

# ATLAS Open Data: coinvolgere il pubblico in Educazione e Ricerca

Luca Clissa<sup>1,2</sup>, [luca.clissa2@unibo.it](mailto:luca.clissa2@unibo.it)  
<sup>1</sup>Università di Bologna, <sup>2</sup>INFN Bologna



## ATLAS e Open Data

### 1. L'esperimento ATLAS

ATLAS è uno dei due esperimenti general purpose di LHC:

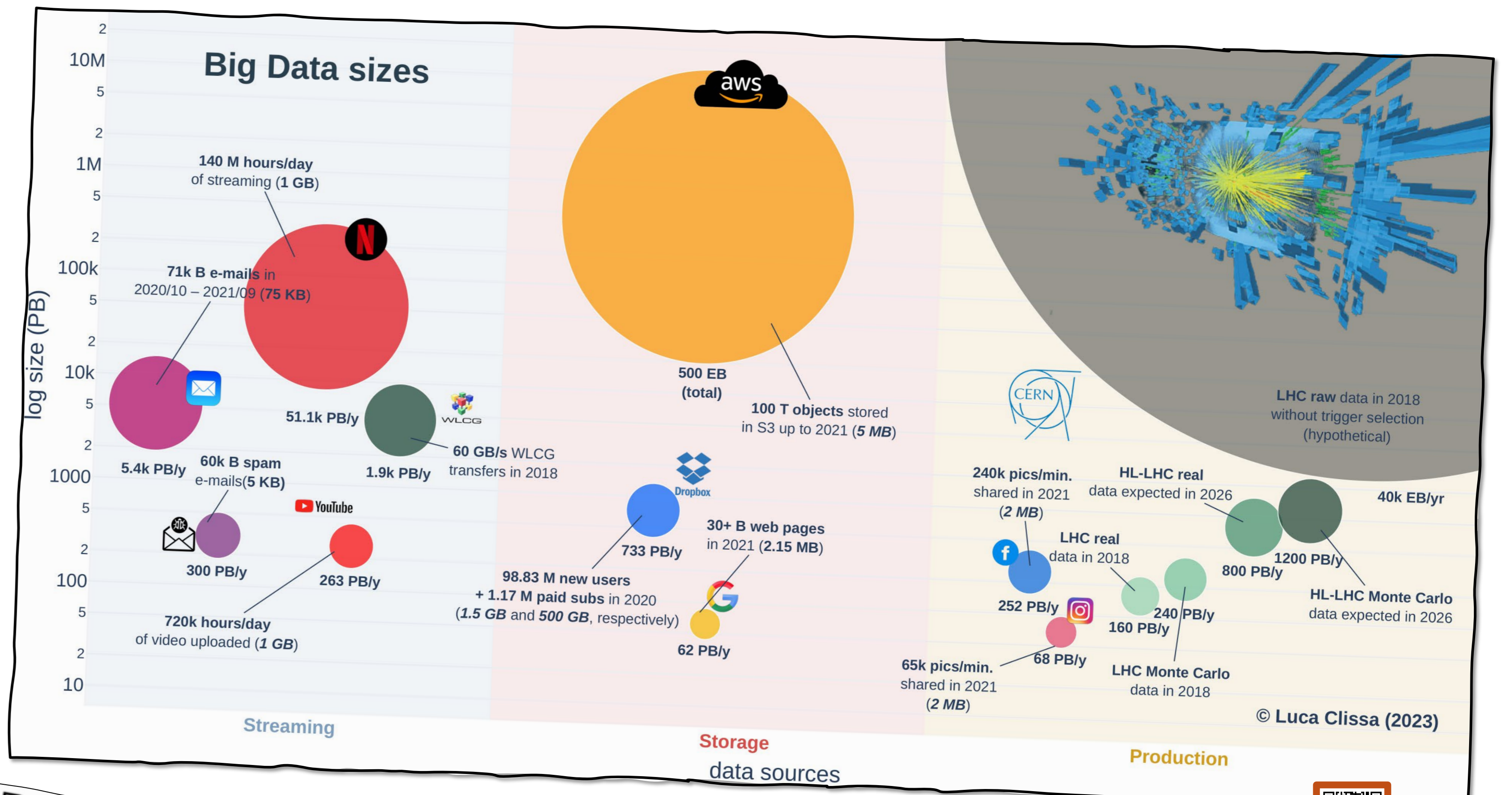
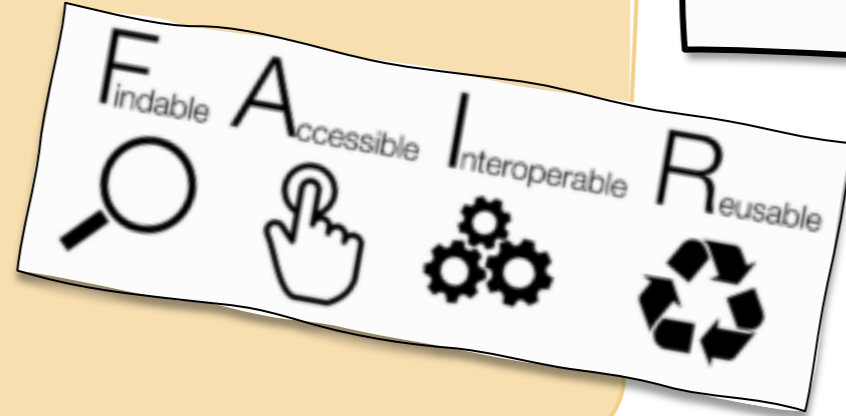
- Studia diversi tipi di processi, dalle misure di precisione alla ricerca di nuova fisica
- > 4 PB di RAW data raccolti
- > 100 PB di disco includendo dati ricostruiti/pre-processati
- Lo stack software è nell'ordine di ~10 GB, 4 M linee di codice

