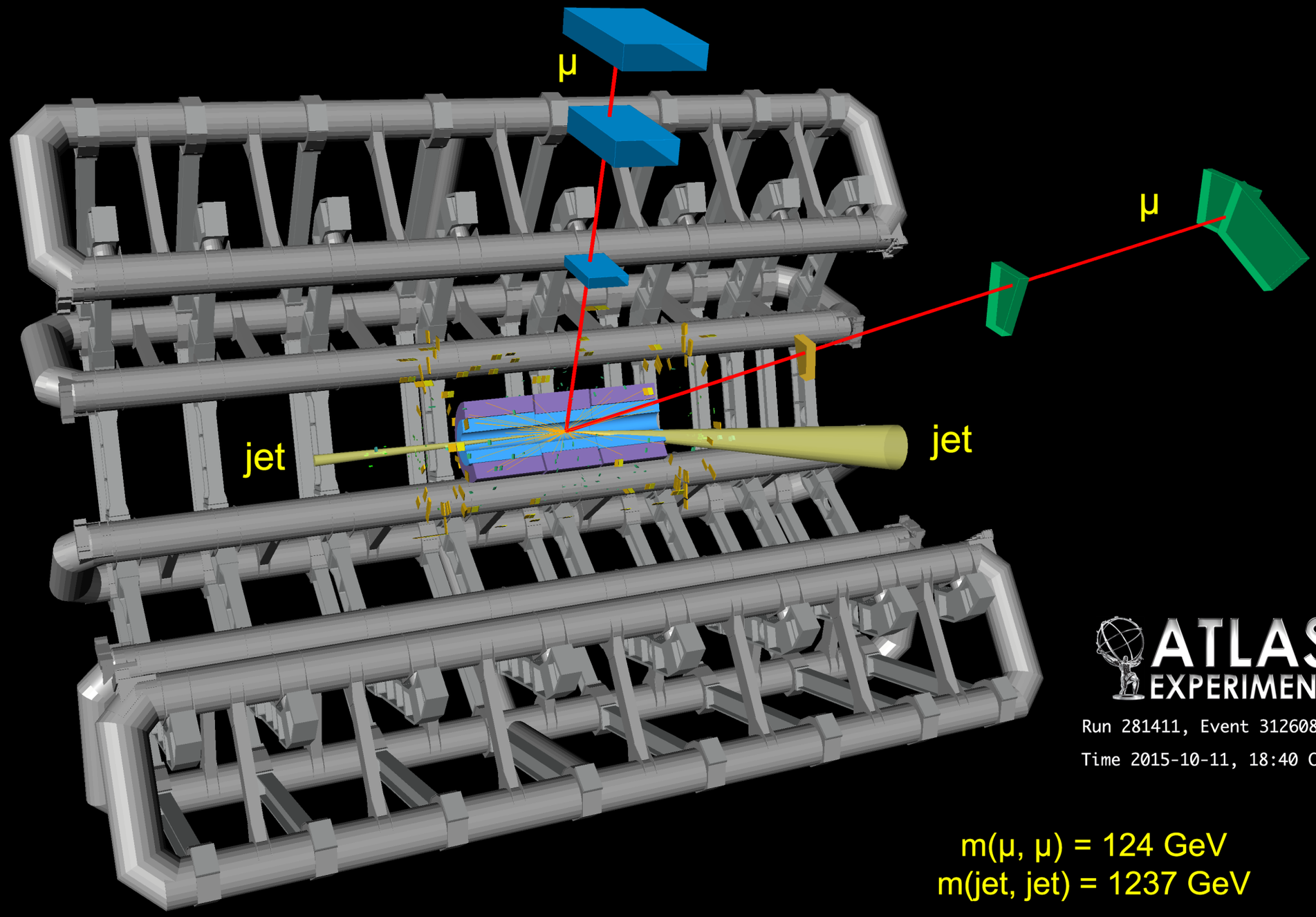
Introduction to the ATLAS Software Tutorial



ATLAS EXPERIMENT

Run 281411, Event 312608026

Time 2015-10-11, 18:40 CEST

m(μ, μ) = 124 GeV

m(jet, jet) = 1237 GeV

James Walder
Lancaster University

Welcome to the Tutorial

- Welcome to the ATLAS software tutorial
  - One packed week of tutorials covering many interesting and essential topics for ATLAS users.

- We hope that by the end of the week you will:
  - Understand different types of data format, what they contain and how to analyse them
    - be comfortable using the ATLAS tools

- Be able to locate and retrieve data / information
  - and understand and use the tools available for Physics analysis.
  - Including cutting-edge knowledge of xAOD and ATLAS Software.

Tutorial Outline

- Today:
  - Preparing your account and running simple jobs
  - The ATLAS **Event Data Model** (Part I & II)
    - what's in ATLAS data
  - Analysing xAOD data (Introduction)
- Wednesday:
  - The ATLAS **Event Data Model** (Part III)
    - what's in ATLAS data
    - **Trigger** for analysis
  - Analysing xAOD data (More details including AthAnalysis)
  - Finding data:
    - **AMI, COMA, run-query**
  - An introduction to the **GRID**
  - **ROOT: New Features**

Tutorial Outline II

- Thursday:
  - **CMake**; version control
  - **git** (for Analysis)
  - **Analysis** tool-kits / central analysis algorithms
  - **Frameworks and ToolKits**
    - **Central analysis algorithms**
      - **Data reduction** tools: the derivation framework
  - **Docker** images for analysis
  - **GitLab** for Analysis

- Friday:
  - **Monte Carlo**
  - Event Displays
  - Intro to statistical analysis with **HistFitter** and **pyhf**

Format of Tutorial

- A lot of material to cover:
  - Ranges from beginner to advanced topics within the sessions.
  - No expectation to be able to cover and digest all exercises.
- You will hear about, and have sessions on, the cutting-edge of xAOD tools
  - Talks and TWiki pages will remain available to refer back to.
- All information can be found from the main Agenda page:
  - https://indico.cern.ch/event/831761/timetable/

Rooms

- These rooms will be used throughout the tutorial:

Date	Room
Tue AM	30-7-018 - Kjell Johnsen Auditorium]
PM	30-7-018 - Kjell Johnsen Auditorium]
Wed AM	222-R-001
PM	222-R-001
Thurs AM	222-R-001
PM	222-R-001
Fri AM	222-R-001
PM	222-R-001

