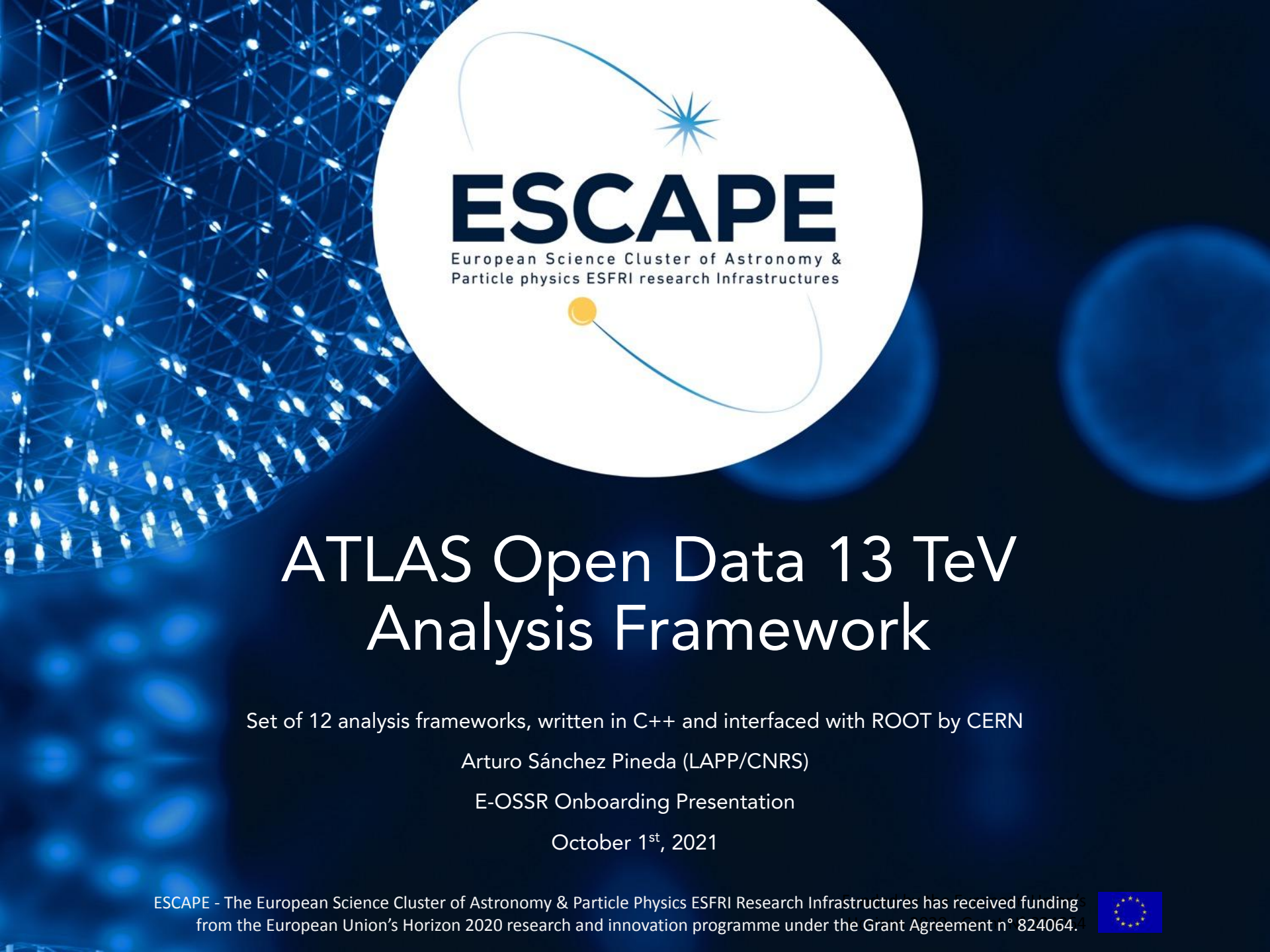


ATLAS Open Data 13 TeV Analysis Framework

Set of 12 analysis frameworks, written in C++ and interfaced with ROOT by CERN

Arturo Sánchez Pineda (LAPP/CNRS)

E-OSSR Onboarding Presentation

October 1st, 2021

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

Request in ESCAPE RedMine → Integration #122: Onboarding: ATLAS Open Data C++ analysis software at 13 TeV - OSSR Onboarding

Funded by the European Union's Horizon 2020 - Grant N° 824064

Introduction

ATLAS is a general-purpose particle physics experiment at the Large Hadron Collider (LHC) at CERN. It is designed to exploit the full discovery potential of the LHC, pushing the frontiers of scientific knowledge. ATLAS' exploration uses precision measurement to push the frontiers of knowledge by seeking answers to fundamental questions such as: What are the basic building blocks of matter? What are the fundamental forces of nature? What is dark matter made of?

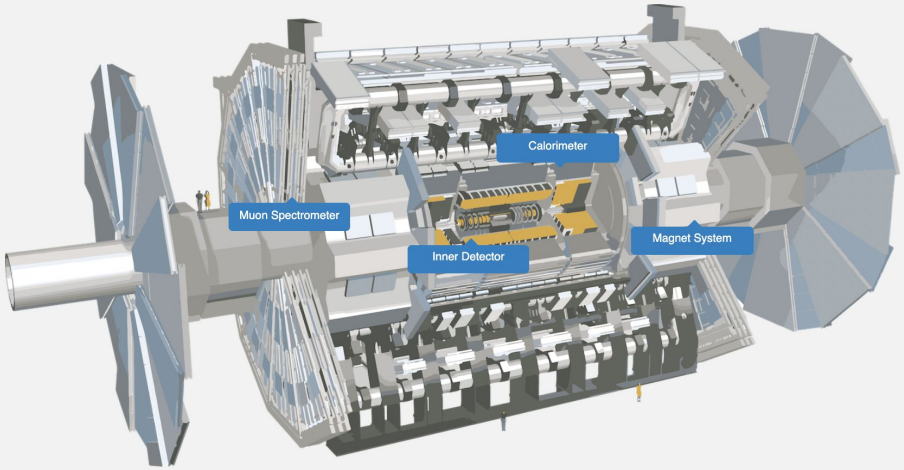
Experimental behemoth

ATLAS is the largest detector ever constructed for a particle collider: 46 metres long and 25 metres in diameter. Its construction pushed the limits of existing technology. ATLAS is designed to record the high-energy particle collisions of the LHC, which take place at a rate of over a billion interactions per second in the centre of the detector. More than 100 million sensitive electronics channels are used to record the particles produced by the collisions, which are then analysed by ATLAS scientists.

More at https://atlas.cern/about

46	100	7000	1000000000
meters long & 25m of diameter	meters below ground	tonnes	collisions / second

Have a closer look



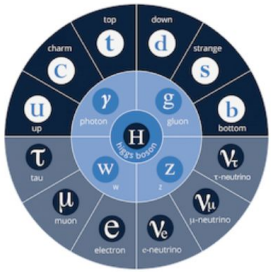
Funded by the European Union's Horizon 2020 - Grant N° 824064

Introduction

ATLAS has a rich physics program, from Standard Model searches and measurements, the quest of Dark Matter, and may others. For more visit https://atlas.cern/discover/physics

What are the basic building blocks of matter?

The Standard Model describes the elementary subatomic particles of the universe which have been experimentally seen. ATLAS studies these particles and searches for others to determine if the particles we know are indeed elementary or if they are in fact composed of other more fundamental ones.



What are the forces that govern their interactions?

The Standard Model also describes the fundamental forces of Nature and how they act between fundamental particles. Possible discoveries at the LHC could validate models, such as those incorporating Supersymmetry, where the forces unify at very high energies.

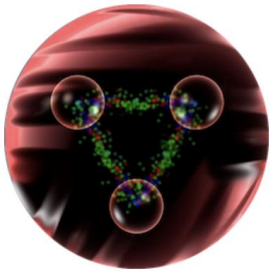
The Higgs boson

What is the Higgs boson and why does it matter?

Physicists describe particle interactions using the mathematics of field theory, in which forces are carried by intermediate particles called bosons. Photons, for example, are bosons carrying the electromagnetic force. In 1964, the only mathematically consistent theory required bosons to be massless. Yet, experiment showed that the carriers of the weak nuclear interaction – the W and Z bosons – had large masses. To solve this problem, three teams of theorists: Robert Brout and François Englert; Peter Higgs; Gerald Guralnik, Carl Hagen, and Tom Kibble independently proposed a solution now referred to as the Brout-Englert-Higgs (BEH) mechanism.

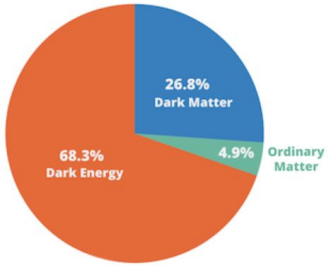
What happened to antimatter?

By searching for imbalances in the production of matter and antimatter, we seek to understand why our universe appears to comprise only matter.



What is "dark matter"?

Astronomical measurements support the existence of matter that cannot be directly seen. The hermetic construction of ATLAS, however, makes it possible to search for this "dark matter".

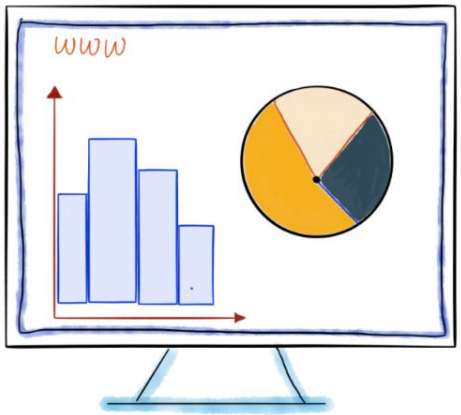
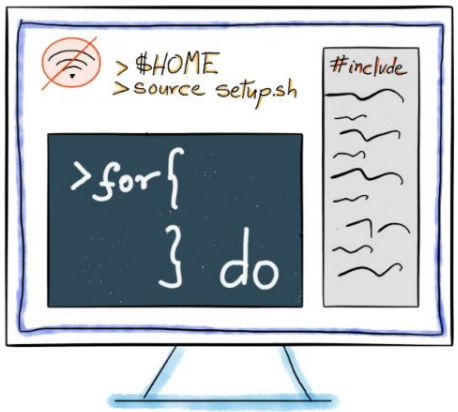
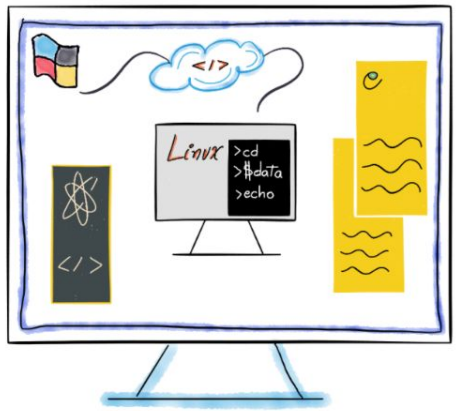


Funded by the European Union's Horizon 2020 - Grant N° 824064

Introduction

The ATLAS Collaboration current approach on the release of datasets is intended for Education, Training and Outreach activities around the World. In order to fulfil that objective, the ATLAS Open Data project was created.

