Integration of Collaborative Data Platform for Large-scale 3D X-ray Data

The extremely high X-ray photon flux density at the KARA synchrotron facility of KIT combined with fast cameras enables researchers to scan biological and material specimen with high resolution in a short time. Currently, it is possible to perform tomographic scan with 3000 projections of size 2016 x 2016 in 90 seconds, yielding 24 GB of raw data. With a sample changer robot and automation of the scanning procedure, it is possible to scan over 1500 samples per week, which is about 40 TB of data [1]. Its processing, including tomographic reconstruction, is realized by the UFO framework [2], which is a GPU-based image processing toolkit. The large amount of scanned samples requires collaboration and automation, which in turn requires a suitable data platform for processing, analyzing, sharing and visualizing large series of data. However, no such platform is currently being used at the KARA synchrotron facility.

Kadi4Mat [3], is a promising platform for management and creation of data processing workflows via various interfaces, including a web-based one. It is minimalistic in the sense of interfacing requirements and thus easily extensible. The API based on REST is available for C, BASH and python.

The task of this project is to integrate UFO into Kadi4Mat as a data processing backend and to customize Kadi4Mat for the use of large synchrotron data series. First, automated on-the-fly (during measurement) data ingestion including metadata storage at some large scale data facility needs to be implemented. To work with the stored data, Kadi4Mat already involves an interface for the creation of data processing pipelines. The next goal is to enable users to define the X-ray image processing pipelines by the Kadi4Mat's interface and to integrate the UFO framework as a backend for their actual execution. Suitable 3D data visualization via the web interface will round up this project.

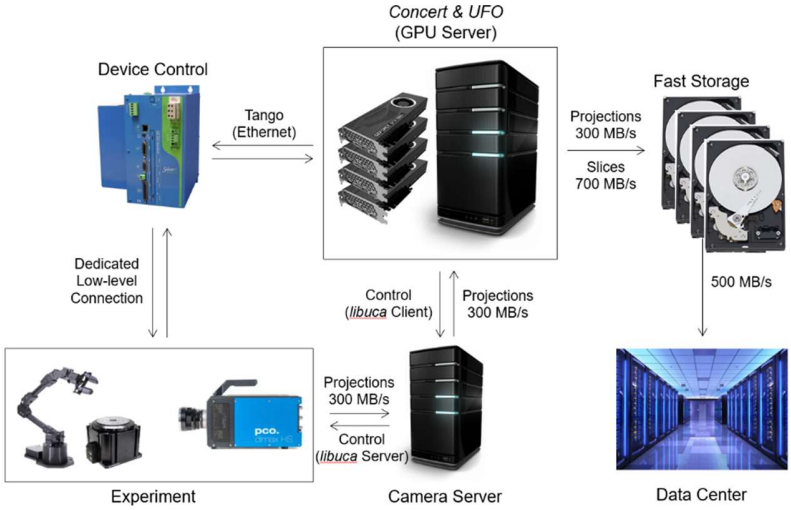

Figure 1: KARA's data acquisition system

[1] van de Kamp, Thomas, et al. "Parasitoid biology preserved in mineralized fossils." *Nature communications* 9.1 (2018): 1-14.
[2] Vogelgesang, Matthias, et al. "UFO: A scalable GPU-based image processing framework for on-line monitoring." *2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems*. IEEE, 2012.
[3] Brandt, Nico, et al. "Kadi4Mat: A Research Data Infrastructure for Materials Science." *Data Science Journal* 20.1 (2021).