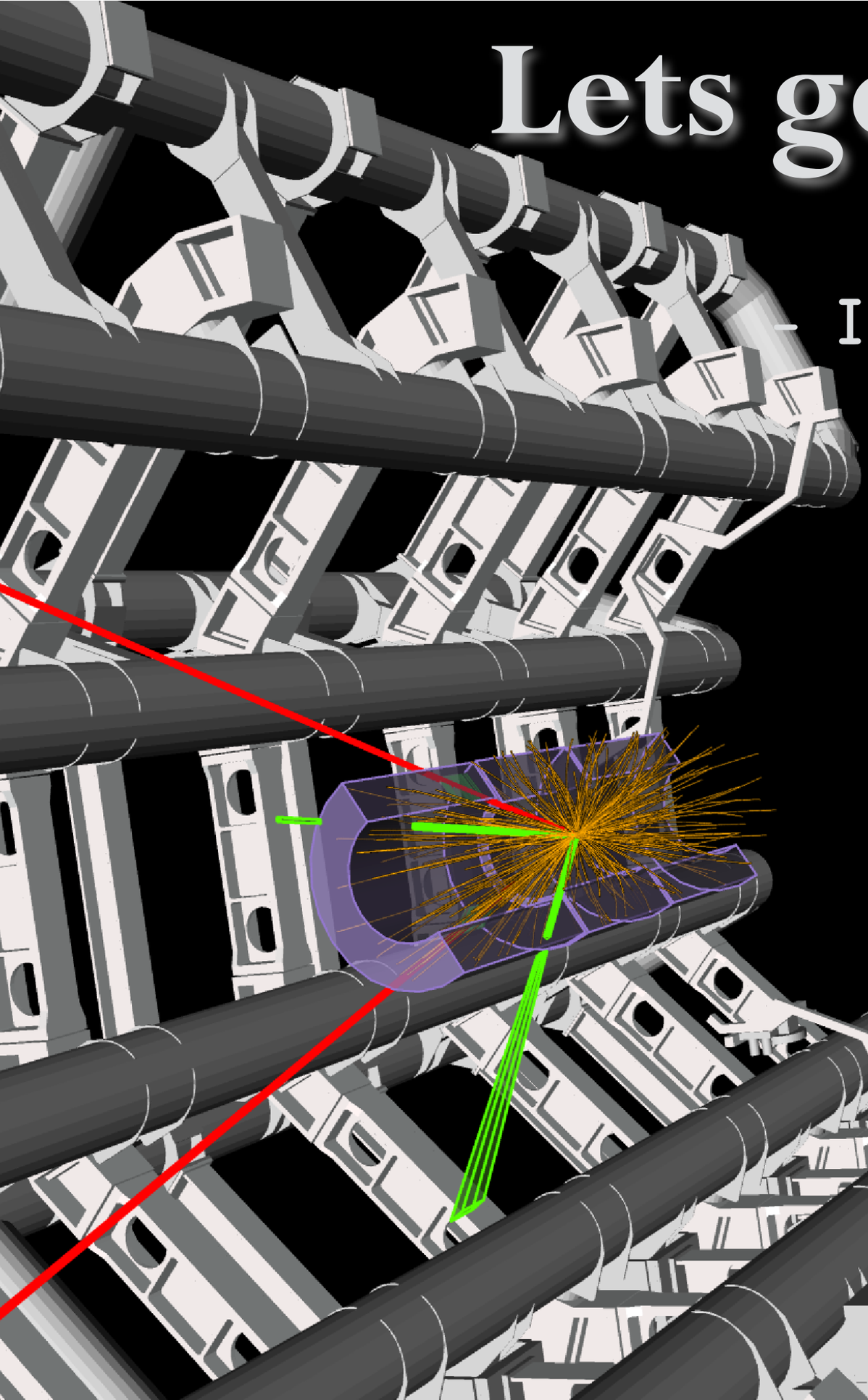
- Please check the agenda for any last-minute alterations.

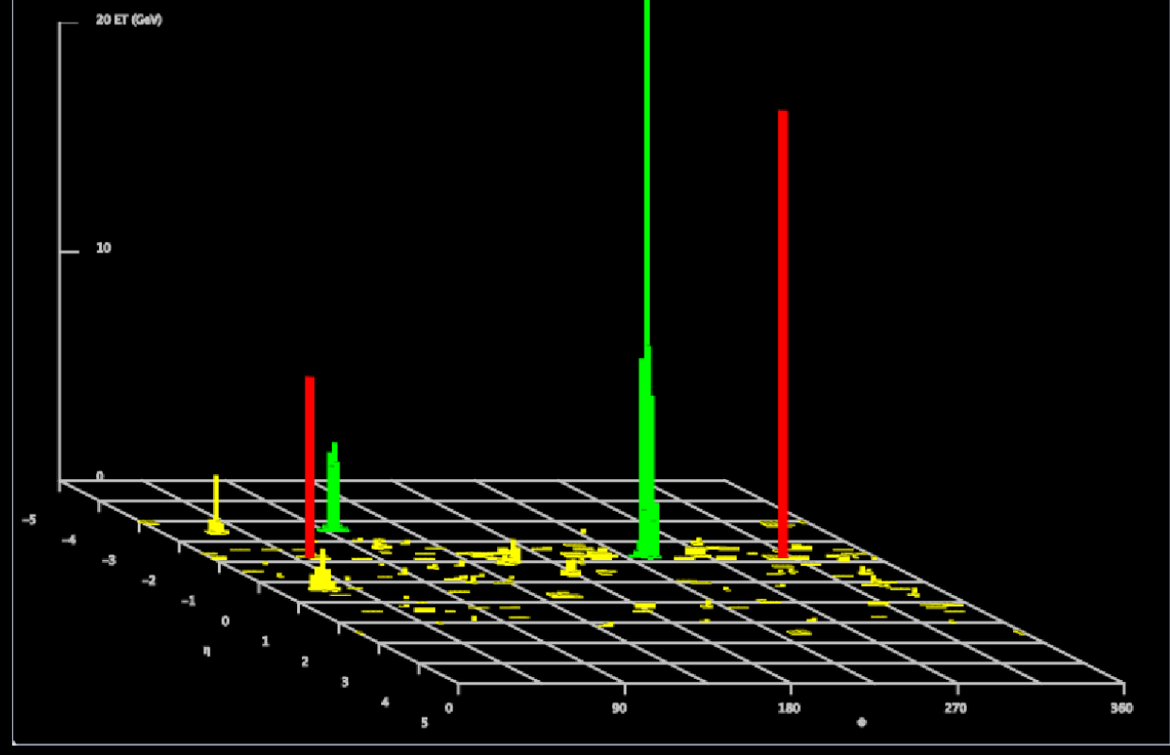
Lets get Started!

