



## LEAPS data strategy

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

<sup>1</sup> European Synchrotron Radiation Facility, Grenoble, France

<sup>2</sup> Institut Laue Langevin, Grenoble, France

<sup>3</sup> Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

<sup>4</sup> Elettra Sincrotrone Trieste, Basovizza, Italy

<sup>5</sup> SESAME, Allan, Jordan

<sup>6</sup> ALBA, Cerdanyola del Vallès, Spain

<sup>7</sup> MAXIV Laboratory, Lund University, Lund, Sweden

<sup>8</sup> SOLEIL, Saint-Aubin, France

<sup>9</sup> European Spallation Source ERIC, Copenhagen, Denmark

<sup>10</sup> DESY, Hamburg, Germany

Received: 17 December 2022 / Accepted: 14 June 2023

© The Author(s) 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 paper 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.

<sup>a</sup> e-mail: [andy.gotz@esrf.fr](mailto:andy.gotz@esrf.fr) (corresponding author)

<sup>b</sup> e-mail: [le-gall@ill.fr](mailto:le-gall@ill.fr)

<sup>c</sup> e-mail: [george.kourousias@elettra.eu](mailto:george.kourousias@elettra.eu)

<sup>d</sup> e-mail: [o.knodel@hzdr.de](mailto:o.knodel@hzdr.de)

<sup>e</sup> e-mail: [salman.matalgah@sesame.org.jo](mailto:salman.matalgah@sesame.org.jo)

<sup>f</sup> e-mail: [darren.spruce@maxiv.lu.se](mailto:darren.spruce@maxiv.lu.se)

facilities and specifically the LEAPS data and computing facilities consume large and increasing amounts of resources in terms of data center running costs and increasing energy consumption that have to be addressed in the data strategy.

This document will outline the main goals and objectives of the LEAPS facilities data strategy to address the challenges of research data for the future to ensure that the cost in terms of energy and its carbon footprint equivalent, and limited budget, are put to best use in a world of finite resources. The strategy proposed is a compromise between complex versus simplistic solutions to optimize the resources available to achieve the main goal of increasing the efficiency of the LEAPS facilities. The strategy is closely aligned with the data strategy of the neutron sources in Europe (represented by the LENS initiative [3]) who are facing similar challenges. The strategy outlined in this document has been developed together with representatives of the LENS community.

## 2 Overall goals

The overall goal of the LEAPS data strategy is to provide a means for participating facilities to achieve their aims of greater efficiency and effectiveness in translating data into knowledge across European synchrotron light sources and neutron sources. The challenge is to identify themes and approaches which can be endorsed and supported within institutions and at the same time be driven across the LEAPS–LENS communities.

To achieve these three main areas of development have been identified:

### 2.1 Increasing efficiency of experiments

It has long been recognized that scientific data from experiments at x-ray user facilities are highly valuable and should be preserved, shared and reused. The Protein Data Bank (PDB) is frequently used as a particularly great example illustrating the value of a distinct class of reduced scientific data. The PDB served as training data for AlphaFold to successfully predict protein structures using a novel machine learning approach. The corresponding raw scientific data, the data volumes exceeding the reduced data stored at PDB by orders of magnitude, can be quite valuable as well not only to validate a result, but also as the unique and comprehensive compilation of information from which to extract knowledge. Experimental data, in addition, provide an invaluable fountainhead for future knowledge: through future re-analysis using data analytics or as a valuable resource for machine learning, deep learning or other artificial intelligence algorithms. Data, in summary, are the outcome of the experiment which needs to be preserved, as there is no sound experimental result without proper data to prove it. Data and metadata are often the only permanent record that remains after an experiment.

To increase the effectiveness of PaN facilities a strong investment by all developers of data reduction, data processing and data management systems to allow easier, faster assessment of data as well as to make it FAIR is required. Sharing of technologies, methods and code among communities and across facilities will help converge the efforts as has been shown from previous EU projects and international collaborations (see “[Appendix](#)”) which give freedom to innovate on standardized shared technical platforms reducing the need to build duplicate existing solutions. In addition, cost-benefit analysis and reviews to find the most prominent issues allowing to focus more effort in those areas leading to larger gains.

