# Leveraging open data from PaN facilities for machine learning

📅 Oct 17 - 18, 2023

🧭 Synchrotron SOLEIL - CNRS - CEA, Paris-Saclay, France

## Exploiting open data for machine learning training: can we do it?

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. Exploitation of experimental training datasets is a key component of machine learning. However, finding the right data to train algorithms is a challenge and one of the motivations for making data FAIR is exactly that: to provide scientists working on AI applications with quality training datasets.

But what does 'quality' mean to PaN science communities? What metadata fields are needed to find the data, to understand if it is suitable for our research, and ultimately to be able to ingest it in our training models? How can we provide sufficiently rich metadata?

## Objectives

With this workshop, we aim to discuss these questions, among staff and users of the LEAPS and LENS facilities, across disciplines and across Europe.

We will present projects and teams that have successfully used open datasets from PaN facilities to train their specific ML application (data consumers), as well as domain scientists (data producers) who have published curated data specifically for ML applications. We will also look at cases where it hasn't worked so well, and discuss why and how it could be better in the future.

## Call for abstracts

During the workshop, we have slots available for 20 minute presentations. We are particularly interested in contributions that can address the following points:

- Have you used open data for machine learning? What has been your experience?
- Do you feel your research could benefit from more and better curated open data? Tell us how.
- Have you already made open datasets available to your community for ML training purposes?
- Have you managed to enrich the metadata of open datasets for better training?

Whatever your discipline, as long as the data originally came from experiments at a photon or neutron source, please submit an abstract in Indico ⏩