

SpinQuest Data Management

Summary

This SpinQuest data management plan details the collaborations plan to responsibly manage the scientific data recorded by the SpinQuest experiment. This document lays out the SpinQuest collaboration plan for the experimental facility at Fermilab NM4 and is intended as a reference for the upcoming experiment (E1039) using the SpinQuest target and detectors. The Collaboration Chair and Spokesperson Dustin Keller (UVA) manages collaboration membership with the help of the SpinQuest Institutional Board and fellow Spokesperson Kun Liu (LANL). The Fermilab liaison manages safety and experimental hall activities. Fermilab badge and ID manages computing accounts. The collaboration is responsible for the software utilities used for reconstruction, calibration, and monitoring and all major aspects of event reconstruction.

Responsibilities

With the assistance of Fermilab IT, the SpinQuest Collaboration is responsible for data management at the NM4 facility including all target, spectrometer, and physics data. The maintenance of this document, the plan that it describes and its implementation is the responsibility of the Software Management team of SpinQuest formed by project leadership. This team is made up of the University of Virginia and Los Alamos National Labs as well as additional institutions that volunteer to take ongoing roles in this regard.

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 at Fermilab and copied to institutional storage in a timely fashion. This is done by copying the data to a UVA based server and then preserving the data on a RAID disk array owned and maintained by the UVA Spin Physics group.

The Fermilab E-1039/SpinQuest experiment is expecting to collect approximately 20 Tb of raw data between commissioning and the end of data acquisition. The raw data is subsequently processed and stored in ROOT files. In addition, there is a substantial volume of simulated, Monte Carlo events produced and stored at collaborating institutions and universities.

The raw data consists of event records from the CODA data acquisition system. These records contain the digitized hit information from the various detectors elements, including, for example, drift times from tracking chambers, hodoscope hits, scaler values, etc. The raw data from the SpinQuest detector are stored on disk, at a rate of about 0.3 TB/week, with information on the particles as they transverse the detector components as well as information on target polarization and target parameters. These data will be stored on-site at Fermilab in the experimental counting-house on a RAID disk array and backed up daily by the Fermilab Scientific Computing Division.

It is the SpinQuest Collaboration's policy that these raw data and processed data are available to collaboration members for use in collaboration-approved scientific studies and analyses. Completed analyses will be submitted for publication and shared with outside researchers. SpinQuest will maintain the ability to access these data indefinitely stored at UVA and as well as Fermilab maintaining a copy archived automatically.

Contribution from Fermilab Scientific Computing Division:

- Provide appropriate networking at NM4 hall including WiFi in both the counting area and detector hall for commissioning, data transfers to mass storage, network access for users' laptops, etc. Provide firewalls/bridges which Fermilab deems necessary to isolate the experiment's network from the general Fermilab network.
- Provide "General Computing" accounts for collaborators. Primary analysis and Monte Carlo computing will be done on LINUX-based PC's provided by the collaboration.
- Provide storage for 50 TB of raw data. The collaboration also plans to keep a second copy of the raw data on a RAID disk array stored and maintained at UVA.
- Support for 4 virtual machines.
- Access to grid resources including Open Science Grid and Fermigrid

Processed Data: After decoding the raw data is considered processed and is then stored into ROOT files at Fermilab and at the University of Virginia (UVA). The decoding takes the information in the raw CODA data records and translates them into a more user-friendly format, for example, assigning specific wire numbers in tracking chambers to digitized drift time information or hodoscope numbers to hits. Further processing then occurs on these data to change the hits into reconstructed tracks and events that are also stored in ROOT files. Processed data is initially stored on disk and migrated to institutional storage as required. The processed data will be approximately twice the size of the raw data after decoding. The processed data will be regularly copied to a UVA owned server and then copied to a RAID disk array owned and maintained by the Spin Physics group. The processed data are also stored on disk for analysis by members of the SpinQuest research community to analyze. Processed data is in a format that will be analyzed with a ROOT based reconstruction and analysis framework.

Run Conditions: Run conditions (machine energy, beam intensity, target polarization, etc.) are stored in the experiment logbook maintained by the Fermilab IT division and the SpinQuest collaboration.

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

Log Books: SpinQuest uses an electronic logbook system [SpinQuest ECL](#) with a database back-end maintained by the Fermilab IT division and the SpinQuest collaboration.

Calibration and Geometry databases: Running conditions, as well as the detector calibration constants and detector geometries, are stored in a database at Fermilab.

Analysis software source: Data analysis software is developed within the SpinQuest reconstruction and analysis package. Contributions to the package are from several sources, university groups, Fermilab 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 the oversight of the Analysis Working Group. Source code repositories and managed third party packages are continually backed up to the University of Virginia [Rivanna](#) storage.

Documentation: Documentation is available online in the form of content either maintained by a content management system (CMS) such as a Wiki in *Github* or Confluence pages or as static web pages. This content is backed up continually. Other documentation for the software is distributed via wiki pages and consists of a combination of *html* and *pdf* files.