### 2.2 Open Science

Open Science, as defined by UNESCO [4], is “**making science more accessible, inclusive and equitable for the benefit of all**”. The European Commission (EC) has adopted an Open Science policy for all EC funded projects and as recommendation for science in Europe and in general [5]. In parallel, UNESCO also recommended its adoption and has highlighted the major benefits Open Science would bring to humanity. The EC has funded the European Open Science Cloud Association as an EC Partnership for promoting Open Science. The LEAPS–LENS communities have participated in implementing the EOSC open science platforms, through the ExPaNDS [6] and PaNOSC [7] projects. These projects have accelerated the implementation and adoption of Open Science practices and platforms in the community to the extent that the tools developed are now generally available and have been used by a selection of use cases and are ready to assist researchers to adopt Open Science practices. Adoption of Open Science practices can be further increased by continuing to build on top of these platforms.

### 2.3 Sustainability of solutions

In a world of finite energy and material resources the LEAPS data strategy aims at continuous improvement in the sustainability of solutions in the following ways:

- more sustainable use of resources by increased efficiency, reduction in duplication and loss of solutions/knowledge by adopting best practices and disseminating these and providing central repository locations and catalogues for making data Findable, Accessible, Interoperable and Reusable (FAIR).

- energy saving practices which can lower the overall carbon footprint of data production, such as improving the quality of remote operation to reduce the number of users travelling to facilities and improving the availability and usability of data processing during experiments to enable easier selection of valuable data and speed up data analysis times.
- early involvement in the other research fields of the Photon and Neutron (PaN) community to help build on existing software solutions gaining more efficiency for both the IT teams and the scientific community.

### 3 Increasing efficiency of experiments

In order to improve the efficiency of scientific experiments, the IT departments of research institutes are implementing various strategies. The first step to dealing with the large amounts of data recorded, is to provide sufficient metadata for workflows to process the data automatically and to offer access to data analysis as a service. This access must be efficient and secure while allowing increased flexibility. In the years to come, a significant increase in the quantity of data requires an update of the physical infrastructures including suitable file systems.

The increase in the amount of acquired data will require some experiments to develop more sophisticated workflows to cope with streamed data acquisition processes integrated with online data processing in order to filter noisy data and provide feedback to the control system adapting the acquisition process on the fly to optimize beamtime and to prevent inefficient data storage consumption.

Collaboration on a standard file format (NeXus/HDF5) [8] and the processing of the data before storage have already taken place improving the analysis capacity of the researchers and therefore the efficiency of the experiments. These collaborations need to be continued and reinforced especially in the growing area of workflows. Sharing of know-how and experience with file systems and native storage access needs to be continued in order to propose the best solution in the near future. Alternate types of access to facilities are also to be expected, in particular the post-COVID has reinforced the need to provide access to remote experiments. This includes providing access to huge datasets that are difficult to transfer efficiently over the Internet. It is therefore necessary to provide remote analysis solutions, which allow scientists access via a browser to a set of visualization and analysis tools, training, collaborative work plus direct access to data (since the platform needs to be hosted as close as possible to the data). Such a solution would make it possible to limit physical travel, to intervene very quickly on an experiment, to instantly share the analysis environment with a group and to very easily increase productivity. VISA [9, 10], one of the outcomes of PaNOSC, is an example of a remote analysis solution. It is already deployed at several LEAPS–LENS facilities. VISA provides safeguards in terms of cyber-security to help prevent hackers gaining access to sensitive resources. Ideally, VISA must be supported by a remote experiment control solution at each facility to enable the scientific community to run experiments without needing to travel. In some cases, for example during a pandemic, this is essential as it allows experiments to continue without scientists needing to travel to the facilities. During non-pandemic periods such a solution helps increase the efficiency of experiments.

Naturally, IT must also support scientific experiments with the provision of new infrastructures adapted to artificial intelligence, machine learning and cutting edge technologies. Some experiments are already familiar with the use of AI technologies, however without strong integration and the provision of sometimes complex tools and equipment, these new methods are not adopted by research teams. Collaborations between research teams and the IT departments of Research Infrastructures could make it possible to determine common needs with the goal of adopting generic tools that encourage researchers to take advantage of these technologies in a more systematic way.

