SNOplus Data Management Policy

version 1, SNO+ Collaboration

2020

SNO+ data is managed centrally by the experiment. Raw data (the direct output of the DAQ), processed data (after reconstruction has been applied) and production Monte Carlo files are stored on grid enabled storage elements located in both Canada and the UK using the same infrastructure as LHC experiments. The LCG File catalogue (LFC) and dedicated couchDB databases store the details and location of each file. The software and procedures for simulating events, processing digitized data, accessing data, and extracting data are well documented and maintained for the collaboration to ensure reproducibility and backward compatibility. Simulations and both raw and processed data are available to all collaboration members, accessed via grid certificate with membership of the `snoplus.ca` virtual organisation (VO), which is managed by collaboration members.

Raw data is written to files that are limited in size to 1 GB each and is stored in triplicate between dCache at Compute Canada, dCache at Fermilab and dCache and tape at Rutherford Appleton Laboratory. As our raw data is invaluable, we have built this redundancy into our storage to ensure no loss of data. Raw data will be maintained on these sites for the active life of the experimental collaboration and will be maintained at the RAL tier 1 Grid storage element indefinitely.

Processed data files are produced in ROOT format from raw data by applying selections and analysis algorithms. Both full rat data structure (ratds) and reduced ntuple files are produced for good physics runs. Since processed data is derived from raw data, this data is considered less precious and can be reconstituted by reprocessing with the same procedures and algorithms. Therefore, processed data is retained with a reduced redundancy factor and processed data sets may be deleted when they become redundant - ie. when a newer and improved processing pass is available and it is confirmed that no collaborator is using those processed files. The ntuple files from any processed data that has been used for a publication will be maintained for the active life of the experimental collaboration.

Simulated events are produced in ROOT formats that replicate the processed data on a run-by-run basis. As with processed data, Monte Carlo simulations are not replicated since the production could be replaced by re-running. Production ntuples used for

publication will be maintained for the active life of the experimental collaboration but redundant Monte Carlo productions can be removed once it is established that they are no longer needed by any member of the collaboration.

Measurements from the SNO+ detector. An extensive analysis effort, which applies calibrations and corrects for changing detector configurations, is required to convert the raw inputs into measurements, with quantified uncertainties, of physical quantities in SI units. The collaboration is committed to make measurements that are complete, as generally useful as possible and available through peer-reviewed archival publications. These data are disseminated to the public through scientific publications, technical reports, and presentations at scientific conferences and organized workshops. All publications will also be shared on the open access ePrint server arXiv.org. The SNO+ collaboration maintains public webpages allowing our faculty, postdocs, students to share completed data analyses worldwide with anyone interested. On such public webpages, our collaboration will make our analysed data available to the public in the form of the numerical content of all plots and tables in published papers, including any relevant correlation matrices. This content will be freely available on our public webpages soon after the time of publication.