ATLAS Open Data project aims to provide data and tools to students, as well as teachers and lecturers, to help educate them in physics analysis techniques used in experimental particle physics. Ideally, sharing data collected by the ATLAS experiment aims to generate excitement and enthusiasm for fundamental research, inspiring physicists of the future.
http://opendata.atlas.cern/about

Jupyter Notebooks	C++/Python frameworks	Virtual Machine(s)
		
Let's run some real code and visualisations on your browser	Perform real HEP analysis as the ATLAS Physicists	Slow Internet? run the analyses with minimal installation

Funded by the European Union's Horizon 2020 - Grant N° 824064

ATLAS PUB Note
PUB-OTRC-2020-01
28th April 2020

The ATLAS Open Data and analysis examples (i.e. the c++ framework) release in 2020

https://cds.cern.ch/record/2707171

- This is the publication that contains the validation of the datasets and the software (framework)

Review of the 13 TeV ATLAS Open Data release

The ATLAS Collaboration

The ATLAS Collaboration is releasing a new set of proton–proton collision data to the public for educational purposes. The data has been collected by the ATLAS detector at the Large Hadron Collider at a centre-of-mass energy $\sqrt{s} = 13$ TeV during the year 2016 and corresponds to an integrated luminosity of 10 fb^{-1}. This dataset is accompanied by simulated events describing both several Standard Model processes, as well as hypothetical Beyond Standard Model signal production. Associated computing tools are provided to make the analysis of the dataset easily accessible. This document summarises the properties of the 13 TeV ATLAS Open Data set and the available analysis tools. Several examples intended as a starting point for further analysis work by users are shown. The general aim of the dataset and tools released is to provide user-friendly and straightforward interactive interfaces to replicate the procedures used by high-energy-physics researchers and enable users to experience the analysis of particle-physics data in educational environments.

ATL-OREACH-PUB-2020-001
24 Jan 2020

ATLAS Note

Report number	ATL-OREACH-PUB-2020-001
Title	**Review of the 13 TeV ATLAS Open Data release**
Corporate Author(s)	The ATLAS collaboration
Collaboration	ATLAS Collaboration
Imprint	24 Jan 2020. - 28 p.
Note	All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-OREACH-PUB-2020-001
In:	24th International Conference on Computing in High Energy and Nuclear Physics, Adelaide, Australia, 4 - 8 Nov 2019
Subject category	Particle Physics - Experiment
Accelerator/Facility, Experiment	CERN LHC ; ATLAS
Free keywords	ATLAS ; outreach ; open access ; open data ; education ; open source ; open science ; public data
Abstract	The ATLAS Collaboration is releasing a new set of proton-proton collision data to the public for educational purposes. The data has been collected by the ATLAS detector at the Large Hadron Collider at a centre-of-mass energy $\sqrt{s} = 13$ TeV during the year 2016 and corresponds to an integrated luminosity of 10 fb^{-1}. This dataset is accompanied by simulated events describing both several Standard Model processes, as well as hypothetical Beyond Standard Model signal production. Associated computing tools are provided to make the analysis of the dataset easily accessible. This document summarises the properties of the 13 TeV ATLAS Open Data set and the available analysis tools. Several examples intended as a starting point for further analysis work by users are shown. The general aim of the dataset and tools released is to provide user-friendly and straightforward interactive interfaces to replicate the procedures used by high-energy-physics researchers and enable users to experience the analysis of particle-physics data in educational environments.

Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

ATLAS EXPERIMENT

Collaboration Site | Physics Results

ABOUT DISCOVER RESOURCES UPDATES SEARCH

All News Briefings Features Portraits Press Blog

Updates > Press statement > ATLAS Experiment releases 13 TeV Open Data for Science Education

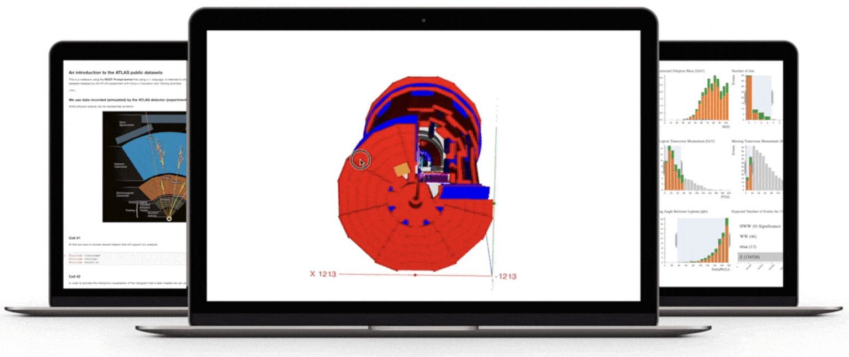
Press Statement

Tags:
open data,
13 TeV

ATLAS Experiment releases 13 TeV Open Data for Science Education

Highest-energy particle-collision data ever released through open access.

10th February 2020 | By ATLAS Collaboration



Animation of ATLAS detector using ROOTJS. (Image: ATLAS Collaboration/CERN)

Geneva, 10 February. The ATLAS Collaboration at CERN has just released the first open dataset from the Large Hadron Collider's (LHC) highest-energy run at 13 teraelectronvolts (TeV). The new release is specially developed for science education, underlining the Collaboration's long-standing commitment to students and teachers using open-access ATLAS data and related tools.

Alongside impressive new open datasets, the ATLAS Collaboration has also released new simulated datasets, web-based and offline analysis software, as well as extensive documentation and tutorials. "These are the tools of a particle physicist's trade, allowing us to go from data-taking to physics measurements and eventually discovery," says Arturo Sánchez Pineda, co-leader of the ATLAS Open Data team (University of Udine, ICTP and INFN, Italy). "Simulated datasets allow physicists to compare theory with real data. They are based on theoretical models of the expected physics processes taking place in the collisions, together with a detailed description of the ATLAS detector. By providing such resources, we hope to empower students, professors and dedicated self-learners worldwide to learn and teach experimental particle physics, as well as the computer science behind the field."

EPJ Web of Conferences **245**, 08023 (2020)
https://doi.org/10.1051/epjconf/202024508023

ATLAS Open Data – Development of a simple-but-real HEP data analysis framework

Farid Ould-Saada*

University of Oslo, Norway

* Farid Ould-Saada: farido@uio.no, on behalf of the ATLAS collaboration.

Published online: 16 November 2020

EPJ Web of Conferences **245**, 08026 (2020)
https://doi.org/10.1051/epjconf/202024508026

The release of the 13 TeV ATLAS Open Data: using open education resources effectively

Leonid Serkin* on behalf of the ATLAS Collaboration

INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine and ICTP, Trieste Strada Costiera 11, Trieste 34151, Italy

* e-mail: Leonid.Serkin@cern.ch

Published online: 16 November 2020

ATLAS Open Data also presented in several conferences and ATLAS Press releases https://atlas.cern/updates/press-statement/13-tev-open-data

Funded by the European Union's Horizon 2020 - Grant N° 824064

C++ Framework

External repository (for release)
https://gitlab.cern.ch/atlas-outreach-data-tools/atlas-outreach-cpp-framework
Latest version (for validation)
https://gitlab.cern.ch/lserkin/atlas-outreach-cpp-framework-13tev/

Funded by the European Union's
Horizon 2020 - Grant N° 824064

The ATLAS Open Data 13 TeV Analysis framework

The framework makes use of the C++ language and is interfaced with ROOT, analysis framework by CERN. After cloning/downloading the repository, the only things you need to setup are: you need to have the ROOT framework and a gcc compiler.

The current version of the framework was compiled using gcc v6.20 and ROOT v6.10.04.

Currently, the framework can access the samples in two ways:

- reading them online directly (be default, they are stored in a GitHub repository);
- reading them form a local storage (the samples need to be downloaded locally).

The framework consists of two main parts:

- the analysis part, located within the "Analysis" directory: it performs the particular object selection and stores the output histograms;
- the plotting part, located within the "Plotting" directory: it makes the final Data / Prediction plots.

Software's description

- Build a framework based on ROOT to analyse high energy physics datasets
- **Use C++ programming language**
  - One of the main programming languages for high energy physics
  - This is one important piece still missing in the set of ATLAS Open Data public analysis codes
  - Improve the speed and the ability or running in multiple CPU cores at once
- **The framework contains**
  - All the needed pieces to run, edit and create physics analyses
  - Seven cut-and-count physics analyses
  - Documentation to guide the user on how to include a new analysis using the same datasets.

Funded by the European Union's Horizon 2020 - Grant N° 824064

Software/Service Requirements

The user can run the complete framework in a Virtual Machine that was prepared for it. It is an Ubuntu OS.

But in terms of OS, any OS that can hosted ROOT (https://root.cern.ch/) will be enough.

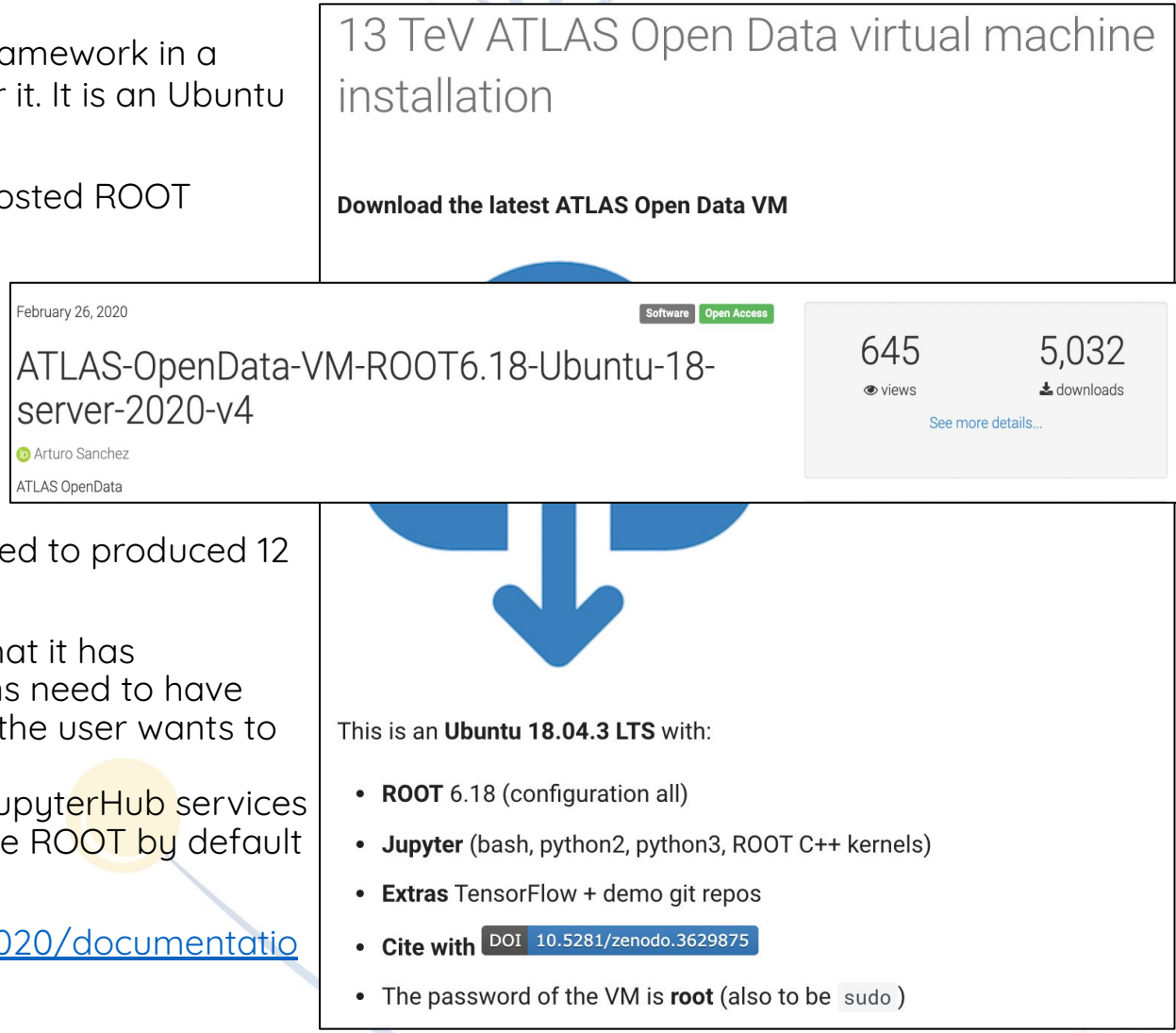
Available also in Zenodo →

https://zenodo.org/record/3687320

This is an stable software that was used to produced 12 HEP analysis.

