Argonne NATIONAL LABORATORY

… for a brighter future

U.S. Department of Energy

UChicago ▶ Argonne LLC

Office of Science U.S. DEPARTMENT OF ENERGY

A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC

TAGs and Early Data

David Malon (malon@anl.gov)
for the ATLAS TAG Team
U.S. ATLAS Analysis Workshop
Argonne National Laboratory
18-20 January 2010

Outline

- Brief introduction to TAGs
- Physics TAGs and commissioning TAGs
- A look at commissioning TAGs and their uses
- A word about Good Run Lists and TAGs
- Some TAG browsing
- Using TAG-based selections

Event TAGs

- TAGs are event-by-event metadata records containing
  - Key quantities that identify and describe the event, intended to be useful for event selection, and
  - Sufficient navigational navigational information to allow access to the event data at all prior processing stages: RAW (including bytestream), ESD, and AOD (and possibly more, e.g., for Monte Carlo data)
- TAG is not an acronym
  - Capitalized (like RAW) because it is one of a discrete list (enum) of recognized data products, and because “tag” has too many other meanings in ATLAS
- Content is standardized in production, but custom TAGs are also definable
  - You could create your own
- Commissioning TAGs began as such custom-made TAGs
  - Now a standard part of ATLAS production
- In ATLAS dataset names,
  - TAG = physics TAG
  - TAG_COMM = commissioning TAG

Standard Physics TAG content

- More than 200 attributes covering
  - Event identification (run, event, lumi block numbers, timestamp, …)
  - Global quantities (e.g., missing eT)
  - Trigger decisions at all three levels (bit encoded)
  - Numbers of electrons, muons, photons, taus, and jets
  - pT, eta, phi for highest-pT objects
  - Detector status and quality words
  - Physics and performance group status words
- https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TagForEventSelection15 for details
- Content evolving
  - Additional trigger information
  - Status words now being filled
  - Stream information being added
- Current content gatekeepers are Physics Analysis Tools (PAT) group and, in principle, Event (Data) Management Board (EMB)

Not much about physics TAGs in today's talk

- Physics TAGs have been described many times, including at earlier Argonne –hosted analysis workshops
- Will not reprise all of this here
- TAGs are a routine component of ATLAS offline software tutorials
  - See https://twiki.cern.ch/twiki/bin/view/Atlas/EventTagTutorials and accompanying slides, e.g., linked to the December 2009 ATLAS tutorial agenda http://indico.cern.ch/conferenceDisplay.py?confId=72458
- Physics TAG content designed for physics runs at higher energy
  - Cannot find the two-muon event Tom showed via physics TAGs
  - Though muon information is part of TAG content
  - TAG thresholds currently too high to show low-pT muons
- Will look at commissioning TAGs today
- … but the same principles apply
  - And physics TAGs should rapidly become more interesting

Commissioning TAGs: TAG_COMM

- Primitive TAGs introduced during cosmic ray commissioning
  - Event identification and L1 triggers only, plus a variety of track and hit counts; no physics objects
  - Useful for finding cosmic events with tracks through multiple detectors, for example
  - As use in commissioning increased, more and more information was added
- Not originally foreseen as a standard data product, but sufficiently useful that ATLAS now produces both TAG and TAG_COMM by default
- Now rivaling physics TAGs in size, with some physics content
  - Usual problems with overlapping content, different names, thresholds
  - Some of these issues are being addressed by developers
- When the detector is well understood, TAG_COMM should fade away, but for now, TAG_COMM may be even more useful than physics TAGs

Commissioning TAG content

- See https://twiki.cern.ch/twiki/bin/view/Atlas/CommissioningTag for an explanation of TAG_COMM content and how to use it
- Event identification (run #, event #, lumi block, timestamp, BCID, …)
- Inner detector track info
- Muon track and segment info
- Tile muon fitter info
- Combined track counts
- Trigger info
  - Initially just L1; more now
- Calorimeter info (cell energy sums, …)
- Hit multiplicities
- Some timing information
- Now also information about (two each) electrons, taus, jets, …
  - Creeping toward physics TAG content …

Identifying collision candidates

- Two timing variables (delta t in ns) were added to TAG_COMM
  - MBTStimeDiff
  - LArECtimeDiff
- The idea is that when these readout times on the A and C sides of the detector are sufficiently small, there is a reasonable chance that the beams meet somewhere near the interaction point
- The cut used in producing the DESD_COLLCAND (collision candidate derived ESD) stream was(?)
  - Abs(LARECtimeDiff)<5 || Abs(MBTStimeDiff)<10
- Alternatively, one might simply look at the MTBS_1_1 trigger (MTBS triggered on both sides)
  - L1 trigger information is in both TAG_COMM and (physics) TAG

Some uses of TAG_COMM for early data to date (thanks to Jamie Boyd for much of this information)