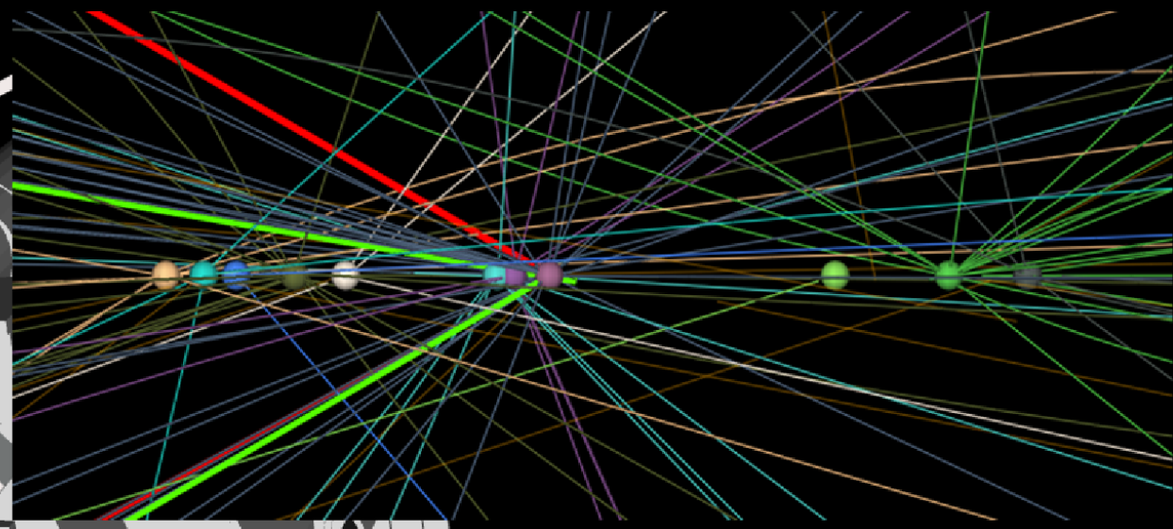
ATLAS EXPERIMENT
http://atlas.ch

Run: 205113
Event: 12611816
Date: 2012-06-18
Time: 11:07:47 CEST

- Introduction to ATLAS software

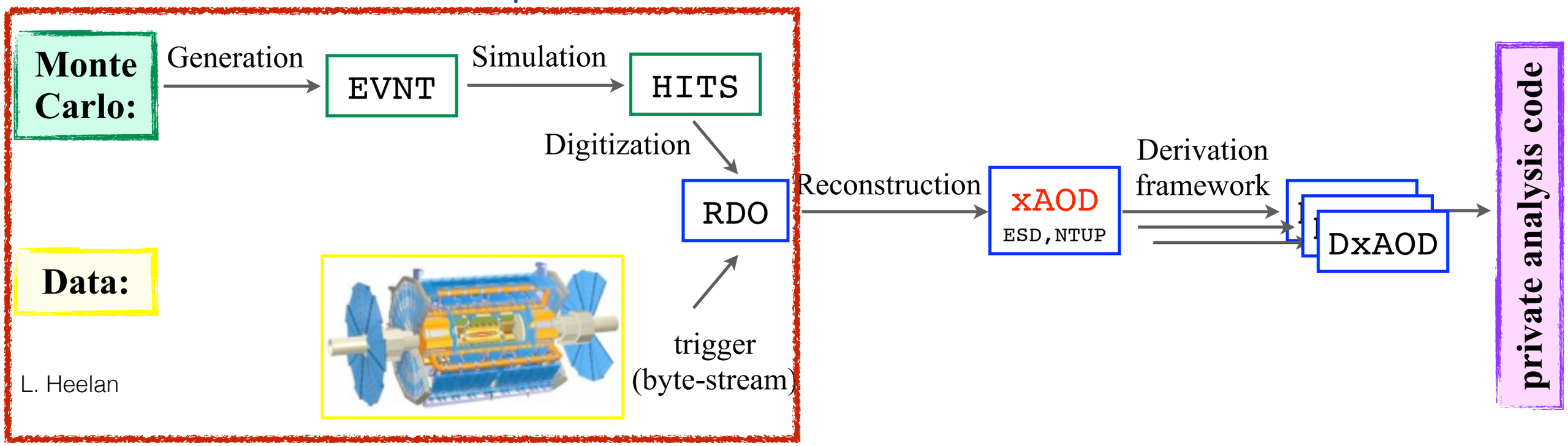






Analysis in ATLAS

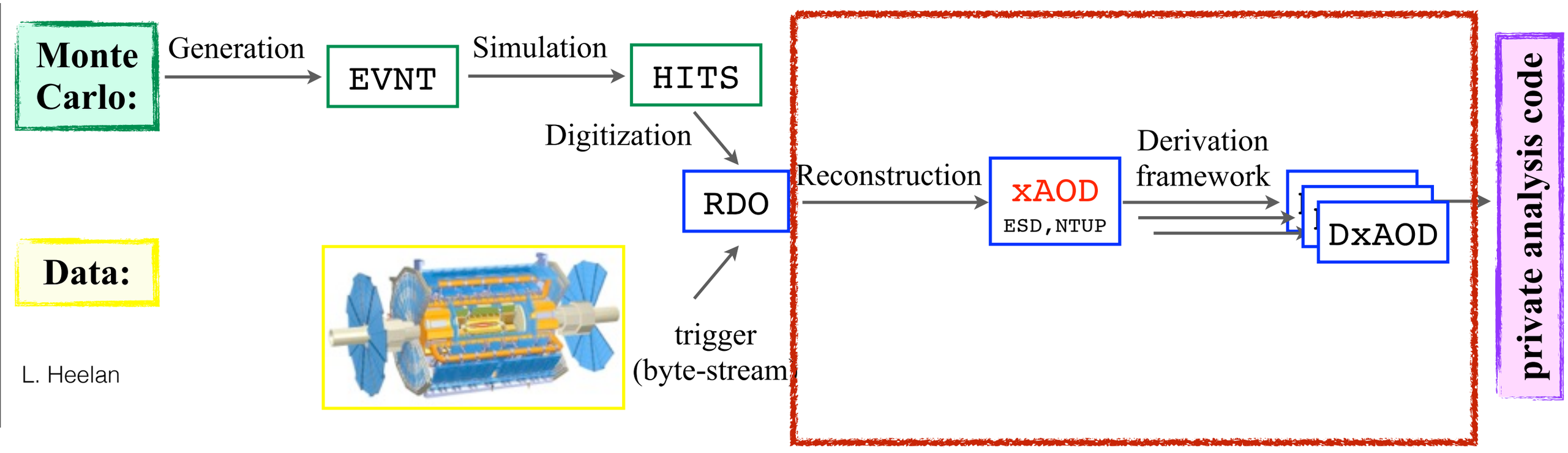
- Complex series of steps to take data from detector / generator,
  - to data format ready for analysis:
    - See Data Preparation talk.



