
ATLAS Open Data - Implementations and tools for the next release of datasets

9th Colombian Meeting on High Energy
Physics (COMHEP)
December 03, 2024

Miguel Ángel García Ruíz
miguel.angel.garcia.ruiz@cern.ch - miagarciar@unal.edu.co
Universidad Nacional de Colombia



Introduction to ATLAS Open Data

Open Data: Used by institutions (schools, universities) and individuals interested in experimental particle physics analysis techniques.

- Previous release: ATLAS Open Data collaboration (8 & 13 TeV) with tools like videos & data visualizers.

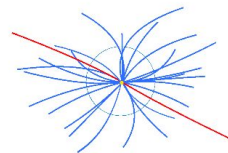
Current Effort:

- Consolidating 8 & 13 TeV tools and documentation in a single, organized, and accessible website.

New Campaign:

- Preparing a 36 fb^{-1} dataset release with improved documentation, new analyses, and expanded tools.

ATLAS Open Data



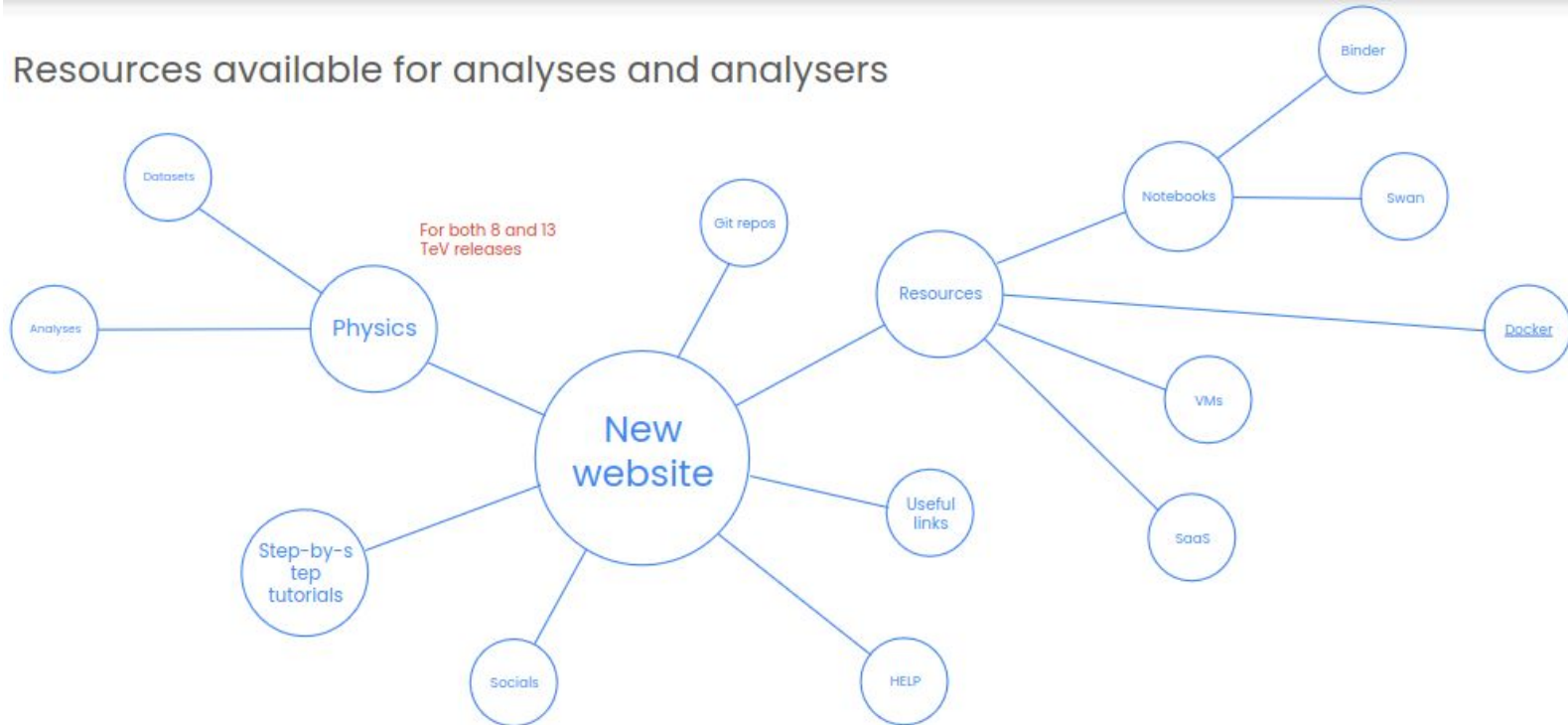
High Energy Physics data for everyone.

