# Evaluating non-relational storage technology for HEP meta-data and meta-data catalog

**M A Grigorieva[1], M V Golosova[1], M Y Gubin[2], A A Klimentov[1,3], V V Osipova[2], E A Ryabinkin[1]**

[1]National Research Centre "Kurchatov Institute", Moscow, Russia
[2]National Research Tomsk Polytechnic University, Tomsk, Russia
[3]Brookhaven National Laboratory, Upton, NY, USA

E-mail: maria.grigorieva@cern.ch, marina.golosova@cern.ch, maksim.gubin@cern.ch, alexei.klimentov@cern.ch, vikosi@tpu.ru, rea@grid.kiae.ru

**Abstract**. Large-scale scientific experiments produce vast volumes of data. These data are stored, processed and analyzed in a distributed computing environment. The life cycle of experiment is managed by specialized software like Distributed Data Management and Workload Management Systems. In order to be interpreted and mined, experimental data must be accompanied by auxiliary metadata, which are recorded at each data processing step. Metadata describes scientific data and represent scientific objects or results of scientific experiments, allowing them to be shared by various applications, to be recorded in databases or published via Web. Processing and analysis of constantly growing volume of auxiliary metadata is a challenging task, not simpler than the management and processing of experimental data itself. Furthermore, metadata sources are often loosely coupled and potentially may lead to an end-user inconsistency in combined information queries. To aggregate and synthesize a range of primary metadata sources, and enhance them with flexible schema-less addition of aggregated data, we are developing the Data Knowledge Base architecture serving as the intelligence behind GUIs and APIs.

## 1. Introduction

The volume of metadata generated at all stages of the mega-science experiment's life cycle in High Energy and Nuclear Physics (HENP) and other compute-intensive scientific applications such as bioinformatics and astro-particle physics, is comparable to the scientific data volume of the XX century experiments. Metadata have become an integral part of the overall architecture of storage and preservation of scientific information. Efficient metadata access and handling is the vital challenge for data-intensive science. Processing and analysis of huge volume of auxiliary metadata, which is constantly grows, is no less crucial challenging task than the management and processing of experimental data itself.

Scientific experiments are using software packages for distributed processing and data management, for instance, PanDA Workload Management System (WMS) in the ATLAS collaboration [1], AliEn[1] Grid framework in ALICE [2], CRAB (CMS Remote Analysis Builder)[2] in CMS. These packages orchestrate workload and data management in a heterogeneous computing environment, including payloads execution on the Grid, academic and commercial clouds, and supercomputers; and store transient and persistent meta-information in various databases needed by the international collaborations.

For the majority of scientific communities the actual issue is the lack of connectivity between metadata, describing data processing cycle, and metadata, representing the life cycle of scientific research in general, including annotation, indexing and publication of the results.

In this paper, we will explore the possible scenarios to solve the above problem by developing methods of integration of heterogeneous and distributed metadata sources, and infrastructure storage of meta-information, relevant to the specific research and experimental activities of scientific collaborations.

Though our initial motivation was related to the ATLAS needs, the proposed solution and architecture is more general and it will be applicable to a wide range of cutting edge large scientific experiments.

## 2. Metadata Sources

LHC Experiments have a set of metadata sources [3,4], with data, recorded at each phase of the experiment lifecycle: from the proposal of hypothesis and detector prototype, the selection of engineering solutions, modeling of experimental installation, to launch processing and data analysis tasks, writing of publications, conference papers and preprints. ATLAS metadata sources are listed below as an example.

