

Draft Data Management Plan for: Hall-A

Author: Jens-Ole Hansen, Hall A

Contact : Jens-Ole Hansen, Hall A

Last reviewed/updated: 20-Sep-2013

Summary: The Jefferson Lab data management plan document details the lab's plan to manage the scientific data generated in connection with the lab's research program. This document sets out the plan of the experimental hall identified above and is intended as a reference for the plans of individual experiments.

Responsibilities: With the assistance of the Scientific Computing group in the 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 hall leader. Each Hall-A research project is required to submit an experiment-specific data management plan, which should reference this document and must address the items specifically delegated to individual collaborations as well as any necessary additional items.

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: Raw data are stored initially on local, redundant disk systems in the counting house and are copied automatically, typically within 24 hours, to the tape library using local scripts and tools provided by IT division. Duplicate copies of raw data are automatically created on archival tapes by IT division facilities.

Processed Data: Management of processed data is to be addressed by each individual experiment's data management plan. It is recommended to duplicate or archive all processed data that are time-consuming to reproduce to the tape library in an automatic fashion (e.g. after completion of farm jobs).

Run Conditions: The primary reference for overall run conditions (machine setup, target, spectrometer configuration, etc.) is the Jefferson Lab electronic logbook. Certain run conditions are also stored as part of the EPICS/slow control data stream in the raw data files.

Logbooks: Hall-A will use the common Jefferson Lab electronic logbook system managed by the IT division.

Databases: Calibration and geometry data are stored either in flat text files or in SQL databases. Management of flat text file databases is to be addressed by experiment-specific data management plans. It is recommended to maintain a revision history of these files in a dedicated repository on the central Jefferson Lab version control system servers, where automatic backup is provided by IT. Experiments should give preference to storing databases in SQL format on the MySQL server operated by Hall-A. This server is managed by IT, and regular snapshots of the database content are stored along with the tools and documentation required for their recovery.

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 git repositories on github.com and SVN repositories on IT division servers. Third party software is managed by software maintainers under oversight of the Software Support Committee. Local source code repositories and managed third party packages are backed up by IT.

Documentation: Documentation is available online in a wiki and as static web pages of individual experiments. All web content is backed up by IT.

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.