Finally, an effort must be made to reduce extra administrative burden without added value carried out by the research teams. Indeed, each Research Infrastructure has its own needs and develops separate interfaces and tools to respond quickly and efficiently to a specific need. The administrative tasks made through several forms from the experiment proposal to the publication are different at each facility. Global harmonization, driven by simplicity and optimal user experience, would allow research teams to better understand their needs prior to the experiment, to better prepare for experiments and to accelerate the steps up to the publication. IT teams must be able to integrate generic developments with local needs without having to develop the entire solution from scratch. Harmonisation between data policies, e.g., length of embargo periods, on the European scale would contribute to defining common governance rules for data from LEAPS facilities.

#### *Objectives: increasing efficiency of experiments*

Prior experience of multiple facilities and outcomes of related projects suggest that the objective of increasing efficiency of experiments may be achieved easier by having:

- 3.1 LEAPS facilities to setup a working group for sharing data processing and analysis software and workflows for existing codes and new algorithms.
- 3.2 LEAPS facilities to share development of new AI/ML-based data processing codes.
- 3.3 LEAPS facilities to continue the development a remote analysis platform.
- 3.4 LEAPS facilities to work closely with user communities to define common metadata and strategies for data formats, compression, and publication.
- 3.5 LEAPS facilities to harmonize their data policies in terms of scope and embargo period of data.

#### 4 LEAPS and the European Open Science Cloud (EOSC)

The European Open Science Cloud (EOSC) is the European Commission financed project for implementing open science principles across the EU [11]. While the vision is clearly defined, the architecture and requirements for how to integrate data repositories and infrastructures into the EOSC are still an ongoing process. LEAPS facilities represent a significant number of research infrastructures and can therefore play a role to ensure that the EOSC is attainable with the available resources. In 2021 the EC established the EOSC as an EC Partnership, with the EOSC Association [12] being the association in charge of defining the EOSC standards and roadmap. It is important that the LEAPS facilities form a PaN scientific cluster and speak with one voice vis-à-vis the EOSC.

The LEAPS–LENS facilities have been strongly involved in the first phase of the EOSC thanks to the PaNOSC and ExPaNDS projects. The main goal of these two projects was to make FAIR data a reality for the photon and neutron facilities in Europe. In order to achieve this ambitious objective the two projects worked on the following outcomes:

1. FAIR data policy and DMPs.
2. FAIR assessment and common PID framework.
3. Standardized metadata (NeXus/HDF5, PaN ontologies).
4. Federated search API for PaN data catalogues.
5. Open data portal for searching + downloading data.
6. Community AAI UmbrellaId [13].
7. JupyterLab notebooks and NeXus/HDF5 files visualization.
8. Remote data analysis with VISA + data analysis pipelines.
9. Simulation software for simulating experiments and data (Oasys [14], SIMEX [15], and McStas for neutrons [16]).
10. PaN-learning platform (pan-training.eu and e-learning.pan-training.eu).

The outcomes support the adoption of FAIR data policies, FAIR data (through standardized metadata and ontologies), data services (Jupyter notebooks and remote analysis), simulation and training. The FAIR data policy framework recommends implementing an embargo period protecting data during the embargo period for the exclusive use of the experimental team who produced the data. The standard embargo period is 3 years which corresponds to the average length of a PhD with the possibility for the experimental team to open data before or request an extension to the embargo in case the analysis is part of a longer study.

The LEAPS–LENS facilities have committed to adopting and sustaining the above outcomes. The LEAPS–LENS facilities are providers of FAIR data and scientific software services which is the core of the European data strategy as implemented by the EOSC and Data Spaces.

*Objectives: European Open Science Cloud*

- 4.1 LEAPS facilities to implement the FAIR principles for data as soon as possible.
- 4.2 LEAPS facilities to commit to adopting and sustaining the outcomes of PaNOSC and ExPaNDS.
- 4.3 LEAPS facilities to participate in building the EOSC with the other EOSC science clusters.
- 4.4 LEAPS facilities to adopt the practices of Open Science.

#### 5 PaN data commons

As already explained above, the LEAPS–LENS facilities are producers of huge amounts of data. The type and variety of experiments cover a wide range of applications for addressing societal challenges. The adoption of open data policies has opened the way to make these data available for reuse. The goal of the PaN Data Commons is to provide a single point of access to all open data from the LEAPS–LENS facilities in order to make open data easy to find, access and reusable. A Data Commons is a concept for making data from different sources available according to a common set of governance rules via a single-entry point which can be a data portal, database, API or a combination of all three. The advantages of the Data Commons are that “by connecting the digital objects and making them accessible, the Data Commons was intended to foster novel scientific research that wasn’t possible before, including hypothesis generation, discovery, and validation” [17].

