JOURNAL OF SYNCHROTRON RADIATION
ISSN 1600-5775

# The design of a data management system at HEPS

Hao Hu,* Fazhi Qi, Hongmei Zhang, Haolai Tian and Qi Luo

Computing Center, Institute of High Energy Physics, 19B Yuquan Road, Shijingshan District, Beijing 100049, People's Republic of China. *Correspondence e-mail: huhao@ihep.ac.cn

Received 18 May 2020
Accepted 13 November 2020

Edited by M. Wang, Paul Scherrer Institute, Switzerland

**Keywords:** data management; high-energy photon sources; metadata ingestion.

According to the estimated data rates, it is predicted that 24 PB raw experimental data will be produced per month from 14 beamlines at the first stage of the High-Energy Photon Source (HEPS) in China, and the volume of experimental data will be even greater with the completion of over 90 beamlines at the second stage in the future. To make sure that the huge amount of data collected at HEPS is accurate, available and accessible, an effective data management system (DMS) is crucial for deploying the IT systems. In this article, a DMS is designed for HEPS which is responsible for automating the organization, transfer, storage, distribution and sharing of the data produced from experiments. First, the general situation of HEPS is introduced. Second, the architecture and data flow of the HEPS DMS are described from the perspective of facility users and IT, and the key techniques implemented in this system are introduced. Finally, the progress and the effect of the DMS deployed as a testbed at beamline 1W1A of the Beijing Synchrotron Radiation Facility are shown.

## 1. Introduction

China's High-Energy Photon Source (HEPS), the first national high-energy synchrotron radiation light source and soon to be one of the world's brightest fourth-generation synchrotron-radiation facilities, has been in construction since 29 June 2019 in Beijing's Huairou District and will be completed in 2025.

With the rapid development of synchrotrons, new techniques increase the amount of raw data collected during each experiment. According to the estimated data rates (shown in Table 1) provided by beamline scientists, we predict that 24 PB of raw experimental data will be produced per month from 14 beamlines at the first stage of HEPS, and the volume of experimental data will be even greater with the completion of over 90 beamlines at the second stage in the future.

Traditionally, data are collected and saved to the hard drive of the local beamline server directly without a data management system (DMS). For example, at the Beijing Synchrotron Radiation Facility (BSRF), which is a first-generation synchrotron radiation facility in the Institute of High Energy Physics (IHEP), users have to write down all the useful experimental environment information about the data on their notebook during beam time, manually copy the large amount of data from the local beamline disk to portable hard drives and then spend several months analyzing the data on personal computers to obtain few results that are suitable for publication, which severely limits the scientific research output of synchrotrons. Instead of writing experimental data on a user notebook, electronic logbooks are routinely used at newer synchrotron facilities.

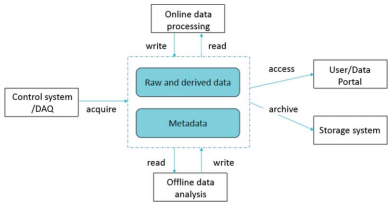

Table 1
Estimated data rates of 14 beamlines of HEPS.

| Beamlines | Burst output (bytes day$^{-1}$) | Average output (bytes day$^{-1}$) |
|---|---|---|
| B1 Engineering materials beamline | 600 TB | 200 TB |
| B2 Hard X-ray multi-analytical nanoprobe (HXMAN) beamline | 500 TB | 200 TB |
| B3 Structural dynamics beamline (SDB) | 8 TB | 3 TB |
| B4 Hard X-ray coherent scattering beamline | 10 TB | 3 TB |
| B5 Hard X-ray high-energy-resolution spectroscopy beamline | 10 TB | 1 TB |
| B6 High-pressure beamline | 2 TB | 1 TB |
| B7 Hard X-ray imaging beamline | 1000 TB | 250 TB |
| B8 X-ray absorption spectroscopy beamline | 80 TB | 10 TB |
| B9 Low-dimension structure probe (LODISP) beamline | 20 TB | 5 TB |
| BA Biological macromolecule microfocus beamline | 35 TB | 10 TB |
| BB Pink small-angle X-ray scattering | 400 TB | 50 TB |
| BC High-resolution nanoscale electronic structure spectroscopy beamline | 1 TB | 0.2 TB |
| BD Tender X-ray beamline | 10 TB | 1 TB |
| BE Transmission X-ray microscope beamline | 25 TB | 11.2 TB |
| BF Test beamline | 1000 TB | 60 TB |
| Total average | | 805.4 TB day$^{-1}$<br>24.16 PB month$^{-1}$ |

