

Metadata aided run selection at ATLAS

To cite this article: R M Buckingham *et al* 2011 *J. Phys.: Conf. Ser.* **331** 042030

View the [article online](#) for updates and enhancements.

Related content

- [Conditions and configuration metadata for the ATLAS experiment](#)
E J Gallas, S Albrand, J Fulachier *et al.*
- [An integrated overview of metadata in ATLAS](#)
E J Gallas, D Malon, R J Hawkings *et al.*
- [Engineering the ATLAS TAG Browser](#)
Qizhi Zhang and the ATLAS Collaboration

Recent citations

- [Lydia Audrey Beresford](#)
- [Conditions and configuration metadata for the ATLAS experiment](#)
E J Gallas *et al*



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Metadata aided run selection at ATLAS

R M Buckingham¹, E J Gallas¹, J C-L Tseng¹, F Viegas² and E Vinek² for the ATLAS Collaboration

¹Department of Physics, Oxford University, Denys Wilkinson Building, Keble Road, Oxford OX1 3HR, United Kingdom

²CERN, CH-1211 Geneva 23, Switzerland

E-mail: Ryan.Buckingham@cern.ch

Abstract. Management of the large volume of data collected by any large scale scientific experiment requires the collection of coherent metadata quantities, which can be used by reconstruction or analysis programs and/or user interfaces, to pinpoint collections of data needed for specific purposes. In the ATLAS experiment at the LHC, we have collected metadata from systems storing non-event-wise data (Conditions) into a relational database. The Conditions metadata (COMA) database tables not only contain conditions known at the time of event recording, but also allow for the addition of conditions data collected as a result of later analysis of the data (such as improved measurements of beam conditions or assessments of data quality).

A new web based interface called “runBrowser” makes these Conditions Metadata available as a Run based selection service. runBrowser, based on PHP and JavaScript, uses jQuery to present selection criteria and report results. It not only facilitates data selection by conditions attributes, but also gives the user information at each stage about the relationship between the conditions chosen and the remaining conditions criteria available.

When a set of COMA selections are complete, runBrowser produces a human readable report as well as an XML file in a standardized ATLAS format. This XML can be saved for later use or refinement in a future runBrowser session, shared with physics/detector groups, or used as input to ELSSI (event level Metadata browser) or other ATLAS run or event processing services.

1. Introduction

The LHC is currently the world’s largest and highest energy particle accelerator. After over 10 years of waiting, thousands of physicists scattered throughout the world are eager to get their hands on this long awaited data. With beam crossings at a rate of 10^9 s^{-1} during regular LHC operation, the ATLAS detector will produce petabytes of data which must be managed and processed in a computing grid which extends world-wide.

As we deal with these new levels of superior data acquisition rates, size and complexity, it is increasingly important to ensure that data, and metadata (i.e. data about data) are stored in an efficient and easily accessible manner which is able to take advantage of the latest database and internet technology available. This is achieved through the use of various database systems with complex relational tables. These systems are interfaced with a selection of web-browser displays - each with a different role and purpose.

2. ATLAS Conditions Database

Over the course of a physics Run¹, there are many variables relating to the operation of the beam and detector which can vary considerably. These ‘conditions’ are not recorded on an event-by-event basis but are instead organized into an IOV (Interval of validity). This can be an interval in time or a period specified by a range of Run Numbers or Luminosity Blocks². ‘Conditions data’ are stored in the ATLAS Conditions Database which is accessed using the COOL (Conditions database Of Objects for LHC) API (Application Programming Interface).

Examples of conditions stored in the ATLAS Conditions Database are: LHC beam conditions, online configuration and operation, calibration & alignment, data quality, luminosity normalization, object reconstruction efficiency and book-keeping data for various cross checks of data completeness/integrity.

The runQuery [1] utility is the web interface to Run-related information in the Conditions Database. Users are able to find Runs and their associated information using a command line interface. This interface does not however allow the user to see what they are selecting from. Output from runQuery is in the form of a human-readable table and an XML output³ containing all Luminosity Blocks from the selected Runs.