The goal of the PaN Data Commons is to build on the PaNOSC Open Data portal [18] to federate open data from all LEAPS–LENS facilities to make them accessible via a single data portal. The portal will allow domain specific data collections like the Human Organ Atlas, COVID-19, Additive Manufacturing, Batteries, etc., to be easily found and reused by researchers. The PaN Data Commons will learn from the data commons developed by other communities to provide a rich and efficient user experience. By making data from LEAPS–LENS facilities easily accessible the PaN Data Commons will promote the adoption of common practices for making data FAIR, e.g., standard metadata, PIDs, download service, etc. The PaN Data Commons will expose data services to users in a common way which will help in their adoption. The goal of the PaN Data Commons is to make it the place to find and publish data from LEAPS–LENS facilities.

*Objectives: PaN data commons*

- 5.1 LEAPS facilities to implement the PaN Data Commons based on a federated solution for searching and accessing open data.

5.2 LEAPS facilities to link their data repositories to the PaN Data Commons to expose open data according to the FAIR principles.

5.3 PaN Data Commons to monitor the number of searches, downloads and citations of data published.

## 6 Training and E-learning

Education and training are increasingly important for helping scientists work on photon and neutron (PaN) sources. The pan-training.eu [19] portal is a result of the two EOSC projects ExPaNDS and PaNOSC and provides a catalogue of training materials, a calendar of upcoming events, as well as access to e-learning courses. The portal is registered as an EOSC service to make it findable through the EOSC eco-system.

The PaN training catalogue registers and shares third-party training materials such as tutorials, videos or repositories. The catalogue is based on an established solution developed by the Elixir community [20]. It also provides a way to describe dependencies between specific datasets and the related analysis software with simple graphical workflows. In future, the real execution of these workflows can be realized through the generation of a relevant VISA instance containing the relevant data/software. In this way, it is possible to introduce the resources and infrastructure used by the LEAPS facilities. The events calendar of the training portal catalogue shares different training events, workshops, or conferences that are automatically harvested from LEAPS partners' websites and other relevant websites. Hence, providing the visitor with an overview of upcoming events.

E-learning courses are provided through an e-learning system (often called pan-learning) based on the widely used Moodle platform [21] and with Jupyter integrated such that Jupyter notebooks can be used as training materials in the context of a moodle course. This is particularly useful for hands-on lessons on programming, data analysis, modelling or simulations. The e-learning platform has been used at several neutron courses and summer schools but also for photon-specific courses.

A number of challenges still need to be addressed for the training portal to become the community standard for sharing, finding and accessing training material. One of the challenges is how to curate the training material shared on the portal to prevent it from becoming polluted with courses that are not maintained that may turn users away. Having proper policies in place may be a way to mitigate this. Moreover, facilities (and universities) should ensure that their training material and events are registered in the training catalogue/calendar and e-learning (or blended) courses should be made available in the PaN training portal's e-learning platform.

*Objectives: training and e-learning*

6.1 LEAPS facilities should adopt the PaN-training portal as the standard community platform for sharing training and e-learning material on photon science. By doing so they will specifically: ensure that training materials and events are registered in the training catalogue/events calendar and adapt pan-learning for e-learning and blended learning.

## 7 Sustainable software

Research software must be developed sustainably and published according to the FAIR principles. Sustainable development requires adopting professional software engineering practices that cover not only the software itself, but also the operating systems, libraries, reference data and application workflows, versioning, professional packaging, ease of deployment, and most importantly: maintenance [22]. Publication in the sense of open science and according to the FAIR criteria means that scientific papers, data, experiments and software reference each other. An example of how both could be implemented at LEAPS facilities is shown by CERN with its integrated approach based on Zenodo (Invenio RDM [23]).

The software engineering at LEAPS–LENS institutions is increasingly sustainable, partly due to the adoption of git repositories like GitLab and GitHub, which support best practices like CI/CD and code review and are also platforms for collaboration among the LEAPS–LENS partners. To promote visibility and reusability, the central LEAPS–LENS software catalogue [24] developed in 2011 will be continued and enhanced with new features, e.g., notebooks, workflows, and badges. The catalogue will be better integrated in the workflow and tools of the LEAPS facilities to ensure stronger involvement of the user community to promote solutions, obtain user feedback and collect useful add-ons that are developed locally.