- ROOT is the only dependency that it has
- Evidently, the machine when runs need to have ROOT dependencies in place, if the user wants to get ROOT independently.
- The ESCAPE DLaaS and other JupyterHub services like SWAN at CERN, already have ROOT by default

More on the VM here
http://opendata.atlas.cern/release/2020/documentation/vm/vm.html

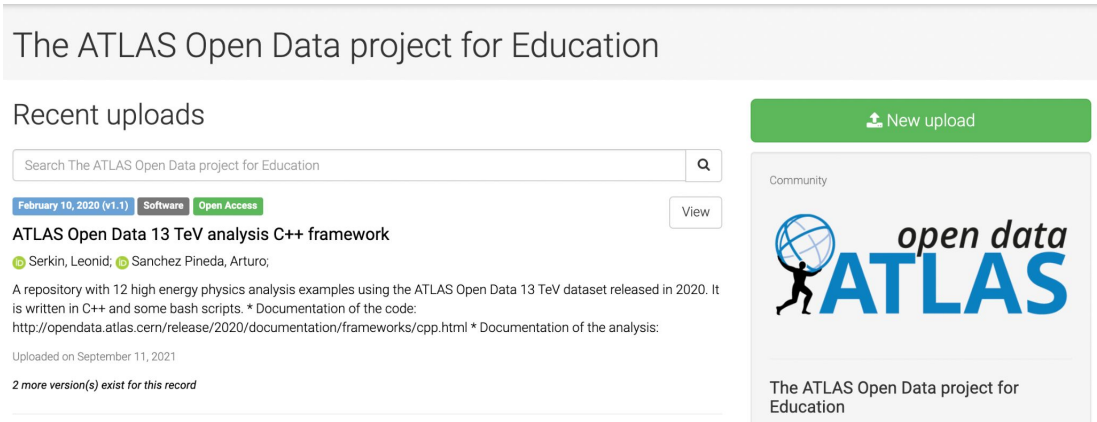


Funded by the European Union's Horizon 2020 - Grant N° 824064

OSSR Integration

- What is available? → **A C++ software**
- What will be onboarded (source code, container, test workflow incl. data)? → **Source code and internal documentation**
- Are there open points and requirements? → **No at least from our perspective. This is a stable piece of software validated inside and outside ATLAS.**

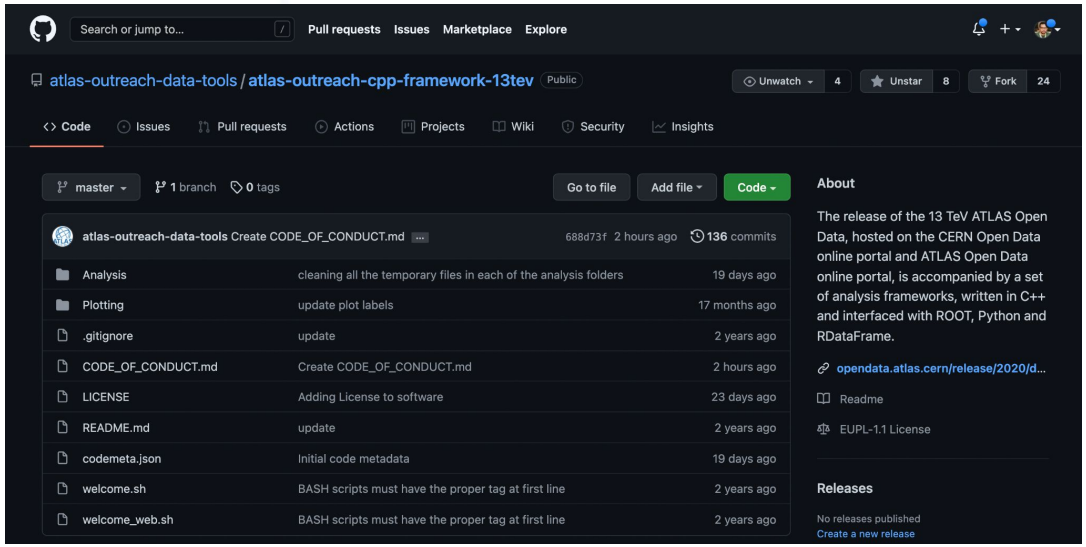
- What is the ”user story” of a EOSC user taking on the software/service?
  - From the data side (what data can be analysed and how)
  - From the OSSR side (how to find data and easy use demos, tutorials, documentation, …)

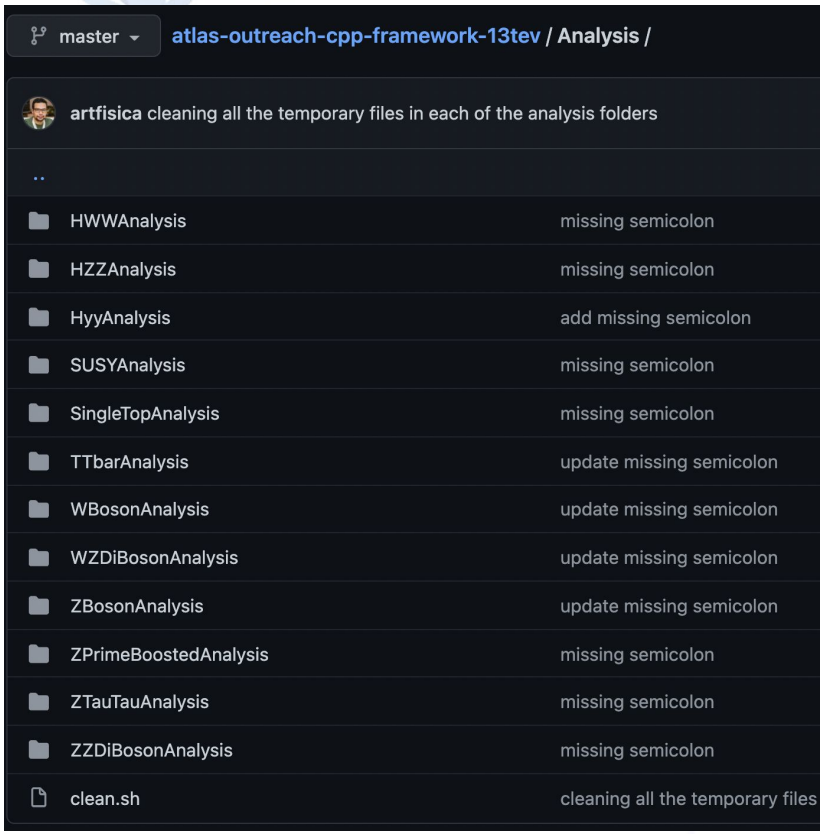


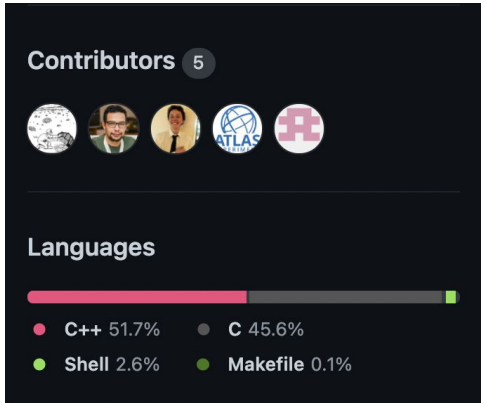
https://zenodo.org/communities/atlas-open-data/

Funded by the European Union’s Horizon 2020 - Grant N° 824064

The Analysis framework in GitHub







The framework is hosted in GitHub since January 2020. We also have a mirror at CERN GitLab for internal developments

Some part of the framework, like the binding scripts, are written in BASH/Shell. More information in the README file of the repository, including user instructions

https://github.com/atlas-outreach-data-tools/atlas-outreach-cpp-framework-13tev

Funded by the European Union's Horizon 2020 - Grant N° 824064

ATLAS Open Data datasets in the Datalake

- ROOT yet need to be importable from a notebook
  - It is deployed for testing in DockerHub
- Add more datasets to the Datalake
  - All the 13 TeV and 8 TeV ATLAS Open Data samples
  - 16 datasets → 940 samples (ROOT files)
  - < 200 GB
  - Scope used: ATLAS_OD_EDU (for ATLAS Open Data for EDUcation)
  - Source of the datasets:

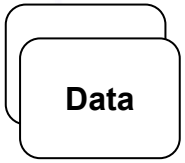
    http://opendata.atlas.cern/samples-13tev/ & http://opendata.atlas.cern/samples-8tev/
    - Another set of 10 ROOT files to come (dedicated Jet MC samples) → 1 dataset, ~21 GB.