2.1. Metadata

Management of the large volume of data collected by any large scale scientific experiment requires the collection of coherent Metadata quantities, which can be used by reconstruction or analysis programs and/or user interfaces, to pinpoint collections of data needed for specific purposes. The production of these metadata has many advantages [3] making it possible to

- know which data does not have sufficient quality to be included in a final analysis,
- improve and/or understand the resolution of various measurements and their efficiencies,
- follow the data through various stages of processing (and know when processing failure occurs) and
- make the connection between the data in the final physics analysis to the conditions existing when those events were recorded online.

In the ATLAS experiment at the LHC, metadata from systems storing non-event-wise data (Conditions) has been collected into a relational database. The Conditions metadata (COMA) database tables not only contain conditions known at the time of event recording, but also allow for the addition of conditions data collected as a result of later analysis of the data (such as improved measurements of beam conditions or assessments of data quality).

3. Event selection using ATLAS TAGS

Building on the experience of previous experiments, ATLAS is attempting to go one step further and build an event-level selection capability allowing the user to select only those events which are of particular interest. The RAW data collected by the ATLAS detector is passed through various stages of processing in order to produce slimmed and compact data sets which contain only those variables essential to the end user. The final smallest files produced, TAGs, are produced for each event and contain all of the variables needed to make an event selection.

¹ A Run is a consecutive interval of time of data recording, usually a few hours in duration, during which many aspects of the detector configuration are constant.

² A Luminosity Block is a period in time, usually lasting a few minutes, for which the instantaneous luminosity is approximately constant.

³ This format is that of a ‘Good Run List’ which can be used within the ATLAS Athena framework allowing users to run their analysis on only those luminosity blocks which have satisfied their selection criteria. This analysis can be run on the Grid where it will automatically locate and run over only the files necessary to include all the specified runs.

When a user runs their analysis code using TAGs on the Grid, the software will find the original larger data set in order to access all of the required data.

The ELSSI (Event Level Selection Service Interface) interface [4] allows users to make an event level selection based on TAGs. Selection can be made using criteria such as physics stream, run number, triggers and physics objects in the event.

4. runBrowser - Dynamic metadata aided Run selection

The ATLAS metadata infrastructure has been designed to facilitate logical and efficient access to all flavors of analysis. The runBrowser interface [4], shown in figure 1, allows users to select Runs based on all available conditions data and conditions metadata. A flow diagram of their dependencies is shown in figure 2. The target audience of the interface ranges from those who are not aware of exactly what data is available (and therefore a dynamic selection interface gives transparency) to the expert user who wants full flexibility to select a precise combination of Luminosity Blocks based on all available information.

The bulk of the browser is coded using PHP, interfaced with HTML and CSS. Faced with the challenge of allowing selection from such a large pool of information, we employ JavaScript and jQuery to develop the ‘ergonomics’ of the interface. Although a first version has now been put into production, there are still many plans for further development - this is discussed in section 5.

The main page consists of multiple expandable sections: one for each selection criteria. Using jQuery, the user may collapse and expand each section to reveal the values which are available for selection. In most cases these values are in the form of a table. Values are retrieved from the database using SQL queries which are built using PHP. For some criteria, a text box alternative is included allowing the use of ranges and wild-cards - for example, a run selection may consist of 152409-152415,15250%⁴. There is also one expandable section which contains all those criteria which have been selected so far. If a user has made a selection using a text box, they may also edit/update the contents of the text box without ‘going back’ to a previous page.