- MC simulation
- RAW data

Analysis in ATLAS

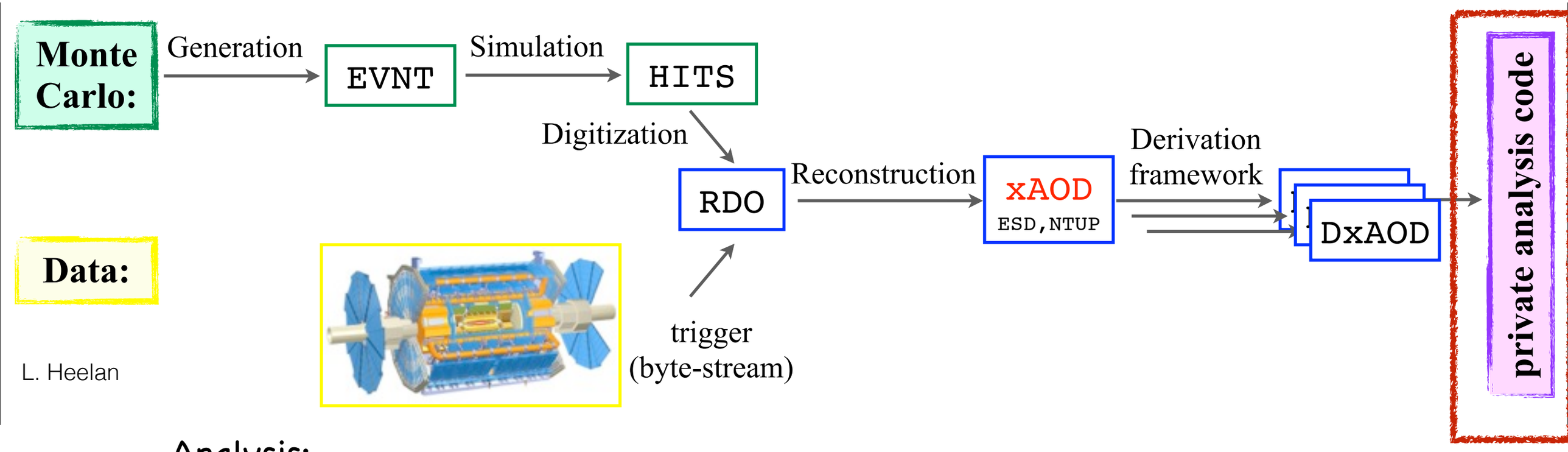
- Complex series of steps to take data from detector / generator,
  - to data format ready for analysis:
    - See Data Preparation talk.



- Reconstruction; creates xAOD (now also called just AOD).
- Derivation format (makes xAODs -> DxAOD)
  For analysis, will interact with (D)xAOD data format:
  - Athena and 'root'-readable
  - DxAODs are further reduced for only the information needed for particular subset of analyses (see Derivation Framework talk).

Analysis in ATLAS

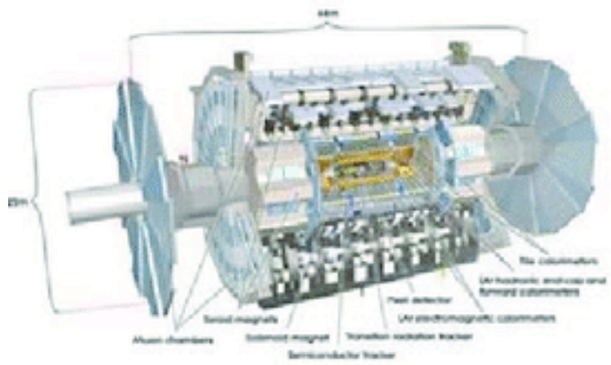
- Complex series of steps to take data from detector / generator,
  - to data format ready for analysis:
    - See Data Preparation talk.