Because of the huge increase in data volume at HEPS, copying data manually and analyzing data on a personal computer are no longer possible. Obviously, an efficient DMS, which is capable of automating the organization, transfer, storage and distribution of the collected data, is quite necessary and essential for HEPS. With a DMS, the raw data produced from each experiment are acquired and saved to the storage located in the central computing centre, the metadata are ingested and stored in the metadata database automatically. The experimental users can access and analyze the data online via a web service or other access protocol interfaces.

After a preliminary investigation of DMSs of other synchrotron facilities at home and abroad, we give the detailed design of a DMS at HEPS, including data policy, data flow, metadata catalogue framework and deployment architecture, and the approaches of metadata acquisition.

## 2. Scientific data policy

Scientific data policy (EuXFEL, 2017) should be brought forward primarily since it is a guideline for the design and implementation of data management. The governments all over the world are strengthening the construction of policies and laws related to data protection. The Chinese government also released the 'Measures Of Science Data Management' (General Office of the State Council, 2018) in March 2018. However, no practices and regulations about data openness and sharing are available for user facilities so far in China.

Despite the lack of legal support to make the data open to other users after an embargo period, functions of DMSs which enable the publicity of data should be provided.

According to the national laws and regulations relevant to scientific data, we finished a draft version of the scientific data policy for HEPS, which will be discussed and approved by the HEPS council. The main elements of the scientific data policy comprise general principles, data classification, data ownership, data curation and archiving, and data access and sharing.

### 2.1. General principles

The scientific data policy pertains to the ownership, curation, archiving, and access to scientific data and metadata collected and stored at HEPS. Of course, this scientific data policy will be governed by and constructed in accordance with the law of China. The acceptance of the scientific data policy must be a condition for the award of HEPS beam time.

### 2.2. Data classification

Scientific data are the data collected from experiments performed on HEPS instruments. Metadata is information collected in relation to the scientific data, including (but not limited to) the information regarding the context of the experiment, the experiment group, experiment conditions and other logistical information. All the scientific data are classified into various types according to the data lifecycle phase.

(*a*) Raw data means a category of the scientific data that is recorded during experiments and registered in the metadata catalogue.

(*b*) Processed data means a category of the scientific data that is derived from raw data after some data processing.

(*c*) Calibrated data means a subcategory of the processed data that is obtained from the raw data by applying detector-specific corrections.

(*d*) Calibration data means a subcategory of the processed data that describes detector correction.

(*e*) Result data is a subset of processed data and other outcomes arising from the analysis of raw data, excluding publications and intellectual property rights based on such analysis.

### 2.3. Data ownership

HEPS is defined as the custodian of raw data and metadata in the data policy. This means HEPS will provide users with automatically collected metadata for all experiments carried out on its beamlines. The metadata will be stored in the metadata catalogue database, which can be accessed online to browse and download data. The experimental team will have sole access to the data during the embargo period. After the embargo, the data will be released with open access to any registered users of the HEPS data portal.

