A new release of ATLAS Open Data for Education

David Koch on behalf of the ATLAS Education and Outreach Group

DPG 8.3.2024

FAIR principles

FAIR principles

ATLAS Open Data for Education

…focuses on implementing FAIR principles with **outreach and education** in mind, i.e. the target audience ranges from high schoolers to teachers to the general public to university students. The released data is **not** suited to produce publication-ready physics results.

FAIR principles

ATLAS Open Data for Education

...focuses on implementing FAIR principles with **outreach and education** in mind, i.e. the target audience ranges from high schoolers to teachers to the general public to university students. The released data is **not** suited to produce publication-ready physics results.

FAIR principles
- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

FAIR principles

ATLAS Open Data for Education

…focuses on implementing FAIR principles with **outreach and education** in mind, i.e. the target audience ranges from high schoolers to teachers to the general public to university students. The released data is **not** suited to produce publication-ready physics results.

FAIR principles
- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

$\longrightarrow$

get the absolute **most** out of our **extremely valuable** data

FAIR principles

FAIR principles and how ATLAS Open Data for Education implements them
- **F**indable
  - data is hosted on the searchable and well-indexed **CERN Open Data Portal**
  - polished **web-appearance** communicates the contents and example use-cases of the released data and software

FAIR principles

FAIR principles and how ATLAS Open Data for Education implements them
- **F**indable
  - data is hosted on the searchable and well-indexed **CERN Open Data Portal**
  - polished **web-appearance** communicates the contents and example use-cases of the released data and software
- **A**ccessible
  - data as well as accompanying software is made **accessable to the public** keeping in mind the range of internet bandwidths, computer OSs, storage and RAM, and access to experts
  - provide material for audiences with **different background and skills** with a wide range of learning objectives and project goals

FAIR principles

- **I**nteroperable
  - data can be analysed with a **variety of open source software** like ROOT, the Python HEP ecosystem or even Julia

FAIR principles

- **I**nteroperable
  - data can be analysed with a **variety of open source software** like ROOT, the Python HEP ecosystem or even Julia
- **R**eusable
  - **high quality documentation** is provided alongside the release of data and software to ensure long-time usability to users
  - software to produce the open data samples is also open source → can be reused for future releases of open data for education

ATLAS Open Data for Education

Overview of ATLAS Open Data for Education

- two campaigns of open data and Monte Carlo simulations were released so far:
  - proton-proton collision data at **8 TeV**: 1fb^{-1} of data, 44 MC datasets
  - **13 TeV**: 10fb^{-1} of data, 120 MC datasets
- Wide range of **analysis tools, videos and data visualisers** provided with the data
- **Datasets, associated information, tools and interactive material** accessible via [website](#)
- extensive **technical documentation** in the [Pub Note](#)

Overview of ATLAS Open Data for Education

- two campaigns of open data and Monte Carlo simulations were released so far:
  - proton-proton collision data at **8 TeV**: 1fb^{-1} of data, 44 MC datasets
  - **13 TeV**: 10fb^{-1} of data, 120 MC datasets
- Wide range of **analysis tools, videos and data visualisers** provided with the data
- **Datasets, associated information, tools and interactive material** accessible via website
- extensive **technical documentation** in the Pub Note

New campaign with 36fb^{-1} of 13 TeV data

- ~ 25% of Run 2 data
- more luminosity, more recent data → allows for better analyses with more advanced techniques, rarer decays and better signal strength
- improved documentation + wider range of tools to reach more people