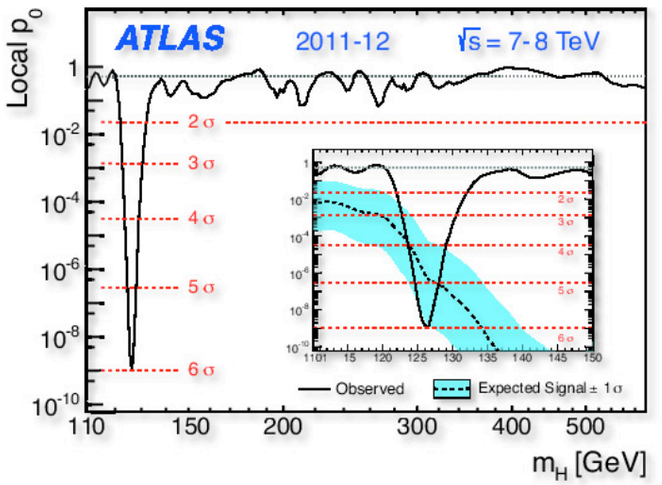


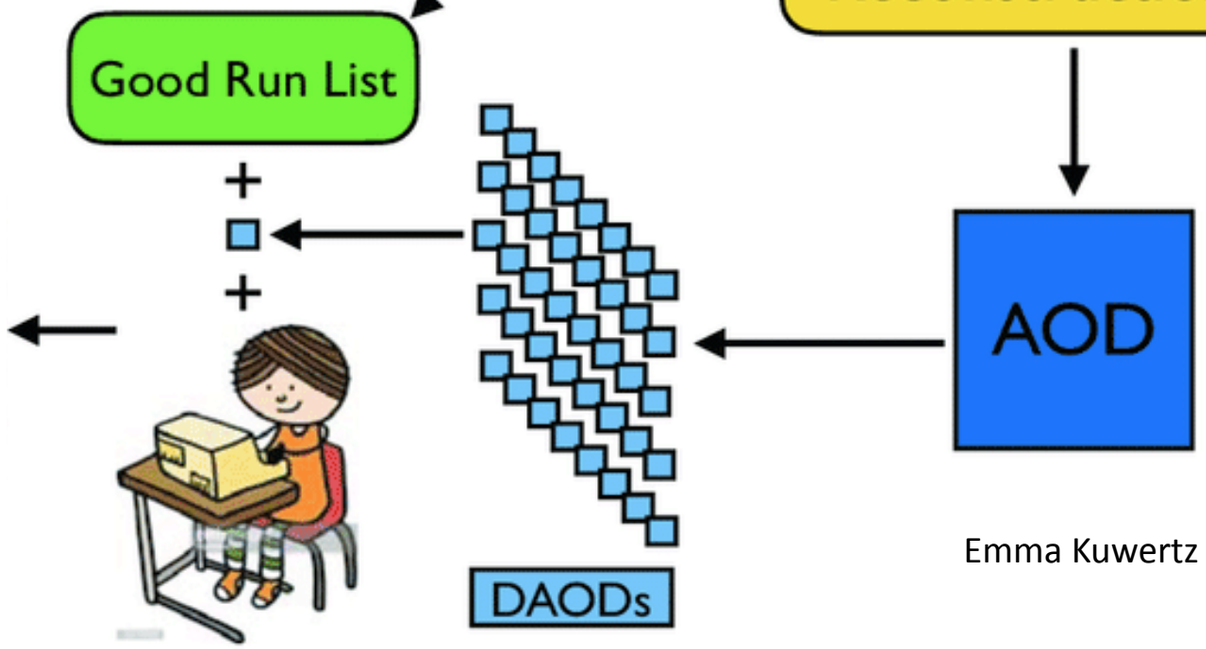
L. Heelan

- Analysis:
  - Performed in Athena (AthAnalysis) or AnalysisBase (e.g. 'EventLoop')
  - Runs (usually) over the outputs of the Derivation formats (e.g. DAOD_PHYS)
    - Make small ntuples or histograms for final analysis

Run-2 Analysis Model







- Chain of steps to convert RAW and Generated data into 'root-readable' AODs
- Analysis concerned typically with: (Trigger); definition of DAOD format, and subsequent steps shown in the tutorial.

The ATLAS Software Model

- Experiment of thousands of collaborators:
  - Need to provide:
    - Common tools
    - Common data formats:
      - Lot of work in data reduction techniques (cf. Derivations)
    - Recommendations on approaches to analyse the data
    - Common Combined Performance tools: e.g. muon quality selections

- Result is that:
  - Help analyses become reproducible,
  - Common tools => reduces likelihood of unnoticed bugs,
  - Definitions of objects standardised across analyses; ie. can compare..
  - Common set of definitions => easier to convince edBoards, etc. of what the analysis is doing.
    - Can utilise achievements of other people's work.
- Development of Frameworks: e.g. Athena (AthAnalysis), 'EventLoop' (AnalysisBase):
  - Something you plug your code into:
- And Toolkits;
  - You plug these into your code (e.g to obtain good quality calibrated physics objects).

Where can I run ATLAS software?

- CVMFS – Cern Virtual Machine File System:
  - Provides homogeneity across ranges of systems:
    - LXPLUS; Grid; Local institute; Tier-3; Laptop
- Laptop:
  - Variety of options available:
    - Standalone (for AnalysisBase)
    - CernVM (+CVMFS), Containers
- Containers;
  - Containers (Docker, Singularity) becoming more actively used (See tutorials)

Where?	Pros	Cons
On LXPLUS	**All set up for you; quickest way to start; ideal for testing**	**Limited quota and sharing nodes with other users.**
At my institute	**Lots of disk space; no limitations on execution time.**	**You/your institute has to maintain it and pay for it,**
On the Grid	**The only way of running over huge amounts of ATLAS data**	**A bit difficult at first; the machine you submit from must have ATLAS software. Bookeeping**
Laptop	**You are in control; many tools now to help**	**You are the maintainer; limited resources**

Setting up ATLAS Software

- Type **setupATLAS** to set up environment (with CVMFS set up):
  - (see Hands-on Session)

```
lsetup
 lsetup agis
 lsetup asetup
 lsetup atlantis
 lsetup eiclient
 lsetup emi
 lsetup ganga
 lsetup lcgenv
 lsetup panda
 lsetup pod
 lsetup pyami
 lsetup rcsetup
 lsetup root
 lsetup rucio
 lsetup sft
 lsetup xcache
 lsetup xrootd
advancedTools
diagnostics
helpMe
printMenu
showVersions
```

```
lsetup <tool1> [ <tool2> ...] (see lsetup -h):
 ATLAS Grid Information System
 (or asetup) to setup an Athena release
 Atlantis: event display
 Event Index
 EMI: grid middleware user interface
 Ganga: job definition and management client
 lcgenv: setup tools from cvmfs SFT repository
 Panda: Production ANd Distributed Analysis
 Proof-on-Demand (obsolete)
 pyAMI: ATLAS Metadata Interface python client
 (or rcSetup) to setup an ASG release
 ROOT data processing framework
 distributed data management system client
 setup tools from SFT repo (use lcgenv instead)
 XRootD local proxy cache
 XRootD data access
advanced tools menu
diagnostic tools menu
more help
show this menu
show versions of installed software
```

  - **lsetup root;** or, say
  - **lsetup 'asetup AthAnalysis,21.2.93,here' panda**
    - many other options available …
- See Tues morning session for more details

SL6 and Centos7

- Lxplus recently moved its default nodes from Scientific Linux 6 to Centos 7.

- ATLAS currently building against SLC6 in most cases see for example:
  - showVersions --show=athena | grep 21.2 | less

- These release will work on lxplus (and other) Centos 7 machines.
- Note - non-analysis releases are *typically* still using slc6 releases.

AnalysisBase-21.2.87-x86_64-slc6-gcc62-opt
AnalysisBase-21.2.88-x86_64-centos7-gcc8-opt
AnalysisBase-21.2.88-x86_64-slc6-gcc62-opt
AnalysisBase-21.2.89-x86_64-centos7-gcc8-opt
AnalysisBase-21.2.89-x86_64-slc6-gcc62-opt
AnalysisBase-21.2.9-x86_64-slc6-gcc62-opt
AnalysisBase-21.2.90-x86_64-**centos7**-gcc8-opt
AnalysisBase-21.2.90-x86_64-**slc6**-gcc62-opt
AnalysisBase-21.2.91-x86_64-centos7-gcc8-opt
AnalysisBase-21.2.92-x86_64-centos7-gcc8-opt
AnalysisBase-21.2.93-x86_64-centos7-gcc8-opt
AnalysisBase-21.2.94-x86_64-centos7-gcc8-opt
AnalysisBase-21.2.95-x86_64-centos7-gcc8-opt

- Current recommendations: For local-only work, lxplus (with Centos 7) is fine:
  - The grid should also now cope with centos7 versions

- If you do have a problem on the grid, then:
  - Use a container:
    - setupAtlas -c SLC6
  - or, login to lxplus SLC6 machine
    - ssh -l <username> lxplus6.cern.ch
- Once transition period is over, this will not be necessary
- **In general, you should not need to worry about this!**

Running Athena

- In the first practical session you will see how to set up Athena
  - in principle no code needs to be checked-out and modified
  - it is all contained in an Athena release which is available after you set it up

- To make modifications to code or get the latest versions you need to check out a version of the code locally from git
  - and compiled using cmake

- setup a version like this:
  - `lsetup 'asetup AthAnalysis,21.2.93,64,slc6,gcc62,opt'`
    - this is a lot to type, but will make assumptions.
      - opt is assumed
      - will use the compiler version of the latest release
      - if no OS is specified, it will infer from your current OS
      - use 'lsetup' if you have other software pieces to setup
    - Use "showVersions" if in doubt.
- (Will see later how to access xAODs outside of Athena)

Which Athena version?