Rendere accessibile questa complessità non è banale!

### 2. Cosa significa Open Data?

Il CERN aderisce ad una politica Open Data secondo i principi FAIR. I dati raccolti sono a disposizione di tutti, senza restrizioni di accesso e utilizzo. Chiunque può:

- Consultare e riutilizzare i dati
- Verificare e riprodurre i risultati di ricerca
- Contribuire al progresso scientifico



## Iniziative

### 3. Perché Open Data?

Condividere i dati col pubblico ha diversi benefici:

- Rendere la società partecipe della ricerca scientifica
- Supportare insegnanti nella formazione degli scienziati del futuro
- Suscitare interesse e appassionare le nuove generazioni
- Rendere tangibile il risultato dei finanziamenti pubblici
- Attrarre expertise da altri settori, e.g. Machine Learning, Computing



### 4. Dataset

Collisioni protone-protone raccolte dall'esperimento ATLAS:

- $1 \text{ fb}^{-1}$  at 8 TeV,  $10 \text{ fb}^{-1}$  at 13 TeV
- Sia dati veri che simulazioni Monte Carlo
- 120+ simulazioni di vari processi fisici
- 100+ variabili per evento
- **Nuova versione in uscita, con focus su incertezza sistematica**



### 5. Eventi

ATLAS usa gli Open Data per diverse iniziative:

