Data Management Plan for Hall-B

Authors: *Veronique Ziegler* **Contact** *Veronique Ziegler*
Last reviewed/updated: August 7, 2015

Summary: The Jefferson Lab data management plan for Hall B document details the lab's plan to responsibly manage the scientific data recorded by the CLAS12 detector. This document sets out the plan of experimental hall B and is intended as a reference for the plans of upcoming experiments using CLAS12 detectors. The Hall Leader is Volker Burkert, the CLAS12 Software Project Manager is Veronique Ziegler. Harut Avakian is managing CLAS accounts. Gagik Gavalian is handling the development of the framework for CLAS12 reconstruction, calibration and monitoring and playing a major role in event reconstruction.

Responsibilities: With the assistance of the Scientific Computing group in IT division the Experimental Nuclear Physics (ENP) division management is responsible for the data management of nuclear physics data. The maintenance of this document, the plan that it describes and its implementation are the responsibility of the Software Management team for Hall B.

Experimental Nuclear Physics Data Management processes: The data management processes are listed as follows according to the broad categories of data that they address:

Raw Data: Newly acquired raw data is stored on disk and moved to the tape library in a timely fashion using tools provided by IT division. IT division also makes a duplicate copy of raw data at a later date on tapes that are removed from the library and stored. The raw data format chosen by HALLB is the EVIO format, which was developed by Jefferson Lab scientists.

Processed Data: Processed data is initially stored on disk and migrated to tape using IT tools as required. The raw data from the CLAS12 detector are stored on tape, at a rate of about 5-10 TB/day, with information on the particles as they transverse the detector components. The processed data are also stored on tape for analysis by members of the Hall B research community to analyze. Processed data is in EvIO format which will be analyzed with Groovy and Java scripts included in the CLAS12 reconstruction and analysis framework.

Run Conditions: Run conditions (machine energy, polarization and intensity, target, etc.) are stored in the experiment logbook and in database called CCDB (Calibration & Constants Database).

Databases: Database servers are managed by IT and regular snapshots of the database content are stored along with the tools and documentation required for their recovery.

Log Books: Jefferson lab uses an electronic logbook system with a database back-end.
Calibration and Geometry databases: Running conditions, as well as the detector

calibration constants and detector geometries are stored in the CCDB database at Jefferson Lab.

Other databases: Other databases may be relevant to data management, for example the JInventory database tool that catalogs which electronic modules were in the online systems.

Analysis software source code and build systems: Data analysis software is developed within the CLAS reconstruction and analysis package. Contributions to the package are from several sources, lab staff and users, off-site lab collaborators and third parties. Locally written software source code and build files, along with contributions from collaborators are stored in a version management system, git. Third party software is managed by software maintainers under oversight of the Software Support Committee. Source code repositories and managed third party packages are backed up by IT.

Documentation: Documentation is available online in the form of content either maintained by a content management system (CMS) such as a Wiki or Drupal or as static web pages. This content is backed up by IT. Source code documentation is part of the software through Doxygen (C++) and Javadocs (Java). Other documentation for the software is distributed via wiki pages, and consists of a combination of html and pdf files. Documentation LaTeX source files are stored in the source code repository under a subheading "docs". Maintenance of the wiki is performed by a small hall-B group.

Quality Assurance: As stated in the lab data management plan document, the data management plan process is overseen by the Deputy Director for Science. Periodic reviews of data management will be made. Quality Assurance of the software is ultimately the responsibility of an Analysis Coordinator and a committee selected from the collaboration to review reconstruction software.