- Bookkeeping checks of event counts (per lumi block / per trigger) - Eric Torrence
- Finding the ESD for nice events for official event displays - Event scanning team + many others (e.g., muon group)
- Trigger efficiency study for Min Bias analysis (some of these studies use TAG functionality to select the very small number of events from the BPTX which have an offline track, others just use the ntuple functionality to select trigger bit counts) - Emily Nurse, Will Bell, Regina Kwee, and David Berge
- Looking for noise in the RNDM stream for all systems - Jamie Boyd
- Looking for cosmic events with large/strange calo activity as a background for stopped gluino analysis - Paul Jackson
- To find muon candidates (both combined and standalone) in the first collision data (maybe using the TAG picking functionality) - Rosy Nikolaidou, Ed Moyse, Thijs Cornelissen, Domizia Orestano and Wolfgang Liebig
- Studying events for debugging the primary vertex finding - Kirill Prokofiev, Urban Bitenc
- Doing detailed bookkeeping (LB / BCID / trigger /) for MinBias analysis - Jamie Boyd, Alison Lister

Browsing commissioning TAGs

- Plan:
  - interactively find a run in the good run list
  - see what data are available
  - browse TAG_COMM for that run
  - look for jet information (cf. Belen's exercises)
  - select events with, say, two or more jets
  - see what we can learn about those events
  - extract corresponding event list
- Following slides show some of this (as backup)

Interlude: Good Run Lists

Run no.	Beam condictions	good run list	Special remarks	Beams	Solenoid	PIX/SCT
all runs untill 142383	900 GeV collisions	collisions_stablebeams_minbias_900GeV.xml	see here for details	Stable		
all runs untill 142383	900 GeV collisions	collisions_stablebeams_magsol_minbias_900GeV.xml	see here for details	Stable	On	
all runs untill 142383	900 GeV collisions	collisions_stablebeams_magsol_pixsctgreen_minbias_900GeV.xml	see here for details	Stable	On	Green
all runs untill 142383	900 GeV collisions	collisions_stablebeams_pixsctgreen_minbias_900GeV.xml	see here for details	Stable		Green
run 142308, 142402	2 TeV collisions	collisions_minbias_2TeV.xml	see here for details	Unstable		
run 142308, 142402	2 TeV collisions	collisions_magsol_minbias_2TeV.xml	see here for details	Unstable	On	

Sample Good Run List Excerpt

```
<!--This document is created by GoodRunsListWriter.-->
- <LumiRangeCollection>
  - <NamedLumiRange>
      <Name>collisions_stablebeams_minbias_900GeV</Name>
      <Version>30</Version>
      <Metadata Name="Query">Generated by LumiCalculator</Metadata>
    - <LumiBlockCollection>
        <Run>141748</Run>
        <LBRange Start="5" End="14"/>
      </LumiBlockCollection>
    - <LumiBlockCollection>
        <Run>141749</Run>
        <LBRange Start="2" End="17"/>
        <LBRange Start="19" End="101"/>
      </LumiBlockCollection>
    - <LumiBlockCollection>
        <Run>141811</Run>
        <LBRange Start="109" End="114"/>
        <LBRange Start="116" End="121"/>
        <LBRange Start="123" End="125"/>
        <LBRange Start="127" End="165"/>
      </LumiBlockCollection>
    - <LumiBlockCollection>
        <Run>141994</Run>
        <LBRange Start="58" End="74"/>
        <LBRange Start="76" End="79"/>
        <LBRange Start="81" End="83"/>
        <LBRange Start="85" End="209"/>
      </LumiBlockCollection>
    - <LumiBlockCollection>
        <Run>141999</Run>
        <LBRange Start="31" End="33"/>
```

AMI: MinBias TAG datasets for run 141749

is='VALID' AND ((logicalDatasetName like 'data09%900gev%141749%MinBias%TAG%'))

logicalDatasetName	nFiles	totalEvents	dataType
data09_900GeV.00141749.physics_MinBias.merge.TAG.f185_m298 DQ2 - GANGA export - Provenance	3	19317	TAG
data09_900GeV.00141749.physics_MinBias.merge.TAG.f187_m304 DQ2 - GANGA export - Provenance	5	19317	TAG
data09_900GeV.00141749.physics_MinBias.merge.TAG.f196_m325 DQ2 - GANGA export - Provenance	10	19317	TAG
data09_900GeV.00141749.physics_MinBias.merge.TAG.r988_p62 DQ2 - GANGA export - Provenance	10	19317	TAG
data09_900GeV.00141749.physics_MinBias.merge.TAG_COMM.f185_m297 DQ2 - GANGA export - Provenance	1	19317	TAG_COMM
data09_900GeV.00141749.physics_MinBias.merge.TAG_COMM.f187_m303 DQ2 - GANGA export - Provenance	1	19317	TAG_COMM
data09_900GeV.00141749.physics_MinBias.merge.TAG_COMM.f196_m324 DQ2 - GANGA export - Provenance	1	19317	TAG_COMM
data09_900GeV.00141749.physics_MinBias.merge.TAG_COMM.r988_p63 DQ2 - GANGA export - Provenance	1	19317	TAG_COMM
data09_900GeV.00141749.physics_MinBias.recon.TAG_COMM.r988 DQ2 - GANGA export - Provenance	99	19317	TAG_COMM

What can you learn from this?