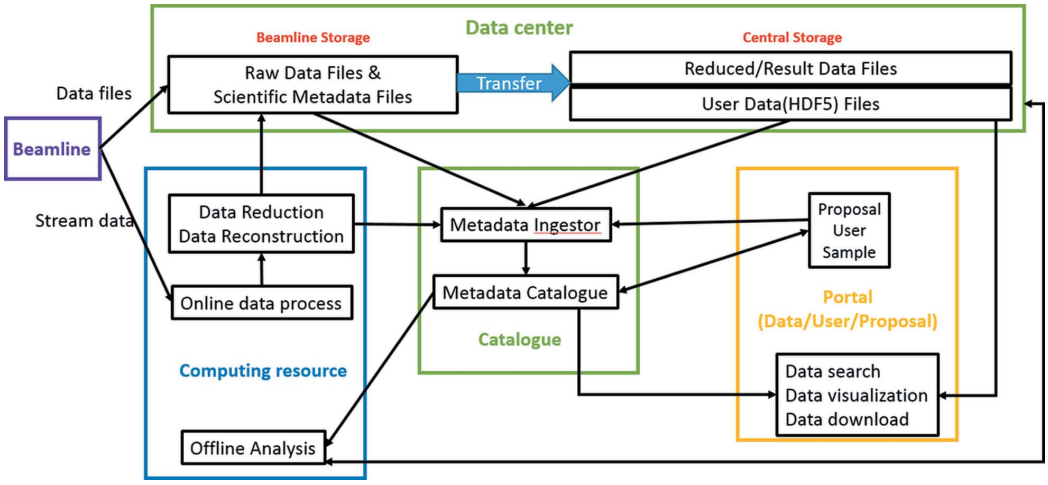


**Figure 2**
Data flow.

### 2.4. Data curation and archiving

(i) It is recommended that long-term storage for data produced from experiments carried out on beamlines should be provided, which includes disk storage for at least three months and permanent tape archive. Certainly, how the storage policy is adjusted depends on the final funding eventually.

(ii) All the data should be curated in well defined formats, and the means of reading the data will be provided.

(iii) Datasets are a collection of files produced during a data-taking run. Each dataset has a unique persistent identifier (PID). HEPS will use the PID service (http://en.pid21.cn) provided by the Computer Network Information Centre (CNIC) of the Chinese Academy of Sciences (CAS), which supports two kinds of data identifiers, PID21 (following the international Handle21 standard) and the Chinese Science and Technology Resource (CSTR). Any publications related to raw data and metadata from HEPS must cite the PID of data.

### 2.5. Data access and sharing

Users of HEPS must register at the user management and service system. Access to the data and metadata catalogue of HEPS will be restricted to registered users. All raw data and the associated metadata obtained as a result of public research will be made available as open access after an embargo period during which access is restricted to the experiment group.

## 3. Overview of HEPS DMS

### 3.1. Interfaces with other systems

As the central and fundamental part of IT systems, a DMS has interfaces with other systems, which are shown in Fig. 1. For data acquisition, the DMS collects data from the control system or DAQ (Data Acquisition) and saves them to the storage system. For data analysis, an online data processing system and an offline data analysis system read data files from storage through the metadata catalogue of the DMS, and write the processed data back to storage when the data analysis is finished. Users can access data files on the data service portal.

### 3.2. Data flow

The data flow is shown in Fig. 2. When an experiment starts, raw data files and metadata files produced from beamlines are saved to beamline storage; meanwhile, the metadata ingestor collects all the metadata from the control system or metadata files, which depends on how metadata are provided. At the same time, raw data are sent to the online data processing system as the input in the form of stream or data files. After data reduction or data reconstruction is carried out, the reduced data files and processed data files are produced. The data transfer service is responsible for monitoring the beamline storage device for new files and transferring new or

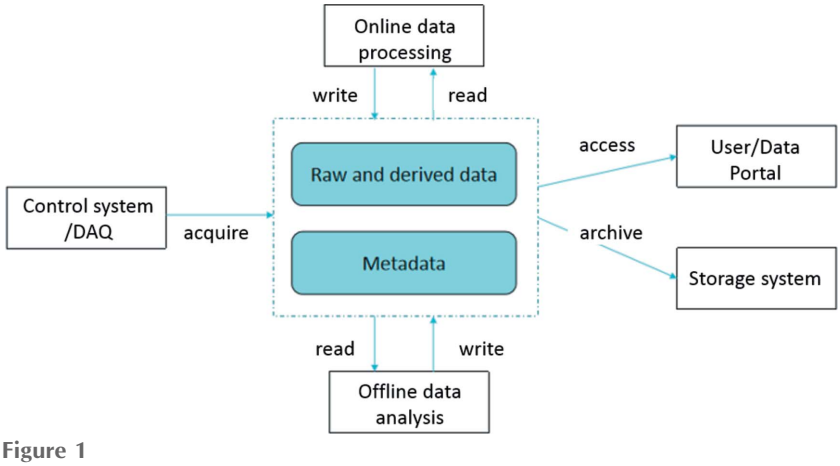


**Figure 1**
DMS interfaces with other systems.

modified data files to the central storage. The metadata ingestor keeps track of the metadata associated with a data file, such as its file size, MD5 checksum value and original location in the file system. After the experiments, data files can be accessed from the central storage under user authorization and access control. Actually, users and administrators can access the DMS via a web portal, a desktop graphical user interface (GUI) or a full set of command-line tools, as well as via Python or other application programming interfaces (APIs).

