

Data Management Plan

1. Types of Data

There are various types of data collected and/or generated during the course of experimental particle and nuclear physics research. These data are nearly exclusively produced electronically by Data Acquisition (DAQ) systems and computer programs and are digitally stored on a computer hard drive. The data are initially collected from various electronic devices, e.g. ADCs, TDCs, etc. which convert an electronic signal from an experimental detector or probe into a digitally store-able format. This initial data is referred to as raw data and is grouped into unique computer files which share common experimental conditions. These files typically have names which include a run-number indicating a chronological progression of collected data. Following the data collection, the raw data is calibrated, mathematically and statistically processed, and then saved into another computer file which also includes the run-number. The processed data is then typically further refined and consolidated to produce plots and/or numerical results which convey the outcome of the experiment. The format of the raw data is typically a custom file format generated by the DAQ system control software (such as CODA or MPANT). The desired data is extracted from the raw data using an analyzer C++ program which also processes and/or calibrates the data and typically stores it in a ROOT file (from a standard physics analysis software application called Root).

Along with the raw and processed data files there are typically other axillary data which are organized in a data base (almost exclusively a mySQL database). The database stores various types of data including calibration constants for the detector systems, translation maps to know which DAQ electronic signal pathways correspond to which detectors, and geometrical or spacial survey information indicating where in physical space the detector systems are located. These data are typically entered into the database automatically using C++/perl programs which are developed by members of the research group.

2. Data and Metadata Standards

The format of the raw data is documented in the DAQ control software documentation which could be a large formal document available online or could be a small simple file in the data repository named README. The ROOT file format is more customized to the particular experiment, but typically organizes the data into histograms and/or matrices called Trees or Ntuples. The format of the ROOT files are also documented in README files or simply detailed in a traditional or electronic logbook. As a general rule, the format of the above file types are designed to be consistent for all run-numbers of a given experiment. In addition, some raw data file formats such as those produced by MPANT include a header in the data file which defines the data columns and other information such as time-stamps, etc. However, the main source of information detailing the specific conditions during the run files is maintained and stored in a logbook.

3. Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy

There are no patent issues bestowing a proprietary nature to the data collected during the outlined research projects, therefore, there are no restrictions to its access. The data will be freely available to other groups internal to Jefferson Lab (JLab) by simply accessing the data repository; to groups outside JLab, the data may be made freely available by verbal or written request. As with most fundamental physics research, groups outside the experiment collaboration may not publish any results without the permission of said collaboration. There are no ethical or privacy issues with any of the data.

4. Policies and Provisions for Re-Use, Re-Distribution

Since there are no patent issues which restrict access to the data, the data will be made freely available to all with no restrictions or provisions for re-use or re-distribution.

5. Plans for Archiving and Preservation of Access

The long term strategy for maintaining, curating and archiving the processed data is to store it on a UNIX computer system with a RAID 1 disk storage system which is regularly backed-up to an additional RAID 1 disk on a central server system maintained by the ISU physics department computer administrator. In addition, all raw data, processed root file data, database data, software codes and scripts, and webpage source code will be stored in a robotic tape mass storage system at JLab and will be preserved for a period not less than 20 years after the data was collected. There are no major transformations that are necessary for preparing the data for preservation or for sharing the data. However, any documentation that is not in electronic form will be scanned into a pdf document and stored alongside the data.