A screenshot of the GitHub repository page for 'ATLAS Outreach data and tools'. The repository is owned by 'ATLAS EXPERIMENT' and has 15 followers. It is located in CERN, Switzerland, and its URL is http://opendata.atlas.cern/. The page displays a grid of popular repositories, including 'notebooks-collection-opendata', 'atlas-outreach-cpp-framework-13tev', 'atlas-outreach-data-tools-framework', 'notebooks', 'atlas-outreach-PyROOT-framework-13tev', and 'atlasoutreach-webpage'. On the right side, there is a section titled 'You can now follow organizations' with a green 'OK, got it!' button, and a 'Top languages' section listing Jupyter Notebook, Python, JavaScript, HTML, and Sass.

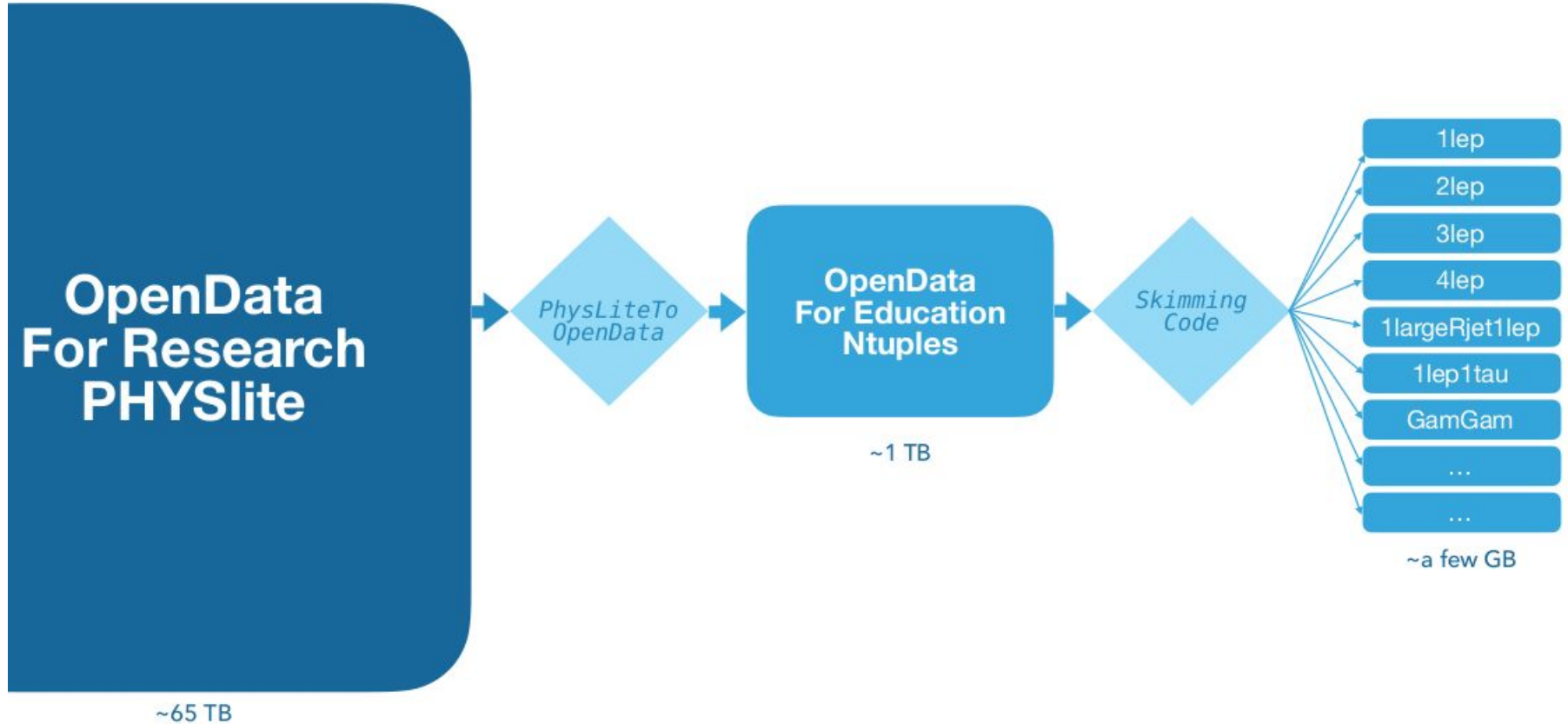
Website - Resources

- The 8 and 13 TeV documentation, analyses and tools have been collected to single website <https://opendata.atlas.cern/>

Resources available for analyses and analysers



New release of datasets to be released - workflow overview



Derivation Framework - PHYSLITE to Open Data

New Dataset Release: 2015/2016 data + MC samples for educational purposes.

- **Format:** Flat ROOT ntuples.
- Usable for existing & new/improved analyses.