## 4. Predefined directory on storage

Several data files, including raw data files, metadata files, processed data files and result data files, are produced after each run of an experiment. To organize all the data files in the storage more efficiently, the directory structure should indicate the beamline, date, beam time ID and the data type. Fig. 3 shows the design of the directory structure for storage, which refers to the design of DESY (Rothkirch, 2018).

## 5. Metadata

Metadata are classified into three categories according to different purposes and sources: administrative metadata, scientific metadata and other metadata. Administrative metadata refers to the information about data ownership, data-management lifecycle and the location of data files in storage. Administrative metadata ingested from the proposal system and the user service system can be registered into the catalogue database. Scientific metadata describes the sample, beamline and experiment condition parameters relevant for the data analysis. Scientific metadata will be registered to the catalogue database and also saved into a standard HDF5 data file for permanent storage. Other metadata are collected from file-transfer tools and data-analysis softwares; these metadata are associated with data files such as MD5 checksum values, analysis software versions and update times.

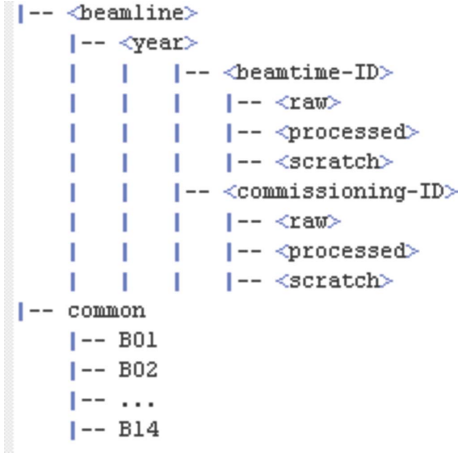


**Figure 3**
The directory structure for storage.

### 5.1. Scientific data schema

As described above, scientific metadata are technique-specific, rich and complex, which is defined by beamline scientists and finally approved by the HEPS Council. To make data interoperable (the I in FAIR data principles, which means making scientific data findable, accessible, interoperable and reusable), metadata are stored following the NeXus conventions (Dimper *et al.*, 2019). All experiments from all beamlines therefore produce HDF5 files using a common NeXus schema. The schema, which contains a generic part and a beamline-specific part, is shared among all beamlines. The generic part applies to all beamlines and describes beamlines and environment configurations. The beamline-specific part strongly depends on the technique used for carrying out the experiment. We hope that this schema can support the following techniques: tomography, fluorescence, Kmap, crystallography, electron microscopy, ptychography and microbeam-radiation therapy.

### 5.2. Metadata catalogue framework SciCat

SciCat (Gwilliams & Egli, 2017; Wang, Steiner & Sepe, 2018), which is developed by the Paul Scherrer Institute (PSI), the Europe Spallation Source (ESS) and MAX IV, is an open source (https://github.com/scicatproject) framework aimed at the management of the whole data lifecycle. Micro-service architecture with the latest technologies is applied in this framework. As metadata models are defined in JSON format, relevant RESTful APIs are exposed via a *NodeJS* web server. Metadata are stored in a MongoDB database and RESTful APIs can be accessed with any modern web browser. From the web front end (based on *Angular8*), data can be accessed with any modern web browser for all metadata. Fig. 4 shows the architecture of SciCat.

SciCat has high scalability and performance because it applies a message queue and a documental database. It is estimated that HEPS will produce ∼100000 records of metadata per day, which means that using a NoSQL database is more beneficial than using a SQL database. Additionally, some experimental metadata are technique specific and complex so a document database could be a better choice. Therefore, we choose SciCat for metadata cataloguing.

### 5.3. Metadata acquisition

Whichever approach we choose to collect the metadata from beamlines depends on how these metadata are provided. Currently, all the projects of HEPS are in the design phase, some detectors are self-designed and others are purchased from commercial companies. Therefore, we give three schemes to collect metadata based on the investigation of the metadata acquisition.

