DATA MANAGEMENT INFRASTRUCTURE FOR NEUTRON SCATTERING EXPERIMENTS IN J-PARC/MLF

Kentaro Moriyama and Takeshi Nakatani

J-PARC Materials and Life Sciences Experimental Facility

Abstract

The role of data management is one of the greatest contributions in the research workflow for scientific experiments such as neutron scattering. The facility is required to safely and efficiently manage a huge amount of data over the long duration, and provide an effective data access for facility users promoting the creation of scientific results. In order to meet these requirements, we are operating and updating a data management infrastructure in J-PAPC/MLF, which consists of the web-based integrated data management system called the MLF Experimental Database (MLF EXP-DB), the hierarchical raw data repository composed of distributed storages, and the integrated authentication system. The MLF EXP-DB creates experimental data catalogues in which raw data, measurement logs, and other contextual information on sample, experimental proposal, investigator, etc. are interrelated. This system conducts the reposition, archive and on-demand retrieve of raw data in the repository. Facility users is able to access the experimental data via a web portal. This contribution presents the overview of our data management infrastructure, and the recent updated features for high availability, scaling-out, and a flexible data retrieval in the MLF EXP-DB.

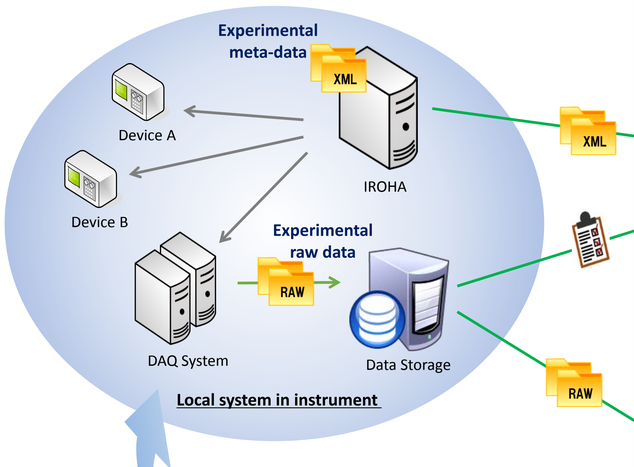
Data Management Infrastructure

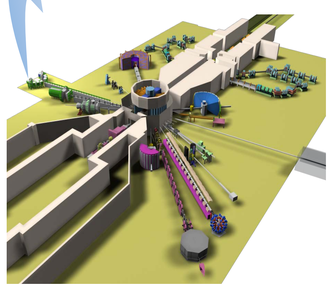
J-PARC/MLF

- J-PARC/MLF is a neutron scattering experimental facility providing one of the highest intensity beam in the world.
- About a thousand users annually perform several hundreds experiments in wide variety of research fields using twenty neutron instruments.
- The total amount of data produced annually in the facility at full performance is on the order of petabyte.

Local system in neutron instrument

- Each neutron instrument is equipped with large-area neutron detectors and various sample environmental devices.
- The MLF control software framework “IROHA” coordinates measurement with controlling data acquisition and various sample environmental devices.
- IROHA creates measurement log as an **experimental meta-data** indicating measurement condition in XML format.
- DAQ system creates **experimental raw data** in the data storage.

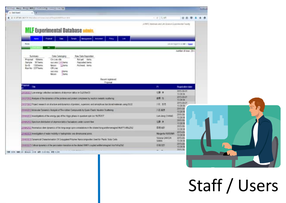




MLF Experimental hall

Web portals

Web portals are the front-end of the data management infrastructure. It supports the data management of staff and data utilization of facility users.



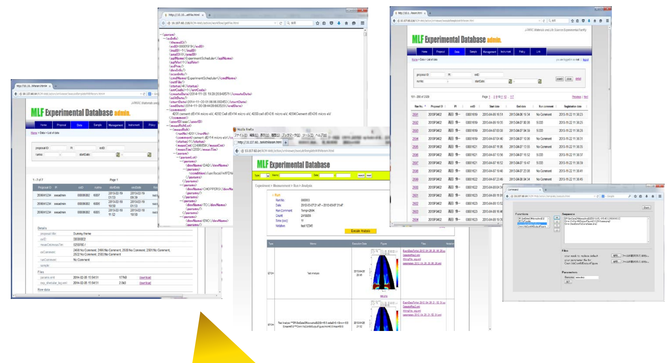
Staff / Users

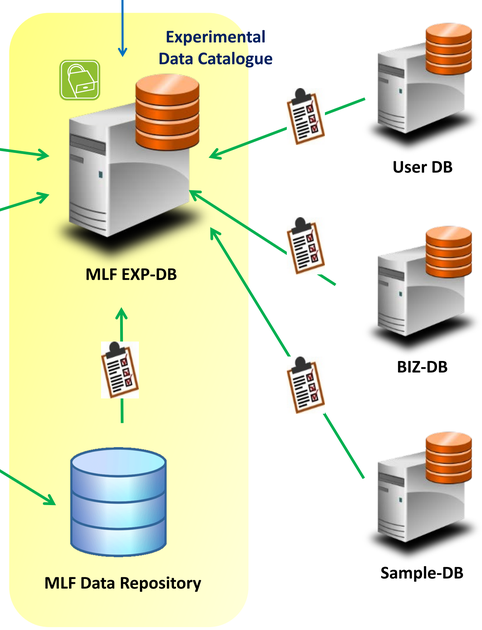
Data management portal for staff

- Data catalogue management
- Data registration (on-line, off-line)
- Raw data management
- Quality determination of measurement
- annotation providing

