Contribution ID: **20** Type: **Oral Presentation**

Feedback on the tools in place to analyze Run3 data at the LHCb experiment

The LHCb experiment at CERN studies the outcome of particle collisions at the Large Hadron Collider at CERN. Since it began operations in 2010, the experiment has collected more than 100 PB of data resulting from proton or ion collisions. As the LHCb detector was upgraded between 2019 and 2022 and now records tens of PB of data per year, another 60 PB are expected before the end of LHC Run 3 in 2026.

Processing and analyzing such amounts of data is a huge challenge. They are distributed on the Worldwide LHC Computing Grid (WLCG) and the DIRAC system allows organising the processing and tracking its output. The Analysis Production system was developed for analysts to extract the information they need in a scalable way. To ensure the preservation and the reproducibility of the results, the LHCb collaboration applies the principles of FAIR data management as much as possible. Recording the provenance of all derived data artefacts, and the code and environments to produce them is paramount. Tools have been developed to do so for experiment-wide transforms as well as to allow data analysts to track their own processing and derived artefacts, using state of the art workflow management tools such as Snakemake, and ensure analysis code continues to be functional in the future. Following the CERN Open Data policy, the experiment also opens part of its data for public access, and is putting in place a dedicated tool, the Ntuple Wizard, to ease this task.

This paper presents the global picture of the tools in place to address those challenges, how they are faring for the analysis of the data recorded in LHC Run 3, and how this informs their evolution in the coming years.

Primary author: COUTURIER, Benjamin (CERN)

Presenter: COUTURIER, Benjamin (CERN)

Track Classification: Track 6: Data Management & Big Data