New release of ATLAS Open Data for Education

What people want to learn
- particle physics
- programming (C++/Python)
- Likelihood fitting
- Machine learning
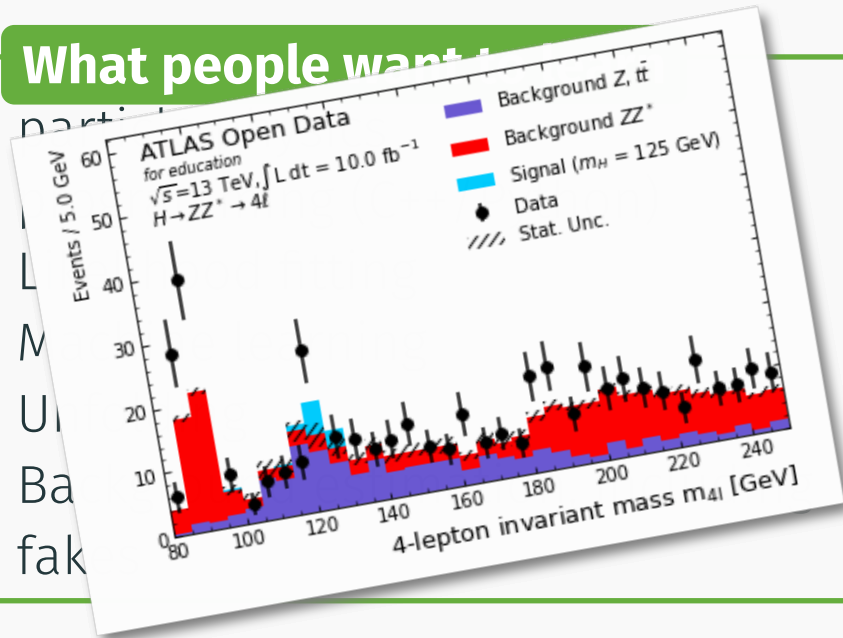- Unfolding
- Background estimation, including fakes

New release of ATLAS Open Data for Education

What people want to learn

- particle physics
- programming (C++/Python)
- Likelihood fitting
- Machine learning
- Unfolding
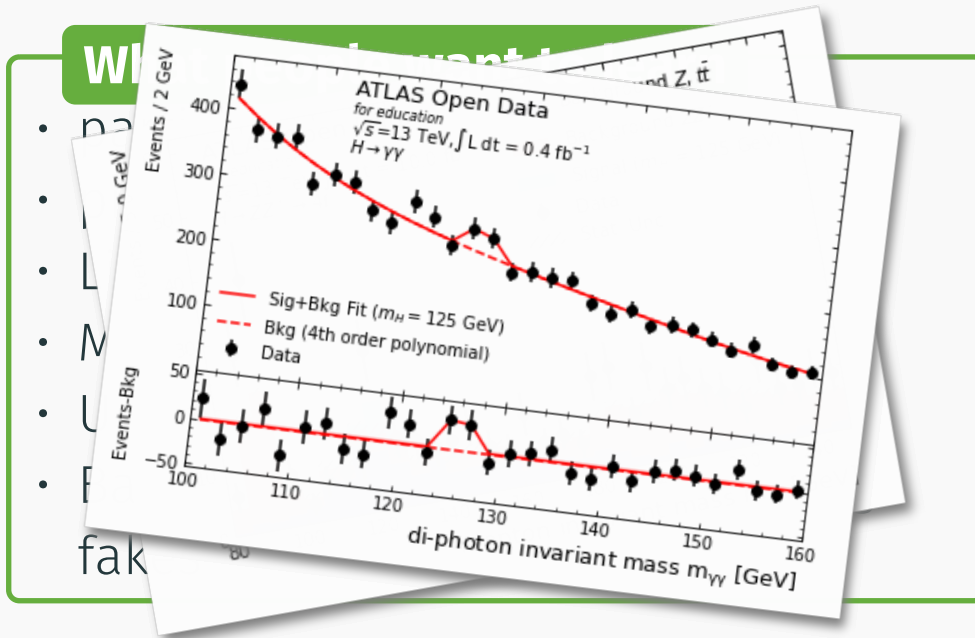- Background estimation, including fakes

Selection of analyses that can be done with open data

- 4 lepton final state: Z, Higgs, ZZ peaks
- Higgs to gamma gamma: Higgs discovery, fits, background estimation
- Higgs to mumu: increase sensitivity (e.g ML), background estimation
- ttbar: forward backward asymmetries, BDTs for signal classification, etc
- … and **many more**

What people want

- particl...
- ...
- L...
- M...
- U...
- Ba...
  fak...

ATLAS Open Data
for education
$\sqrt{s} = 13$ TeV, $\int L\,dt = 10.0$ fb^{-1}
$H \rightarrow ZZ^* \rightarrow 4\ell$

Events / 5.0 GeV

Background Z, $t\bar{t}$
Background ZZ^*
Signal ($m_H = 125$ GeV)
Data
Stat. Unc.