Funded by the European Union's Horizon 2020 - Grant N° 824064

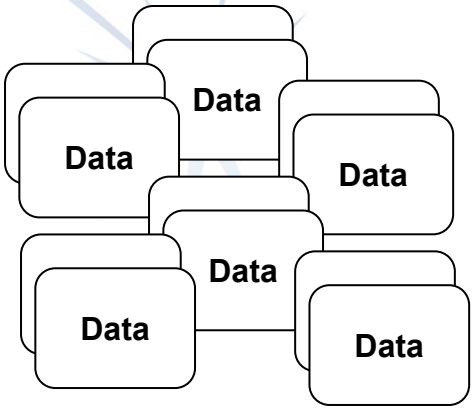
Input Data

13 TeV release

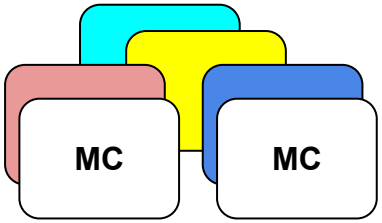
1 fb^{-1}



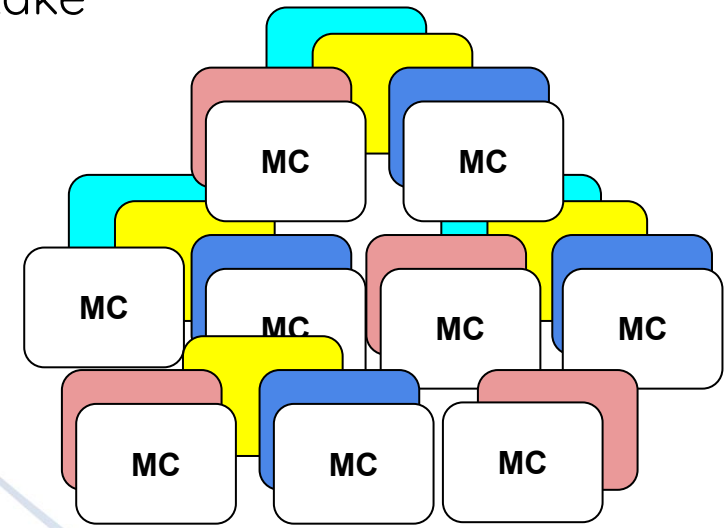
10 fb^{-1}



ATLAS Open Data for this framework is already hosted in the ESCAPE DataLake



44 samples

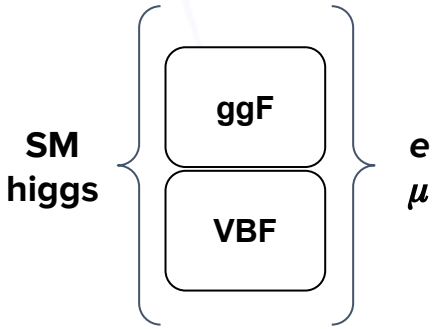


~120 samples

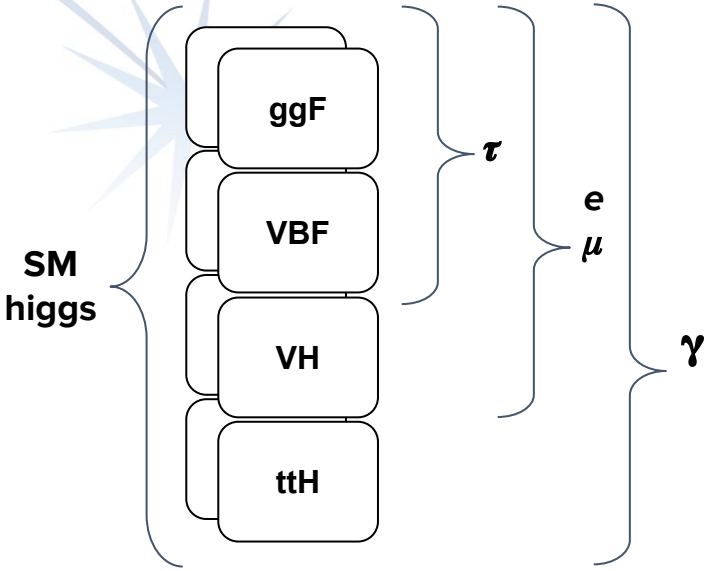
≥ X 7 Collections based in final states

Funded by the European Union's Horizon 2020 - Grant N° 824064

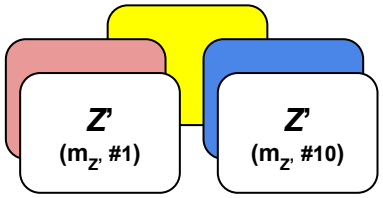
Input Data



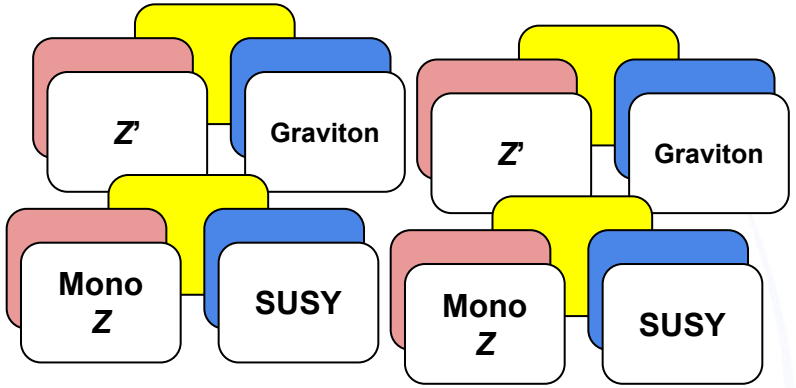
13 TeV MC signals



ATLAS Open Data for this framework is already hosted in the ESCAPE DataLake



14 BSM samples



≥ **50 BSM** samples

Funded by the European Union's Horizon 2020 - Grant N° 824064

ATLAS Open Data → C++ examples framework

To run C++ analyses

More computational-complex particle physics analysis examples using the existing publicly available data

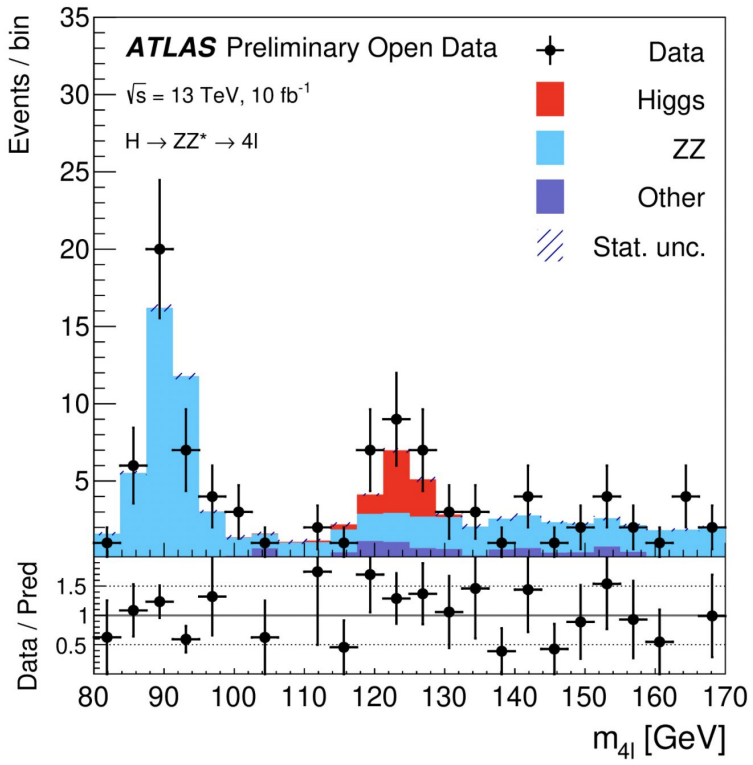
More in [Opendata.atlas.cern - documentation 13 TeV - physics]

Also use PROOF, adding a parallel component to the examples.

SM Higgs boson production in the H → ZZ decay channel in the four-lepton final state

Physics analysis examples

- Overview of physics analysis examples
- Brief introduction to the physics of the Higgs boson
- SM W-boson production in the single-lepton final state
- Single-top-quark production in the single-lepton final state
- Top-quark pair production in the single-lepton final state
- SM Z-boson production in the two-lepton final state
- SM Higgs boson production in the H → WW decay channel in the two-lepton final state
- Search for supersymmetric particles in the two-lepton final state
- SM WZ diboson production in the three-lepton final state
- SM ZZ diboson production in the four-lepton final state
- SM Higgs boson production in the H → ZZ decay channel in the four-lepton final state
- SM Z-boson production in the two-tau-lepton final state
- Search for BSM Z' → tt in the single-lepton boosted final state
- SM Higgs boson production in the H → yy decay channel in the

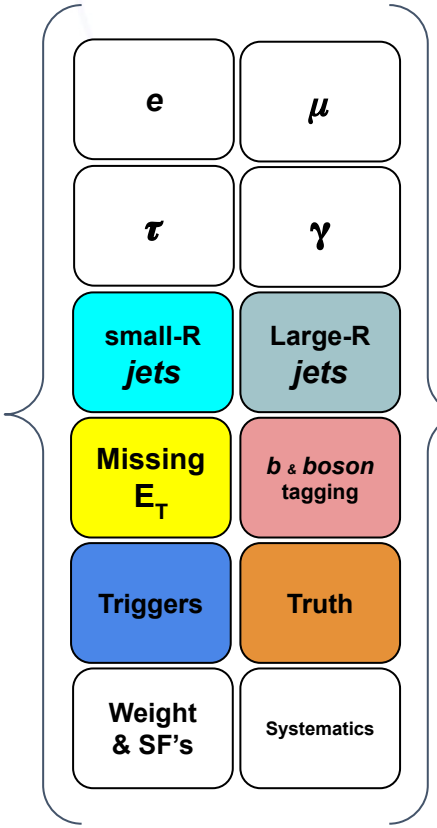


Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

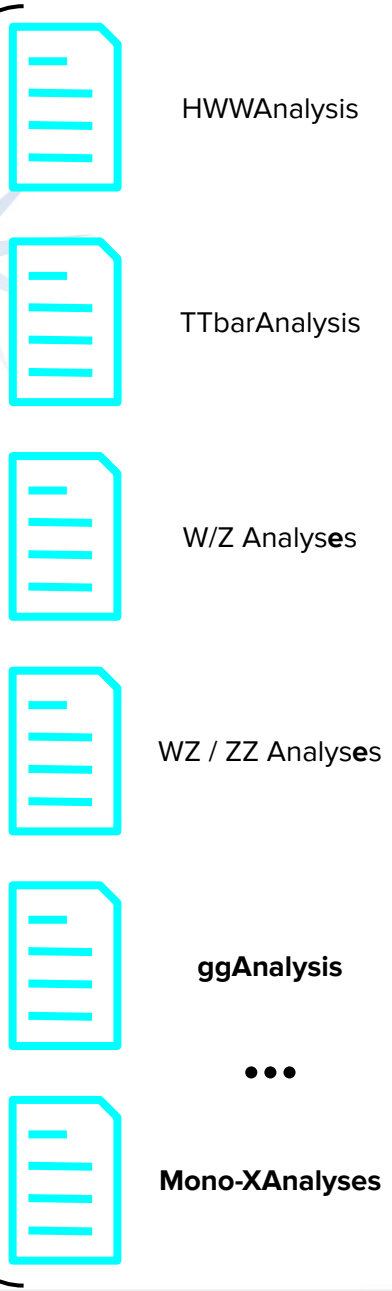
13 TeV TTree

https://gitlab.cern.ch/meevans/dataset-proposal/blob/master/mydocument.pdf



~100 variables

New analysis (will) use the objects available like photons, muons, electrons, small and large-R jets and missing transverse energy