Selection criteria can be split into two types: those which allow the user to select whole runs; and those which allow the user to make a Luminosity Block level selection. The selection criteria which are currently included in the the interface are:

- **Data Source** - i.e. Is the data simulation or ‘real’ data?
- **Run Type** - Allows more specific selection for simulation data: e.g. runs based on different Monte Carlo releases and the ATLAS Full Dress Rehearsal (FDR) simulated Runs.
- **Date Range** - Allows users to select the dates between which a selections of Run have taken place. It also includes a dynamic pop-up calendar.
- **Period Selection** - Periods are periods in time for which LHC configuration was approximately stable. The periods are human defined and allow a convenient division of the data into smaller, more manageable chunks.
- **Filename (Project) Names** - A further tag given to data. Examples include ‘data09_900GeV’ and ‘data10_7TeV’.
- **Trigger Master Key**⁵ - Each set of unique logical conditions imposed by the online (or simulated) trigger system is assigned a unique integer ‘Trigger Super Master Key’. runBrowser allows selection of one or more of these keys.
- **DAQ Configuration** - This is an online configuration tag used by the Tier 0 - may be useful to experts in a detector and TDAQ (Trigger & Data Acquisition) commissioning.

⁴ In the SQL environment, the percentage sign (%) is the symbol for a wild-card.

⁵ Selection of a Trigger Master Key then allows further selection of the Level 1 and High Level Trigger (HLT) prescales.

runBrowser - Run Metadata Selection V2

Click [here](#) to find out about future development plans...

- Click here for Purpose & Instructions

- SELECTION SUMMARY -

COOL_INDEX IN DetStatusLSUMM-repro05-01
Project Period IN ('data10_7TeV.period')

Remove all COOL_INDEX criteria
Remove all PROJECT_PERIOD criteria

7 runs left to choose from

Run based selection...

Uploaded Runs - ☐ Yes ☐ No

- Date range selection -

- Period Selection -

- Data Source -

- Run type -

- Filename (Project) Names -

AMI Tag Name Selection

This optional textbox allows wildcards (%) in AMI Tag selection:
Examples: '%m496%' (then press return)

- DAQ Configuration -

This optional textbox allows wildcards (%) in DAQ Configuration selection:
Examples: 'Schema=176:Data=454,Schema=206:Data=%' (then press return)

- Run number (7 values to choose from) -

This optional textbox allows multiple value and range selection:
Example: '152409,152405-152407' (then press return)

Luminosity Block based selection...

- LHC Stable Beams -

- ATLAS Ready Status Flag Selection -

- Trigger Master Key -

- Detector Status Conditions Tag -

This optional textbox allows multiple value and range selection:
Example: '682,667-669' (then press return)

Selected Detector Status Conditions Tag(s) found:

Run Type | Detector Status COOL Tag | Run Range | Run Count
Physics | DetStatusLSUMM-repro05-01 | 152166-167844 | 126

Combined performance virtual DQ flags are available selection here.
This selection will automatically filter through to flag below.

Detector Status (DQ) Selection: - The first 2 rows are Detector Groups: a Group is the set of Detectors in the table below that Group Name. Choices at the Group level are used unless their blue button is clicked.

Summary: All Detectors

Group	Detector	Flag
Inner Detector	PIXB	Good
	PIXO	Good
	PIXEA	Good
	PIXEC	Good
Calorimeters	EMBA	Good
	EMBC	Good
	EMECA	Good
	EMECC	Good
Muons	MDTBA	Good
	MDTBC	Good
	MDTEA	Good
	MDTEC	Good
Forward	LCD	Good
	LCDA	Good
	LCDC	Good
	ALFA	Good
Trigger	L1CAL	Good
	L1MUB	Good
	L1MUE	Good
	L1CTP	Good
Other	LUMI	Good
	LUMIONL	Good
	RCOPL	Good
	ATLGL	Good
Physics Objects	EIDB	Good
	EIDCR	Good
	EIDE	Good
	PIDB	Good
	PIDCR	Good
	PIDE	Good
	EIDF	Good
	EIDSOFT	Good
	MSTACO	Good
	BTGEA	Good
	BTGEC	Good
	TAUB	Good
	TAUEA	Good
	TAUEC	Good

Notes on difference between Group and individual ID criteria:

When using Detector Groups, if individual DQ assignments for an ID do not exist for that DQ Tag and Run, then the criteria for that ID is ignored to form the result. To demand particular assignments for particular IDs, set the criteria at the detector (not the group) level (disable that Group with the blue button). Group and individual ID assessments will be equivalent for a Run if all IDs have non-'unknown' assignments within that DQ Tag for that Run.

