

# ATLAS Event Metadata Records as a Testbed for Scalable Data Mining

*Monday 23 March 2009 08:00 (20 minutes)*

At a data rate of 200 hertz, event metadata records ("TAGs," in ATLAS parlance) provide fertile grounds for development and evaluation of tools for scalable data mining. It is easy, of course, to apply HEP-specific selection or classification rules to event records and to label such an exercise "data mining," but our interest is different.

Advanced statistical methods and tools such as classification, association rule mining, and cluster analysis are common outside the high energy physics community. These tools can prove useful, not necessarily for discovery physics, but for learning about our data, our detector, and our software.

A fixed and relatively simple schema makes TAG export to other storage technologies such as HDF5 straightforward. This simplifies the task of exploiting very-large-scale parallel platforms such as Argonne National Laboratory's BlueGene/P, currently the largest supercomputer in the world for open science, in the development of scalable tools for data mining. Using a domain-neutral scientific data format may also enable us to take advantage of existing data mining components from other communities.

There is, further, a substantial literature on the topic of one-pass algorithms and stream mining techniques, and such tools may be inserted naturally at various points in the event data processing and distribution chain.

This paper describes early experience with event metadata records from ATLAS simulation and commissioning as a testbed for scalable data mining tool development and evaluation.

**Authors:** Dr MALON, David (Argonne National Laboratory); Dr VAN GEMMEREN, Peter (Argonne National Laboratory)

**Presenters:** Dr MALON, David (Argonne National Laboratory); Dr VAN GEMMEREN, Peter (Argonne National Laboratory)

**Session Classification:** Poster session

**Track Classification:** Software Components, Tools and Databases