- Most information in a dataset name is (usually) self-explanatory
  - "Project" is data09 (real data)
  - Sub-project is 900GeV running
  - Run number is 141749
  - Stream name is MinBias
  - Processing stage is recon or merge (in this case)
  - Data product is TAG or TAG_COMM
  - Configuration tag is …
- N.B.: merge does not have a consistent meaning
  - Physics TAGs are produced when many AOD files are merged into fewer files, hence "merge" in the processing stage
    - *Physics TAG files are not always further merged*
  - TAG_COMM are produced when ESD are created, which are not (to date) routinely merged, hence "recon" in the processing stage
    - *TAG_COMM files merged in a separate stage later*

Configuration tags

- Look up in AMI (AMI is your friend)

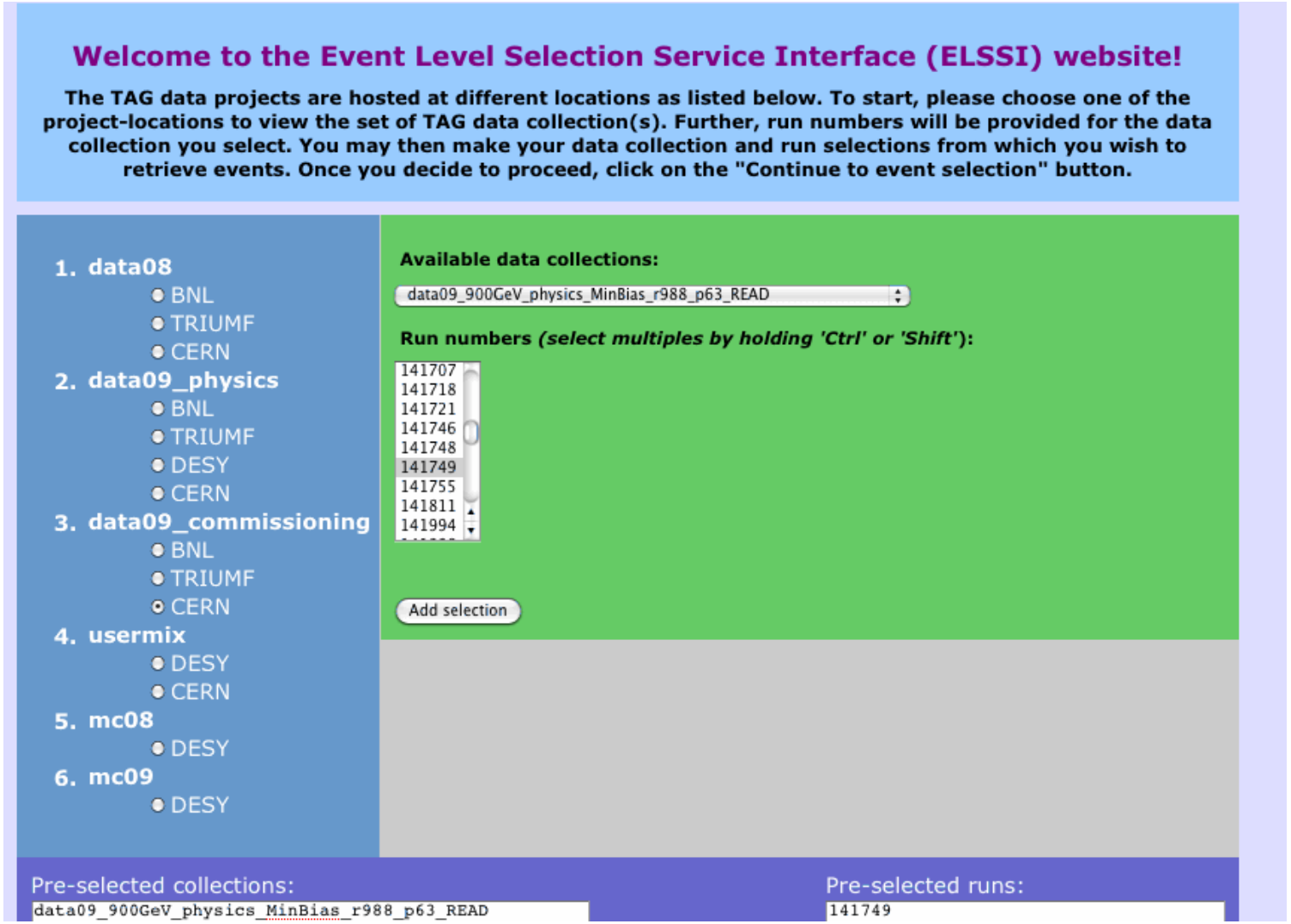
productionStep	tag	SWReleaseCache	DBRelease
recon	r988	15.5.4.10	8.1.3.1

productionStep	tag	SWReleaseCache	DBRelease	Geometry
recon	f196	AtlasTier0-15.5.4.9	none	Auto-config. See dataset info.