In addition to this catalogue, support and training services for the shared software can be offered. An example of this is the open source software VISA, which was rolled out in several facilities as part of the PANOSC and ExPaNDS projects (see “Appendix” for more examples). Similar developments took place for the Tango and EPICS control toolkits and the metadata management systems ICAT [25] and SciCat [26], which are also widely used in the community. These project-driven developments and shared processes need to be systematically developed for the LEAPS and LENS communities for more common software solutions to provide researchers with software to process and analyze data. The software catalogue will be supported and accompanied by the common pan-training.eu platform; this can both be supported by courses in the training catalogue and in the elearning (moodle/Jupyter) platform.

*Objectives for sustainable software*

7.1 LEAPS facilities will continue to maintain the common software catalogue of software for PaN facilities to actively promote common software as well as data processing methods and workflows.



- 7.2 LEAPS facilities will set up official collaborative projects supported by formal MoUs to foster collaboration on specific outcomes of the PanOSC and ExPands projects, e.g., VISA platform, FAIR data management, search APIs, data portal, data catalogues, DMPs, etc.
- 7.3 LEAPS facilities will continue the work launched in the frame of LEAPS-INNOV WP7 for data compression and reduction to develop common validated algorithms to help address the data deluge.

## 8 Climate change and green IT

Since the consumption of energy is vital in scientific research, the objective for Research Infrastructures is to optimize any use of matter or energy in order to control its footprint on the environment. The impact of IT is not negligible, both on purely scientific aspects (HPC, Data, Backup, Network...) as well as user aspects (travels, workstations, screens, printers, computers...). There are many levers and several IT departments are already doing selective efforts, e.g., limiting the clock frequency of processors.

Nevertheless, several aspects merit a more global strategy common to all the Research Infrastructures. To reduce the carbon footprint of IT, it is necessary to define key parameters and to measure those in a framework common to Research Infrastructures in order to be able to compare the environmental cost of IT for each Research Infrastructures. A collaboration allowing the creation of a framework defining the measurement methods and the results of these measurements will make it possible to assess the energy consumption of the Research Infrastructures.

Examples for those indicators would be:

- Percentage of remote/mail-in experiments (travel footprint).
- Data center efficiency (PUE).
- Energy/carbon footprint of Data Analysis Workflows.
- Energy/carbon footprint of Data Storage including data compression/decompression.

A second step would be the creation of best practices, recommendations and a common green IT policy. This would allow the Research Infrastructures to encourage all involved to propose “Green IT” innovations responding to the laboratory strategy on energy optimization, without decreasing the quality of research output.

Finally, collaboration on developments taking advantage of low-power processors and energy optimized methods (e.g., for AI, compression, and acceleration) would also allow a change of paradigm on the optimization of computing resources, directly reducing electricity consumption compared to currently used solutions.

*Objectives for Climate Change and Green IT*

- 8.1 LEAPS facilities to reduce the volumes of data using innovative data compression schemes, e.g., by continuing the work started in the LEAPS-INNOV EC funded project.
- 8.2 LEAPS facilities to improve the remote operation solutions and to encourage their use to reduce travel to the facilities.
- 8.3 LEAPS facilities will measure their carbon footprint equivalent and take measures to reduce it with HW and SW suppliers.

## 9 Sharing know-how

Being a member of the research communities of LEAPS and LENS delivers many benefits, including the ability to share expertise and best practices through channels of collaboration among member facilities (see examples in “[Appendix](#)”). The goal is to address common technical challenges in IT. Collaboration with partners covers a wide range of topics, however, currently establishing and defining a FAIR-based data policy that describes, controls and constitutes the experimental data being collected from beamlines is a hot topic and a common aim of the community. On the other hand, through many projects and activities, community efforts and developments have led to solutions that effectively address issues by standardizing and resolving their particular requirements and special data challenges.

Collaborations allow facilities to share success stories and best practices among themselves, and to launch a structured dialogue for their IT works and future projects. A good example of having a common web portal which gives the visitor an overview about all technical specifications for each beamline across the facilities is Way for Light [27]. Implementing such an approach will be highly valuable for IT experts and opens the collaboration channels between LEAPS and LENS facilities.

