

The 8th International Conference "Distributed Computing and  
Grid-technologies in Science and Education" (GRID 2018)



Contribution ID: 316

Type: **Sectional reports**

## Data Knowledge Base for the ATLAS collaboration

*Tuesday, 11 September 2018 15:30 (15 minutes)*

ATLAS experiment at the CERN LHC is one of the most data-intensive modern scientific apparatus. To help managing all the experimental and modelling data, multiple information systems were created during the experiment's lifetime (more than 25 years). Each such system addresses one or several tasks of data and workload management, as well as information lookup, using specific sets of metadata (data about data). Growing data volumes and the computing infrastructure complexity require from researchers more and more complicated integration of different bits of metadata from different systems using different conditions.

A common problem are multi-system join requests, which are not easy to implement in timely manner and, obviously, are less efficient than a request to a single system with integrated and pre-processed information would be. To address this issue, a joint team of researchers and developers from Kurchatov Institute and Tomsk Polytechnic University has initiated the Data Knowledge Base (DKB) R&D project in 2016. This project is aimed at knowledge acquisition and metadata integration, providing fast response for a variety of complicated requests, such as finding articles, based on same or similar data samples (search by links between objects), summary reports and monitoring tasks (aggregation requests), etc.

In this report we will discuss main features and applications of the DKB prototype implemented by now, its integration with the ATLAS Workflow Management, and future perspectives of the project.

**Primary author:** Mrs GOLOSOVA, Marina (National Research Center "Kurchatov Institute")

**Co-author:** Dr GRIGORIEVA, Maria (National Research Center "Kurchatov Institute")

**Presenter:** Mrs GOLOSOVA, Marina (National Research Center "Kurchatov Institute")

**Session Classification:** 10. Databases, Distributed Storage systems, Datalakes

**Track Classification:** 10. Databases, Distributed Storage systems, Datalakes