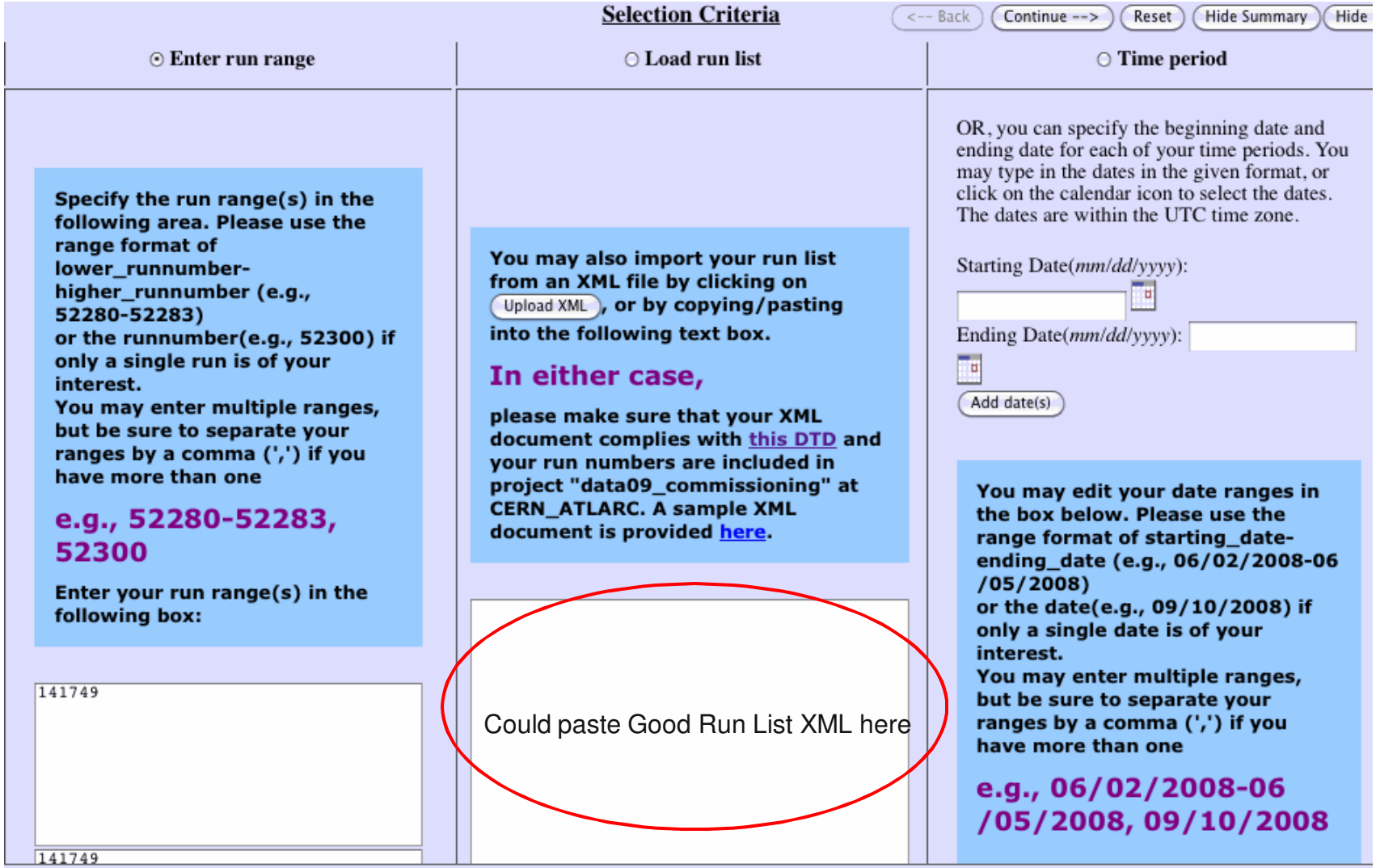
Other information

- Configuration tag hint:
  - f signals first-pass (Tier 0) processing
  - r signals reconstruction generally (in this case, a hint that these are the reprocessed data)
- Note that the event counts on all of these datasets agree(!)
  - Reassuring news, not to be taken for granted
  - ATLAS has some metadata “issues” here
  - When they do not agree, you may wish to compare event and file counts to the corresponding upstream data (ESD and AOD datasets, for example)
    - *Ensure that you know what you are and are not getting*

TAG browser (ELSSI): data source selection



ELSSI: further run selection



Some sample collision candidate selections

Your logical query is :
SELECT 'data09_900GeV_physics_MinBias_r988_p63_READ' as stream_name,count(runnumber) as count FROM ATLAS_TAGS_COMM_2009.POOL_COLLECTION_DATA_585_R WHERE (RunNumber = 141749) and (abs(LArECtimeDiff)<5)

STREAM_NAME	# OF EVENTS FROM YOUR QUERY	TOTAL # OF EVENTS
data09_900GeV_physics_MinBias_r988_p63_READ	7842	2619863

Your logical query is :
SELECT 'data09_900GeV_physics_MinBias_r988_p63_READ' as stream_name,count(runnumber) as count FROM ATLAS_TAGS_COMM_2009.POOL_COLLECTION_DATA_585_R WHERE (RunNumber = 141749) and (abs(MBTStimeDiff)<10)

STREAM_NAME	# OF EVENTS FROM YOUR QUERY	TOTAL # OF EVENTS
data09_900GeV_physics_MinBias_r988_p63_READ	7584	2619863

Your logical query is :
SELECT 'data09_900GeV_physics_MinBias_r988_p63_READ' as stream_name,count(runnumber) as count FROM ATLAS_TAGS_COMM_2009.POOL_COLLECTION_DATA_585_R WHERE ((RunNumber=141749 AND (bitand(CTPWord7, power(2,4)) > 0)))

