

TAG and iELSSI Progress



Elisabeth Vinek, CERN & University of Vienna
on behalf of the TAG developers group

TAG data management news

TAG TYPE	# events (M)	DB Storage (GB)	Avg. space/event (kB)
First Pass 2010	1155	3978	3.8 ⁽¹⁾
September Reprocessing	557	940	1.6 ⁽²⁾
Monte Carlo 2009	865	2552	2.9 ⁽³⁾

- (1) Average for latest events (first pass is a mix of merged and unmerged TAG and TAG_COMM data)
- (2) We are now applying vertical partitioning to reprocessed data.
 - Details can be found here (ATLAS S&C Week July):
<http://indico.cern.ch/getFile.py/access?contribId=38&sessionId=16&resId=0&materialId=slides&confId=76895>
- (3) MC data is not compressed (>255 columns)

iELSSI news: hiding data distribution

- TAG DB sites: CERN, DESY, PIC, BNL, TRIUMF
- Previously, the user had to select a site
- Now, this selection is done automatically, based on:
 - data locality
 - proximity to iELSSI browser

The screenshot shows a web interface with a blue header containing navigation tabs: 'data10', 'data09', 'usermix', and 'mc09'. The main content area is split into two panels. The left panel has a green background and displays the breadcrumb 'data10 » data10_7TeV » TAG:'. Below this, it shows 'September10 TAG reprocessing TAG data collections (sub_total: 189)' with a dropdown menu currently set to 'data10_7TeV_physics_Egamma_f280_m568_p250_p160_READ'. Underneath is 'May10 reprocessing TAG data collections (sub_total: 19)' with a dropdown menu set to 'Select one'. At the bottom of the green panel is a button labeled 'Show other passes' with the text 'Click on' to its left and 'to see/hide the other passes' to its right. The right panel has a blue background and is titled 'Run numbers (select multiples by holding 'Ctrl' or 'Shift'):' above a scrollable list box containing the numbers 160387, 160472, and 160479. Below the list box, it says 'Conditions Metadata for ALL run(s) in the above list is available at [COMA runBrowserReport](#)'. At the bottom of the right panel is a button labeled 'Select all (3) runs'.