(1) It is recommended that the DMS provides interfaces for the DAQ/control system to register metadata. This is suitable for a self-designed detector since one can develop the function modules of the control system, which are implemented to send JSON-format metadata to the DMS by making use of open-

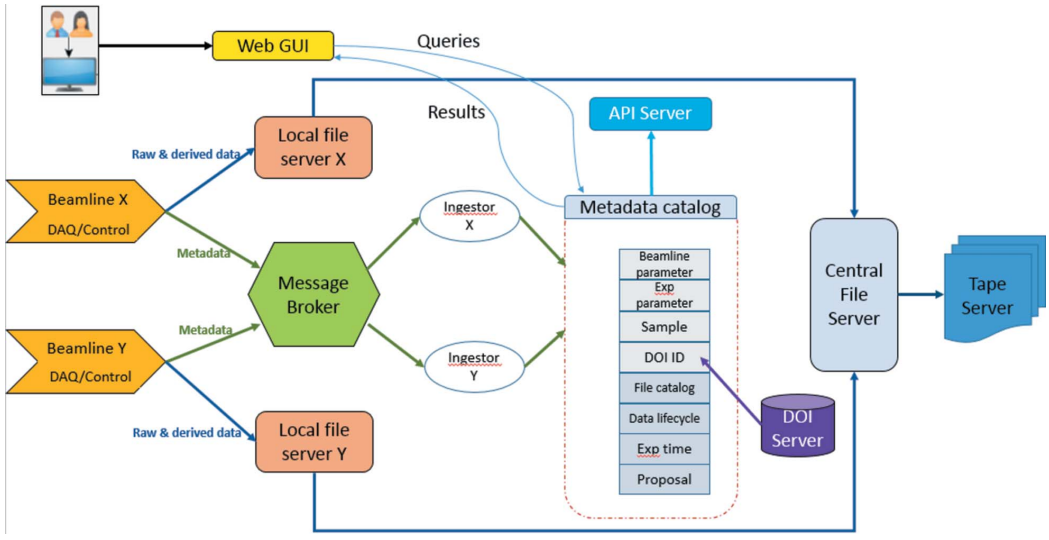


**Figure 4**
The architecture of the metadata catalogue framework SciCat.

source software like *Epics* (https://epics.anl.gov/) or *Bluesky* (https://nsls-ii.github.io/bluesky/).

(2) If the detector is a commercial product and the metadata are collected and saved to files in TXT or NXS format, we will develop a metadata ingestor which will be responsible for extracting all the metadata from these files. The metadata ingestor will run in a flexible plugin mode (Schwarz *et al.*, 2019). In addition to user and experiment information, the plugin will also collect information such as file size, checksums, processing time and file replicas.

(3) If the DAQ has an independent database, another metadata ingestor can ingest metadata from the DAQ database directly and save them to the metadata catalogue database.

### 5.4. Metadata cataloguing architecture

Each metadata ingestor is dedicated to each detector; however, we have a unified architecture to realize the metadata acquisition and cataloguing functions (Fig. 5). No matter how metadata are collected, a *Kafka* cluster, which is a message broker, is applied to receive messages from the metadata ingestor or control system. The metadata creator has two functions, one is interpreting messages received from *Kafka* into metadata items, the other is complementing other missing metadata items from the user database, proposal database and sample database. Finally, all the metadata needed for datasets is saved to MongoDB through APIs of SciCat.

## 6. File-format standardization

There exist many formats of data files produced from experiments carried out at beamlines because of the discipline preference and the limit of detector outputs. We hope to supply users with data files in a standard data format for long-term storage in the future.

After investigation and research, HDF5 (HDF Group, 2018) is chosen as the standard data-file format. HDF5 is a file format used for large quantities of numerical data, which is suitable for the management of extremely large and complex data collections. Furthermore, the HDF5 file format follows