*Objectives for Sharing know-how*

- 9.1 Establishing a centralized sustained web portal to present and list all LEAPS IT collaborations, official development websites, brief about developments and projects, news, categorized community code repositories, important links, job announcements, IT conferences, workshops which could be extended later to a technical forum.

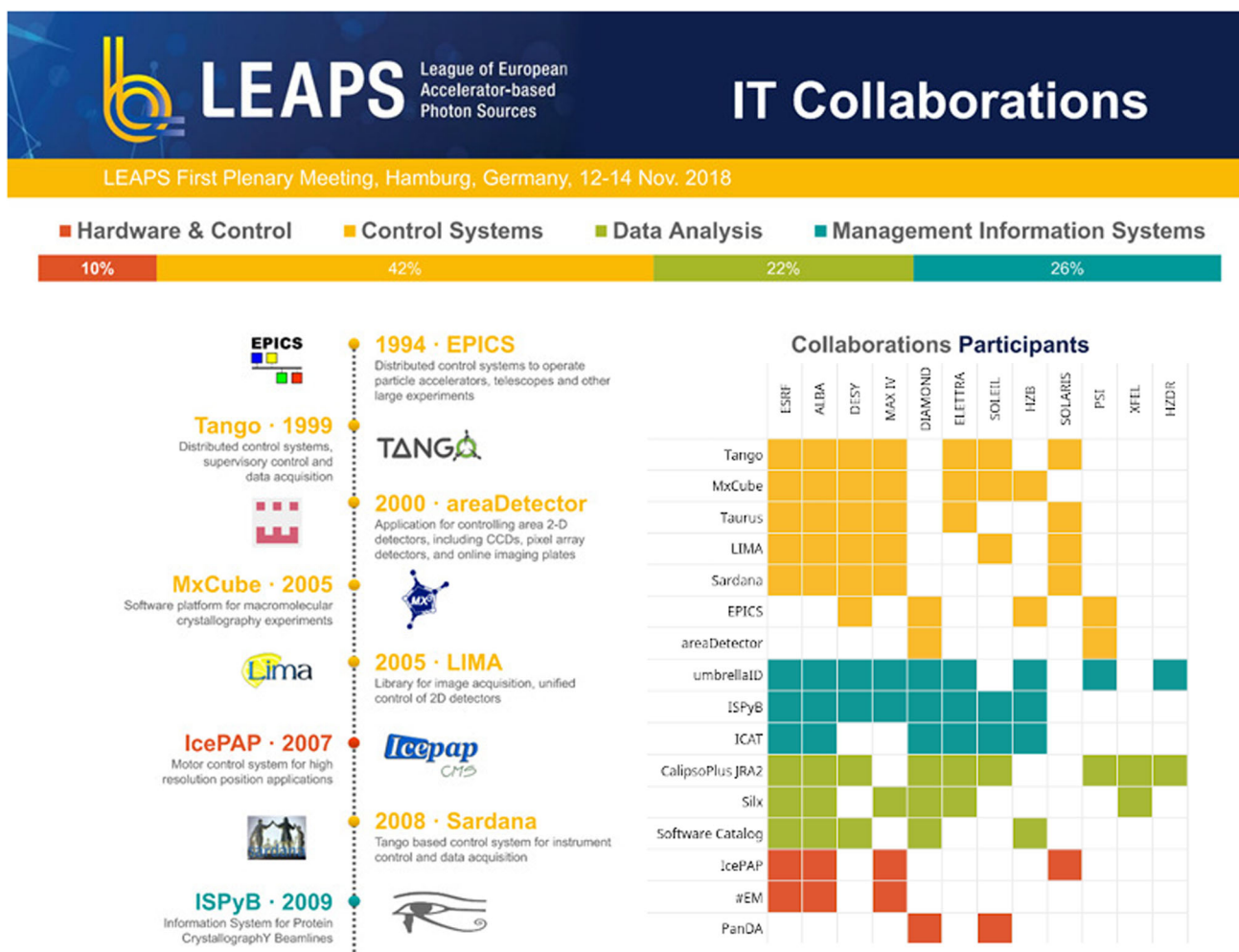


Fig. 1 LEAPS IT Collaborations 1999–2009

### 10 Conclusion

Data are essential to science. Therefore, for producers of huge data like the LEAPS facilities, it is of paramount importance to have a strategy to address the data challenges in the future. This document outlines twenty objectives in seven areas that the LEAPS facilities management should endorse and adopt in order to deal with the huge amounts of data being produced and to ensure that best use is made of the data. First and foremost implementing the objectives will serve the creators of the data, i.e., experimental teams, to deal more efficiently with their data to do better science while adhering to the principles of Open Science. To achieve this, the LEAPS management is strongly committed to implementing the FAIR principles for data in order to share them with the scientific community and to maximize their reuse. Implementing the objectives outlined in this document will contribute to minimizing the amount of data which are not published and thereby contribute to improving the efficiency in terms of scientific publications and energy consumption of the LEAPS facilities.