4-lepton invariant mass $m_{4\ell}$ [GeV]

Selection of analyses that can be done with open data

- 4 lepton final state: Z, Higgs, ZZ peaks
- Higgs to gamma gamma: Higgs discovery, fits, background estimation
- Higgs to mumu: increase sensitivity (e.g ML), background estimation
- ttbar: forward backward asymmetries, BDTs for signal classification, etc
- … and **many more**

New release of ATLAS Open Data for Education

ATLAS Open Data
for education
$\sqrt{s}$ =13 TeV, $\int L$ dt = 0.4 fb^{-1}
$H \to \gamma\gamma$

— Sig+Bkg Fit (m_H = 125 GeV)
-- Bkg (4th order polynomial)
• Data

di-photon invariant mass $m_{\gamma\gamma}$ [GeV]

Selection of analyses that can be done with open data

- 4 lepton final state: Z, Higgs, ZZ peaks
- Higgs to gamma gamma: Higgs discovery, fits, background estimation
- Higgs to mumu: increase sensitivity (e.g ML), background estimation
- ttbar: forward backward asymmetries, BDTs for signal classification, etc
- … and **many more**

New release of ATLAS Open Data for Education

Selection of analyses that can be done with open data

- 4 lepton final state: Z, Higgs, ZZ peaks
- Higgs to gamma gamma: Higgs discovery, fits, background estimation
- Higgs to mumu: increase sensitivity (e.g ML), background estimation
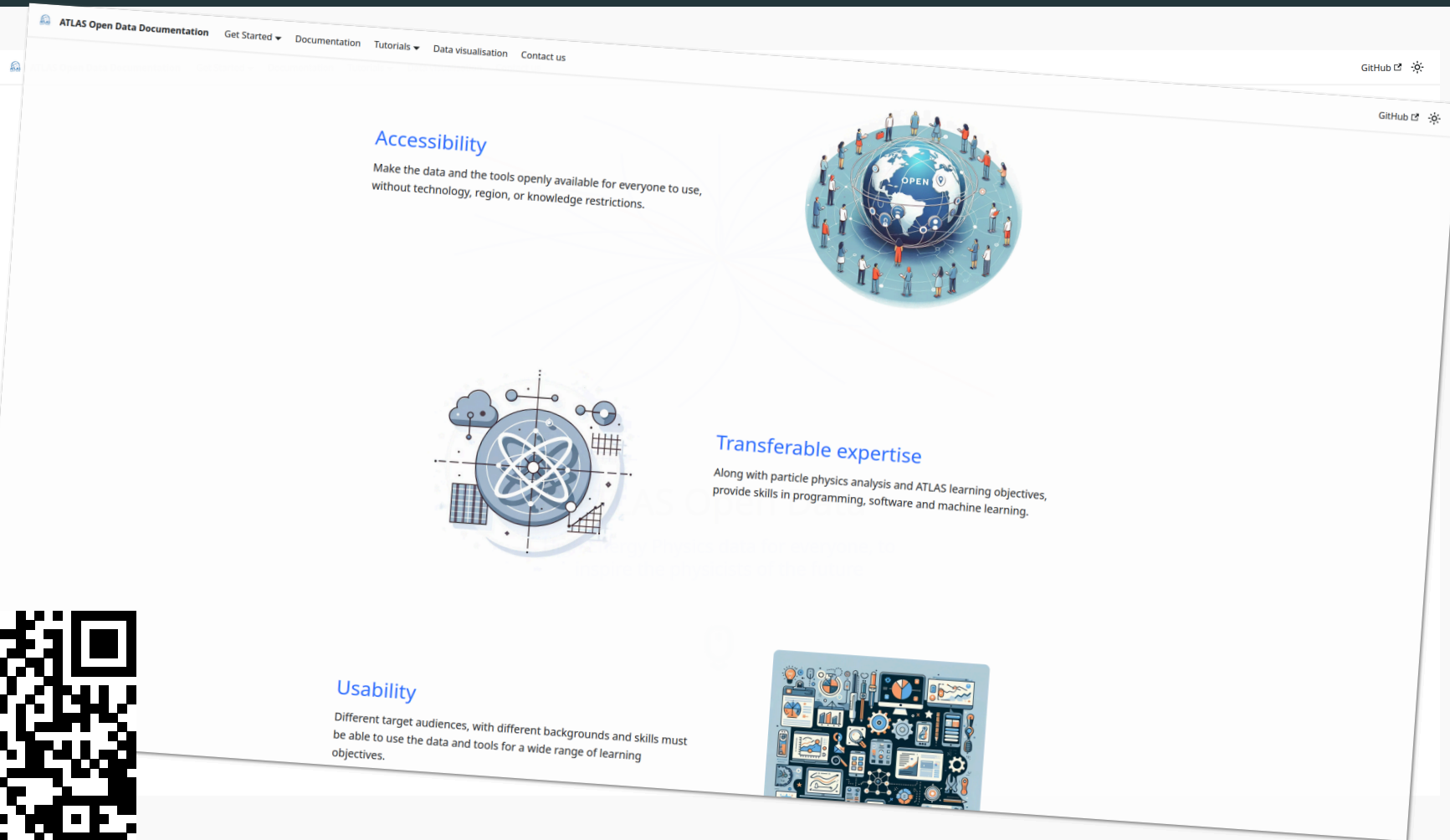- ttbar: forward backward asymmetries, BDTs for signal classification, etc
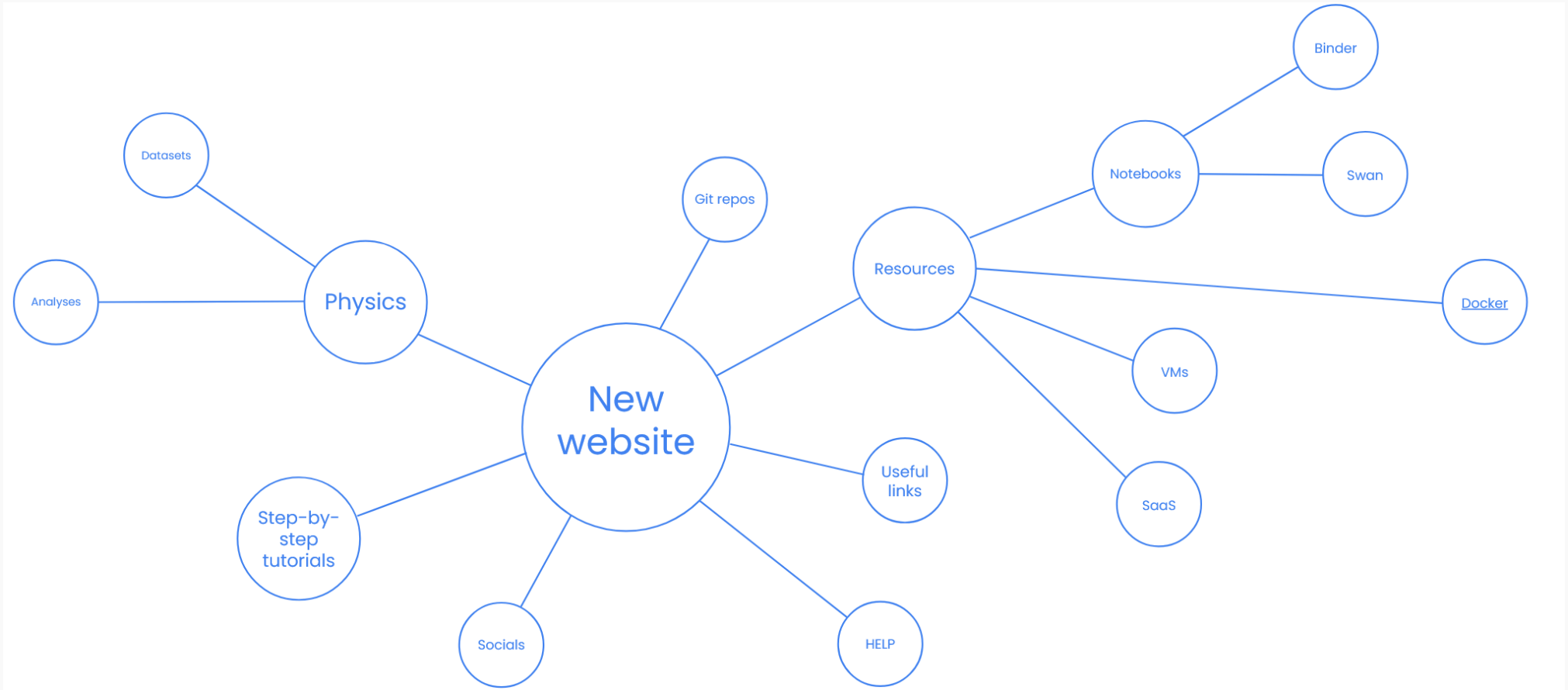- … and **many more**