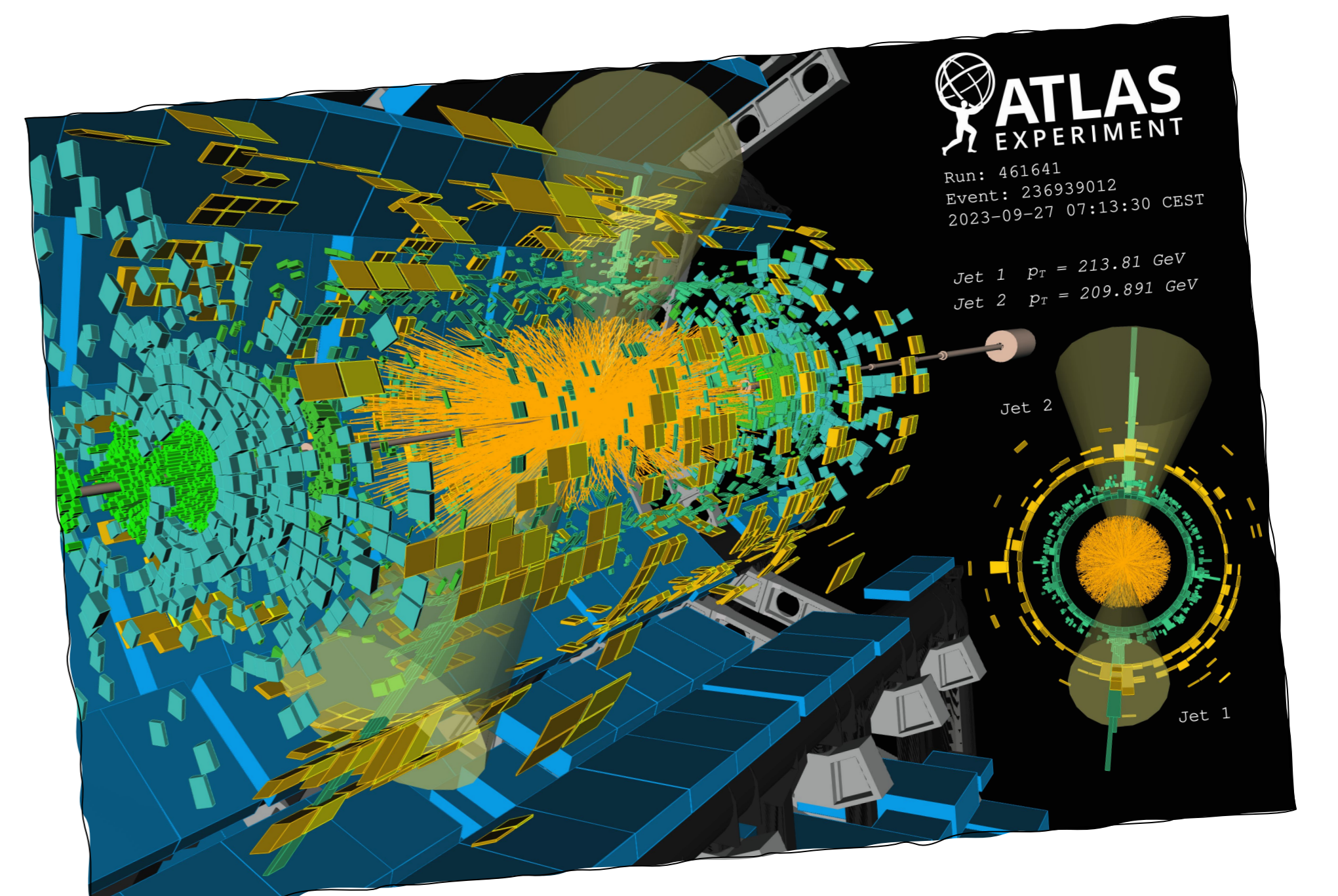
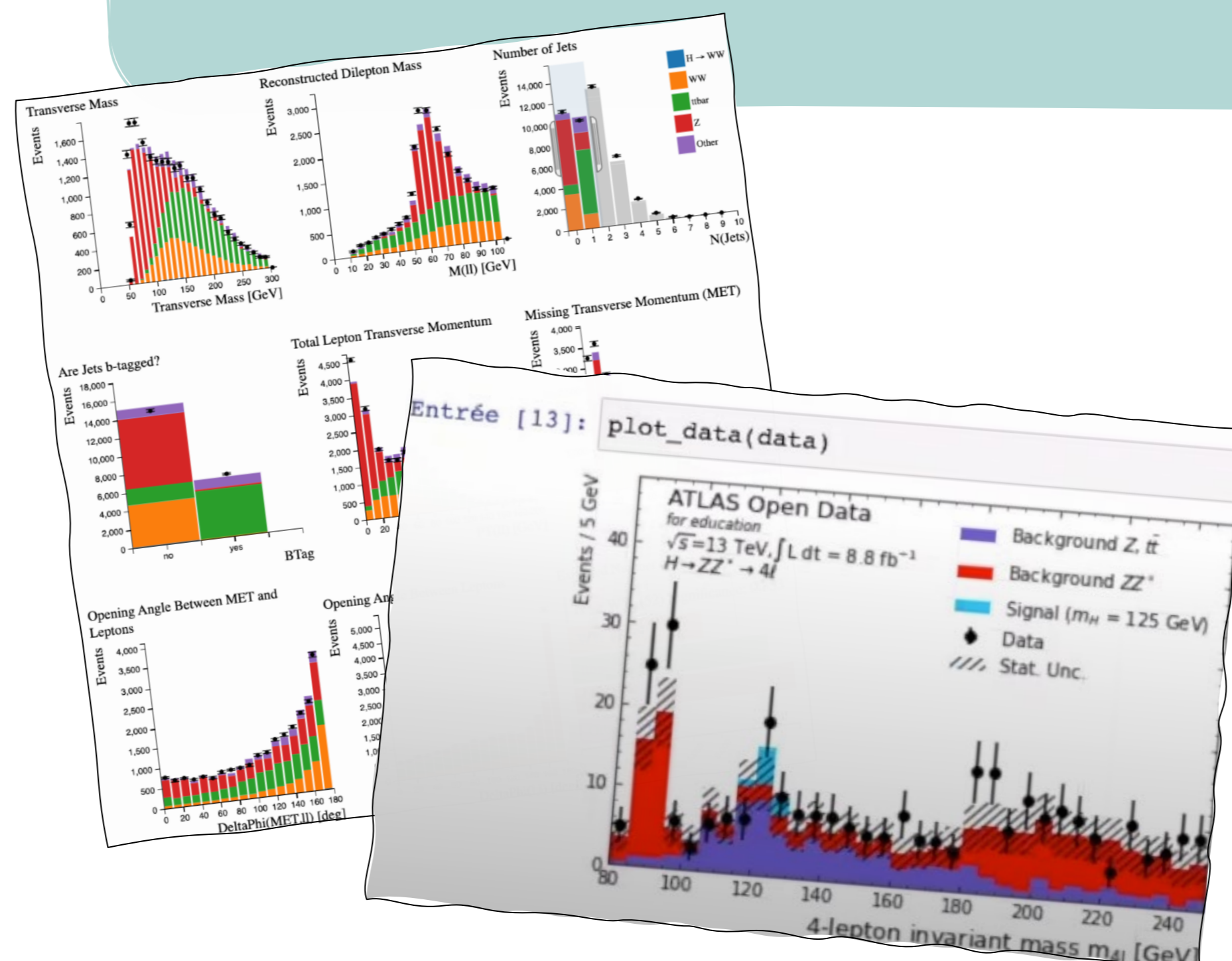
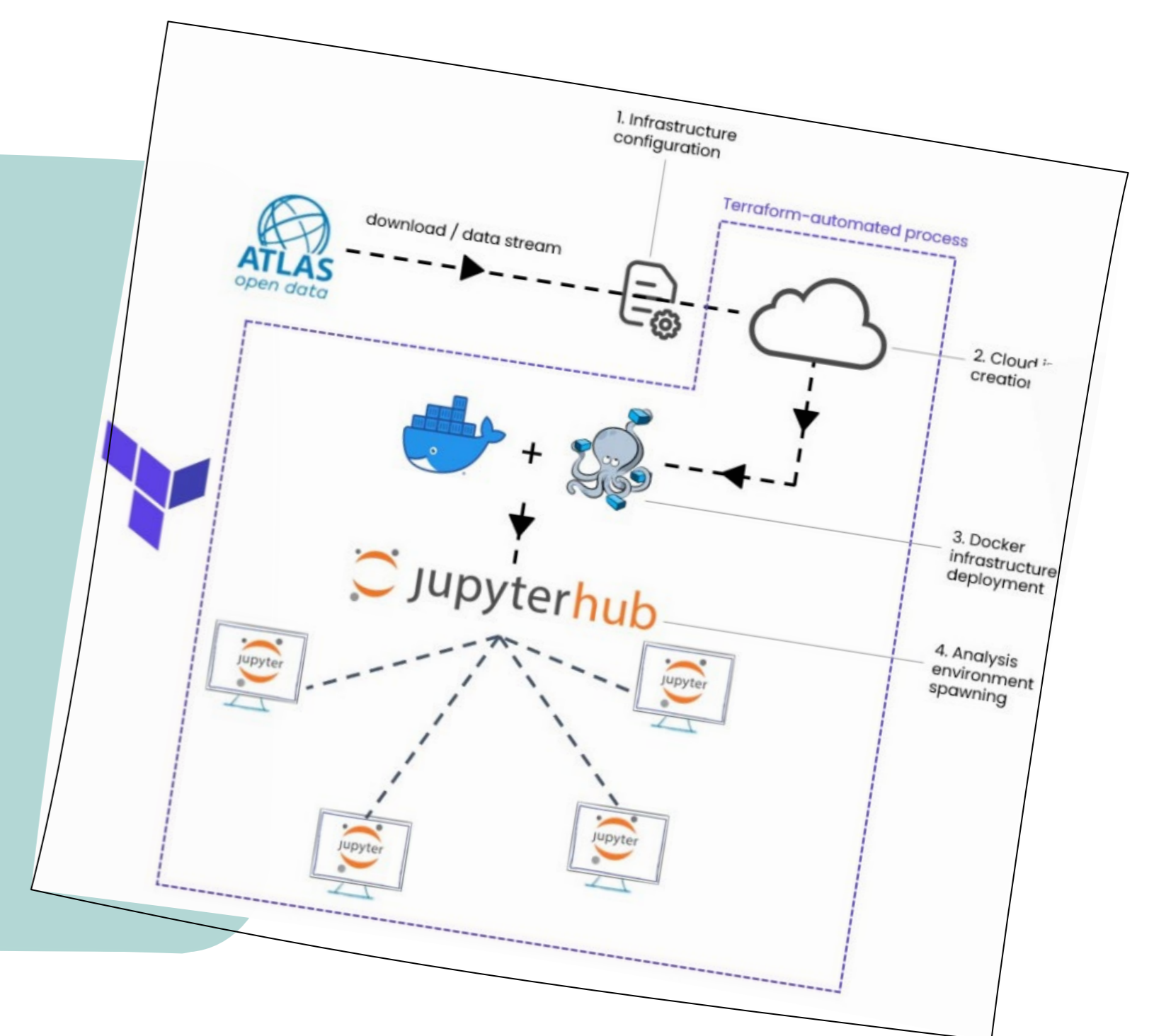
- Masterclass per studenti
  - 225 istituti di ricerca coinvolti
  - 13k studenti in 60+ Paesi
- Challenge per ricercatori:
  - [TrackML](#), [HiggsML](#)
  - [Uncertainty-aware HiggsML](#)



### 6. Risorse

Grande varietà di risorse per accedere ed utilizzare i dati:

- Ispezione istogrammi direttamente su browser
- Visualizzazione event-display
- Jupyter notebooks (Python/C++)
- Docker containers, Virtual Machines, istanze Cloud
- Documentazione online, esempi e tutorial



## About me

RTDa @ Unibo & PhD in Data Science  
 Technical writer @ Towards Data Science  
 Ricerca in ML/AI, Computer Vision & Quantum ML