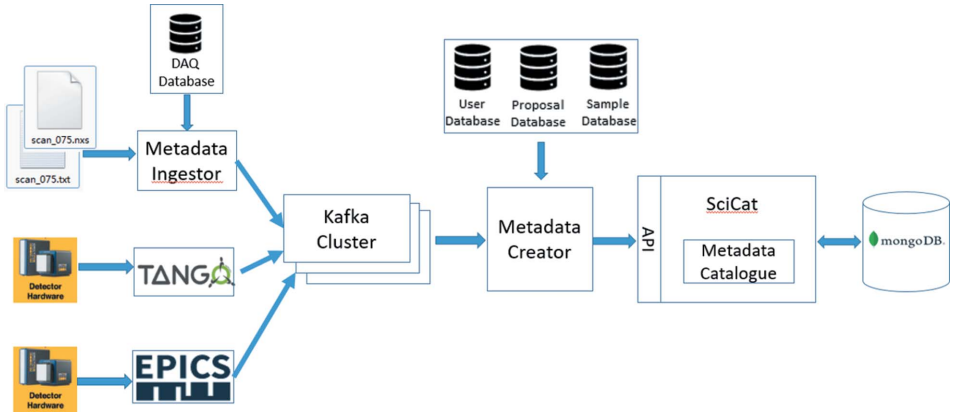


**Figure 5**
The metadata cataloguing architecture.

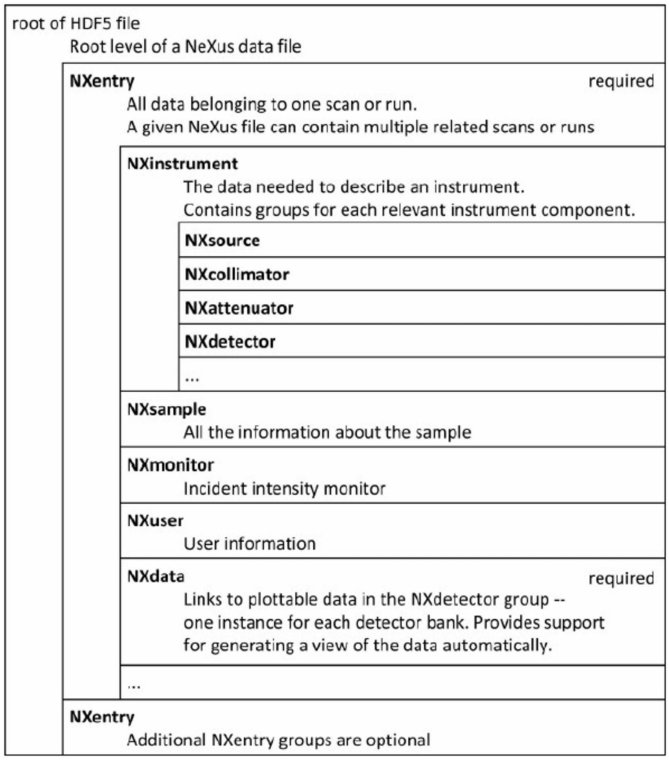


**Figure 6**
HDF5 file schema reference.

NeXus conventions which include NeXus base class definitions and NeXus application definitions. The detailed schema design will refer to the HDF5 and NeXus data format used at the Swiss Light Source (Watts, 2019) (shown in Fig. 6).

## 7. DMS testbed at 1W1A of BSRF

BSRF is a running facility for synchrotron radiation experiments, and also provides the technology R&D and test platforms for HEPS. In order to verify the metadata catalogue functions, we have set up a DMS testbed at 1W1A, which is a diffuse X-ray scattering beamline at BSRF.

The Eiger area detector of 1W1A produces ~170 GB of raw data per day. Although the data volume is not big, we hope to build a testbed environment close to the design, such that one can verify not only the DMS but also the whole process of data acquisition, data transfer, data storage and data access. The testbed architecture is shown in Fig. 7 with key parts explained below.

(*a*) To obtain a stable data-transfer rate, the network bandwidth is upgraded from 100 Mb $s^{-1}$ to 1 Gb $s^{-1}$.

(*b*) We use a 32 TB network attached storage (NAS) with a network file system as local beamline storage in addition to a hard disk on the beamline control server. For central storage, we use four servers and one 80 TB disk array with a Lustre file system located at Computing Center.

(*c*) Two plugins running on a server located at the beamline are responsible for metadata ingesting and data transferring. One plugin, named 'metadata ingestor', collects metadata associated with datasets and data files, including the proposal ID related to the datasets, data file size, data location of the file system, checksum and so on. Another plugin monitors for new data files in a designated folder of NAS while moving the produced data files to the central storage.

(*d*) Metadata associated with datasets and data files are stored in the metadata database, which is backed by MongoDB. The metadata can be accessed, updated and deleted via REST interfaces of the DMS cataloguing service.

(*e*) A simple data-service web portal is deployed for authorized users and administrators to access data.

The testbed is very meaningful for the design and implementation of the DMS. From this practice, we can understand the meaning of each metadata item in SciCat very well and discuss with in-house scientists of BSRF which scientific metadata really needs to be catalogued. At present, however, the DMS does not integrate with any computing resource or facilities – users can only process data using their own computers offsite.