STREAM_NAME	# OF EVENTS FROM YOUR QUERY	TOTAL # OF EVENTS
data09_900GeV_physics_MinBias_r988_p63_READ	10941	2619863

Some sample jet queries

Your logical query is :
SELECT 'data09_900GeV_physics_MinBias_r988_p63_READ' as stream_name,count(runnumber) as count FROM ATLAS_TAGS_COMM_2009.POOL_COLLECTION_DATA_585_R WHERE (RunNumber = 141749) and (jet1_pt>0)

STREAM_NAME	# OF EVENTS FROM YOUR QUERY	TOTAL # OF EVENTS
data09_900GeV_physics_MinBias_r988_p63_READ	344	2619863

Your logical query is :
SELECT 'data09_900GeV_physics_MinBias_r988_p63_READ' as stream_name,count(runnumber) as count FROM ATLAS_TAGS_COMM_2009.POOL_COLLECTION_DATA_585_R WHERE (RunNumber = 141749) and (jet2_pt>0)

STREAM_NAME	# OF EVENTS FROM YOUR QUERY	TOTAL # OF EVENTS
data09_900GeV_physics_MinBias_r988_p63_READ	45	2619863

Your logical query is :
SELECT 'data09_900GeV_physics_MinBias_r988_p63_READ' as stream_name,count(runnumber) as count FROM ATLAS_TAGS_COMM_2009.POOL_COLLECTION_DATA_585_R WHERE (RunNumber = 141749) and (jet2_pt>0 AND abs(LArECtimeDiff)<5)

STREAM_NAME	# OF EVENTS FROM YOUR QUERY	TOTAL # OF EVENTS
data09_900GeV_physics_MinBias_r988_p63_READ	37	2619863

A closer look at multijet candidates

Rows selected (total count = 45):

STREAM_NAME	LArECtimeDiff	MBTStimeDiff	Njet	jet1_pt/1000.0	jet2_pt/1000.0
data09_900GeV_physics_MinBias_r988_p63_READ	-999	2.28666234	0	4.14417041	3.86780225
data09_900GeV_physics_MinBias_r988_p63_READ	1.87626982	-1.27625775	0	5.38649463	3.84280981
data09_900GeV_physics_MinBias_r988_p63_READ	.25298804	-1.16886091	0	9.07076953	7.83537939
data09_900GeV_physics_MinBias_r988_p63_READ	1.87795901	-1.10370541	0	5.40882275	3.91762695
data09_900GeV_physics_MinBias_r988_p63_READ	1.99590623	.612971246	0	5.34178271	3.16694751
data09_900GeV_physics_MinBias_r988_p63_READ	-2.5293386	.787785709	0	5.19507764	3.9247478
data09_900GeV_physics_MinBias_r988_p63_READ	.470768332	.690248072	0	3.95103223	3.46424561
data09_900GeV_physics_MinBias_r988_p63_READ	-3.05138588	-.742467403	0	11.2119043	3.81054102
data09_900GeV_physics_MinBias_r988_p63_READ	-2.0127275	-.253798038	0	8.78374902	5.96993799
data09_900GeV_physics_MinBias_r988_p63_READ	2.00180411	.0568612143	0	3.88486963	3.51509497
data09_900GeV_physics_MinBias_r988_p63_READ	-.473535538	-2.68249416	0	5.18315723	3.38182446
data09_900GeV_physics_MinBias_r988_p63_READ	1.41498232	-1.09247649	0	3.5434082	3.29732495
data09_900GeV_physics_MinBias_r988_p63_READ	.426819921	-.28631711	0	6.24510254	4.88078662
data09_900GeV_physics_MinBias_r988_p63_READ	1.01733756	-.981472254	0	5.10387598	4.63971582
data09_900GeV_physics_MinBias_r988_p63_READ	-999	-999	0	3.17579053	3.15900439
data09_900GeV_physics_MinBias_r988_p63_READ	5.41346645	.778448999	0	5.72449707	3.23516211
data09_900GeV_physics_MinBias_r988_p63_READ	-.0988512039	-.170625001	0	5.99644482	5.98341797
data09_900GeV_physics_MinBias_r988_p63_READ	1.02360833	.760057211	0	4.70923828	4.08097803
data09_900GeV_physics_MinBias_r988_p63_READ	.822659016	-1.43340385	0	10.5424834	3.78923193
data09_900GeV_physics_MinBias_r988_p63_READ	.521181464	-.342449307	0	4.79362646	3.28303442
data09_900GeV_physics_MinBias_r988_p63_READ	3.12715578	-1.884812	0	6.12298584	4.2383291
data09_900GeV_physics_MinBias_r988_p63_READ	2.87915802	-1.14235795	0	9.08439355	4.07767139
data09_900GeV_physics_MinBias_r988_p63_READ	1.69764996	-.614749253	0	4.31867871	3.067729
data09_900GeV_physics_MinBias_r988_p63_READ	1.88617122	.577213228	0	4.33161328	4.20719238
data09_900GeV_physics_MinBias_r988_p63_READ	-999	-999	1	69.2043047	10.8486748
data09_900GeV_physics_MinBias_r988_p63_READ	1.68111777	1.59831357	0	5.15617822	3.53890381
data09_900GeV_physics_MinBias_r988_p63_READ	3.6219759	.570297241	0	3.75264526	3.26298267
data09_900GeV_physics_MinBias_r988_p63_READ	1.28013086	-.672589004	0	7.02211768	5.4329126
data09_900GeV_physics_MinBias_r988_p63_READ	-.0126501024	.203486204	0	6.3735874	4.50564697
data09_900GeV_physics_MinBias_r988_p63_READ	2.61427236	-4.32838154	0	4.88528271	4.40450146
data09_900GeV_physics_MinBias_r988_p63_READ	2.10074425	.111123681	0	5.1328374	3.89952686
data09_900GeV_physics_MinBias_r988_p63_READ	-2.24911737	-.16276522	0	4.54315771	4.01741577