HWWAnalysis

TTbarAnalysis

W/Z Analyses

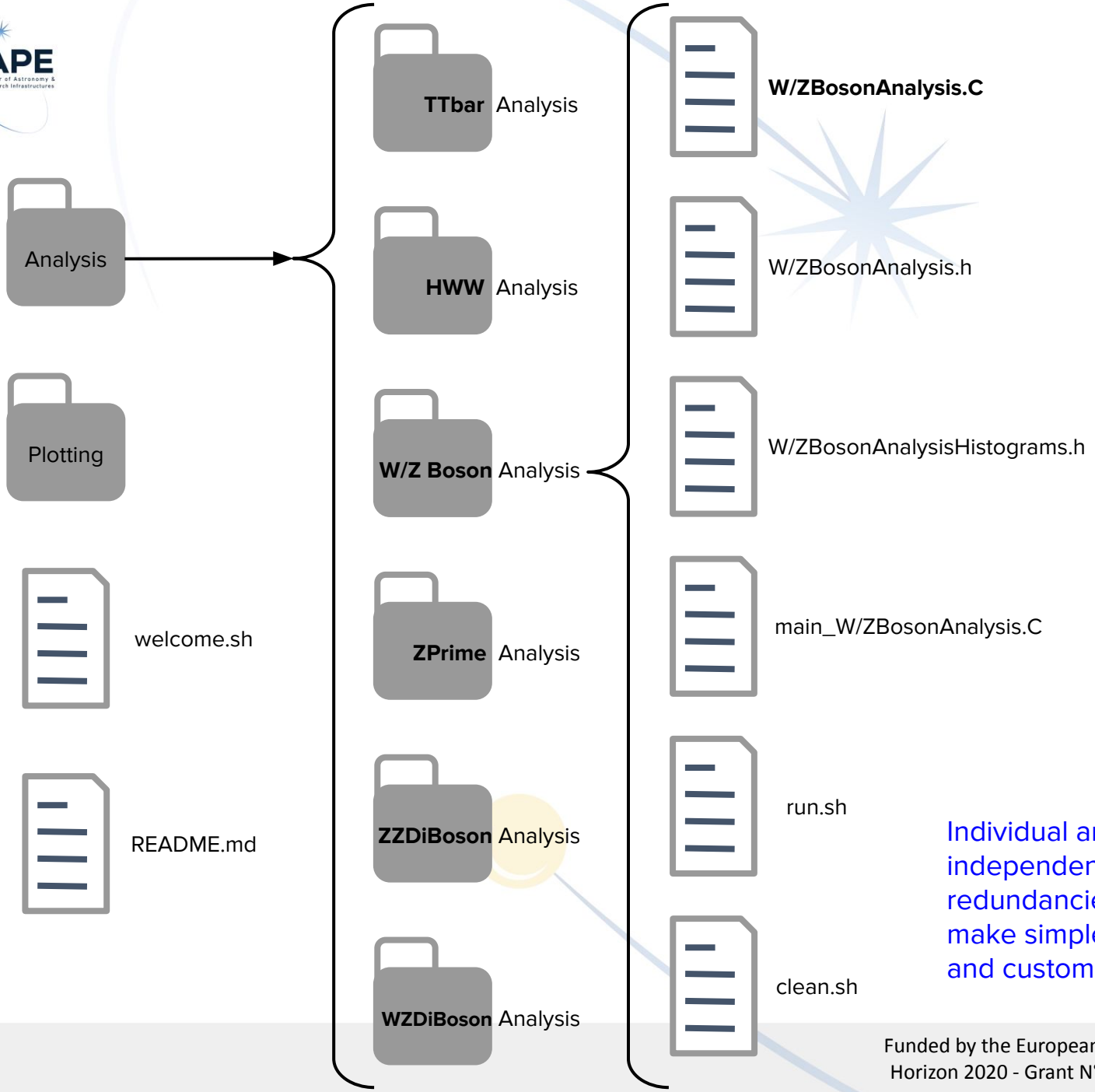
WZ / ZZ Analyses

ggAnalysis

•••

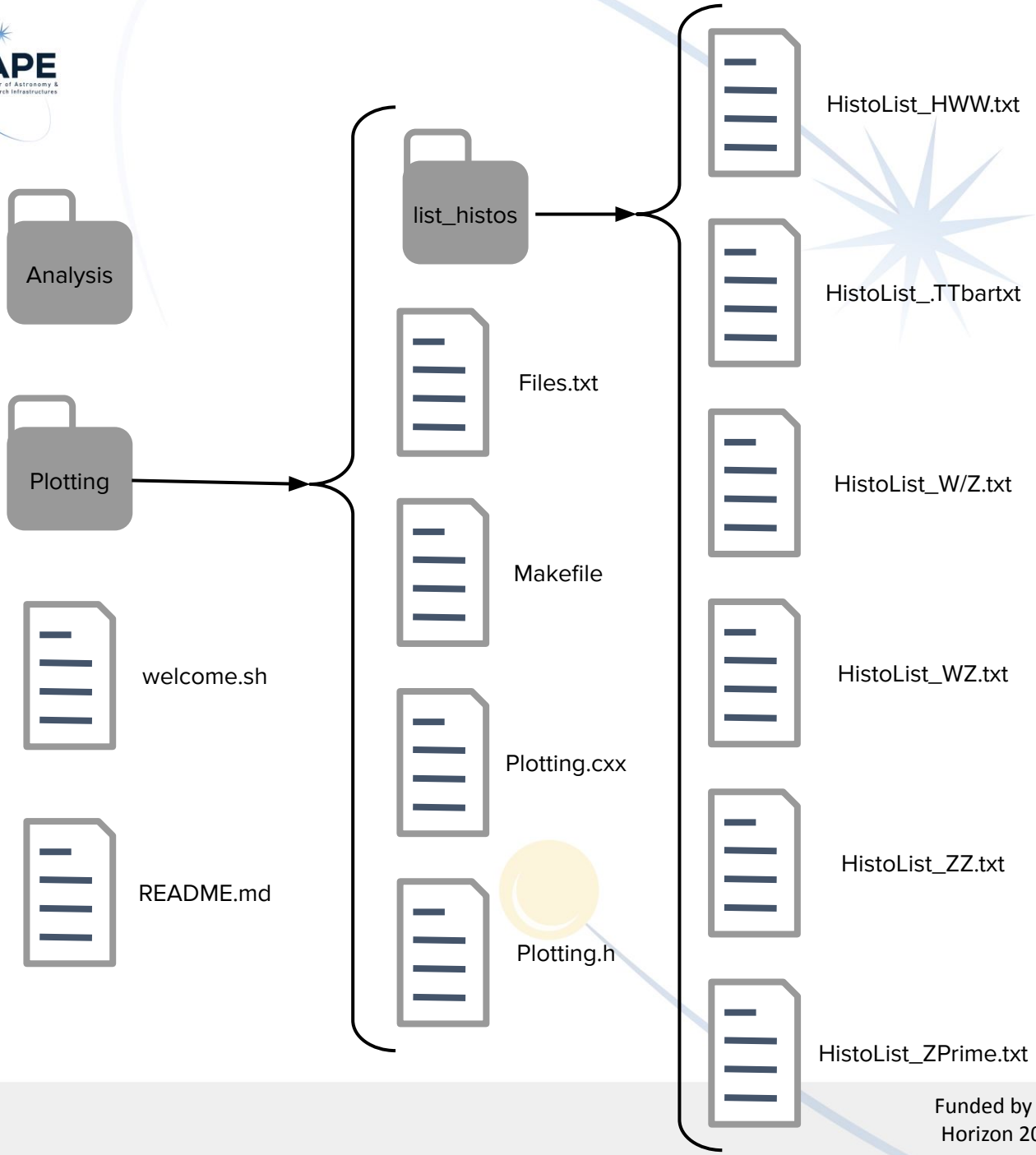
Mono-XAnalyses

Funded by the European Union's Horizon 2020 - Grant N° 824064



Individual analysis are hosted in independent folders. Some redundancies are present, but it make simpler the reproduction and customisation of new code

Funded by the European Union's Horizon 2020 - Grant N° 824064



Text files allow to define collections of histograms per analysis

i.e.

name of the histogram in the files
hist_etmiss
hist_vxp_z
hist_pvxp_n
hist_mt
hist_mLL
hist_threeleptpt
hist_threelepteta
hist_threeleptE
hist_threeleptphi
hist_threeleptch
hist_threeleptID
hist_threelept_ptc
hist_threeleptetc
hist_threelepz0
hist_threel

Funded by the European Union's Horizon 2020 - Grant N° 824064

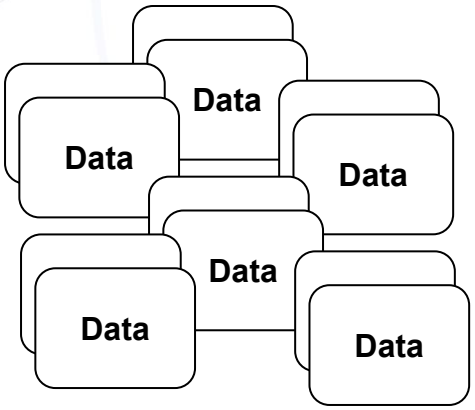
Framework Outputs

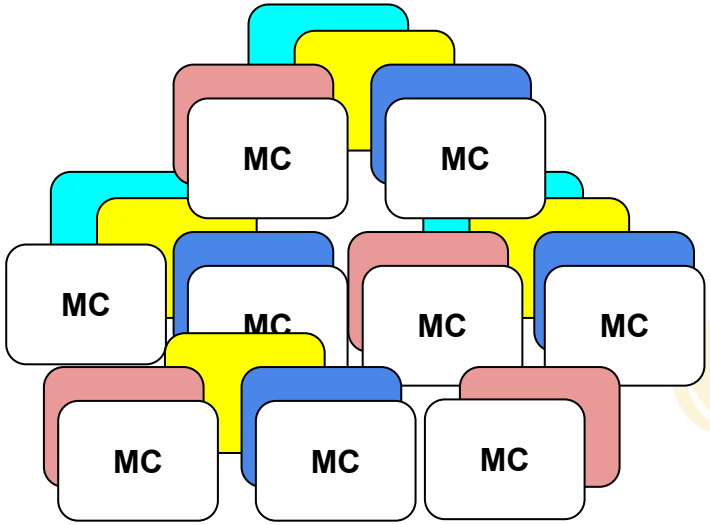
Funded by the European Union's
Horizon 2020 - Grant N° 824064

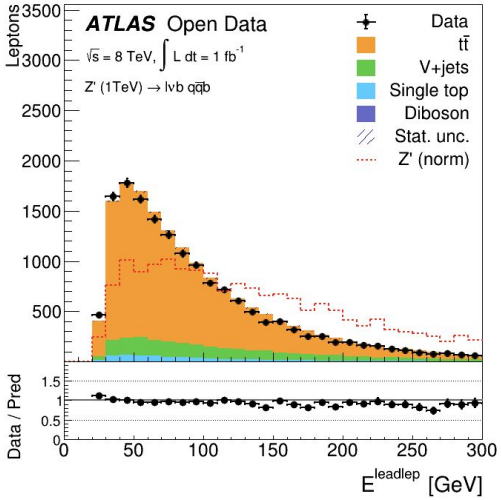
13 TeV release

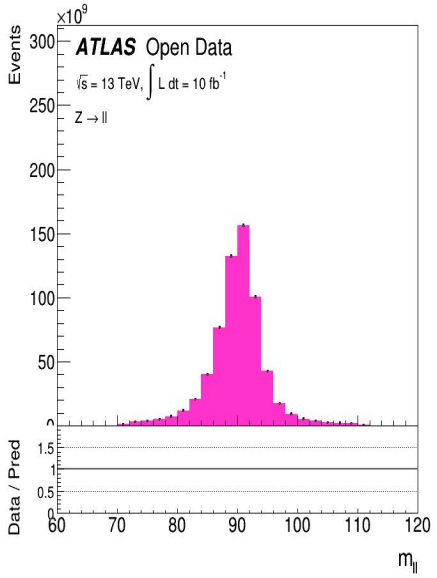
Analysis framework in C++ (~10 examples)

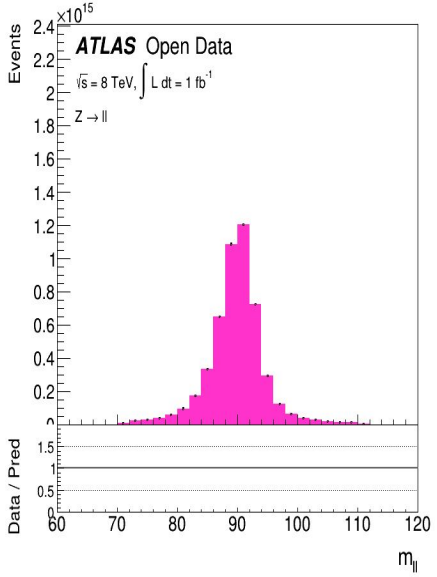
10 fb^{-1}











The analysis code can be run in single-core and multiple-core modes

Funded by the European Union's Horizon 2020 - Grant N° 824064

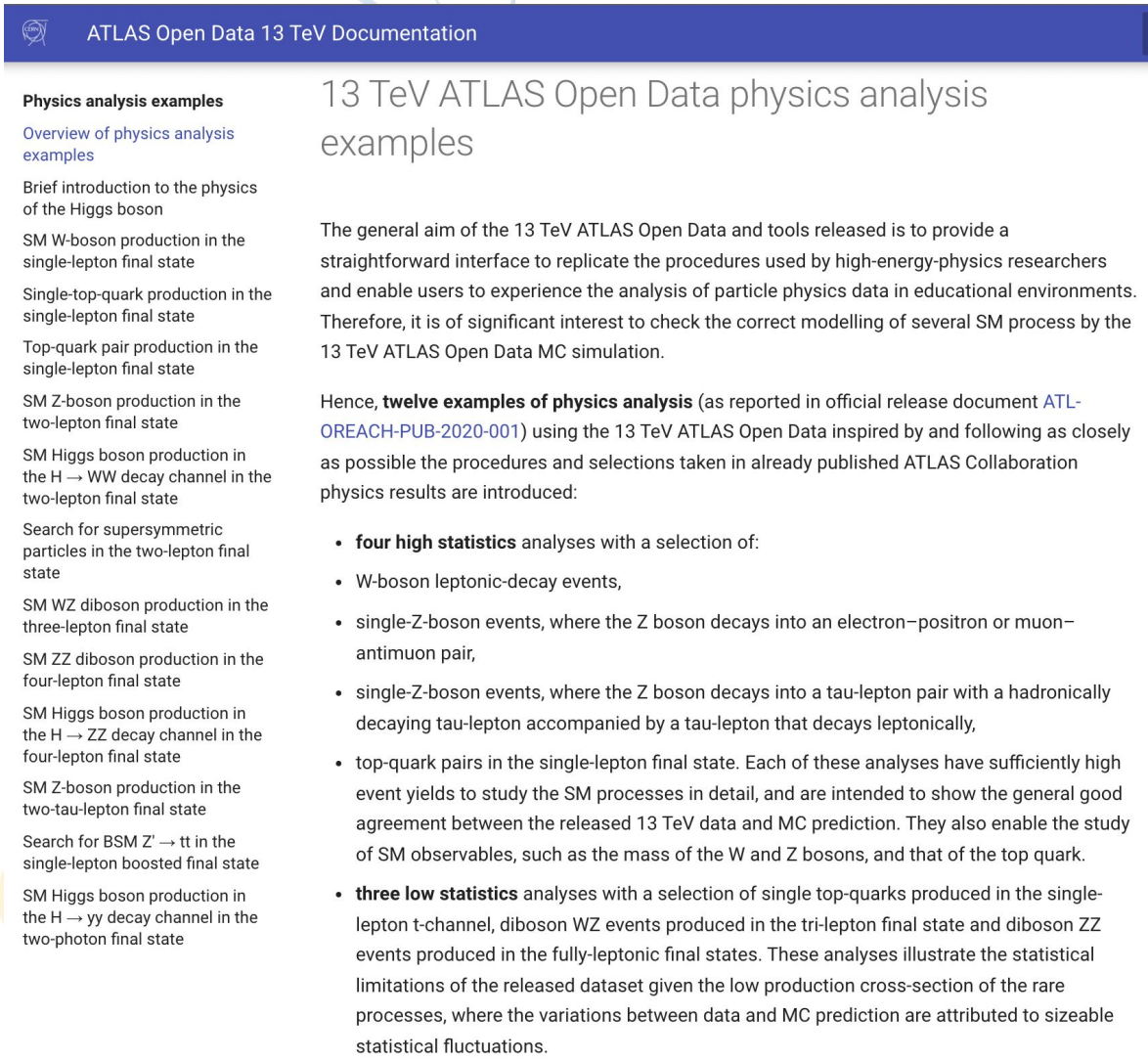
Short demo

- Show how the software is used and what is the outcome
- What should and can a EOSC user do with the software?