Data access portal for users

- Data catalogue browsing
- Searching and Downloading dataset
- Standard data reduction and visualization
- Annotation providing
- Data sharing



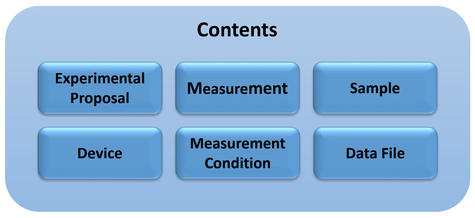


MLF Experimental Database

MLF EXP-DB is the core system of the infrastructure, which is responsible for data managing and access.

Experimental Data Catalogue

Collecting the experimental data and associated data, the system creates the experimental data catalogue.



Raw Data Management

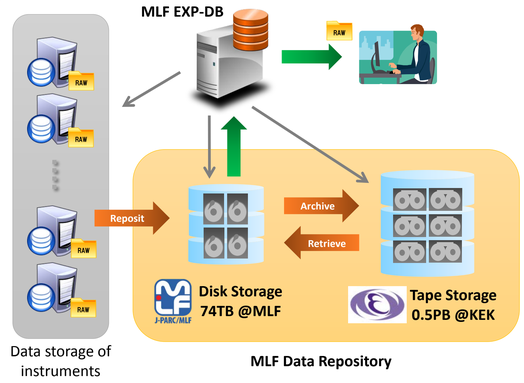
Raw data is centrally stored in the MLF Data Repository. This repository is composed of two data storage separately located. **MLF EXP-DB** manages raw data within this repository in a hierarchical way. It performs “**Reposit**”, “**Archive**”, and “**Retrieve**” of raw data. The system transfers raw data to users on-demand via a web portal. This management scheme enables effective utilization of limited storage resources.

Web portals

MLF EXP-DB provides web portals for facility staff and users. All of data management and access is performed via web portals.

System Architecture

- A Java-based commercial database software. “R&D Chain Management System Software (RCM)”
- XML-based Database **Flexibility**
- An integrated 3-Layer Web system
- XML-based Workflow Engine
- HTML Engine with XSLT
- Secure Access Control **Security**



Recent Updated Features

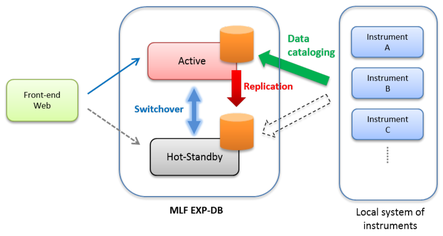
Requirements

Recently, we have improved **MLF EXP-DB** responding to the expansion and growth of experimental environment based on an improvement plan in the facility.

- **High availability** avoiding a service outage owing to the system failure and maintenance.
- **Scalability** of the data cataloging performance responding to the enhancement of data production rate.
- **Flexible retrieval function** for data with experimental condition from a large amount of data, promoting the effective data utilization in the web portal.

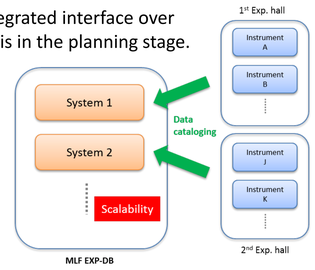
High Availability

- Improved to the Redundant distributed system
- Switch over relationship between two systems
- Active / Hot-Standby
- DB replication



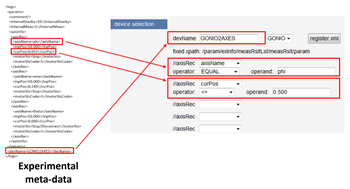
Scaling-out

- Scaling-out for the load distribution of the data cataloguing.
- Current system is composed of two redundant system.
- Static allocation of the data cataloging.
- Development of an integrated interface over the distributed system is in the planning stage.



Flexible Data Retrieval

- Experimental condition is recorded in experimental meta-data in XML-format. The structure of experimental meta-data can be changed depending on the sample environmental devices.
- Flexible data retrieval by specifying the search conditions for each tag of experimental meta-data.



XML measurement experiment Database Archive management WEB RAW DATA METADATA XSLT