[1] European Synchrotron Radiation Facility, Grenoble, France
[2] Institut Laue Langevin, Grenoble, France
[3] Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany
[4] Elettra Sincrotrone Trieste, Basovizza, Italy
[5] SESAME, Allan, Jordan
[6] ALBA, Cerdanyola del Vallès, Spain
[7] MAXIV Laboratory, Lund University, Lund, Sweden
[8] SOLEIL, Saint-Aubin, France
[9] European Spallation Source ERIC, Copenhagen, Denmark
[10] DESY, Hamburg, Germany

Received: 17 December 2022 / Accepted: 14 June 2023

Abstract The continuous evolution of photon sources and their instrumentation enables more and new scientific endeavors at ever increasing pace. This technological evolution is accompanied by an exponential growth of data volumes of increasing complexity, which must be addressed by maximizing efficiency of scientific experiments and automation of workflows covering the entire data lifecycle, aiming to reduce data volumes while producing FAIR and open data of highest reliability. This papers briefly outlines the strategy of the league of European accelerator-based photon sources user facilities to achieve these goals collaboratively in an efficient and sustainable way which will ultimately lead to an increase in the number of publications.

1 Introduction

The LEAPS facilities [1] are producers of ever increasing volumes of valuable data for science, often referred to as the "data deluge" [2]. The upgrades planned over the next few years at most of the facilities will see a substantial increase in brilliance and in other beam characteristics, which are expected to result in more photons on the sample per surface area and solid angle, leading to faster experiments generating more data. More photons per surface area on the sample means shorter data acquisition times and faster experiments. Faster experiments means more data to be acquired, managed and processed so that users can extract useful results. At the same time the LEAPS facilities are attracting new communities, welcoming more and more users who are new to photon sources, and new to dealing with photon science data. This means more user groups being challenged by the data volumes, processing and interpretation of raw data. The LEAPS facilities were already facing a data deluge before embarking on upgrading their sources. After the upgrades the facilities will be even more challenged to manage the petabytes of raw data being produced while their users will be mostly limited by either lack of experience or data processing bottlenecks. Failure to address the data deluge will impact the scientific results negatively as more and more data are not published and end up as "dark data" on tape archives or being deleted and lost forever. The scientific community as a whole has been evolving over the past decades with government, scientific funding and international bodies including publishers promoting and sometimes even mandating adopting the principles of Open Science. Open Science includes producing FAIR data for research data management and software. It is timely therefore to develop a data strategy for the LEAPS facilities which addresses the data challenges of the current and future photon sources in Europe. The LEAPS

Andy Götz, Erwan le Gall, Uwe Konrad, George Kourousias, Oliver Knodel, Salman Matalgah, Oscar Matilla, Darren Spruce, Ana Valceril Orti, Majid Ounsy, Thomas H. Rod and Frank Schluenzen have contributed equally to this work.

[a] e-mail: andy.gotz@esrf.fr (corresponding author)
[b] e-mail: le-gall@ill.fr
[c] e-mail: george.kourousias@elettra.eu
[d] e-mail: o.knodel@hzdr.de
[e] e-mail: salman.matalgah@sesame.org.jo
[f] e-mail: darren.spruce@maxiv.lu.se

Published online: 17 July 2023

doi:10.1140/epjp/s13360-023-04189-6

Topics include:

- Increasing efficiency of experiments
- The European Open Science Cloud (EOSC)
- PaN data commons
- Training and E-learning
- Sustainable software
- Climate change and green IT
- Sharing knowledge

FAIR data is a common theme running through this paper.

LEAPS League of European Accelerator-based Photon Sources

PaNOSC ML workshop

A two half-day event, featuring community talks and working groups.

Key point: the need for FAIR data.

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

LEAPS League of European Accelerator-based Photon Sources

Continuing the work

OSCARS
Open Science Clusters' Action for Research & Society

2024-01-01 / 2027-12-31 (4 years) 25 M€

Project activities and outcomes:

- **Open Science practice**
- **Community-based Competence Centres (CCC)**
- **Composable Open Data and Analysis Services (CODAS)**
- An **established inter-cluster web-based "scientific social network"** in Europe.
- **Seed funding** (100-250k€ projects; 1–2 years). Two calls: first call 2024-03-15/2024-05-15, second call 2024-11/2025-01.

panosc
photon and neutron open science cluster

ESRF
The European Synchrotron

DESY.