The framework produces analysis and plots of 12 different HEP searches →

ATLAS will use this software to deploy and perform the DAC12 in November 2021

This is the first piece of software that ATLAS Open Data wants to include in the OSSR for multiple tests and pipelines development



ATLAS Open Data 13 TeV Documentation

Physics analysis examples

Overview of physics analysis examples

Brief introduction to the physics of the Higgs boson

SM W-boson production in the single-lepton final state

Single-top-quark production in the single-lepton final state

Top-quark pair production in the single-lepton final state

SM Z-boson production in the two-lepton final state

SM Higgs boson production in the H → WW decay channel in the two-lepton final state

Search for supersymmetric particles in the two-lepton final state

SM WZ diboson production in the three-lepton final state

SM ZZ diboson production in the four-lepton final state

SM Higgs boson production in the H → ZZ decay channel in the four-lepton final state

SM Z-boson production in the two-tau-lepton final state

Search for BSM Z' → tt in the single-lepton boosted final state

SM Higgs boson production in the H → yy decay channel in the two-photon final state

13 TeV ATLAS Open Data physics analysis examples

The general aim of the 13 TeV ATLAS Open Data and tools released is to provide a straightforward interface to replicate the procedures used by high-energy-physics researchers and enable users to experience the analysis of particle physics data in educational environments. Therefore, it is of significant interest to check the correct modelling of several SM process by the 13 TeV ATLAS Open Data MC simulation.

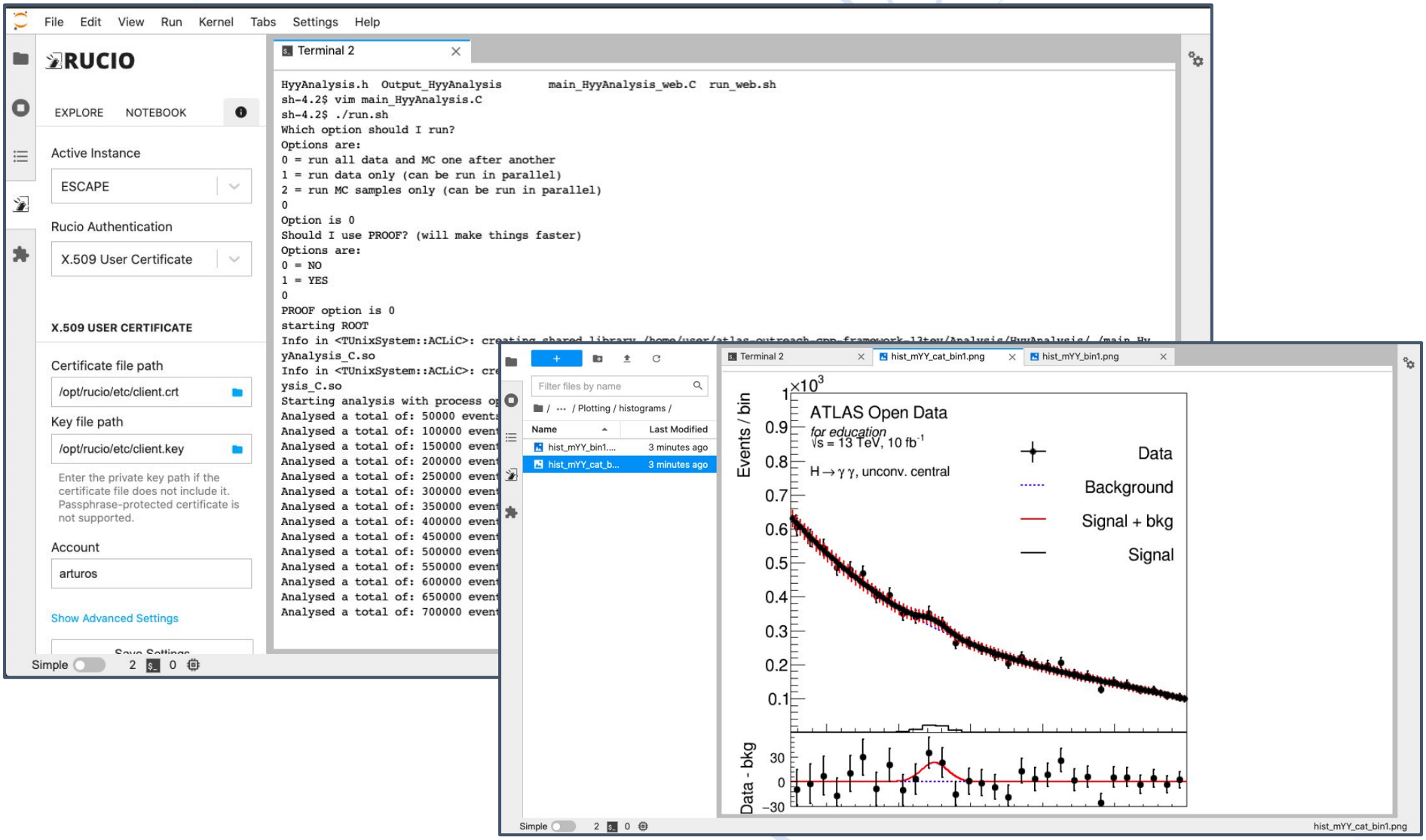
Hence, **twelve examples of physics analysis** (as reported in official release document ATL-OREACH-PUB-2020-001) using the 13 TeV ATLAS Open Data inspired by and following as closely as possible the procedures and selections taken in already published ATLAS Collaboration physics results are introduced:

- **four high statistics** analyses with a selection of:
- W-boson leptonic-decay events,
- single-Z-boson events, where the Z boson decays into an electron–positron or muon–antimuon pair,
- single-Z-boson events, where the Z boson decays into a tau-lepton pair with a hadronically decaying tau-lepton accompanied by a tau-lepton that decays leptonically,
- top-quark pairs in the single-lepton final state. Each of these analyses have sufficiently high event yields to study the SM processes in detail, and are intended to show the general good agreement between the released 13 TeV data and MC prediction. They also enable the study of SM observables, such as the mass of the W and Z bosons, and that of the top quark.
- **three low statistics** analyses with a selection of single top-quarks produced in the single-lepton t-channel, diboson WZ events produced in the tri-lepton final state and diboson ZZ events produced in the fully-leptonic final states. These analyses illustrate the statistical limitations of the released dataset given the low production cross-section of the rare processes, where the variations between data and MC prediction are attributed to sizeable statistical fluctuations.

http://opendata.atlas.cern/release/2020/documentation/physics/intro.html

Funded by the European Union's Horizon 2020 - Grant N° 824064

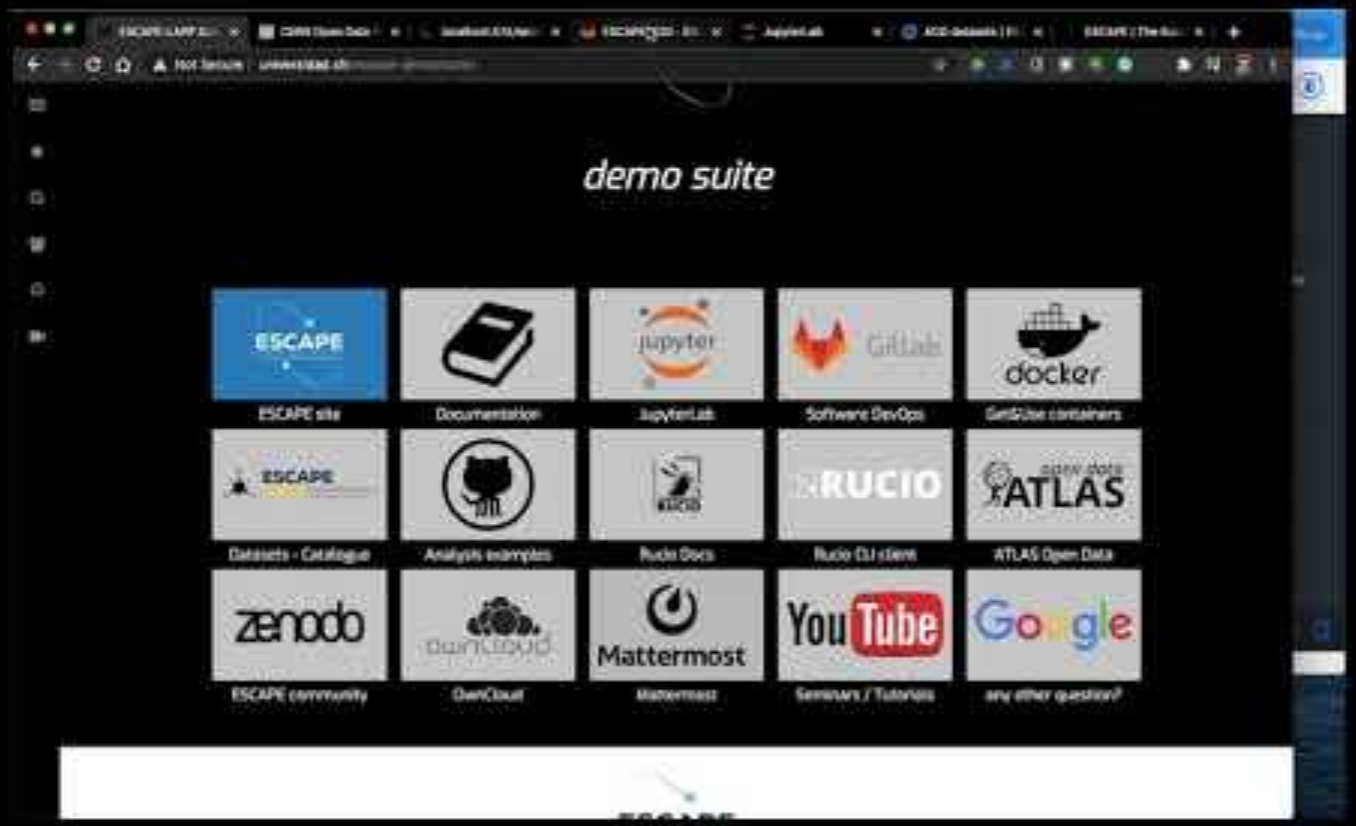
Current usage in the context of ESCAPE



Funded by the European Union's Horizon 2020 - Grant N° 824064

Ongoing developments with **JupyterLab & RUCIO extension**

(a 150 sec video)



http://universidad.ch/escape-demo/suite/

More tools to finish to integrate in the container, like more kernels, PROOF, CVMFS

Open Points and Discussion Time

- Which of your questions have not been covered so far?
- What do you want to discuss?

Funded by the European Union's Horizon 2020 - Grant N° 824064

TOC of Tech Report

- Introduction
  - ATLAS, HEP experimental data analysis software in C++
  - The ATLAS Open Data 13 TeV analysis framework
- Software/Service Development Strategy → And standalone Git repo
- Software/Service Requirements → CEN ROOT framework
- OSSR Integration
  - Status
  - Content
  - User Story

THANKS!