**Data availability** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

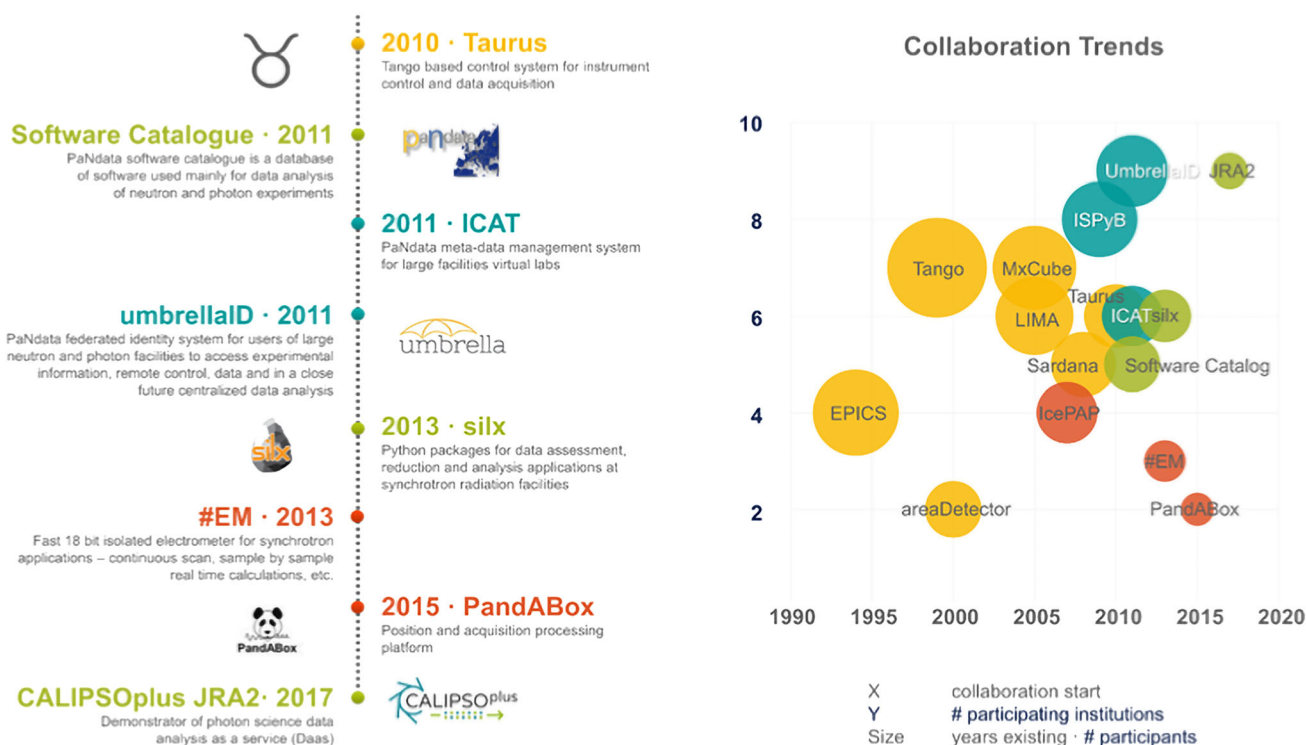


Fig. 2 LEAPS IT Collaborations 2010–2019

## Appendix. LEAPS IT collaborations 1999–2019

IT departments, belonging to many light sources and laboratories in Europe, have a long collaborating history in place as shown in the survey taken for the first LEAPS Plenary meeting in 2019 as depicted in Figs. 1 and 2. Collaborations on control system frameworks, such as EPICS and Tango started in the late '90s, followed by software for detectors, acquisition systems (LIMA) and macromolecular crystallography (MXCube). IcePAP, a project for a high resolution motion control system, started in 2007. Exploiting Tango, Sardana, a framework for instrument control, and Taurus, for dynamic graphical user interfaces, projects started between 2008 and 2010. In 2009 ISPyB, a collaboration for an information system dedicated to protein crystallography beamlines, started. During 2011 a number of projects started under the PaN umbrella, such as Software Catalogue, a database of software used for data analysis, ICAT, for metadata management, and umbrellaID as a federated identity system for users. A collaboration on a Python package for data reduction and analysis at synchrotron sources, named silx, started in 2013. Two hardware projects, to design a fast electrometer (Em#), and an integrated positioning, acquisition and processing platform (PandABox) belong to 2013 and 2015, respectively. The CALIPSOplus JRA2 project started in 2017, to deliver a demonstrator platform for science data analysis as a service.

Despite not all facilities being involved in all the projects, there are obvious advantages in collaborating. Each facility can benefit from the contribution of the other participants, in terms of faster time to deliver and improved quality of the outcome. Moreover, collaborating on the same project, enables building up critical mass and helps reducing single point-of-failures, which benefits reliability and sustainability. Finding common solutions, which can be adopted at different facilities benefit especially users, and make common, already known, software and tools more widely available. The LEAPS and LENS initiatives can clearly contribute to consolidating existing collaborations and creating new ones between the IT departments of their facilities.

