Contribution ID: **375** Type: **Contributed talk**

Data Reduction Strategy at the European XFEL

Friday 30 August 2024 15:40 (15 minutes)

The European XFEL is a megahertz repetition-rate facility producing extremely bright and coherent pulses of duration of the order of few femtoseconds or less. Owing to its X-ray imagers, specifically built to operate at these repetition rates (AGIPD, DSSC and LPD), the amount of data generated in the context of user experiments can exceed hundreds of gigabits per second, resulting in tens of petabytes stored every year. These rates and volumes pose significant challenges both for the facility and its users. In fact, if unaddressed, extraction and interpretation of scientific content is hindered, and investments and operational costs quickly becomes unsustainable.

We are working, in close collaboration with our users, to address the above-mentioned topics at different levels. On the administrative level, we largely revisited our scientific data and retention policies so as to establish a framework for data reduction, and includes provision of open and FAIR data. Then, we are also introducing comprehensive data management plans, to streamline and enhance facility-users communication, including requirements and agreements on services and methods. On a technical and scientific level, we are upgrading our data systems to incorporate data reduction tools and methods [1,2], which are either developed inside or outside the facility. Finally, we are developing extensive metrics to assess reduction quality, and to corroborate automation of reduction processes.

In this talk I will highlight challenges and solutions implemented to date, and detail our vision for user-centric data reduction [2].

[1] Schmidt P, *et.al.* (2024) Turning European XFEL raw data into user data. *Front. Phys.* 11:1321524. doi: 10.3389/fphy.2023.1321524
[2] Sobolev E, *et.al.* (2024) Data reduction activities at European XFEL: early results. *Front. Phys.* 12:1331329. doi: 10.3389/fphy.2024.1331329

I plan to submit also conference proceedings

No

Primary authors: SOBOLEV, Egor (European XFEL); SCHMIDT, Philipp (European XFEL); MALKA, Janusz (European XFEL); HAMMER, David (European XFEL); BOUKHELEF, Djelloul (European XFEL); MOELLER, Johannes (European XFEL); AHMED, Karim (European XFEL); BEAN, Richard (European XFEL); BERMUDEZ MACIAS, Ivette Jazmin (European XFEL); BIELECKI, Johan (European XFEL); BOESENBERG, Ulrike (European XFEL); CARINAN, Cammille (European XFEL); DALL'ANTONIA, Fabio (European XFEL); ESENOV, Sergey (European XFEL); FANGOHR, Hans (European XFEL); FERREIRA DE LIMA, Danilo Enoque (European XFEL); MAIA, Luis (European XFEL); FIROOZI, Hadi (European XFEL); FLUCKE, Gero (European XFEL); GESSLER, Patrick (European XFEL); GIOVANETTI, Gabriele (European XFEL); KOLIYADU, Jayanath (European XFEL); MADSEN, Anders (European XFEL); MICHELAT, Thomas (European XFEL); SCHUH, Michael (European XFEL); SIKORSKI, Marcin (European XFEL); SILENZI, Alessandro (European XFEL); SZTUK-DAMBIETZ, Jolanta (European XFEL); TURCATO, Monica (European XFEL); TURKOT, Oleksii (European XFEL); WRIGLEY, James (European XFEL); APLIN, Steve (European XFEL); HAUF, Steffen (European XFEL); WRONA, Krzysztof (European XFEL); GELISIO, Luca (European XFEL)

Presenter: SOBOLEV, Egor (European XFEL)

Session Classification: Mikrosymposium 3/4: Data, Automation and the Use of AI

Track Classification: 3. Data, Automation and the Use of AI