Code Development: Available on GitLab.

- Input: DAOD_PHYSLITE → Output: Flat ntuple.
- Post-production: Metadata added to output ntuples.
- Open-sourced.

Validation:

- Run test analyses with “clean” processes & selections.

atlas-outreach-data-tools / PhysLiteToOpenData

PhysLiteToOpenData

main physlitetoopendata

History

Find file

Code



Merge branch 'docker-instructions' into 'master'
 Antonio Manuel Mendes Jacques Da Costa authored 1 month ago

6b4faa30



Name	Last commit	Last update
📁 .docker	dockerfile: fix entrypoint	2 months ago
📁 PhysLiteToOpenD...	Update code to R25 version	4 months ago
📁 Root	Fill jet jvt	2 months ago
📁 extras	rename extras/CMakeLists...	1 year ago
📁 hist_branches_to_...	Some changes were adde...	8 months ago
📁 notebooks	Merge branch 'master' int...	9 months ago

Validating Framework & Updating Analysis Examples

Validation process:

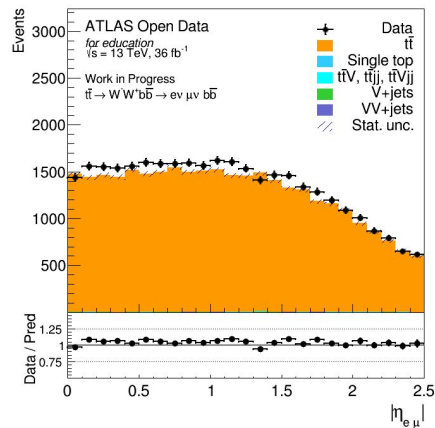
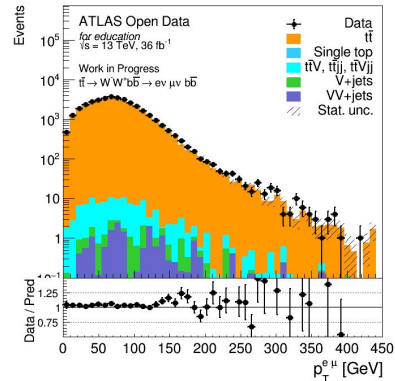
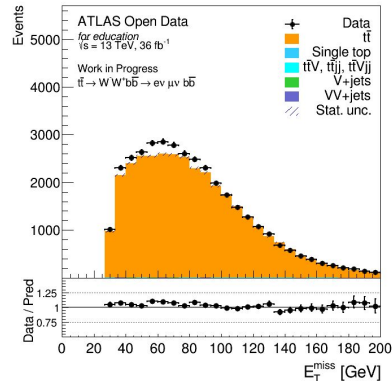
- Updating and running the analysis examples to validate the samples/code.
- Check we have all branches/objects needed.
- Check data/MC in various analyses to validate that SFs/etc are implemented.

Testing physics objects:

- Currently updated examples test electrons, muons, taus, jets (including b-tagging, Scale Factors and JVT) and MET.

Updating all framework versions:

- ROOT, PyROOT, Uproot, cpp

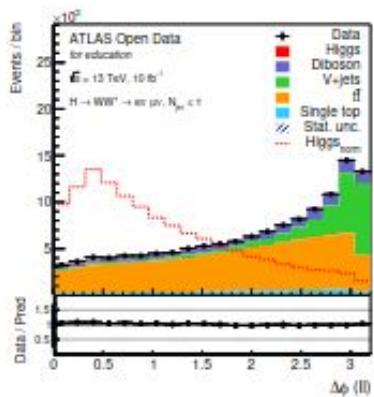


TTbarDilep Analysis:

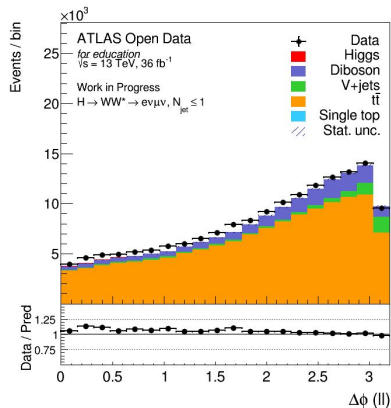
- Clear final state.
- Perfect to test bjets, electrons and muons objects.