- lots of Athena 'flavours' (aka projects), and lots of trailing numbers
  - `showVersions athena`
    - Example: AthAnalysis-21.2.93-x86_64-slc6-gcc62-opt
  - Project naming changes in Release 20→21 update

https://gitlab.cern.ch/atlas/athena

Branch	Purpose	Main Project, Release Series
21.0	Tier0 and MC16c	Athena, 21.0.X
21.0-mc16a	MC16a bugfixes	Athena, 21.0.20.Y
21.0-mc16d	MC16d bugfixes	Athena, 21.0.20.Y
21.0-TrigMC	MC production with new trigger developments	Athena, 21.5.X
21.1	HLT and P1 monitoring	AthenaP1, 21.1.X
21.2	Derivations and Analysis	AthDerivations + (Ath)AnalysisBase, 21.2.X
21.3	MC18 Simulation	Athena + AthSimulation, 21.3.X
21.6	Event Generation	AthGenerators, 21.6.X
21.9	Upgrade Phase-2 developments	Athena + AthSimulation, 21.9.X
master	AthenaMT development	Athena, 22.0.X

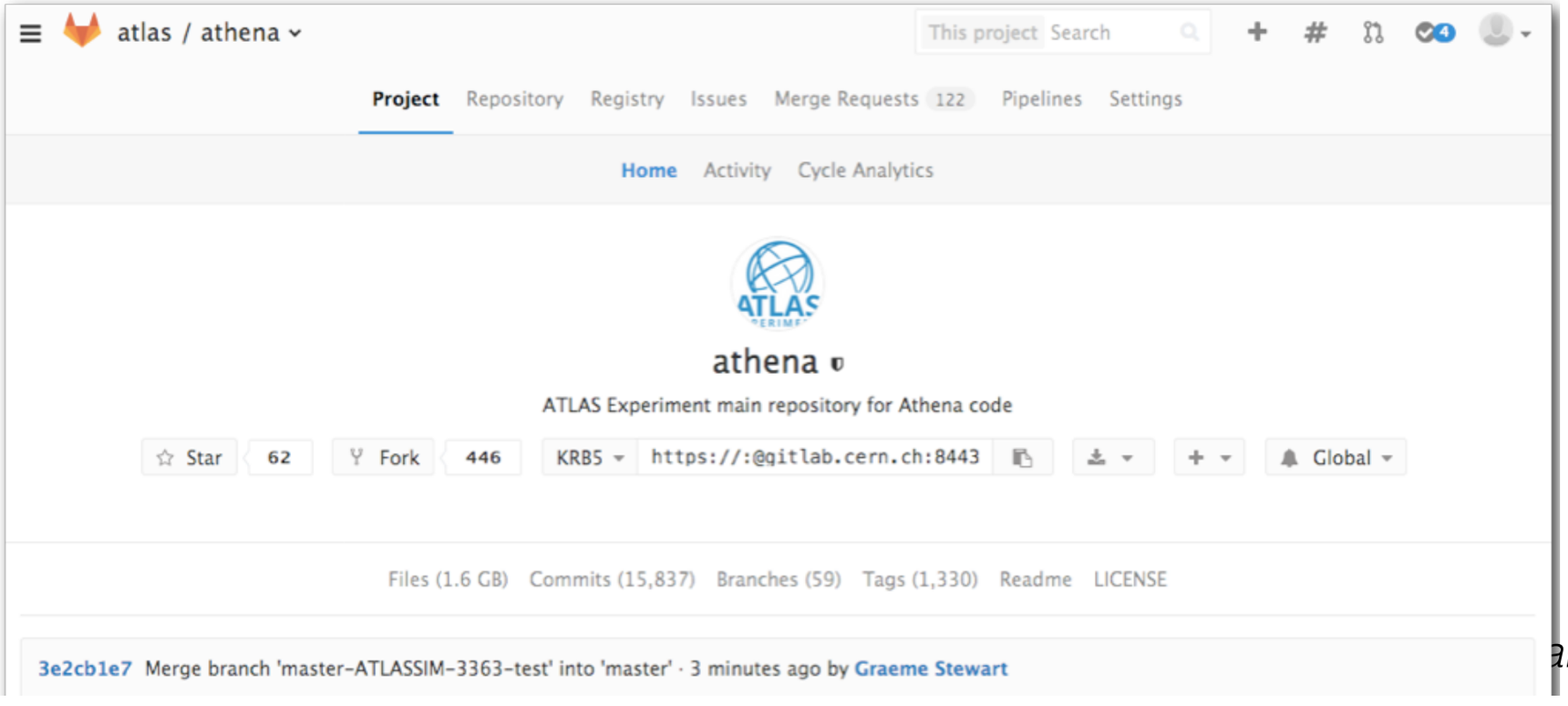
athena/athena-releases/

ATLAS Documentation Links

- Software tutorial TWiki
- Doxygen
  - Generated documentation for the Athena Codebase.
    - How to contribute .
- LXR
  - Code cross reference search tool
- Twiki
  - Wiki containing useful information written by physicists.
    - How to contribute
- ATLAS Software Documentation
  - Pages containing various guides (include GIT and AnalysisBase)
  - These links and others are present on the left side of the ATLAS computing wiki

How to view code

- Web-based git page:
  - https://gitlab.cern.ch
  - For Athena software:
    - https://gitlab.cern.ch/atlas/athena
- Similar interface for personal / group software
  - Web-page also used to configure your projects
  - More examples in Tutorial
- Remember also LXR, Doxygen and TWiki pages.