Funded by the European Union's Horizon 2020 - Grant N° 824064

Backup

Funded by the European Union's
Horizon 2020 - Grant N° 824064

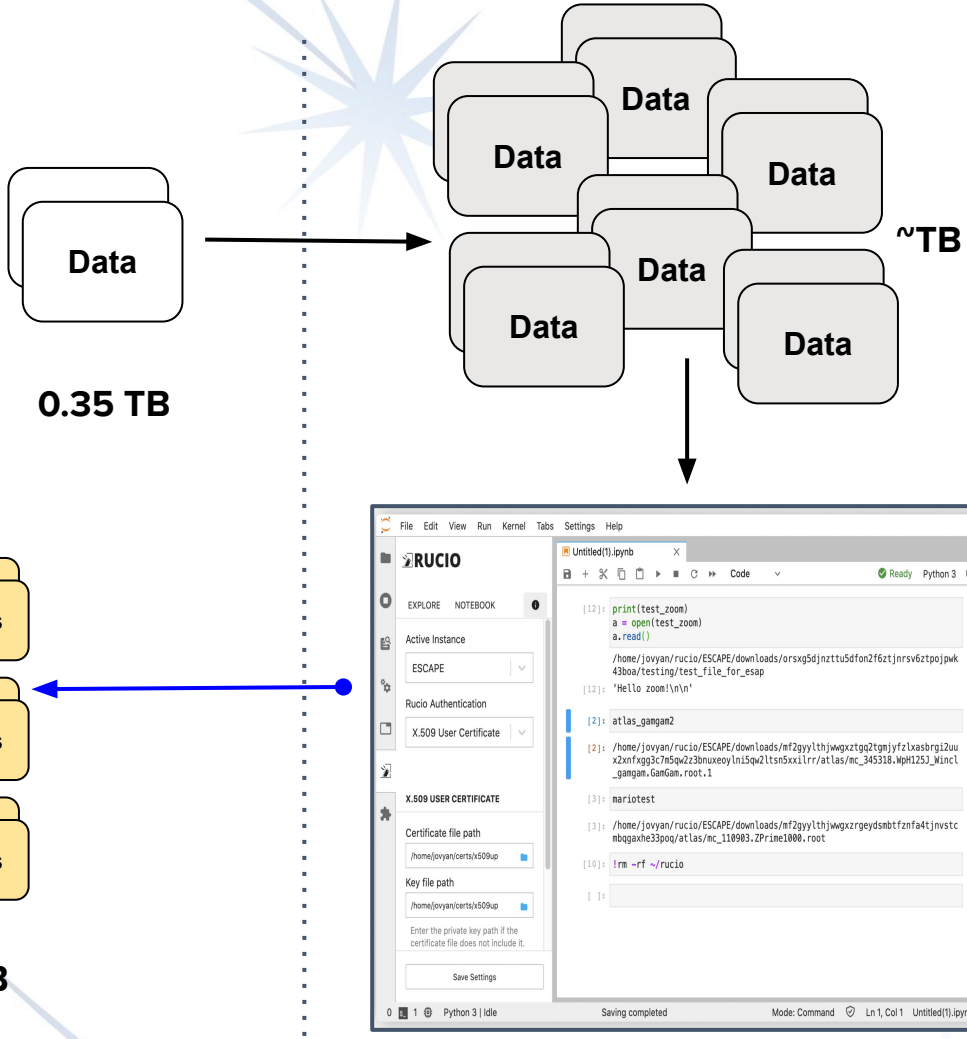
Data Analysis

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Analysis examples

We can also run the analysis examples over the “multiplied” data

- This can help to simulate longer analysis that can last several hours
- In case this kind of “stress” is useful in this challenge



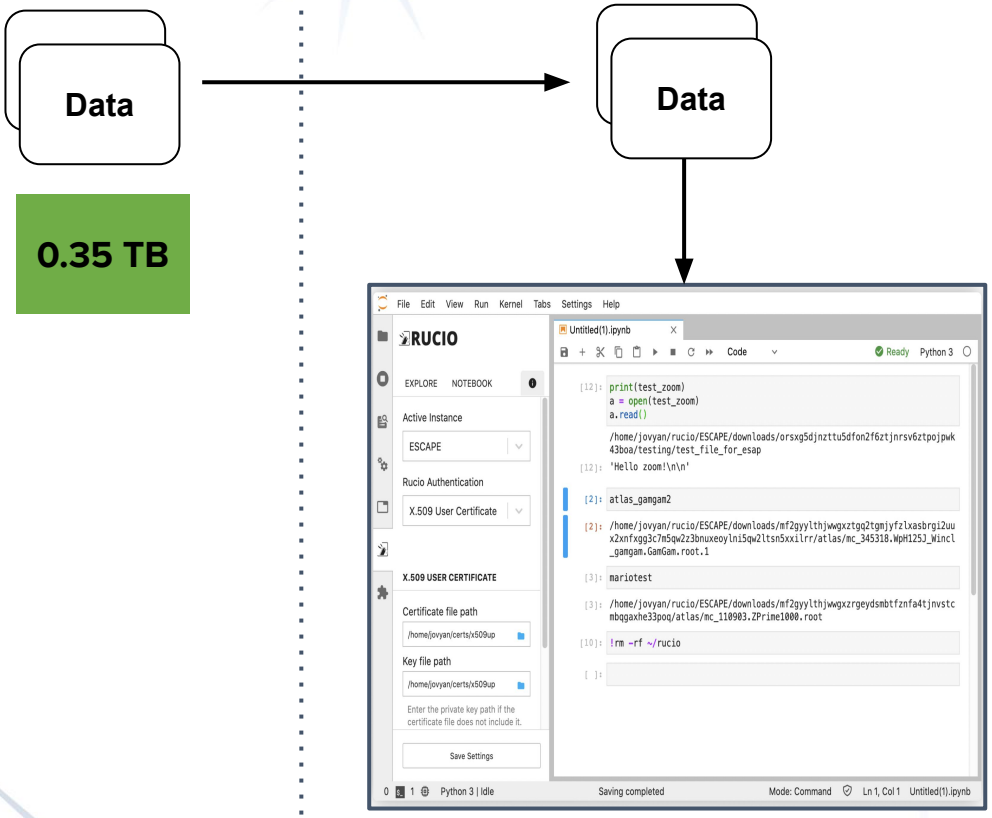
Funded by the European Union’s Horizon 2020 - Grant N° 824064

Analysis examples

We can use the current ATLAS Open Data analysis examples to retrieve and use datasets from the Datalake

- Analysis can be notebooks or analysis frameworks
- They can take from a few minutes (e.g. 5-30 min)
- To several (e.g. 4) hours

Also, write back the outputs



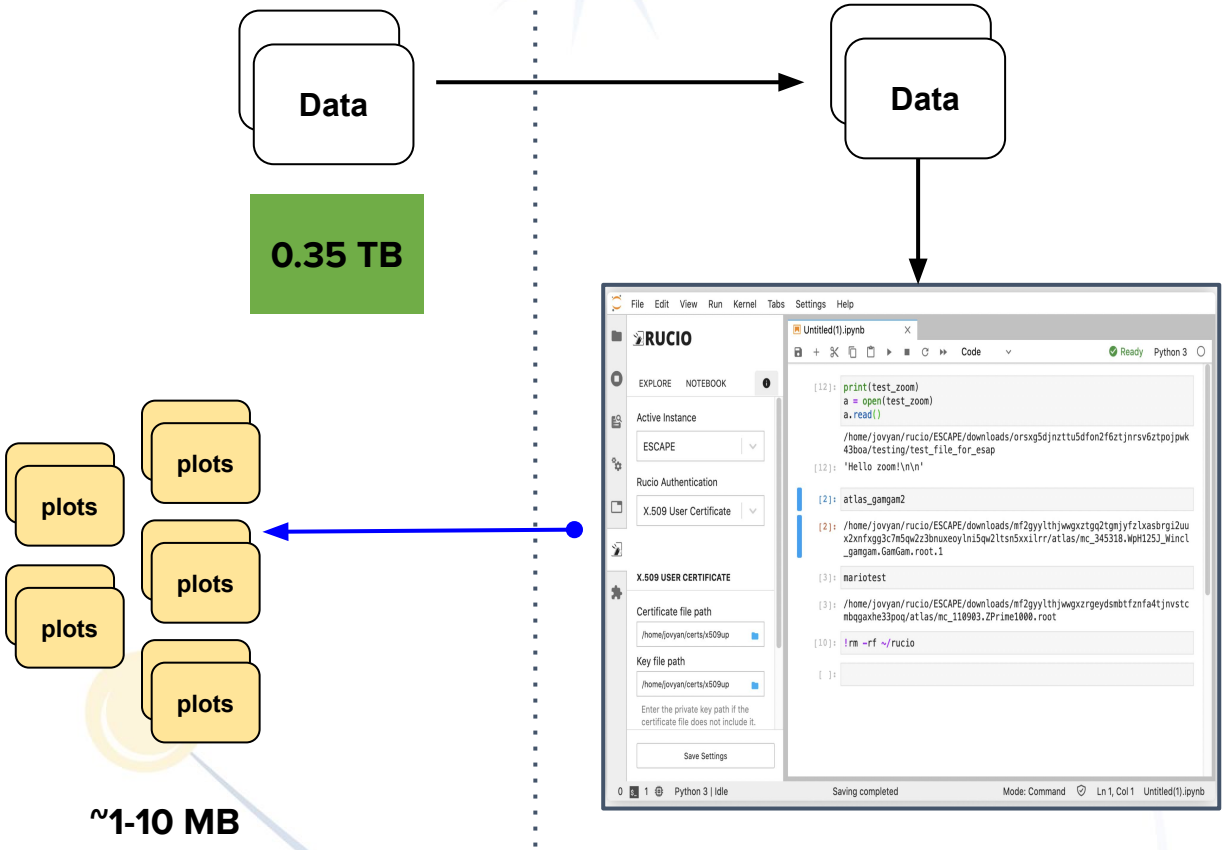
Datalake **Computer Element**

Funded by the European Union's Horizon 2020 - Grant N° 824064

Analysis examples

The outputs of the analysis can be upload to the Datalake

- The outputs are small; they are plots that can also be store in ROOT files
- No intention to upload single PNG files to The Datalake