CERIC

SCIENCE-CLUSTERS.eu
Research Infrastructures for Open Science

LEAPS League of European Accelerator-based Photon Sources

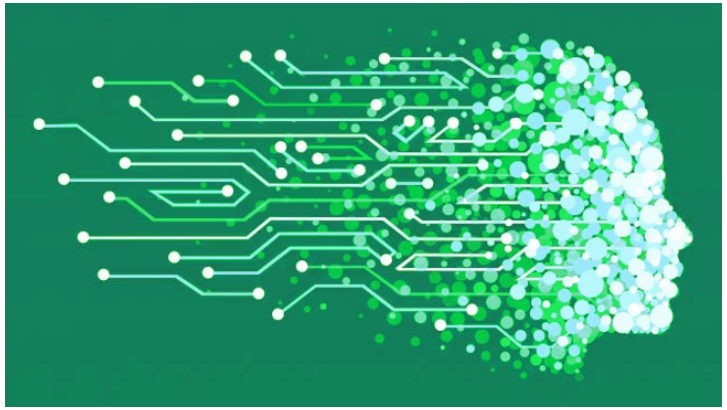
LEAPS WG3 and SIGs: what is a SIG?

- A SIG is a "**Special Interest Group**"
  - A place to "get work done"
  - A natural continuation of projects (PaNOSC, ExPaNDS, LEAPS-INNOV, …)
- Proposed by one facility with three others agreeing to contribute.
  - Bottom-up approach: what topics should we discuss?
  - Four facilities provides some sustainability
  - Relatively easy to establish.
- Collaborative in nature.
  - Open to any to join: LEAPS or non-LEAPS.
- SIGs need to do work, though.

Links:
- What is a SIG?
- Current SIGs

LEAPS League of European Accelerator-based Photon Sources

AI/ML SIG



- New SIG, just getting started.
  - Kick-off meeting to be scheduled.
- Home for potentially many aspects:
  - Machine control
  - Data for ML applications
  - DAC and related applications
  - …
- Potential continuation of the ML workshop
  - Development of a LEAPS ML strategy

PAUL SCHERRER INSTITUT

PSI

Markus Janousch

LEAPS League of European Accelerator-based Photon Sources

Core Trust Seal SIG

- Originally triggered by ESRF obtaining CTS certification
  - Presentation from Andy garnered interest from facilities
- Relatively low overhead process.
- Increasingly recognised within EU
- ESRF renewal of CTS status in 2025
  - Use this to help other facilities through the process.

Andy Götz

EOSC Node SIG

- Coordination of LEAPS activity involving the new EOSC Node concept.
- Application of PaNOSC (thematic) node as a candidate node
  - Candidate nodes will be the first nodes to be integrated.
  - Various LEAPS facilities contributed to the node survey.
- Currently awaiting decision whether PaNOSC node has been selected.
  - Should find out later this year.
  - If successful, we will need to accelerate our plans.

ESRF
The European Synchrotron

Andy Götz

LEAPS League of European Accelerator-based Photon Sources

Direct DOI Data Access SIG

- Researching how to improve the ‘A’ in FAIR.
- Dataset DOIs resolve to a landing webpage.
  - Provides humans with useful information
- Access to the data is usually provided as a link.
  - Requires a human to read the page
  - Each facility is different.
- Can we support automated access to data?

Paul Millar

LEAPS League of European Accelerator-based Photon Sources

Data Reduction SIG

- Continuation of the LEAPS-INNOV project's WP7
- Dedicated to managing the high data flows generated by the new generation of x-ray photon detectors. Focus includes all processes that facilitate the reduction of stored data. Key activities encompass:
  - receiving detector data
  - streamlining and accelerating data reduction workflows
  - high-throughput automated data processing
  - data compression techniques
- SIG is currently in early stages
  - Awaiting kick-off meeting

Zdeněk Matěj

LEAPS League of European Accelerator-based Photon Sources

Final remarks

- Strong FAIR theme running the LEAPS WG3 activity.
- The PaNOSC and ExPaNDS projects started the journey.
- OSCARS and the SIGs give us a vehicle to continue this work.
  - Please join in!

LEAPS League of European Accelerator-based Photon Sources