



Contribution ID: 420

Type: Talk

## LHCb Open Data Ntupling Service: On-demand production and publishing of custom LHCb Open Data

*Monday 21 October 2024 13:48 (18 minutes)*

The Large Hadron Collider Beauty (LHCb) experiment offers an excellent environment to study a broad variety of modern physics topics. Its data from the major physics campaigns (Run 1 and 2) at the Large Hadron Collider (LHC) has accumulated over 600 scientific publications. In accordance with the CERN Open Data Policy, LHCb announced the release of the full Run 1 dataset gathered from proton-proton collisions, amounting to approximately 800 terabytes. The Run 1 data was released on the CERN Open Data portal in 2023. However, due to the large amount of data collected during Run 2, it is no longer feasible to make the reconstructed data accessible to the public in the same way.

We have, therefore, developed a new and innovative approach to publishing Open Data by means of a dedicated LHCb Ntupling Service which allows third-party users to query the data collected by LHCb and request custom samples in the same columnar data format used by LHCb physicists. These samples are called Ntuples and can be individually customized in the web interface using LHCb standard tools for saving measured or derived quantities of interest. The configuration output is kept in a pure data structure format (YAML) and is interpreted by internal parsers generating the necessary Python scripts for the LHCb Ntuple production job. In this way, the LHCb Ntupling Service serves as a gateway for third-party users for preparing custom Ntuple jobs eliminating the need for real-time interaction with the LHCb database and solving potential access control and computer security issues related to opening LHCb internal tools to the public.

The LHCb Ntupling Service was developed as a collaborative effort by LHCb and the CERN Open Data team from the CERN Department of Information Technology. The service consists of the web interface frontend allowing users to create Ntuple production requests, the backend application processing the user requests and storing them in the GitLab repositories, offering vetting capabilities to the LHCb Open Data team, and automatically dispatching user requests to the LHCb Ntuple production systems after the approval. The produced Ntuples are then collected and exposed back to the users by the frontend web interface.

This talk is a joint presentation by LHCb and CERN IT and will elaborate on the LHCb Ntupling Service system infrastructure as well as its typical use case scenarios allowing to query and study the LHCb open data.

**Primary authors:** MORRIS, Adam (CERN); AIDALA, Christine (University of Michigan (US)); ROSENDAL, Daan; FITZGERALD, Dillon (University of Michigan (US)); RODRIGUES, Eduardo (University of Liverpool (GB)); KRAMER, Franz Ludwig (University of Bonn (DE)); HABERMANN, Kai Sebastian (University of Bonn (DE)); DONADONI, Marco (CERN); NOGGA, Piet (University of Bonn (DE)); NEUBERT, Sebastian (University of Bonn (DE)); SIMKO, Tibor (CERN)

**Presenter:** NOGGA, Piet (University of Bonn (DE))

**Session Classification:** Parallel (Track 8)

**Track Classification:** Track 8 - Collaboration, Reinterpretation, Outreach and Education