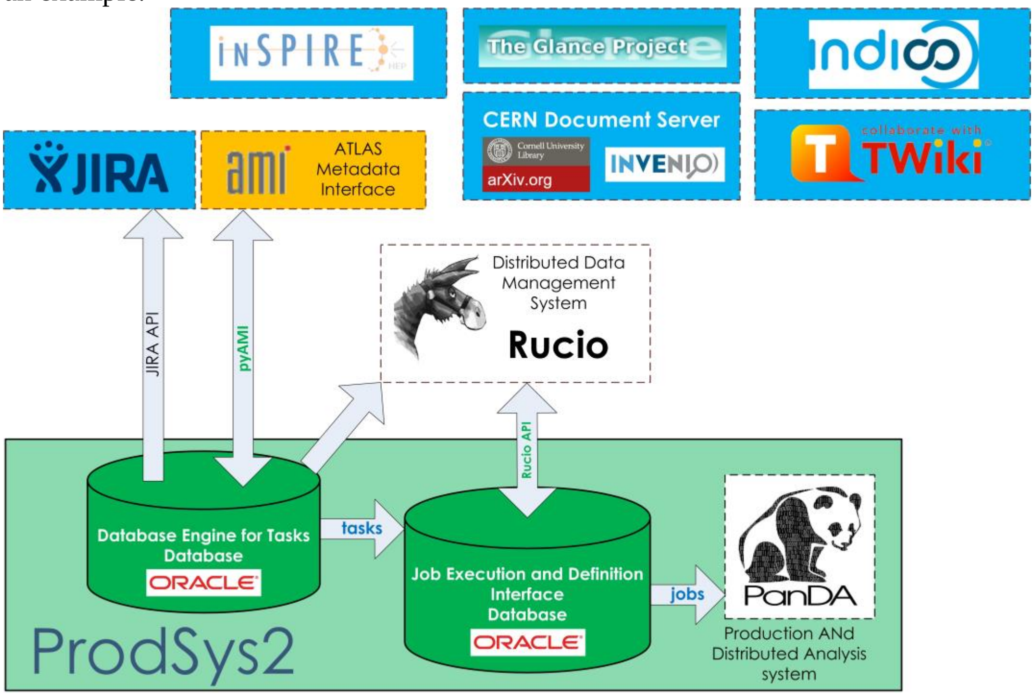


Figure 1 – ATLAS Metadata Sources

*Data processing metadata sources:*

- AMI[3] - Metadata Interface and database
- Rucio[4] - Distributed Data Management system

[1] http://alien.web.cern.ch/
[2] http://home.fnal.gov/~gutsche/computing/crab/crab_old.html
[3] http://ami.in2p3.fr/index.php/en/
[4] http://rucio.cern.ch/

- ProdSys2: DEFT, JEDI, PanDA [5,6,7] - Production system and its components, it is a workflow driven system and it's used to deal with the whole variety of ATLAS data processing and analysis activities: Monte-Carlo simulation, data (re)processing, High Level Trigger[5] Grid processing, physics groups derivation[6] and train production, user's analysis.

*Information resources (publications, conference proceedings, meetings notes and actions items, etc.) metadata sources:*

- GLANCE[7] - search engine for the ATLAS Collaboration publications.
- Indico[8] tools (allows to manage complex conferences, workshops and meetings)
- Inspire[9], containing 500,000 HEP documents
- CDS (CERN Document Server)[10] containing 200,000 CERN documents
    - arXiv[11] – openly accessible electronic archive and distribution server for research articles.
    - Invenio[12] - is an open source software package that provides the tools for management of digital assets in an institutional repository.
- CERN Twikis (a tool for web page collaborative writing)
- JIRA ITS[13] (Issue Tracking Service)

## 3. ATLAS Metadata sources integration challenge

As it is shown on Figure 1, these metadata sources are loosely coupled and information may not be timely synchronized, so they potentially may provide to an end-user inconsistency in requested information.

To aggregate and synthesize a range of primary metadata sources, and enhance them with flexible schema-less addition of aggregated data, we are developing the Data Knowledge Base (DKB) serving as the intelligence behind GUIs and APIs. We will present our current accomplishments to aggregate and to store LHC experiments meta-information and as a backend for future knowledgeable data storage.

### *3.1 Scientific Research Lifecycle*

We have carefully investigated the ATLAS experiment metadata and the relationship between primary and aggregated metadata sources. These analyze allowed us to identify the basic components of a scientific research lifecycle.