Aside: metadata for eventual luminosity and cross-section calculations

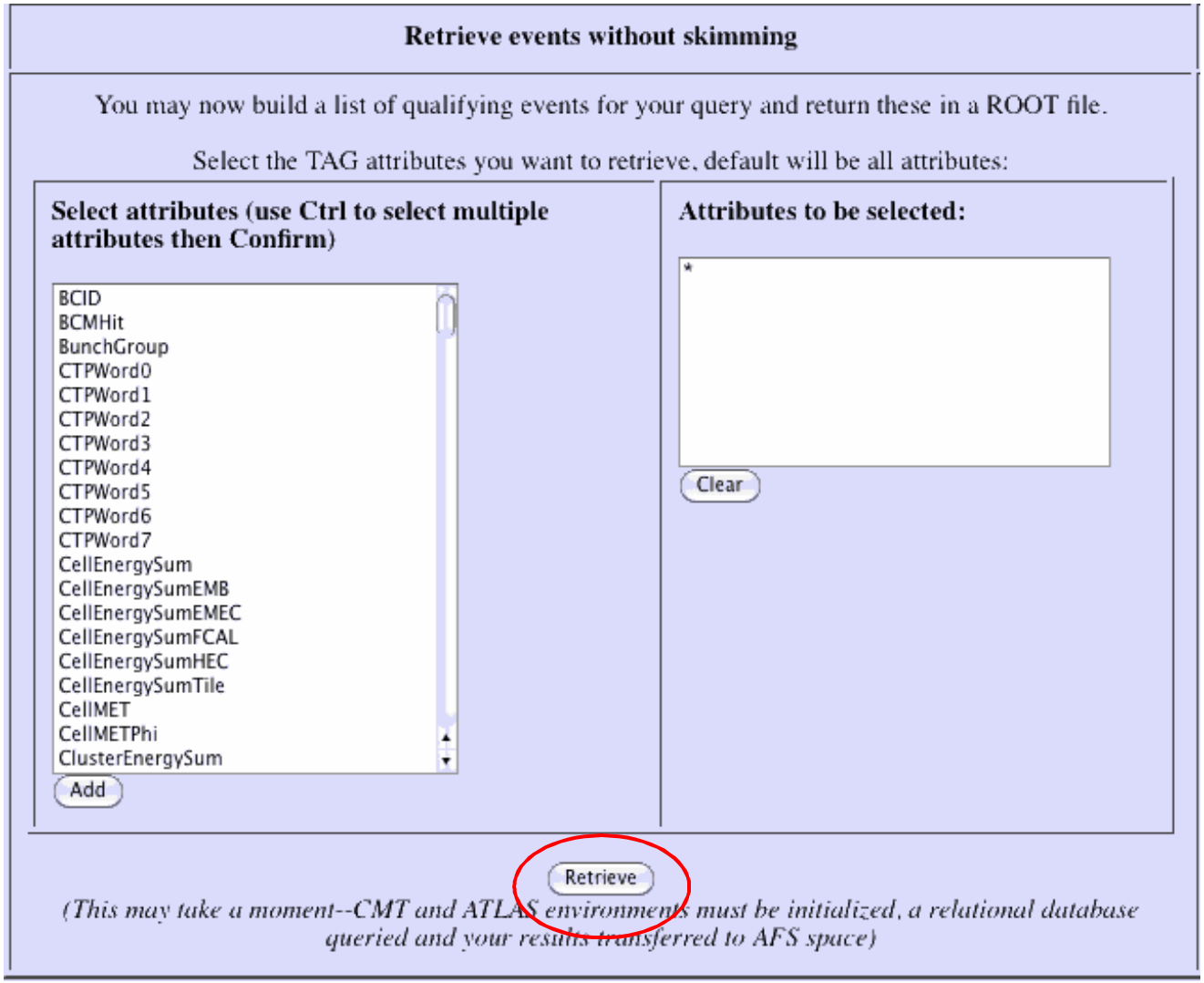
RunNumbers with their LumiBlockN's where the selected trigger(s) () is(are) active. If no triggers are selected, all LumiBlockN's will be counted.
(total count = 99)

(1) data09_900GeV_physics_MinBias_r988_p63_READ:

RunNumber	LumiblockN	Subtotal
141749	2~17,19~101	99

Note that in for this particular run (special case), events from the “bad” lumi blocks (in the data quality sense) were never reconstructed