Validating Framework - Analysis Examples (HWW and ZTauTau)

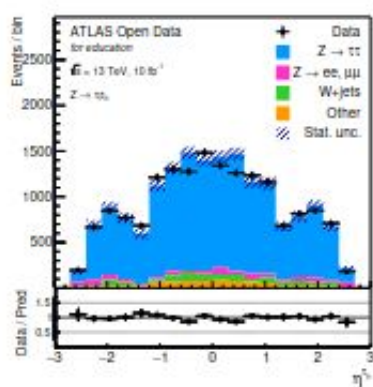
Release 2



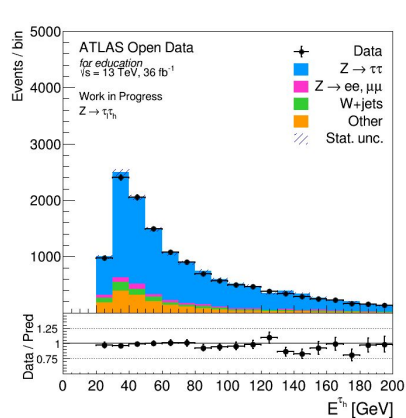
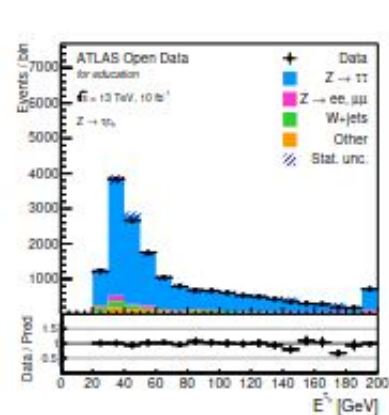
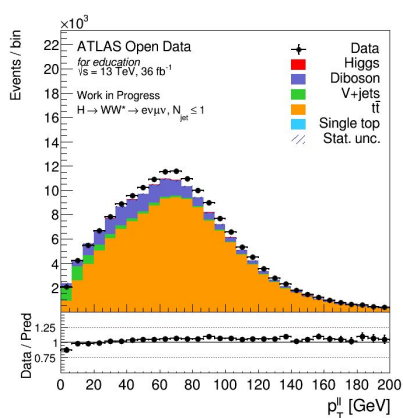
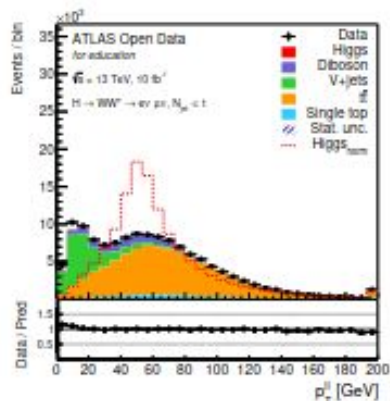
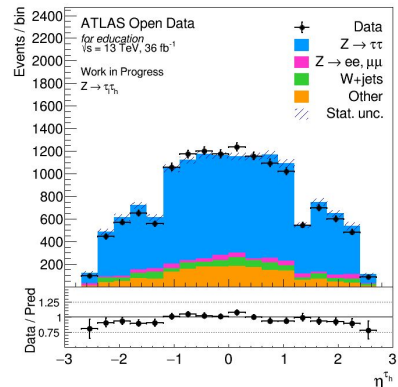
Release 3



Release 2



Release 3



Data Collection and public release

Data Collection

- Previously, the release 2 had a number of “skimmed” data collections.
- Some quite lightweight and easy to download.
- Should consider what splittings are wanted and if any new ones should be produced

Current tasks to do for releasing the new datasets to the public

- Tutorials
- Instructions on how to run the derivation framework.
- Some analyses examples that validates the different ntuples.
- Updates the different repositories/notebooks and codes to use the third release.

Final-state categories	Leading object p_T (min) [GeV]	Collection name
$N_l = 1$	25	1lep
$N_l \leq 1$	25	2lep
$N_l = 3$	25	3lep
$N_l \leq 4$	25	4lep
$N_{\text{largeRjet}} \leq 1 \ \& \ N_l = 1$	250 (large-R jet), 25 (lepton)	1largeRjet1lep
$N_{\tau\text{-had}} = 1 \ \& \ N_l = 1$	20 (τ_h), 25 (lepton)	1lep1tau
$N_\gamma \leq 2$	35	GamGam

1 **INSTALLING VIRTUALBOX** Installing a VirtualBox - ATLAS Open Data Tutorial
ATLAS Experiment • 653 views • 3 years ago

5 **INSTALL A VIRTUAL MACHINE** Installing a Virtual Machine - ATLAS Open Data Tutorial
ATLAS Experiment • 1.4K views • 3 years ago

5 **MAKE SELECTION CUTS WITH PYROOT** Making Selection Cuts with PyROOT - ATLAS Open Data Tutorial
ATLAS Experiment • 1.6K views • 3 years ago

7 **CREATE HISTOGRAMS WITH PYROOT** Create Histograms with PyROOT - ATLAS Open Data Tutorial
ATLAS Experiment • 2.4K views • 3 years ago

3 **OPEN DATA** Getting Started with ATLAS Open Data
ATLAS Experiment • 1.1K views • 3 months ago

Thanks!