Save the current data selection in your cookies:

All submit buttons on this page submit all information on this page:

You have currently selected 7 runs.

Click on the FINISH button to use all criteria to generate results:

Purge options:

Links to Documentation

- Main Menu: [Start Over \(without deleting cookies\)](#)
- Attributes by Release: [TagForEventSelection](#)
- PAT: [PhysicsAnalysisTools](#)

Comments / Questions?

ATLAS_TAGS_METADATA @ INT8R

Figure 1. Detailed screen shot from runBrowser

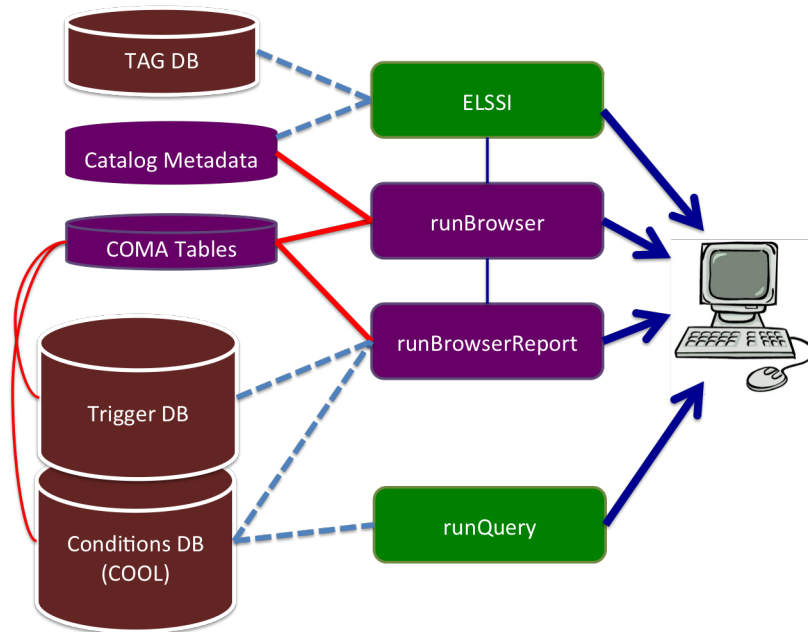


Figure 2. Flow diagram of the relational dependencies of the Data Bases (DB), COMA tables and the various interfaces. Solid (red) lines signify transfer of metadata and dashed (blue) lines signify transfer of data.

- **Detector Status Conditions Tag⁶** - Detector Status (Data Quality) is assessed on a Run/Luminosity Block level by experts. When a set of DQ assessments becomes official, that set is assigned a unique tag in the ATLAS Conditions Database called a ‘Conditions Tag’.
- **Run Number** - This table shows the runs which fit all of the selection criteria so far. With each Run are associated links to run specific information on AMI⁷ [5], runQuery (see section 2) and a separate interface to the trigger database [6]. The table also includes a link to the runBrowser Run Report for each run. This report contains a concise compilation of the information contained within the COMA tables. This information currently includes general Run conditions, load status, debug information, DQ assessments, prescale evolution and trigger information.
- **Current status of TAG availability** - This Yes/No selection allows the user to select Runs which are available, via the ELSSI browser, to the TAG DB: Current policy is to keep only the 2 latest processing of Runs available in the TAG database- this directs the user to the latest/best data and conserves database resources to optimize that access.

Help and documentation is available throughout the interface in the form of pop-up windows, an expandable section containing ‘Purpose and Instructions’ and short strings of text which appear when the cursor hovers over certain objects.

⁶ Selection of a Detector Status Condition Tag then allows further selection of individual detector flags (shown in figure 1). It is also possible to choose from a list of pre-prescribed ‘Virtual Flags’ which, using JavaScript, automatically set each of the individual detector flags to the correct position.

