Data Management Plan for: Hall-C

Author: Brad Sawatzky
Contact : Brad Sawatzky, Steven Wood
Last reviewed/updated: Sept. 6, 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 the experimental hall or collaboration identified above and is intended as a reference for the plans of individual experiments.

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

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 will also make 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.

Run Conditions: Run conditions, (machine energy, polarization and intensity, target, etc.) are stored in the experiment logbook, experimental data files, 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 for online operations. Collaborations are strongly encouraged to use this system for Offline analysis logbooks as well. External logbooks containing information relevant to analysis should be periodically snapshotted and/or mirrored to an IT controlled system.

Calibration and Geometry databases: These databases are operated by ENP staff and users but the servers are managed and backed up by IT.

Accelerator EPICS archiver: This database contains slow-controls logging data covering accelerator, beamline, and some experimental devices. It is managed and backed up by Accelerator IT staff. Data critical to experimental analysis should be incorporated into the Raw Data files or otherwise duplicated and stored along with the tools and documentation required for their recovery.

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. Such systems should be managed on IT servers and subject to their backup protocols.

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 form CERN. Locally written software source code and build files, along with contributions from collaborators are stored in a version management system, SVN or 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.

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.