Contribution ID: **331**                                                        Type: **Poster**

# A New Visual Analytics toolkit for ATLAS metadata

The ATLAS experiment at the LHC has a complex heterogeneous distributed
computing infrastructure, which is used to process and analyse exabytes of data. Metadata are collected and stored at all stages of physics analysis and data processing. All metadata could be divided into operational metadata to be used for the quasi on-line monitoring, and archival to study the systems'behaviour over a given period of time (i.e., long-term data analysis). Ensuring the stability and efficiency of functioning of complex and large-scale systems, such as those in ATLAS computing, requires sophisticated monitoring tools, and the long-term monitoring data analysis becomes as important as the monitoring itself. Archival metadata, containing a lot of metrics (hardware and software environment descriptions, network state, application parameters, user account data, errors) accumulated for more than decade, can be successfully processed by various machine learning (ML) algorithms for classification, clustering and dimensionality reduction. However, the ML data analysis, despite the massive use, is not without shortcomings: the underlying algorithms are usually treated as "black boxes", as there are no effective techniques for understanding their internal mechanisms, and the domain-experts involvement in the process of ML data analysis is very limited. As a result the data analysis suffers from the lack of human supervision. Moreover, sometimes the conclusions made by the algorithms with a high accuracy may have no sense regarding the real data model.
In this work we will demonstrate how the interactive data visualization can be applied to extend the routine ML data analysis methods. Visualization allows to actively use human spatial thinking to identify new tendencies and patterns found in the collected data, avoiding the necessity of struggling with the instrumental analytics tools.
The architecture and the interface prototype of visual analytics platform (VAP) for the multidimensional data analysis of ATLAS computing metadata will be presented. The general data processing and visualization methods of the VAP prototype will be implemented and tested on the slice of ATLAS jobs metadata. As a result, a web-interface will provide ATLAS jobs interactive visual clusterization and search for non-trivial behaviour and its possible reasons. Furthermore, we will demonstrate the prototype of dynamic interactive visualization, providing the possibility of the observation of changing clustering structure in different points in time.

**Authors:**   ALEKSEEV, Aleksandr (National Research Tomsk Polytechnic University (RU)); KLIMENTOV, Alexei (Brookhaven National Laboratory (US)); Mr MILMAN, Igal (National Research Nuclear University "MEPhI"); GRIGORYEVA, Maria (Institute for Theoretical and Experimental Physics (RU)); TITOV, Mikhail (National Research Centre Kurchatov Institute (RU)); PADOLSKI, Siarhei (BNL); KORCHUGANOVA, Tatiana (National Research Tomsk Polytechnic University (RU)); Mr GALKIN, Timofei (National Research Nuclear University "MEPhI"); Dr PILYUGIN, Victor (National Research Nuclear University "MEPhI")

**Presenter:**   GRIGORYEVA, Maria (Institute for Theoretical and Experimental Physics (RU))

**Session Classification:**   Poster Session

**Track Classification:**   Track 2: Data Analysis - Algorithms and Tools