

# Data Management Plan for Hall D

*Author: Mark Ito*

*Contacts : Mark Ito and David Lawrence*

*Last reviewed/updated: November 5, 2013*

**Summary:** The Jefferson Lab data management plan document details the lab's plan to responsibly manage the scientific data generated in connection with the lab's research program. This document sets out the plan of Experimental Hall D and is intended as a reference for the plans of individual experiments conducted in the Hall.

**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 scientific staff of Hall D.

**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.
- **Processed Data:** Processed data is initially stored on disk and migrated to tape using IT tools as required. Intermediate data files may or may not be archived on tape at the discretion of the researcher(s). It is not standard practice to duplicate processed data except in cases where it is moved offsite or as requested by the data owner.
- **Processing Meta-Data:** Histograms and log files created as a by-product of processing will also be archived to tape. Other meta-data will be resident in databases (see relevant section below).

**Run Conditions:** Run conditions, (machine energy, beam polarization and intensity, target type, etc.) are stored in the experiment logbook and/or other databases.

**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. Examples are :

- **Log Books:** Jefferson Lab uses an electronic logbook system with a database back-end.
- **Calibration databases:** These databases are operated by ENP staff and users but the servers are managed and backed up by IT.
- **Other databases:** There are several databases used by the online systems (for example the one recording run conditions) which are physically in the Hall D Counting House. Management practices follow closely those used by IT to maintain the calibration databases, using the same tools and backup regime.

**Analysis software source code and build systems:** Data analysis software is a combination of packages from several sources, lab staff and users, off-site lab collaborators and third parties. Examples of third party software are the ROOT and GEANT packages from CERN. Locally written software source code and build files, along with contributions from collaborators are stored in a version management system (Subversion). 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. Data analysis software is tagged and

released on a regular schedule and before major data analysis efforts to assure reproducibility of the code base in the future. Binaries from production runs are archived and stored in the tape library.

**Documentation:** Documentation is available online in several locations depending on the type of content.

- **Formal documents.** These are stored in a document database. DocDB is used. It is maintained by collaboration users at the University of Regina.
- **Web pages.** Most web content is contained in a wiki maintained and backed up by IT. Mediawiki is used. It is maintained by IT at JLab.

**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.