Website

Website

https://opendata.atlas.cern/

Website

https://opendata.atlas.cern/
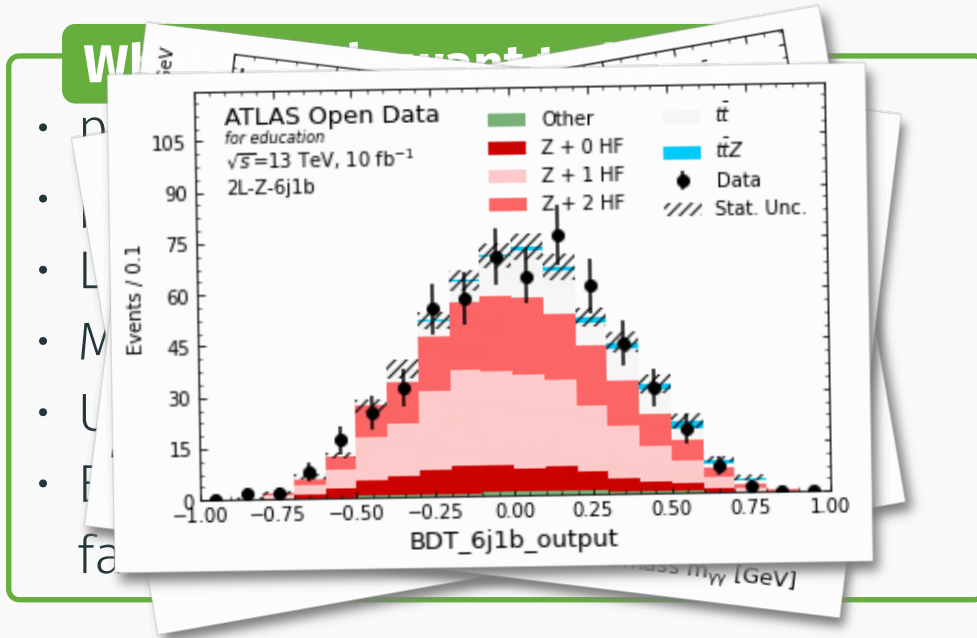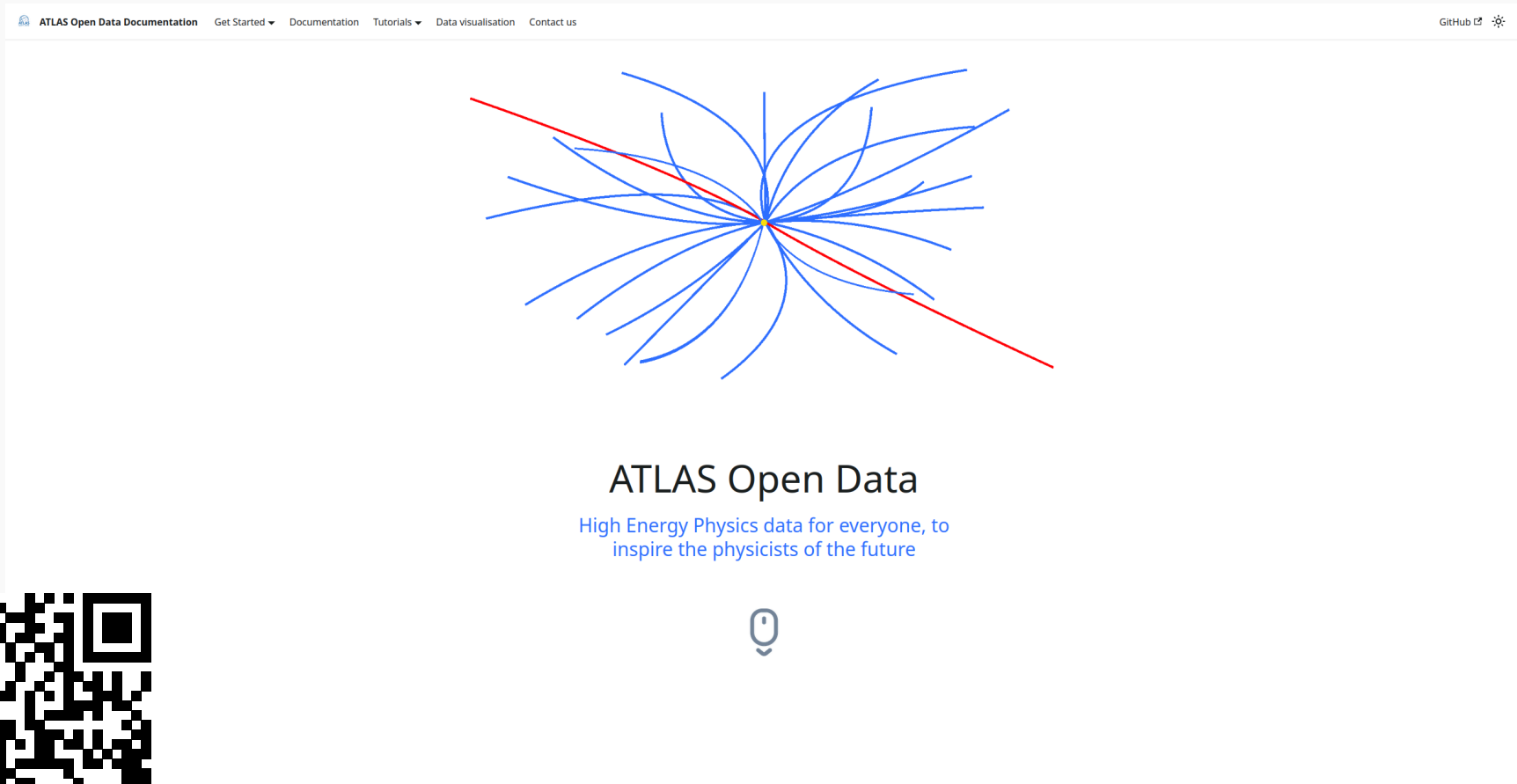
Website

