CLAS12 data processing documentation guidelines

This document provides guidelines on the documentation that is requested according to Bylaw F.III for each pass of data processing of a given data set.

The documentation should contain all relevant information on the data set and on the processing pass that is needed to facilitate the analysis and the preservation of the processed data.

The topics to be covered are listed below.

Experiment configuration

- CLAS12 configuration: summarize the configuration adopted for the CLAS12 sub-detectors and magnets.
- Beam condition: list the beam conditions and corresponding parameters such as beam energy, beam currents, beam helicity configuration.
- Data acquisition and trigger configuration: summarize the data acquisition configuration, listing the detectors included in the readout, and the trigger configuration(s) adopted during data taking. Particular attention should be given to any significant change that can affect the detector performance and calibrations, such as changes to DAQ parameters (readout windows and delays) and trigger configurations. In listing the different triggers, include both the 32-bit level-1 triggers and the special triggers such as helicity or Faraday Cup triggers.

Run list information

- Provide a list of the relevant run ranges for the different experiment configurations.
- Provide a list of special runs, such as empty target runs, random trigger runs, alignment or calibration runs.
- Provide a link to the corresponding run list in RCDB and ensure all information in RCDB is correct.

Data calibration

- Provide a link to the final pass0 timelines.
- Include a summary of the calibration results for all subsystems. This could be in the form of a table listing relevant parameters, such as average resolution and calibration accuracy, and indicating whether the calibration quality is consistent with the subsystem specification.
- Note any anomaly in the calibration quality and stability. For instance, indicate clearly run ranges that show an anomalous resolution since that could have an impact on the data analysis.
- Provide a table with run numbers used for calibration of different subsystems.

Provide information on the alignment constants for all relevant detector systems. Include plots demonstrating their efficacy in correcting for misalignment and the spatial resolution that was achieved.Data processing

- Specify the software version(s) used for data processing in the different stages (calibration, pass0, passX)
- Specify the data processing configuration. This could be done by providing a link to the data processing workflow configuration and README files.

- Provide a clear description of the skims produced as part of the data processing, with the definition of the filter conditions relating to the main analysis channels. Include preliminary results from the analysis based on skims to demonstrate their appropriateness. For skims produced with custom wagons, links to the relevant filtering code should be provided.List clearly the location of the data processing outputs, DSTs and skims.

Data quality

- Provide a list of golden runs/files according to the quality assurance analysis performed upon completion of the data processing pass. The list could be provided as a link to the relevant database.
- Include a summary of the quality assurance findings, with clear description of any anomaly and the corresponding run range.

Simulations

- Provide a link to the settings (GEMC gcard) that should be used for simulations of this data set.
- Provide information on the background events that should be used for background merging: location of the files, event filtering conditions, list of data files.
- Include the results of luminosity scan analyses and comparison with simulations merged with background.

Ancillary information

- If applicable, include information about beam polarization as determined from Moller run analysis.
- Include a description of the ancillary information available in helicity, scaler and EPICS banks and any necessary instruction on how to correctly utilize it.

The information listed above should be provided in concise format, including all but only what is relevant for data analysis. It should not include general information about standard CLAS12 procedures documented elsewhere, as for example calibration procedures, but focus on the specifics of the data set and data processing under exam, reporting clearly any relevant anomaly or special condition beyond what expected from a CLAS12 data set. It should be accessible from a single location, whether provided in the form of a written document or web/wiki page. In all cases this documentation should be under version control.