Activities are conducted by *Physics and Combined Performance Groups* (for instance: Top, ETmiss, BPhysics). A set of initial data processing parameters such as *Projects* (for example, "mc14_8TeV"), information, associated with Software Release versions, and conditions data are recorded as auxiliary metadata. To start the production cycle, which includes a pipeline of data transformations and data analysis tasks, *Group Manager* defines *a Request* to be executed by the *Production System*. Request is translated to the set of tasks and jobs to be brokered by the PanDA

[5] The ATLAS High Level Trigger (HLT) is a distributed real-time software system that performs the final online selection of events produced during proton-proton collisions at the LHC

[6] There are more than 12 physics groups in the ATLAS Collaboration, bulk data processing and simulation is done centrally, but then each group runs group specific code to produce 'events collection' it is called derived data in ATLAS, the process is called derivation production

[7] https://atglance.web.cern.ch/atglance/

[8] https://indico.cern.ch/

[9] http://hep-inspire.net

[10] https://cds.cern.ch

[11] http://arxiv.org/

[12] http://invenio-software.org/

[13] http://information-technology.web.cern.ch/services/JIRA-service

WMS on Grid and opportunistic resources. Each task is accompanied by input data - organized in *Containers* - collection of files, produced or processed under the same conditions (software release, transformation step, conditions data, etc.).

The processes of hypothesis formulation, determination data transformation and data analysis parameters, adoption of the data samples, and other activity related to the organization and conducting the research, are registered in Twiki Pages and Indico. Interim results are published as Conference Notes. The final step of the experimental research lifecycle is a *Publication* in scientific magazines.

### 3.2 *Analysis/paper registry supporting the recording of paper-associated containers/datasets*

Specificity of scientific data is necessity of the eternal storing the RAW data, obtained directly from the detector, and have not been processed and transformed. Any RAW data changes leads to a scientific data, which scientists use in experiments. With the development of science and technology, the gap between the RAW and scientific data is constantly increases, and the procedures of RAW data processing have become more sophisticated. A complete data handling cycle includes RAW data preservation, data processing and analysis and information exchange. The requirement for eternal storage of the RAW data (as well as software version) and the access to it determines the basic principles of scientific research - reproducible scientific results.

The result of a chain of scientific data transforms are publications. Moreover, each publication can be based on the analysis of different versions of the data or use the data, obtained through various transformations. All of this meta-information must be fixed to ensure repeatability and reproducibility.

Currently meta-information, describing the lifecycle of the data processing and catalogs publications, operates separately. It would be useful to have some method of associating, in a searchable way, analysis teams and publications with datasets on the Grid.

## 4. Data Knowledge Base Architecture

DKB is designed to integrate information from many sources in a meaningful way and to perform meta-data analysis and meta-data mining. The structure and data flow of DKB is shown on Figure 2.