Datalake

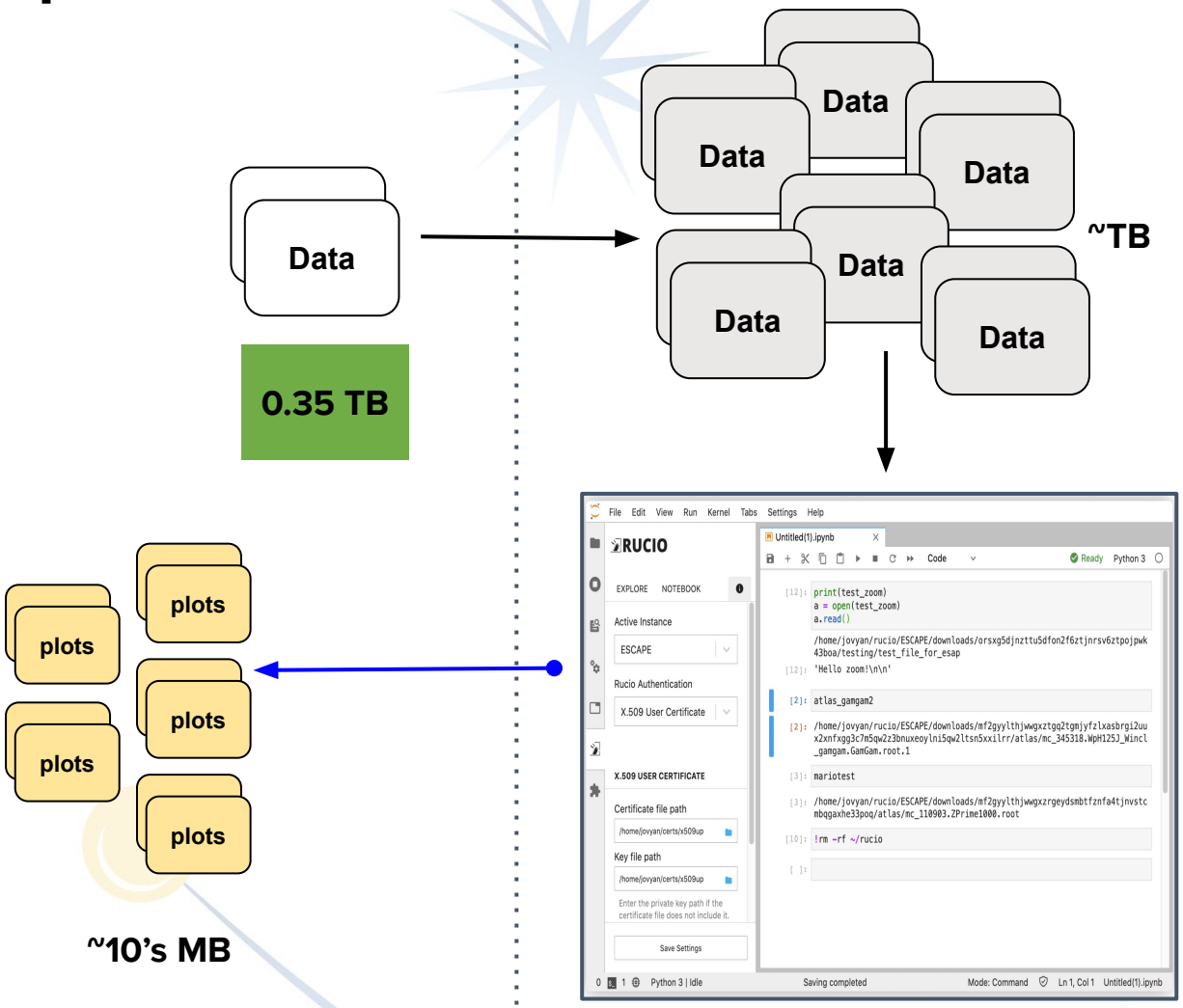
Computer Element

Funded by the European Union's Horizon 2020 - Grant N° 824064

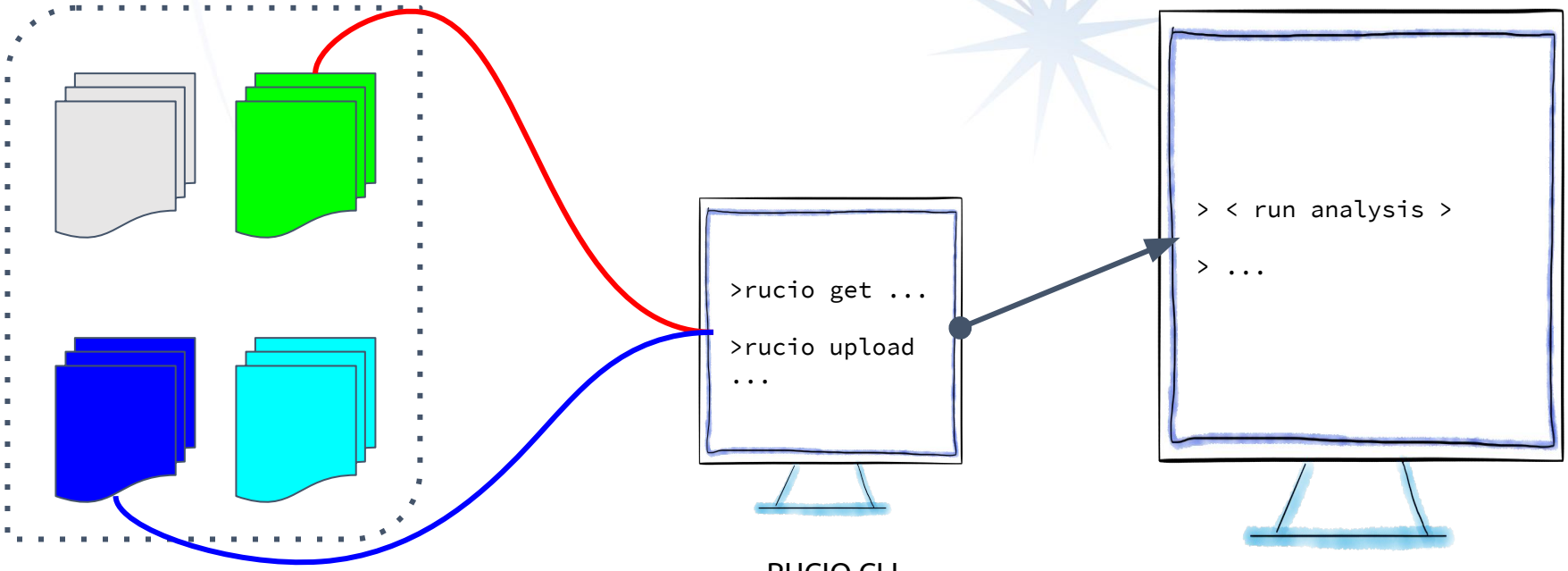
Analysis examples

We can also run the analysis examples over the “multiplied” data

- This can help to simulate longer analysis that can last several hours (e.g. ~8-12 hours)
- In case this kind of “stress” is useful in this challenge

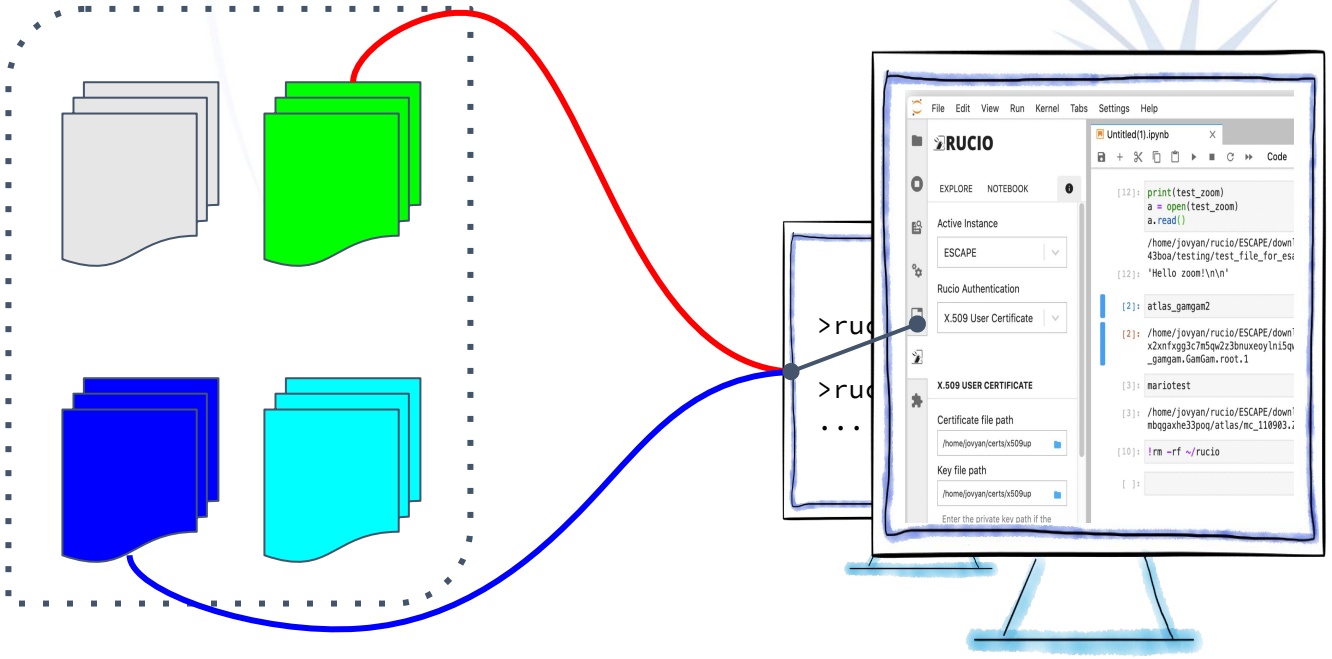


Funded by the European Union’s Horizon 2020 - Grant N° 824064



CLI interaction with samples

Funded by the European Union's Horizon 2020 - Grant N° 824064



Authentication, discovery and download of the samples using the Web UI

Data Lake and stored Open Data datasets

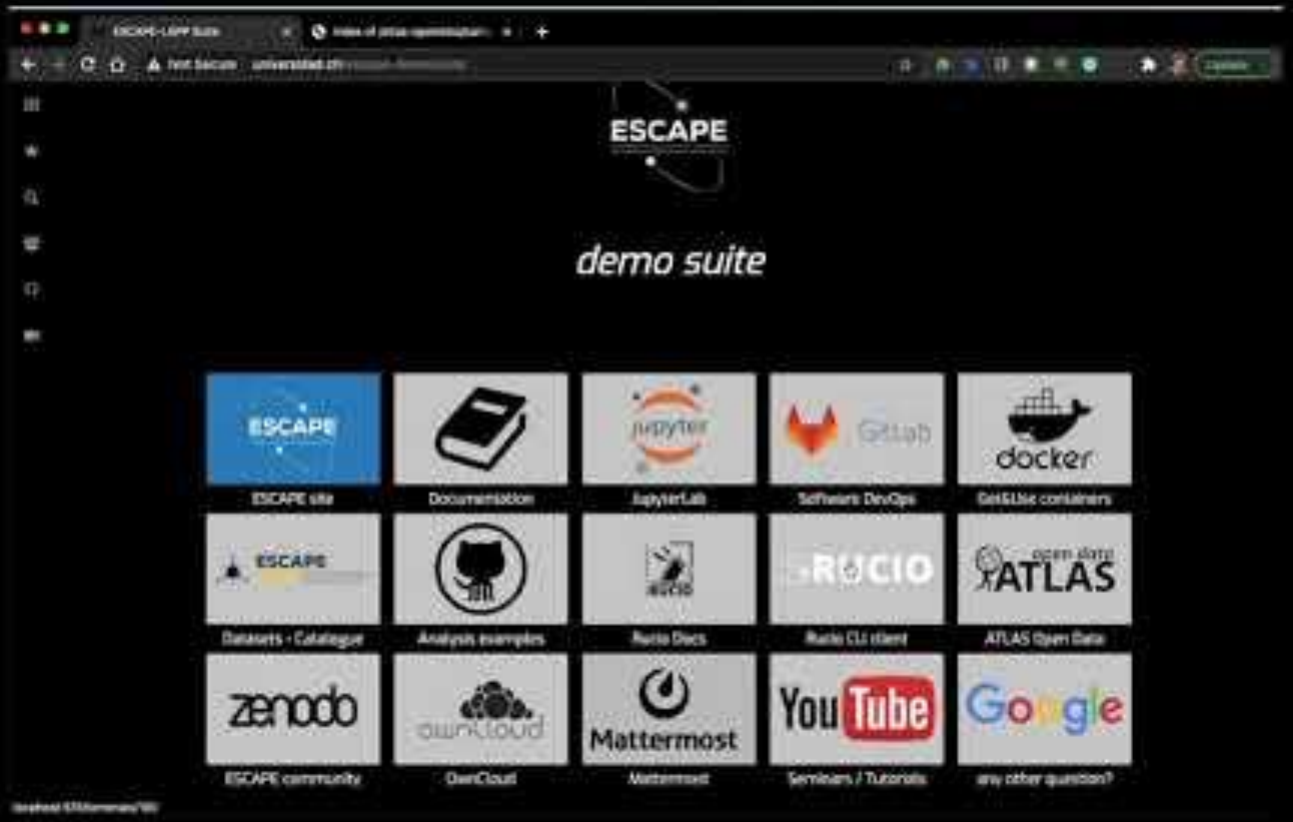
Container with RUCIO + JupyterLab extension

RUCIO+JupyterLab (container) interaction for users

Funded by the European Union's Horizon 2020 - Grant N° 824064

The RUCIO CLI client

(a 90 sec video, mainly for new users)



http://universal.ch/escape-demo/suite/