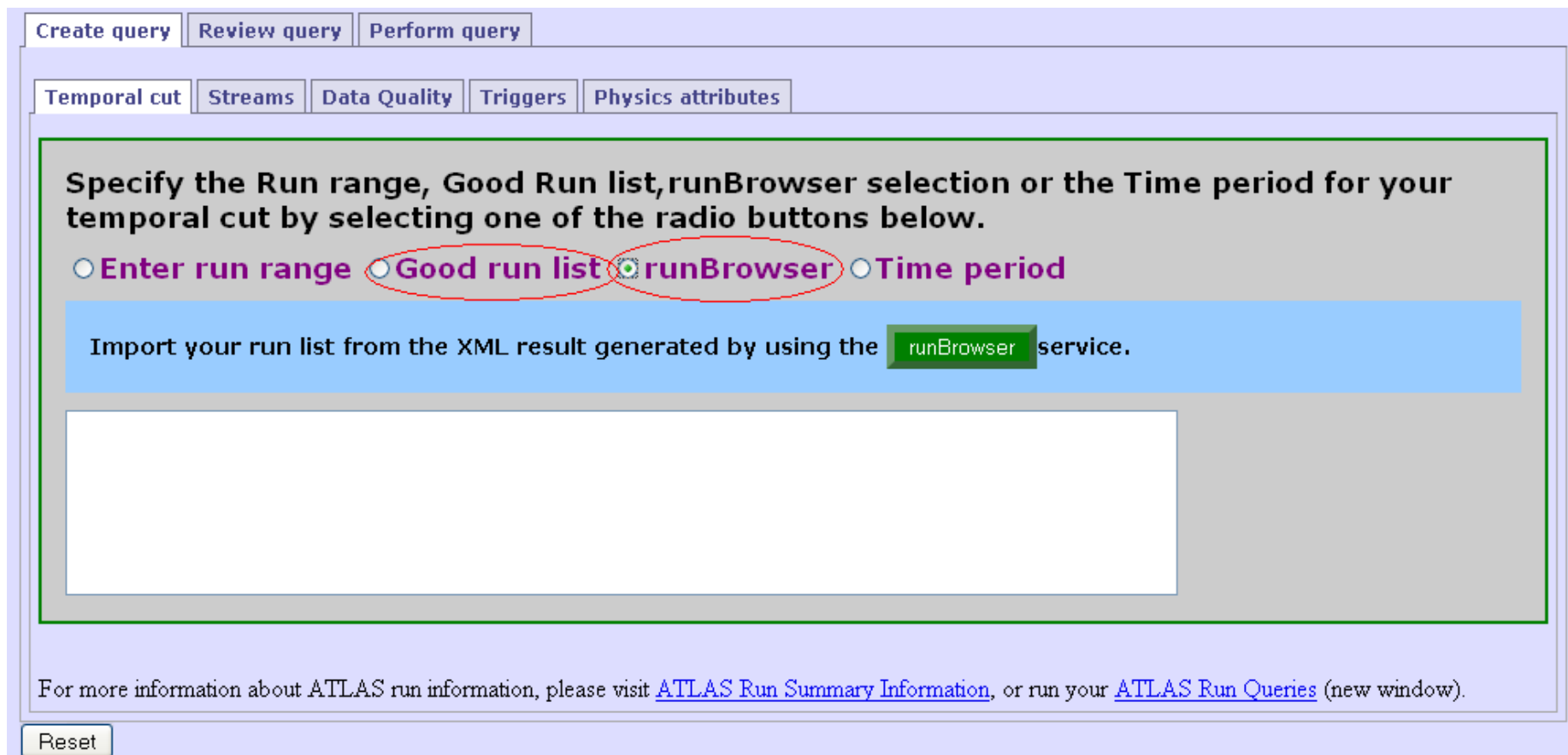
iELSSI news: period-based selection

- TAG data selection based on periods and sub-periods
- PERIOD -> RUN mapping in COMA tables
 - based on information from Data Preparation

The screenshot shows a web interface with a navigation bar at the top containing tabs for 'data10', 'data09', 'usermix', and 'mc09'. The main content area is split into two panels. The left panel, titled 'Periods of data10 (select multiples by holding 'Ctrl' or 'Shift'):', has a green background and contains two dropdown menus. The 'Primary' dropdown lists letters C through I, with 'D' selected. The 'Secondary' dropdown lists sub-periods C1, C2, D1, D2, D3, D4, and D5, with D1, D2, and D3 selected. A 'Select periods' button is positioned to the right of these dropdowns. The right panel, titled 'Run numbers (select multiples by holding 'Ctrl' or 'Shift'):', has a blue background and contains a vertical list of run numbers: 158045, 158116, 158269, 158299, 158392, 158443, and 158466. Below this list is a 'Conditions Metadata for ALL run(s) in the above list is available at [COMA runBrowserReport](#)' and a 'Select all (21) runs' button.

iELSSI news: RunBrowser Integration

- For details on RunBrowser, see talk „Conditions Metadata and RunBrowser Update” by E. Gallas



Create query | Review query | Perform query

Temporal cut | Streams | Data Quality | Triggers | Physics attributes

Specify the Run range, Good Run list, runBrowser selection or the Time period for your temporal cut by selecting one of the radio buttons below.

Enter run range Good run list runBrowser Time period

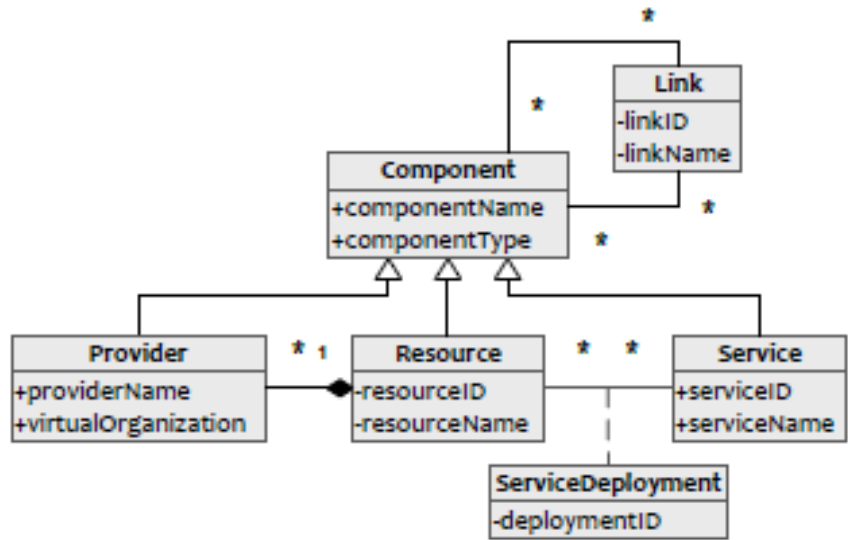
Import your run list from the XML result generated by using the **runBrowser** service.

Reset

For more information about ATLAS run information, please visit [ATLAS Run Summary Information](#), or run your [ATLAS Run Queries](#) (new window).

Architecture news: service catalog

- Site-transparency as first step towards a distributed system - achieved with the „data catalog“.
- Next steps: put service catalog into production catalog in order to:
 - Make service distribution transparent
 - Make the system more robust as regarding responding to outages
- Need to catalog the TAG services in terms of:
 - Functionality (service groups)
 - Available resources (at sites)
 - Compatibility
 - Several non-functional attributes
- Components of a distributed system:
 - Sites
 - Resources
 - Services
 - Deployments
 - Links



Architecture news: service catalog (2)

- Further service catalog use cases:
 - Improve software management by:
 - Defining dependencies between deployments/versions
 - Keeping configuration parameters centrally
 - Use the service catalog schema to log service activity:
 - Gather statistics
 - Analyze and use them to:
 - Optimizing caching strategies
 - Optimizing data and services distribution
 - Ultimately, optimize request distribution

Conclusion

- First step (data distribution transparency) completed and changes implemented on iELSSI.
- Second step (services distribution transparency) currently in development, to make the TAG system:
 - Modular
 - Transparent
 - easily configurable
- Continuous effort to improve on covering the TAG use cases. Thanks to the modular architecture new use cases are easy to implement.
- TAG database at RAL is ready and waiting for data.
- Now waiting for December TAG reprocessing.