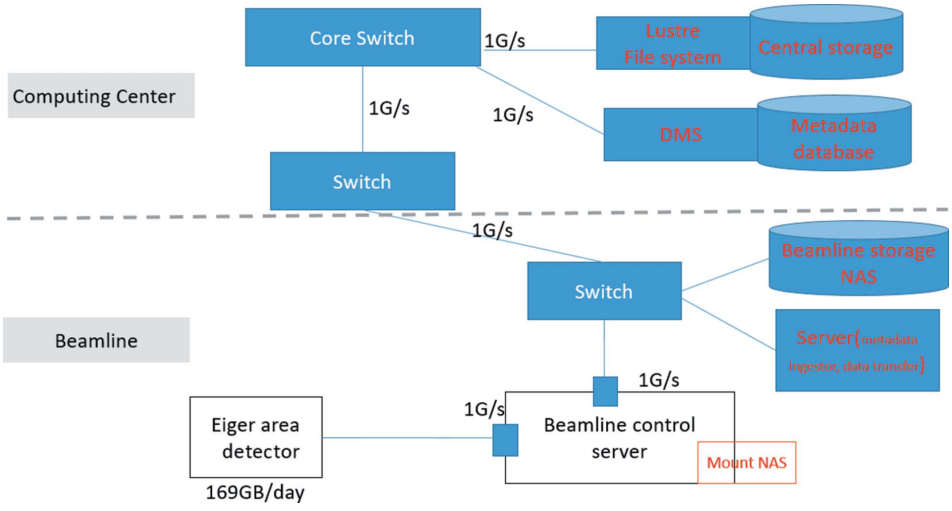


**Figure 7**
The testbed architecture at 1W1A.

## 8. Conclusions and future plan

Metadata cataloguing functions of the DMS designed for HEPS have been implemented on the testbed. For the next step, we need to improve and optimize the design scheme of the DMS, including ensuring the high availability and reliability of the system, as well as specifying the HDF5 data-format schema for each experiment technique.

Furthermore, another testbed at beamline 3W1 of BSRF dedicated to high-throughput instruments has been set up. Integrating with the data-analysis framework and computing system, it will verify the design and performance of the IT infrastructure. The HEPS data-analysis framework, using a *Jupyter-Spark-Docker* technology stack, can conduct *in situ* data reduction and data reconstruction. A dynamic load balance of computing resources is achieved by using a *Kubernetes*-based computing system. To obtain preferable performance with low latency and large-scale serving capabilities, the work on this testbed will directly contribute to HEPS in the future.

Considering that HEPS will produce ∼24 PB data per month, which is really a huge challenge for IT infrastructure and DMSs, we will use modular and scalable architecture in IT systems design. For the testbed, system functions and work flows are implemented and verified. In the future production environment, we hope to scale up the physical resources to meet the performance requirement of HEPS.

## Funding information

This work was supported by the National Natural Science Foundation of China under grant No. 12005247.

## References

Dimper, R., Götz, A., de Maria, A., Solé, V. A., Chaillet, M. & Lebayle, B. (2019). *Synchrotron Radiat. News*, **32**(3), 7–12.
EuXFEL (2017). *Scientific Data Policy of European X-ray Free-Electron Laser Facility GmbH*, https://in.xfel.eu/upex/docs/upex-scientific-data-policy.pdf.
General Office of the State Council (2018). *Notice of the General Office of the State Council on Issuing the Measures for the Management of Scientific Data*, No. 17 of the General Office of the State Council, China.
Gwilliams, C. & Egli, S. (2017). Paul Scherrer Institute, SciCat Project: Data Catalog System, https://icatproject.org/wp-content/uploads/2017/12/ICAT_F2F_2017_PSI.pdf.
HDF Group (2018). *HDF5*, http://www.hdfgroup.org/HDF5/, accessed on May 2018.
Rothkirch, A. (2018). *12th NOBUGS Conference*, 22–26 October 2018, Brookhaven National Laboratory, Upton, New York, USA.
Schwarz, N., Veseli, S. & Jarosz, D. (2019). *Synchrotron Radiat. News*, **32**(3), 13–18.
Wang, C., Steiner, U. & Sepe, A. (2018). *Small*, **14**, 1802291.
Watts, B. (2019). *HDF5 and the NeXus Data Format*, *HDF5 European Workshop for Science and Industry*, 17–18 September 2019, ESRF, Grenoble, France, https://indico.esrf.fr/indico/event/33/.