Extracting selected events



Extracted events

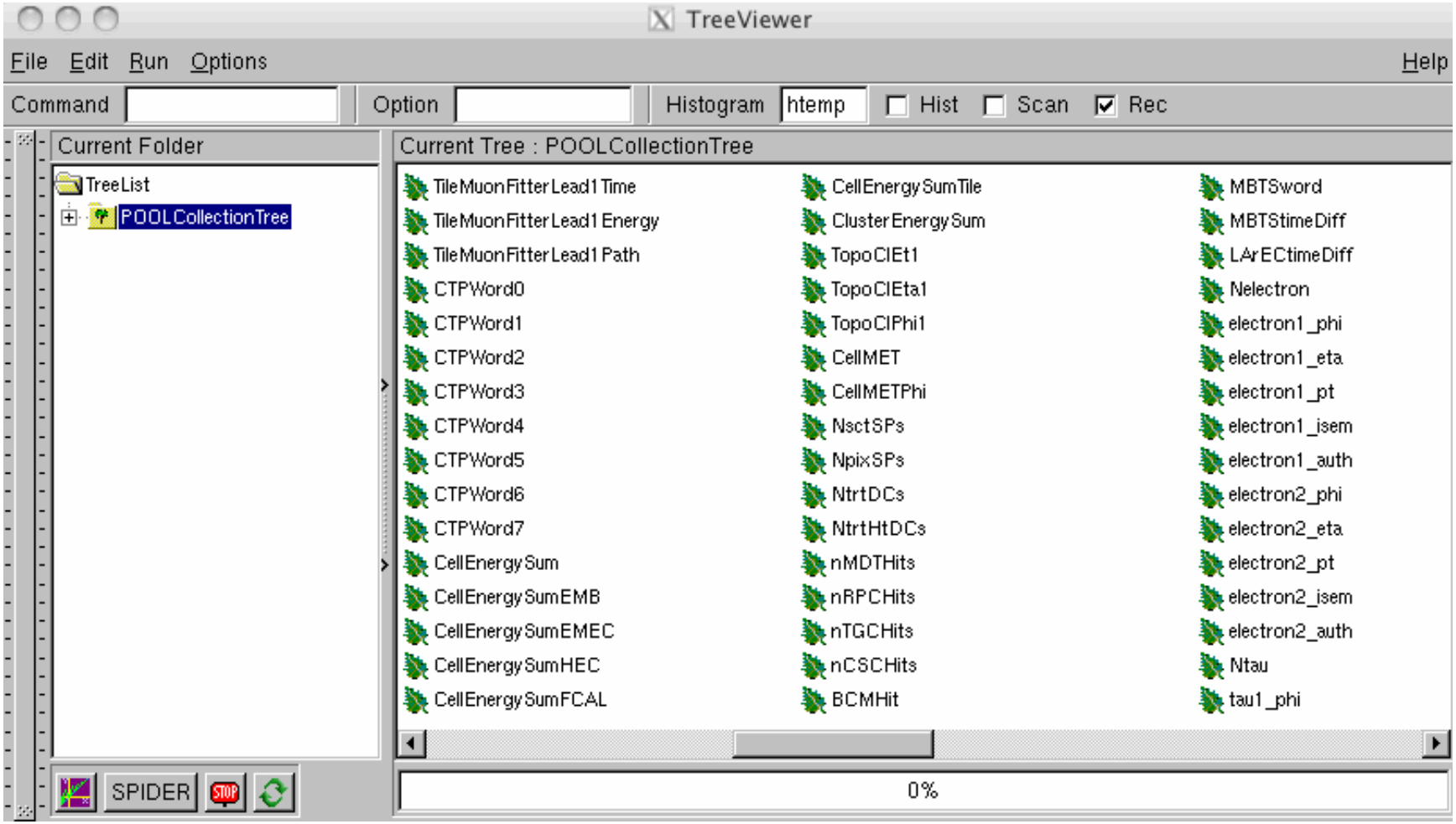
- Extraction retrieves the list of events that passed your cuts, along with references to upstream data
  - For TAG_COMM, this means RAW and ESD
  - Returned as a relatively simple ROOT file
- Can use this returned file as input to any job that reads RAW or ESD
  - With proper job option changes, and *if* the pointed-to data files are reachable
- If you needed ESD for these 37 events only and didn't want to move the entire MinBias ESD for this run to your local site, you could run an ESD-to-ESD skim with this event list as input
- For instructions for reading ESD or RAW via TAG_COMM on the grid, see, for example, Paul Jackson's https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TagCommGridAccess
- Work to integrate skimming into the TAG browser is in progress
  - ESD→ESD, AOD→AOD
  - Or upload your personal favorite AOD→myDPD, for example