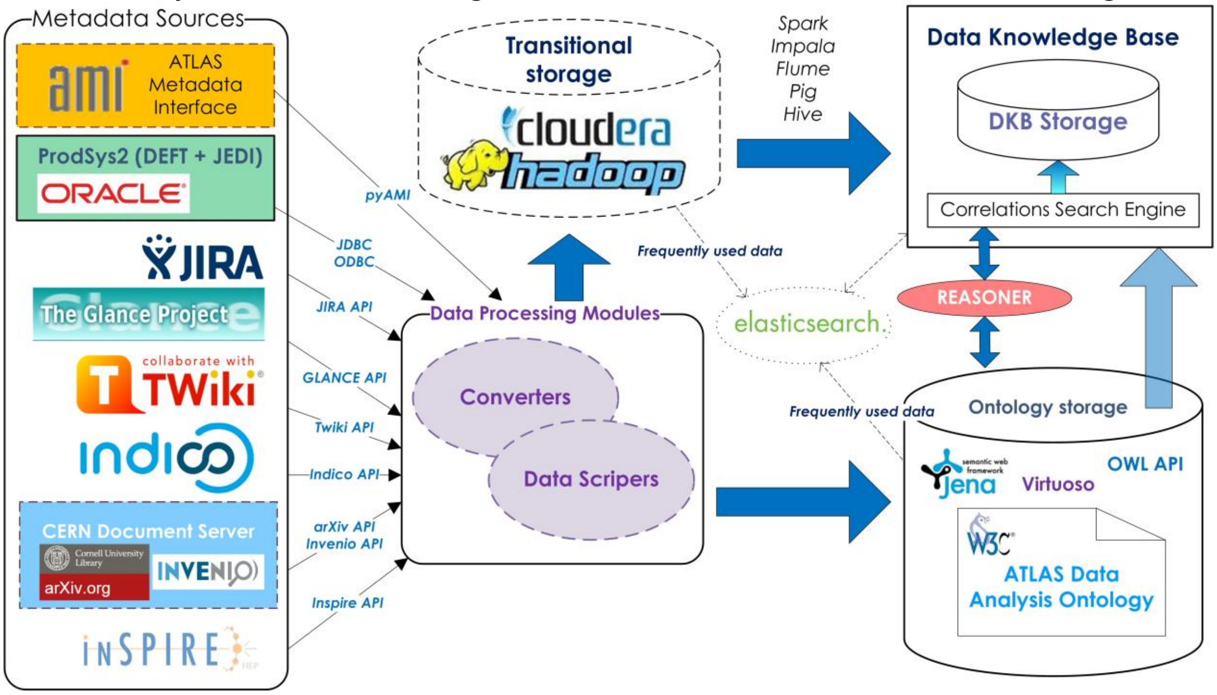


Figure 2 – Data Knowledge Base architecture

*Ontology Storage, particularly OWL (Web Ontology Language) DB,* is used to store information from multiple sources and to transform it into interconnected dataset suitable for further analysis. At this moment the exact platform for implementing OWL DB is under consideration. Possible candidates include Apache Jena[14], Virtuoso[15], and OWL API[16]. The data that is supplied to the OWL DB from various sources (physics papers metadata from GLANCE, meeting agendas from Indico, documents from CDS and arXiv) is usually initially available in the form of human readable texts. Before being stored into OWL DB, all this data is preprocessed by scrapers and converters that produce OWL metadata from natural language documents. After the preprocessed data is stored in the OWL DB, it is transformed into single interconnected dataset with reasoner employing domain ontologies and inference rules to create connections between heterogeneous data.

Each data source that is used in conjunction with OWL DB has an ontology associated with it, describing objects that can be obtained from the data source and their relationships with objects of other domains, most importantly rules for determining identity of objects and determining when separate objects in database refer to a single concept.

The bulk of ProdSys2 and AMI, after minor transformation, is uploaded to transitional storage - Hadoop cluster. There, it is processed to reveal, which data would be accessed by physics groups at various dates, and the ways to process the data.

Due to long response times of Hadoop and OWL DB, they are unsuitable for interfacing with users directly. The most frequently requested data is uploaded to ElasticSearch[17] cluster, providing a distributed, multitenant-capable full-text search engine, based on Lucene. It will allow short response times and generally improve user experience when interacting with DKB.

DKB serves as central part of the database structure, accepting user requests and transforming them into queries aimed at OWL DB, ElasticSearch or Hadoop, respectively. It is also responsible for combining query results and presenting them to user.

**Acknowledgements**
We wish to thank all our colleagues from ATLAS experiment, PanDA team and CERN IT. This work was funded in part by the Russian Ministry of Science and Education under contract № 14.Z50.31.0024.

**References**
[1] G Aad et al. ATLAS Collaboration 2008 The ATLAS experiment at the CERN large hadron collider *Journal of Instrumentation* **3** S08003
[2] K Aamodt, A Abrahantes Quintana, R Achenbach, S Acounis, D Adamová, C Adler, M Aggarwal, F Agnese, G Aglieri Rinella, Z Ahammed, et al. The ALICE experiment at the CERN LHC 2008 *Journal of Instrumentation* **3** S08002
[3] The Large Hadron Collider http://lhc.web.cern.ch/lhc/
[4] J Shiers 2007 The worldwide LHC computing grid (worldwide LCG) *Computer physics communications* **177** 219–223
[5] The PanDA Project http://www.usatlas.bnl.gov/twiki/bin/view/PanDA/WebHome
[6] T Maeno, P Nilsson, K De, A Klimentov and T Wenaus 2009 PanDA Production and Analysis backend *Journal of Physics* **219** 210
[7] M Borodin, K De, D Golubkov, T Maeno, D Oleynik, A Petrosyan, A Klimentov, T Wenaus, P Nilsson, S Jha, et al. 2014 PanDA Beyond ATLAS: A Scalable Workload Management System For Data Intensive Science Technical report https://cds.cern.ch/record/1670021?ln=ru

[14] https://jena.apache.org/
[15] http://semanticweb.org/wiki/Virtuoso.html
[16] http://owlapi.sourceforge.net/
[17] https://www.elastic.co/products/elasticsearch