

# Development and Application of the HEPS Scientific Data Processing Software Framework

*Thursday, 20 March 2025 16:54 (18 minutes)*

The High Energy Photon Source (HEPS) is a new fourth-generation high-energy synchrotron radiation facility, scheduled to become fully operational by the end of 2025. Compared to previous generations, it features significant advancements in brightness and detector performance. In its phase I, HEPS plans to construct 14 beamlines, with an estimated annual experimental data volume exceeding 300 PB. The total data scale is expected to surpass the EB level in a short period. HEPS supports a wide range of experimental techniques, including imaging, diffraction, scattering, and spectroscopy, each with significant differences in data throughput and scale. Meanwhile, the emergence of increasingly complex experimental methods poses unprecedented challenges for data processing.

To address the future EB-scale experimental data processing demands of HEPS, we have developed DAISY (Data Analysis Integrated Software System), a general scientific data processing software framework. DAISY is designed to advance integration, standardization, and high-performance in HEPS experimental data processing. It provides key capabilities, including high-throughput data I/O, multimodal data parsing, and multi-source data access. It supports elastic and distributed heterogeneous computing to accommodate different scales, throughput levels, and low-latency data processing requirements. It also offers a general workflow orchestration system to flexibly adapt to various experimental data processing modes. Additionally, it provides user software integration interfaces and a development environment to facilitate the standardization and integration of methodological algorithms and software across multiple disciplines.

Based on the DAISY framework, we have developed multiple domain-specific scientific applications, covering imaging, diffraction, scattering and spectroscopy, while continuously expanding to more scientific domains. Furthermore, we have optimized key software components and algorithms to significantly improve data processing efficiency. At present, several DAISY-based scientific applications have been successfully deployed on HEPS beamlines, supporting online data processing for users. The remaining applications are scheduled for fully deployment within the year, further strengthening HEPS's data analysis capabilities

**Presenter:** HU, Yu (Chinese Academy of Sciences)

**Session Classification:** Data Management & Big Data