Resources available on the new website

Datasets

Datasets

- Data and MC samples will be released as **ROOT NTuples**
- around **100 variables**—mostly the same as in the previous release—using updated recommendations from ATLAS

Datasets

- Data and MC samples will be released as **ROOT NTuples**
- around **100 variables**—mostly the same as in the previous release—using updated recommendations from ATLAS
- a framework to produce the NTuples is currently being developed and will be released as **open source** software alongside the NTuples
  - the framework converts DAOD_PHYSLITE to NTuples
  - DAOD_PHYSLITE is a commonly used dataformat in ATLAS → the ATLAS Open Data for Research campaign will publish data in PHYSLITE format so users will be able to use the framework to **produce their own NTuples** from the Open Data for Research release

Datasets

- Data and MC samples will be released as **ROOT NTuples**
- around **100 variables**—mostly the same as in the previous release—using updated recommendations from ATLAS
- a framework to produce the NTuples is currently being developed and will be released as **open source** software alongside the NTuples
  - the framework converts DAOD_PHYSLITE to NTuples
  - DAOD_PHYSLITE is a commonly used dataformat in ATLAS → the ATLAS Open Data for Research campaign will publish data in PHYSLITE format so users will be able to use the framework to **produce their own NTuples** from the Open Data for Research release
- samples will be hosted on the **CERN Open Data Portal** alongside the previous releases

Wrap up
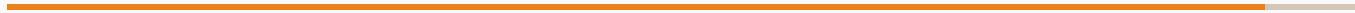
Wrap up

- ATLAS will release $\mathbf{36fb^{-1}}$ of 13 TeV pp-collision data plus Monte Carlo simulation as a new Open Data for Education campaign
- the release focuses on **outreach and education**: target audience is high-schoolers, teachers, the general public, lecturers, university students, ...
- the release of samples is accompanied by new **educational material** (notebooks, documentation, visualization tools, ...)
- reworked web appearance
- release of an open source **NTuple production framework** to produce customized NTuples