## References

1. LEAPS, *League of Accelerator-Based Photon Sources Initiative*, <https://leaps-initiative.eu>
2. C. Wang, U. Steiner, A. Sepe, Synchrotron big data science. Small (2018). <https://doi.org/10.1002/sml.201802291>
3. LENS, *League of Neutron Source Initiative*, <https://lens-initiative.eu>
4. UNESCO Recommendations on Open Science, <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>
5. EC and Open Science, [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en)
6. EOSC PaN Data Services H2020 Project, <https://expands.eu>
7. Photon and Neutron Open Science Cloud H2020 Project, <https://panosc.eu>
8. NeXus/HDF5 Repository, <https://github.com/nexusformat>



9. VISA Repository, <https://github.com/ILLGrenoble>
10. VISA Documentation, <https://visa.readthedocs.io/en/latest/>
11. European Open Science Cloud (EOSC), [https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en)
12. EOSC Partnership, <https://www.eosc.eu/partnership>
13. UmbrellaID: confidentiality for a lifetime, <https://umbrellaaid.org>
14. L. Rebuffi, M. Sanchez del Rio, OASYS (OrAnge SYnchrotron Suite): an open-source graphical environment for x-ray virtual experiments. Proc. SPIE **10388**, 103880S (2017). <https://doi.org/10.1117/12.2274263>
15. SimEx, Software platform for simulation of advanced photon experiments, <https://github.com/PaNOSC-ViNYL/SimEx>
16. P.K. Willendrup, K. Lefmann, J. Neutr. Res. (2020). <https://doi.org/10.3233/JNR-190108>
17. NIH Data Commons, <https://commonfund.nih.gov/commons>
18. PaNOSC Open Data Portal, <https://data.panosc.eu>
19. O. Knodel, A. Padovani, J. Schwabe, Training-catalogue for the photon and neutron data services. Zenodo (2022). <https://doi.org/10.5281/zenodo.7482266>
20. ELIXIR's Training e-Support Service, <https://github.com/ElixirTeSS/TeSS>
21. The open source learning platform, Moodle LMS, <https://moodle.org>
22. J. Wuttke, S. Cottrell, M.A. Gonzalez, A. Kaestner, A. Markvardsen, T.H. Rod, P. Rozyczko, G. Vardanyan, J. Neutr. Res. (2022). <https://doi.org/10.3233/JNR-220002>
23. InvenioRDM, <https://inveniosoftware.org/products/rdm/>
24. PaN Software Catalogue, <https://software.pan-data.eu/>
25. ICAT Project, <http://data.datacite.org/10.5286/SOFTWARE/ICAT>
26. SciCat Project, <https://scicatproject.github.io>
27. Way for Light, <https://wayforlight.eu>