TAG files

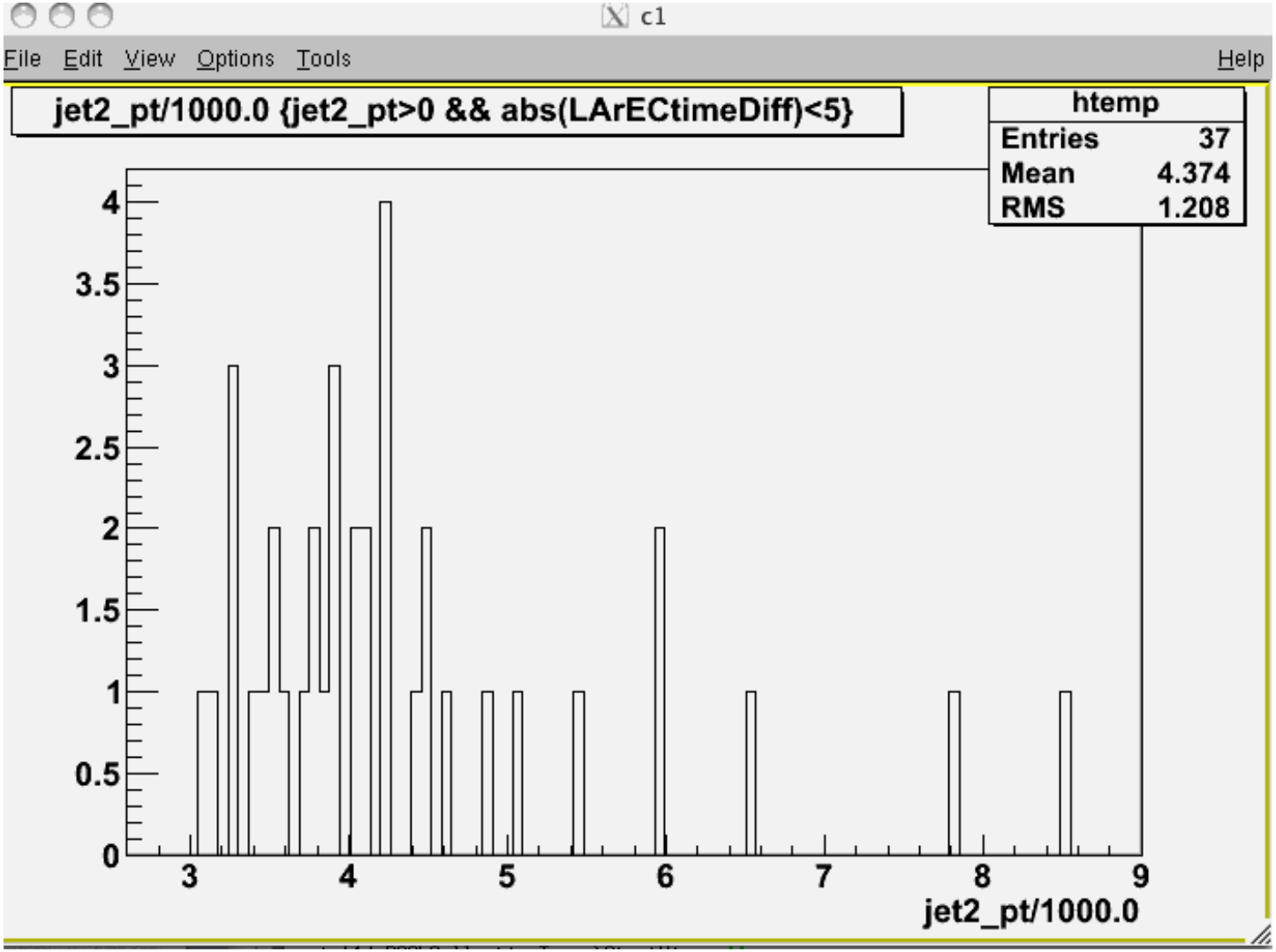
- Sergei Chekanov has put two TAG datasets from the Christmas reprocessing of this run on data1
  - Physics TAGs and TAG_COMM
  - data09_900GeV.00141749.physics_MinBias.merge.TAG_COMM.r988_p63_tid104906_00 (one file)
  - data09_900GeV.00141749.physics_MinBias.merge.TAG.r988_p62_tid102040_00 (ten files)
- Can be used as simple ROOT files, or as input to Athena jobs
- Belen's ESD-reading examples can be modified to use TAG files as input as follows (thanks to Belen for the recipe; direct questions to Belen or Jack)

```
################################
readTAG=Trueif not "PoolTAGInput" in
dir():PoolTAGInput=["/data/nfs/users/belen/DATA09_900GeV/data09_900GeV.00142042.physics_
MinBias.merge.TAG_COMM.f187_m303/data09_900GeV.00142042.physics_MinBias.merge.TAG
_COMM.f187_m303._0001.1"]from AthenaCommon.AthenaCommonFlags import
athenaCommonFlagsathenaCommonFlags.PoolTAGInput=PoolTAGInput# logical
selectionathenaCommonFlags.PoolInputQuery="EventNumber == 36868"from
IOVDbSvc.IOVDbSvcConf import IOVDbSvcIOVDbSvc.GlobalTag='COMCOND-ES1P-001-
00'  ################################
```

TAG files look like ROOT “ntuple” files



Cuts on TAG_COMM contents in ROOT



Feel free to experiment

- Jack Cranshaw and Qizhi Zhang (Argonne), key TAG developers, are here and can answer most questions
- Belen may be able to help with integration with her ESD-reading exercises
- As a last resort, you could, I suppose, ask me …