

Using HDF5 to Store Scientific Data at the European XFEL

Monday 26 May 2025 09:25 (25 minutes)

The **European XFEL** is an X-ray laser research facility that produces extremely short and intense X-ray flashes, enabling investigations across a wide range of fields—from the structure of matter to the dynamic evolution of molecular systems. A typical experiment can generate petabytes of data within a day, originating from diverse detectors and in multiple formats. Managing this high-volume, heterogeneous data, along with metadata, in real- or near real-time poses unique challenges.

HDF5 provides a robust solution to these requirements. Its ability to define both data and its structure enables consistent storage across various data sources. HDF5 also supports parallel, real-time writing, by distributing experimental data across multiple files, and linking them coherently. Features such as sparse chunking and compression further optimize storage, which is critical given the facility's data output.

In this talk, I will introduce **karaboHDF5**, a library that enables real-time, space-efficient storage of experimental data, while ensuring the data is well-structured and easy to access for analysis and processing. Developed by the European XFEL on top of the HDF5 core library, it is tailored to the facility's demanding performance and usability requirements, and fully leverages the capabilities of HDF5.

Keywords: HDF5, real-time data acquisition, high-volume data, heterogeneous data, X-ray data acquisition.

May we record your session?

Yes

Primary author: VAZQUEZ-GARCIA, Jose (European XFEL)

Co-authors: SILENZI, Alessandro (European XFEL); GOERIES, Dennis (European XFEL); BOUKHELEF, Djeloul (European XFEL); FLUCKE, Gero (European XFEL); WRONA, Krzysztof (European XFEL); GELISIO, Luca (European XFEL); ESENOV, Sergey (European XFEL); HAUF, Steffen (European XFEL)

Presenter: VAZQUEZ-GARCIA, Jose (European XFEL)