⁷ AMI is a cataloging system using the Tier 0 database and production database in order to display information about all of the registered ATLAS datasets.

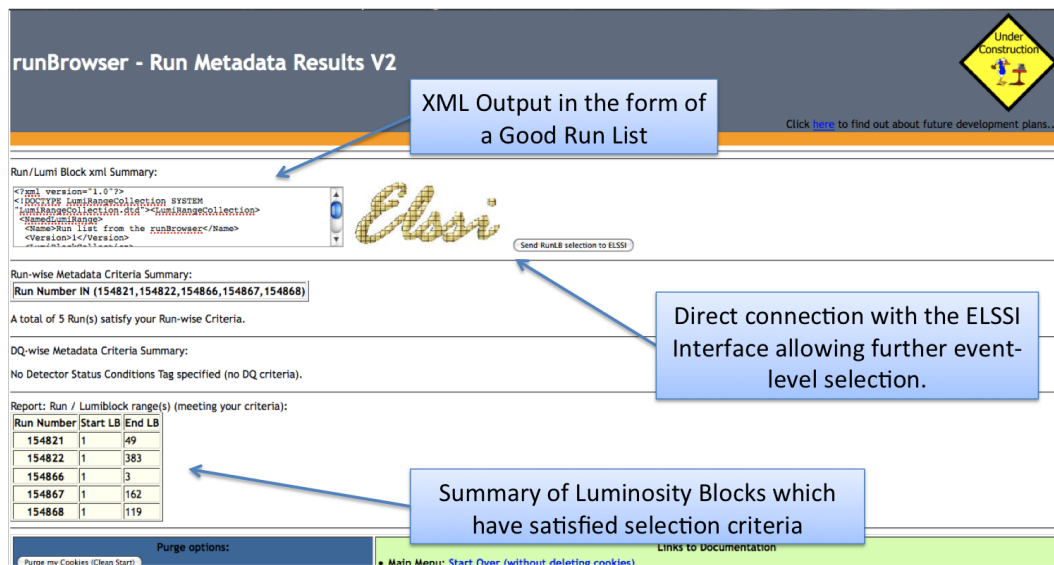


Figure 3. Screen shot from runBrowser results page

4.1. runBrowser Results

When a set of COMA selections are complete, runBrowser produces a results page (shown in figure 3) containing a human readable report as well as an XML file in a standardized ATLAS format (see footnote 3). This XML can be saved for later use or refinement in a future runBrowser session, shared with physics/detector groups, or used as input to ELSSI or other ATLAS run or event processing services.

5. Conclusions, Future Work & Other Work

At present the runBrowser interface facilitates logical, dynamic and efficient access to COMA information, allowing users to choose a selection of runs (or Luminosity Blocks) for further analysis. The interface, as well as the COMA tables themselves, are still in development and there are plans to add:

- **Stable Beams** - A boolean selection allowing the user to include only those Luminosity Blocks, within a run, for which the collisions in the detector are in a stable state which is good for data taking.
- **ATLAS Ready** - A boolean selection allowing the user to include only those Luminosity Blocks, within a run, for which the ATLAS detector has declared itself 'ready'. This flag will only be true during a period of stable beams.
- **Specific Magnet Information** - the status of the ATLAS toroid and solenoid.
- Also other information concerning Luminosity, triggers and data streams.

References

- [1] The runquery interface URL <http://atlas-runquery.cern.ch/>
- [2] Gallas E J, Malon D, Hawkings R J, Albrand S and Torrence E 2010 *Journal of Physics: Conference Series* **219** 042009 URL <http://stacks.iop.org/1742-6596/219/i=4/a=042009>
- [3] The runbrowser and elssi interfaces URL <https://atlddm10.cern.ch/tagsservices/index.htm>
- [4] The ami interface URL <http://ami.in2p3.fr/opencms/opencms/AMI/www/index.html>
- [5] Interface to trigger database URL <http://atlas-trigconf.cern.ch>