Getting information: Metadata

- Data 'about' Data:
  - Several web-based pages to access information at various levels:
- Web-pages:
  - Main collaboration interface : https://atlas-collaboration.web.cern.ch
  - **Indico** : https://indico.cern.ch/category/6733/overview?period=week:
  - **CDS**: https://cds.cern.ch
  - **TWikis**:
    - https://twiki.cern.ch/twiki/bin/view/AtlasProtected/DataMCForAnalysis Specific information eg which MC campaigns for which sets of Data.
    - https://twiki.cern.ch/twiki/bin/view/AtlasPublic Public results
    - https://twiki.cern.ch/twiki/bin/view/Atlas/AtlasGlossary
    - https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PubComHome
      - Includes links to atlas style guide (plots) and latex note template
- **AMI**: https://ami.in2p3.fr/
  - Datasets and files for official MC and Data
- **COMA**: https://atlas-tagservices.cern.ch/tagservices/RunBrowser/index.html Run and trigger information
- **Run-Query**: https://atlas-runquery.cern.ch/ ; run / LHC level information
  - And many others …

Getting information: E-groups / Jira

- **Mailing list groups** (also manages permissions, policies)
  - Access to archive messages
    - Groups can be set up with differing restrictions

- Search, request, admin membership of e-groups:
  - https://e-groups.cern.ch/e-groups/EgroupsSearchForm.do

- View archives, also search facilities:
  - https://groups.cern.ch/Pages/default.aspx

- **Jira:**
  - Track bugs, make requests, etc.
    - Organised by 'projects'
    - https://its.cern.ch/jira/secure/Dashboard.jspa
      - e.g. Trigger

Storing Code / Data

- Code:
  - Some version control system (e.g. cern's gitlab)
- Data
  - the Grid; holds official datasets produced by ATLAS for you to use
    - you can also run jobs on grid sites, that will store your output for a limited time
  - LXPLUS:
    - home: /afs/cern.ch/user/<letter>/<username> max of 10 GB
    - workspace:/afs/cern.ch/work/<letter>/<username> max of 100 GB
  - CERNBox: eos/CERNBox: up to 1 TB
    - activate online: https://cernbox.cern.ch
      eos storage area: root://eosuser//eos/user/<letter>/<username>
      - https://twiki.cern.ch/twiki/bin/view/AtlasComputing/ATLASStorageAtCERN#Users_area_on_EOS
      - https://cern.service-now.com/service-portal/article.do?n=KB0001998

Coding Guidelines

- ATLAS has a set of coding guidelines:
  - http://atlas-computing.web.cern.ch/atlas-computing/projects/qa/draft_guidelines.html

- thousands (!) of people have and will contribution to millions (!) of lines of ATLAS code
  - you will very likely improve packages written by others, other people will take over your work
  - hence guidelines ease diving into other people's code (other people could be you one month later)

Coding Guidelines: Examples

- separate class declaration <MyClass>.h from implementation <MyClass>.cxx (not <MyClass>.cpp, not <SomeOtherClass>.cxx)
- Capitalization:
  - class names start with a capital and have camel case capitalization function names start with a non-capital letter, and follow with camel case

- member variables of a class have to have a name prefixed by "m_", followed by a non-capital letter, then camel case used

- use (doxygen) comments ... lots of comments!

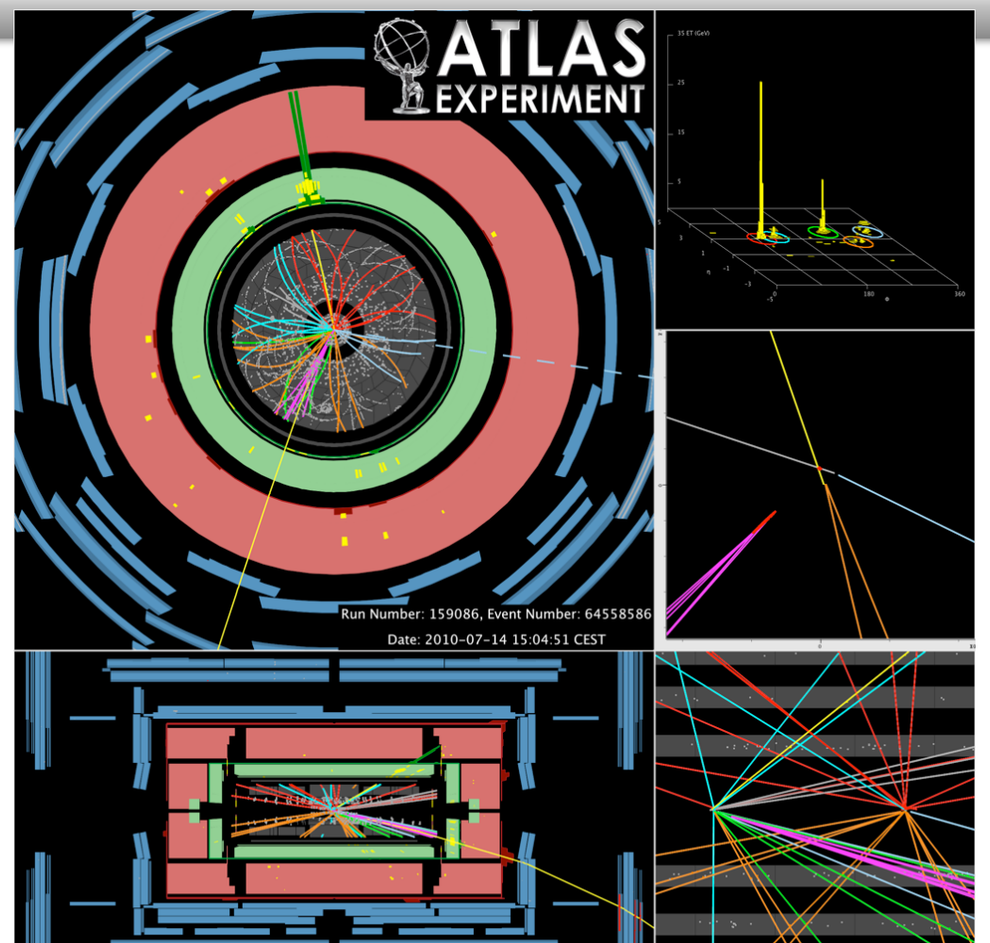
- use explicit braces with if/for/etc. statements, to clearly mark what lines of code belong together
- code should compile with no warnings (if you take over a package which has compilations warnings - fix them)
- ...

