The main HL-LHC will consist in storing and managing the required data volume (x7 larger w.r.t. “flat budget” extrapolation). Largest consumers are Monte Carlo samples (AODs) and their derived formats (DAODs).

The Run-2 analysis model proves to be very effective in terms of analysis organization and turnaround. It requires x2 less CPU resources w.r.t. the “disorganized” Run-1 model. It is however very demanding in terms of storage: full DAOD size = full AOD size; 2 copies of the DAODs are stored on disk to facilitate analysis and ensure integrity, while AODs are archived on tape.

The proposed ATLAS HL-LHC data processing model

1) Reproducibility: removes the need to store multiple copies for integrity reasons. The number of replicas is dictated by processing needs rather than archiving strategy

2) Virtual Data: derived formats are defined in terms of the meta information needed to access the events rather than by replicating the event information itself

A Data Knowledge Base system provides a scalable and extensible service to store and retrieve event, file and dataset level meta-information

An Event Streaming Service is capable to prepare event collections asynchronously from the data processing system and serve them to the processing client. It leverages the DBK information and complements the data management system