

# Dynamic and on-demand data streams

*Tuesday 10 July 2018 16:45 (15 minutes)*

Replicability and efficiency of data processing on the same data samples are a major challenge for the analysis of data produced by HEP experiments. High-level data analyzed by end-users are typically produced as a subset of the whole experiment data sample to study interesting selection of data (streams). For standard applications, streams may be eventually copied from servers and analyzed on local computing centers or user machine clients. The creation of streams as copy of a subset of the original data results in redundant information stored in filesystems and may be not efficient: if the definition of streams changes, it may force a reprocessing of the low-level files with consequent impact on the data analysis efficiency.

We propose a flexible solution for implementation of an analysis framework based on a database of lookup tables intended for dynamic and on-demand definition of data streams. This enables the end-users, as the data analysis strategy evolves, to explore different definitions of streams with minimal cost in computing resources. We also present a prototype demonstration application of this database for the analysis of the AMS-02 experiment data.

**Primary authors:** VAGELLI, Valerio (Universita e INFN, Perugia (IT)); DURANTI, Matteo (Universita e INFN, Perugia (IT))

**Presenter:** FORMATO, Valerio (Universita e INFN, Perugia (IT))

**Session Classification:** Posters

**Track Classification:** Track 4 - Data Handling