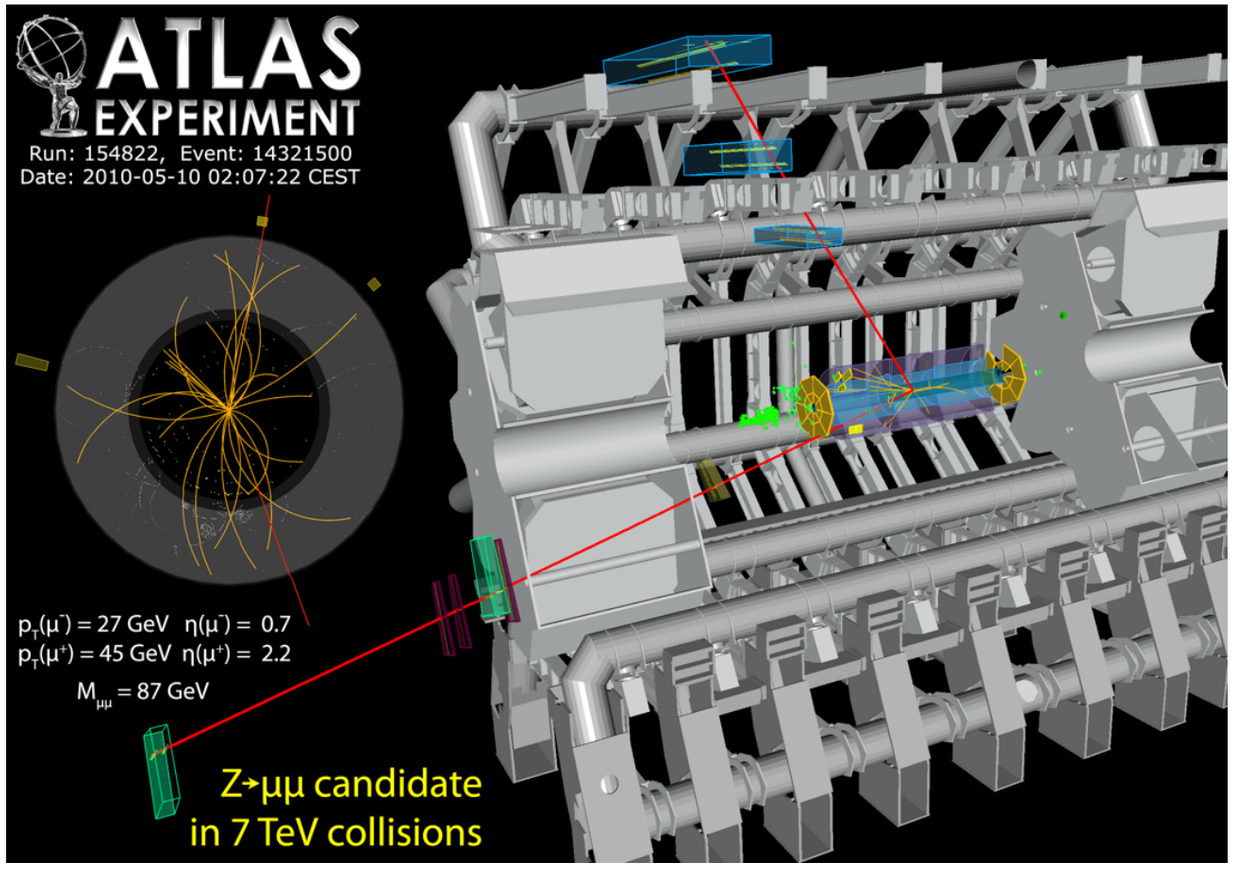
Event Displays

- Several displays available
  - good tool to inspect your candidate events
  - helps to debug reconstruction and detector

- ATLANTIS
  - works in clever 2D projections
  - JAVA software runs on any laptop
  - need to convert events into JiveXML

- VP1 / VP1Light
  - 3D event display in C++
  - it is an ATHENA application
  - makes use of full power of and Services
- PERSINT





Last points

- These tutorials are meant to be **informal**.
  - General structure will be in the form of an introductory talk,
    - followed by a practical session:
    - Best way to learn is by working through the exercises at your own pace.
    - Ask the experts about any problems / queries.

A survey will be sent via email at the end of the week. Please take the time to complete it. Your feedback will help shape the future of these tutorials.

Extra Info

Documentation and help

- Main computing page (including set-up information for non-LXPLUS platforms):
  - https://twiki.cern.ch/twiki/bin/view/AtlasComputing/AtlasComputing
- Documentation for beginners:
  - WorkBook: https://twiki.cern.ch/twiki/bin/view/AtlasComputing/WorkBook
  - Physics analysis WorkBook: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PhysicsAnalysisWorkBook
  - Tutorials: https://twiki.cern.ch/twiki/bin/view/AtlasComputing/ComputingTutorials
- Setting up a desktop at CERN:
  - http://linux.web.cern.ch/linux/install/
- Help forums:
  - https://espace.cern.ch/atlas-forums/Lists/Atlas%20forums/By%20category.aspx
  - This one in particular: https://groups.cern.ch/group/hn-atlas-offlineSWHelp/default.aspx
- Advanced Pointers on C++ and Athena usage including Performance Profiling methods:
  - https://indico.cern.ch/event/290337/
- CMake: https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/CMakeTestProjectInstructions
  - https://twiki.cern.ch/twiki/bin/view/AtlasComputing/SoftwareDevelopmentWorkBookCMakeInAtlas

How *could* physics analyses proceed?

- One possible (contrived) approach to run your analysis - you will see other approaches
- Setup your **environment** using configuration manager, (**CVMFS**) **CMT(/CMake) / RootCore**
- Choose your data:
  - Identify and examine runs satisfying specific criteria, **Run Query**
  - Look for datasets to analyse using a metadata browser, **AMI**
- **Download** a few files locally using a data management tool, **Rucio**
- Inspect a small subset of interesting events **visually** with **ATLANTIS or VP1**
- Build **Derivation** code(/python) in **Athena** using the local data to check that it is doing what you expect
  - Run **Derivation** code on the **Grid** to slim the xAOD datasets to small useful size.
- Obtain **GoodRunsList (GRL)** ; use https://atlas-lumicalc.cern.ch to obtain luminosity estimate.
- Write, test and send your analysis jobs to the **Grid** to process **(D)xAODs.**
  - Using **Athena (AthAnalysisBase)** or **RootCore (AnalysisBase)** releases.
- Download the results (**rucio**)
- Make histograms / fits using ROOT / RooStats / **HistFitter** / ...
- **Fortunately; many of these steps are done already centrally or by your physics group.**
- **Good practice to learn how these step work, and how to generate your own data.**

ATLAS Data Flow

- Many Steps from Detector / MC-generation: