

# SMARTHEP

REAL-TIME ANALYSIS FOR  
SCIENCE AND INDUSTRY

An European Training Network

CHACAL school,  
20/01/2024

**Caterina Doglioni,  
University of Manchester  
(coordinator)**

Photos by CD and M. Montella (OSU, SMARTHEP Partner)  
Slide template by D. Wilson-Edwards (PhD student, UofM)



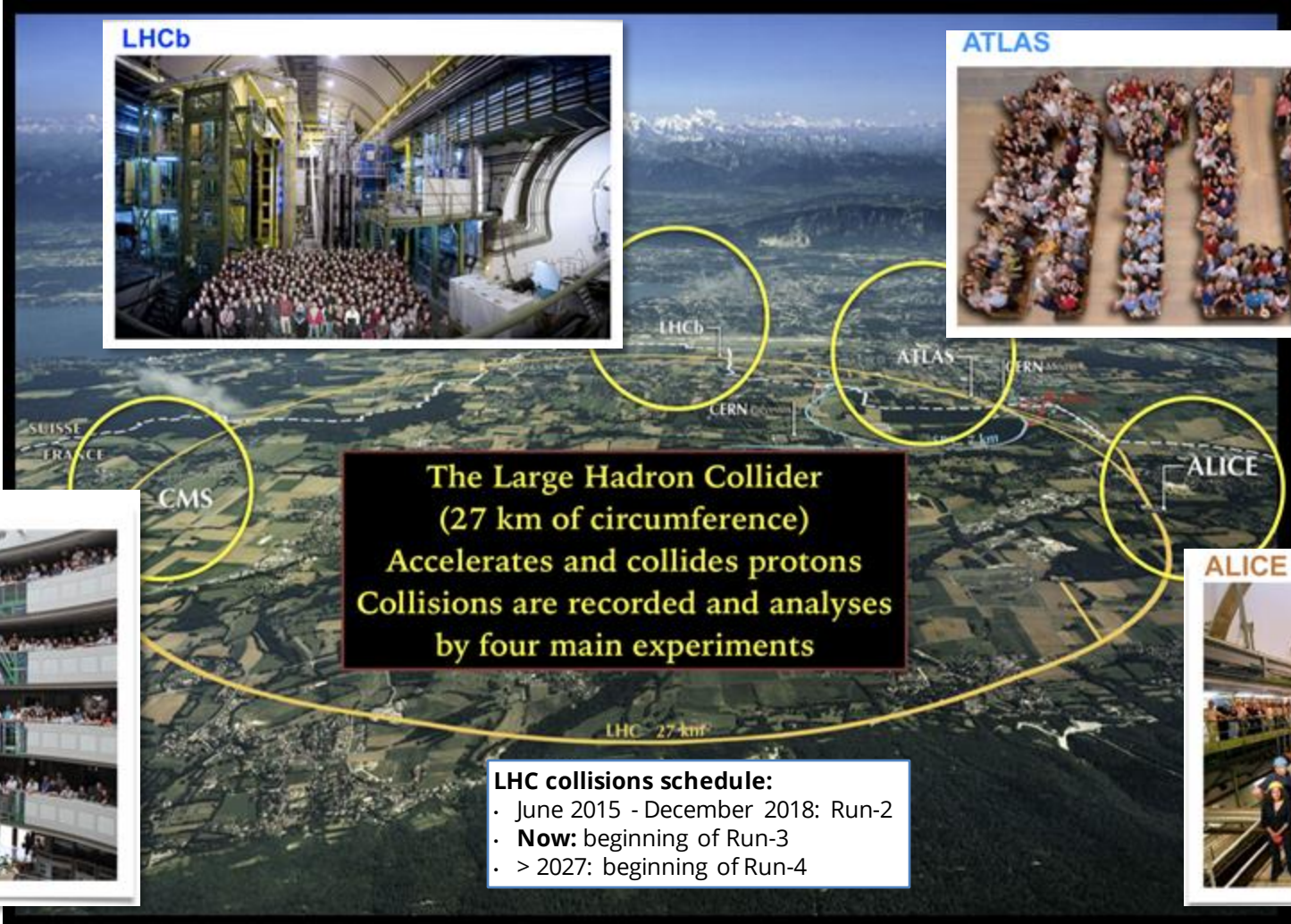
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086



# Outline

- Scientific introduction
- Description of the network and its set-up
- How to find other networks like this

# The Large Hadron Collider and its experiments



**LHCb**

**ATLAS**

**CMS**

**ALICE**

**The Large Hadron Collider**  
(27 km of circumference)  
Accelerates and collides protons  
Collisions are recorded and analysed  
by four main experiments

**LHC collisions schedule:**

- June 2015 - December 2018: Run-2
- **Now:** beginning of Run-3
- > 2027: beginning of Run-4

**LHCb** photo: A large group of people standing in a tunnel.

**ATLAS** photo: A large group of people standing in a tunnel.

**CMS** photo: A large group of people standing in a tunnel.

**ALICE** photo: A large group of people standing in a tunnel.

# Data at the LHC

- The LHC **collides protons** every 25 ns, and can **produce other/new particles**
- **Research goals:** measure fundamental properties of matter, discover new phenomena not included in the current theory
- The **signals** we are looking for are **rare**  
→ need enormous number of collisions to produce them
- Their **backgrounds** often look the same and are **much larger**
- **Problem:** recording all LHC data takes 400000 PB/year [\[Ref\]](#)
  - up to 40 million proton-proton collisions/second (MHz)
  - ~ 1-1.5 MB/data per collision event, including raw data



after selection of "interesting" data

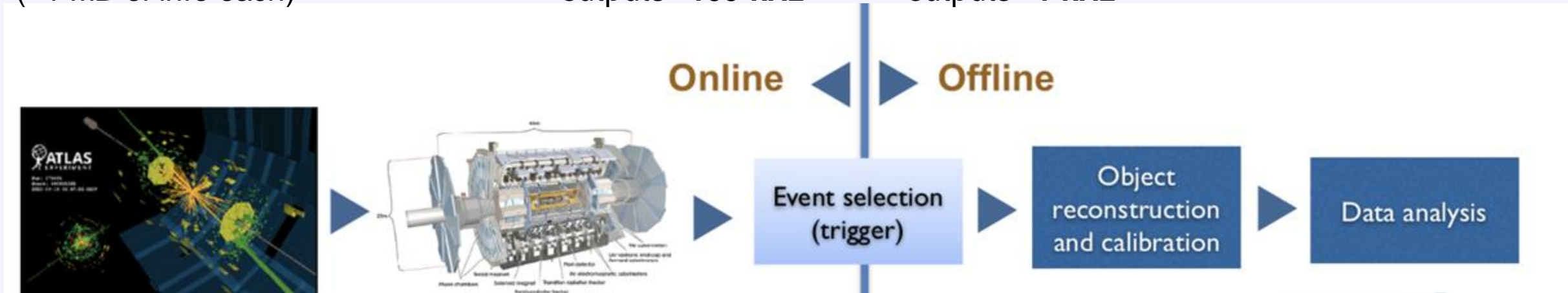
# Triggering on data

LHC experiments have to select "interesting" events (=trigger) in **real-time** (milli/microseconds)  
 Example for the ATLAS trigger:

Collisions at ~40 MHz  
 (~1 MB of info each)

**Hardware trigger**  
 outputs ~**100 kHz**

**Software trigger**  
 outputs ~**1 kHz**



Most data gets discarded...are we discarding some kinds of **signal** as well?

# Main focus of the SMARTHEP project

Traditional data analysis is  
**asynchronous:**

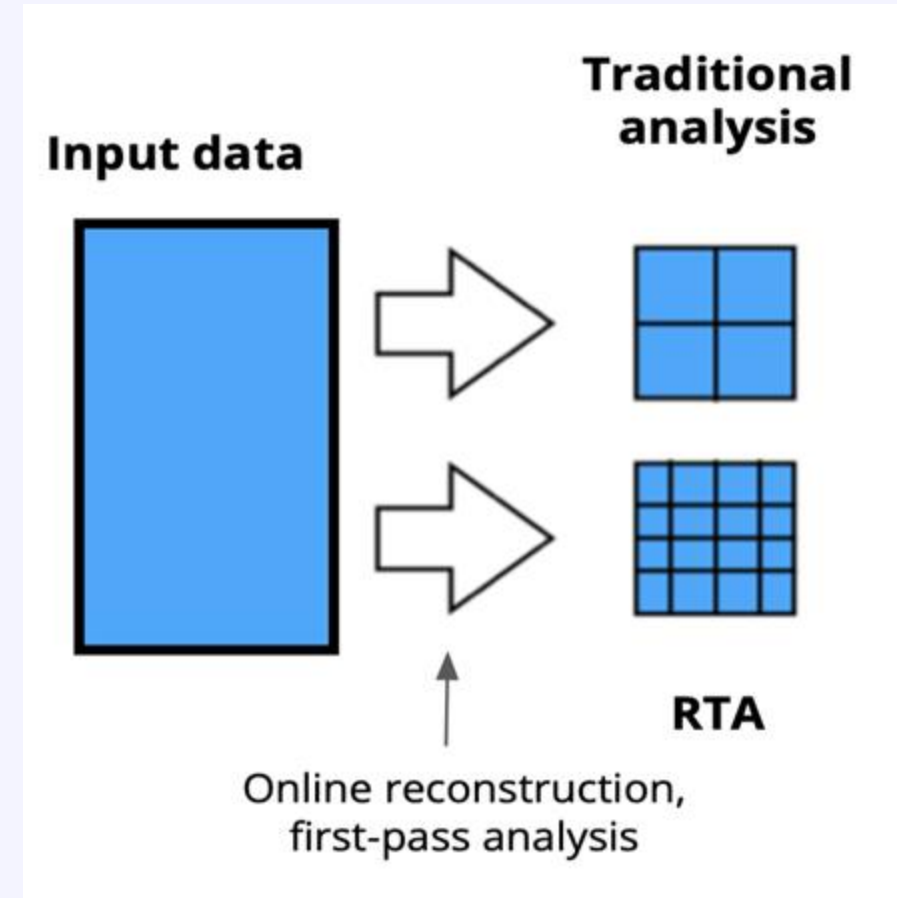
*First record and store data,  
then reconstruct/analyze it*



**Real-time** data analysis

*Analyse data as soon as it is collected*

- only store (smaller) final-state information
- reduce time-to-insight
- accelerate decision making



*only store (smaller) final-state information:  
useful for saving more data in LHC experiments*

# RTA in science and beyond

All four main **LHC experiments** use **Real-Time Analysis** techniques

**ALICE:** [online reconstruction \(O2\)](#) **ATLAS:** [Trigger Level Analysis](#)

**CMS:** [Data Scouting](#) **LHCb:** [Turbo stream](#)

+ the *trigger* system is a real-time decision making system

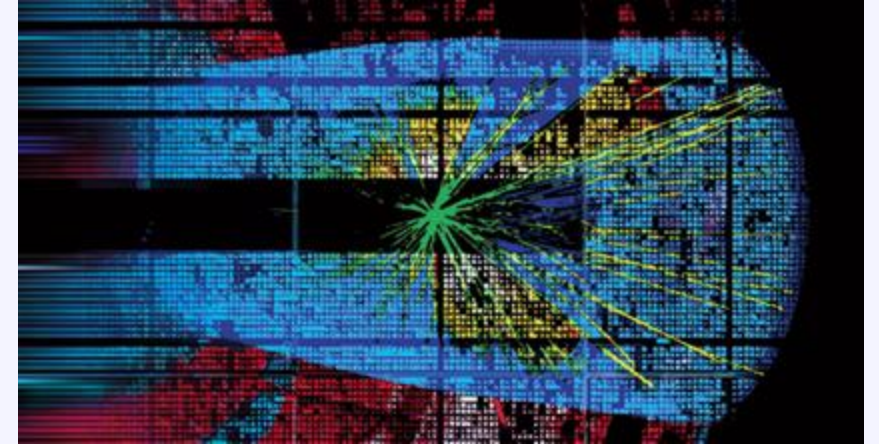
**“Too much data” & “need to analyse data ASAP”** problems  
**not unique** to particle physics

+ use cases in financial transactions, fleet & traffic management, predictive maintenance...



Given these common needs,  
how do we **collaborate** to advance RTA at the LHC and beyond?

# Tools: machine learning



*Artist's impression of an FPGA in the level-one trigger scanning for anomalies. Credit: S.. Summer/CMS-PHO-EVENTS-2021-004-2/M Rayner CERN Courier*

- **Machine learning** is revolutionising high energy physics, industry and society
  - Use of ML is ubiquitous in all of these
- Advantage for RTA: **decisions** based on **large, complex datasets** can be taken on a very **short timescale**
- Particular interest in **unsupervised methods**
  - Algorithms that "learn from the data" (including *rule induction*)
  - Necessary to remove theoretical prejudices on how new physics can look like



# Tools: hybrid computing architectures (accelerators)

- CPU-based architectures (=“computers” as we know them) are not the only option on the market, e.g.:
  - Field Programmable Gate Arrays (FPGA) for fast custom operations
  - Graphical Processing Units (GPUs) for parallel operations
- Advantage for RTA: **hybrid computing architectures** can significantly **accelerate** decision-making

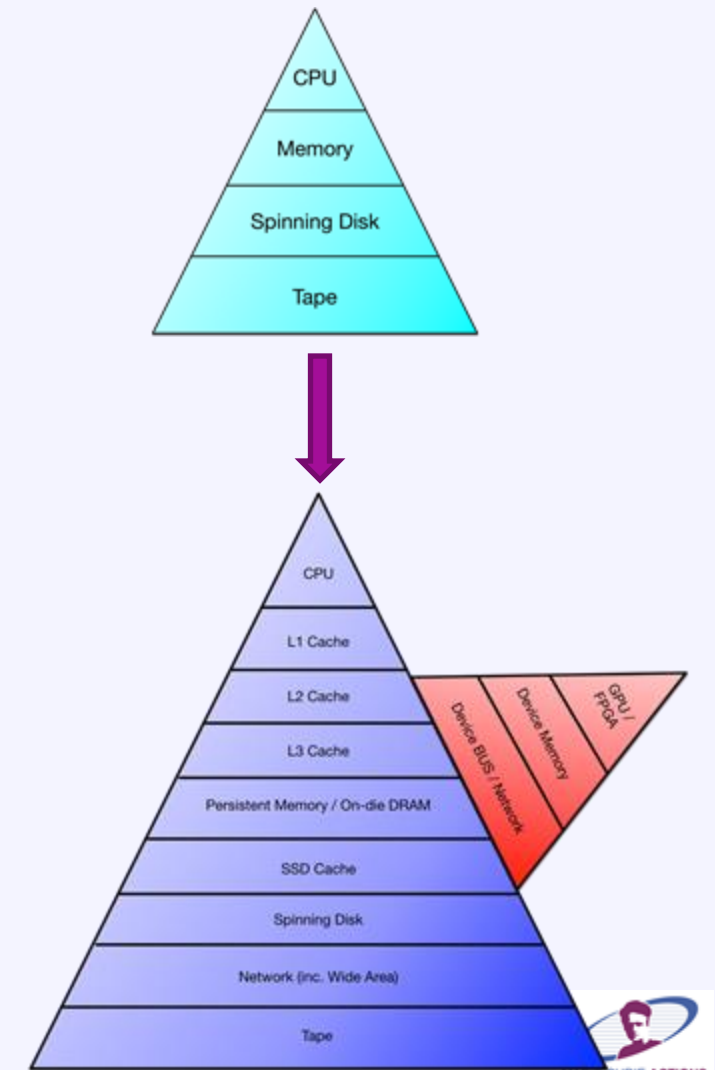


Diagram by G. Stewart

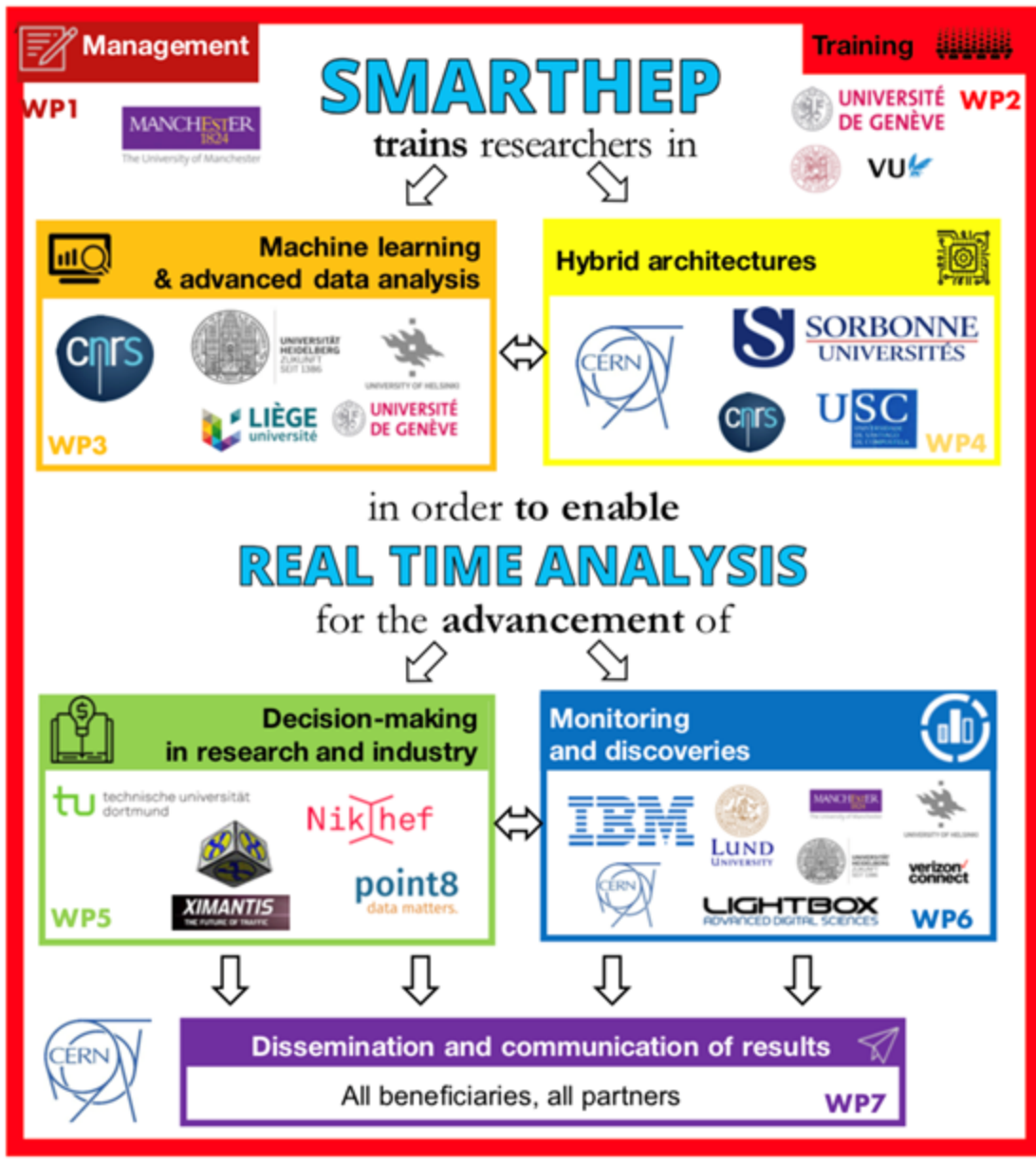


**SMARTHEP**

REAL-TIME ANALYSIS FOR  
SCIENCE AND INDUSTRY

trains **Early Stage Researchers**  
in **Machine Learning**  
and **hybrid computing architectures**  
to advance **real-time analysis**  
in **science and industry**



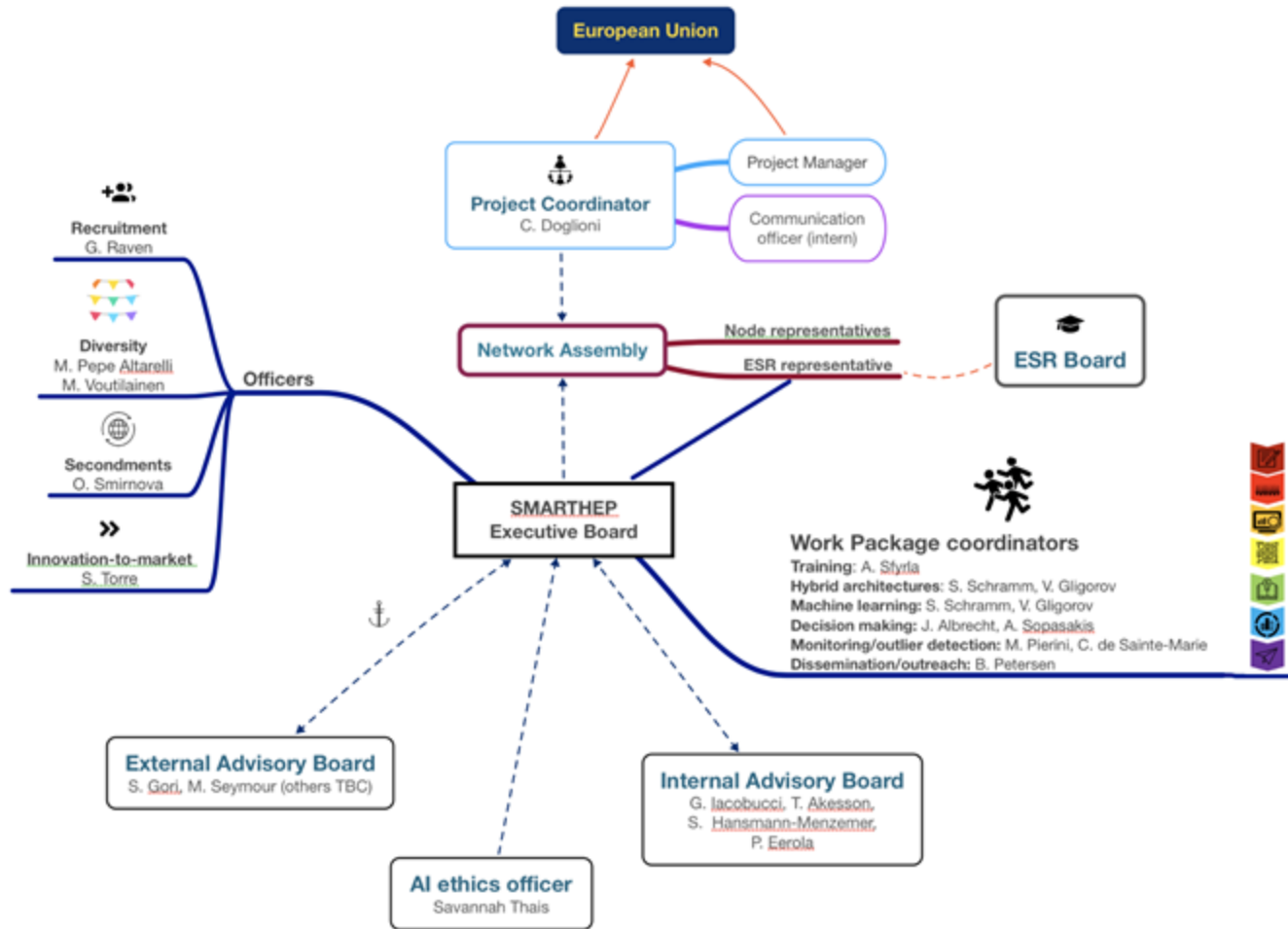


## Synergies between LHC & industry:

- Different use cases
  - Different dataset size/ complexity  
→ Collaborate on **common tools**:
- 1. Machine learning (Work Package 3)**  
→ enables fast and efficient inference
  - 2. Hybrid computing architectures (WP4)**  
→ accelerate RTA w/ FPGA, GPU, multithreading
- Concrete outcomes in **decision-making (WP5)**, **monitoring and discoveries (WP6)**



# Management



## Network officers:

Experts who can advise the network and the ESRs on specific topics (including ethics)

## Work package coordinators:

SMARTHEP supervisors who follow and coordinate the work of each of the topics in the network (work packages also for management, training and dissemination)

## Network assembly:

Decision-making/voting body, includes ESR representation

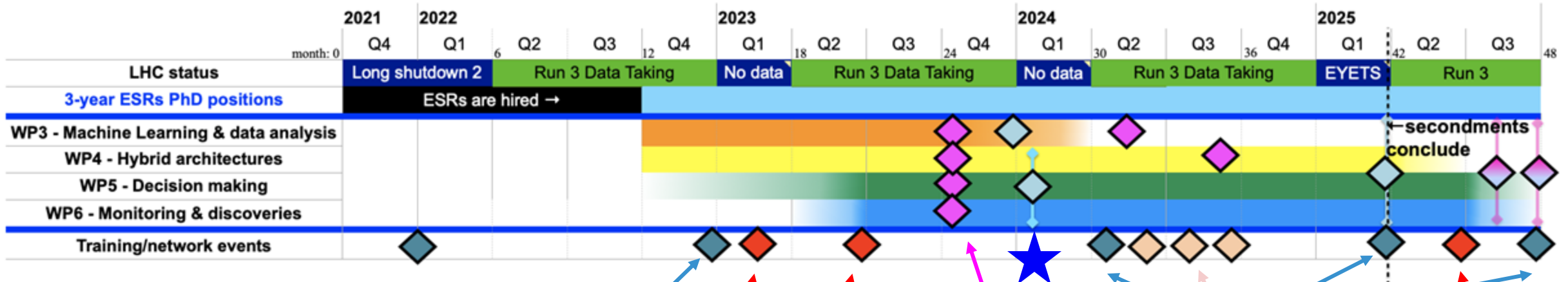
## Executive board (EB):

Unless otherwise specified, EB has open meetings for anyone who wants to help with running the network

# Our 4-year plan

## SMARTHEP Activity Plan

◆ Main research milestones   
 ◆ Main commercial milestones   
 ◆ Network-organized schools (all ESR)   
 ◆ Network-organized schools (ESR choose 1)   
 ◆ Network events and conferences



Virtual kick-off meeting

In-person kick-off meeting

Mtd-term check with the EU Project Officer

UniGe ML + Physics school

Advanced ML school (postponed)

Whitepapers on state-of-the art [further research/commercial milestones: collections of papers and algorithms from individual projects]

We are here

Yearly network meetings

Accelerator boot-camps Being organized, may be able to open to external students

Commercial applications school



## Sample physics outcomes

- Calibration of ALICE TPC for heavy-ion physics
- Improvements & optimization of the experiment trigger systems for Run-3 and High-Luminosity LHC
- Data analysis with real-time analysis workflows, e.g.
  - New resonance searches
  - Lepton flavour violation analyses
  - New physics searches

*Check out the [ESR](#) talks from our December meeting*



## Sample industry outcomes

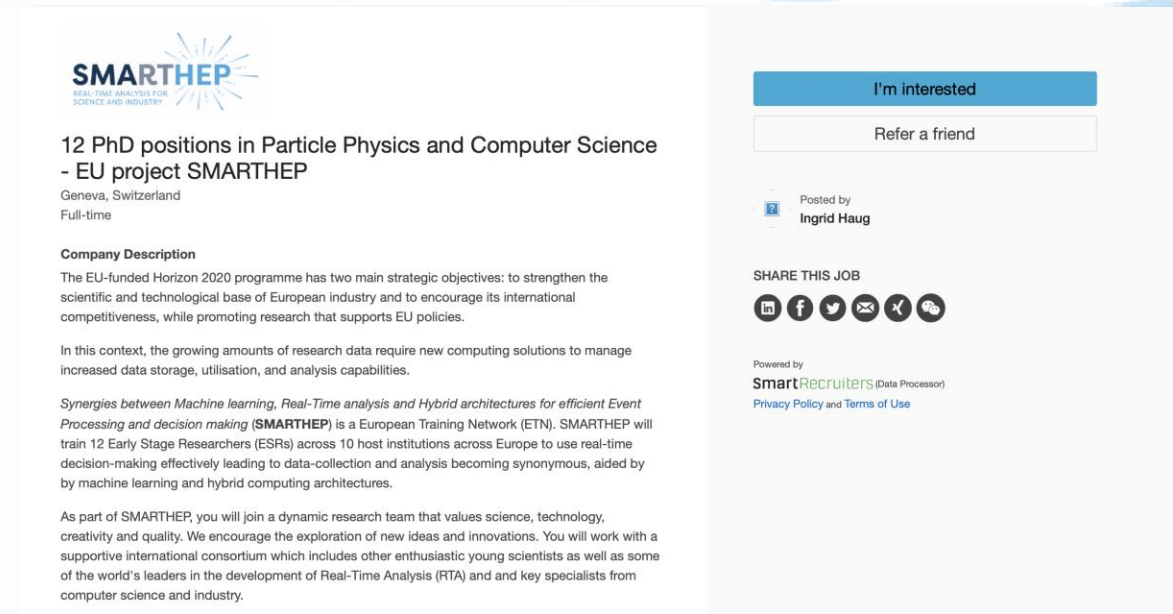
- Algorithms for real-time traffic prediction (Ximantis)
- Real-time analysis of videos and sensor data collected by dashcams (camera on vehicle, Verizon Connect)
  - Running fast analysis in embedded (on-car) system
- Automating decision-making for fraud detection (IBM)
- Analysis of time series for suggesting next-step financial decisions to investors (Lightbox)

*Check out the [ESR talks](#) from our December meeting*



# How did we recruit students?

- Network-wide candidate selection, organized by CERN HR
  - Eligibility checks are important for this kind of network, e.g. you shouldn't get a job in the same country as your previous career step
- Procedure for selection agreed upon in network meetings:
  - Each institute made a shortlist of 5-10 candidates, then proceeded to interview
  - Network-wide recruitment meeting to disclose results and make offers
    - If two institutes wanted to make offer to the same candidate, candidate's preference would be used for decision
    - If institute ran out of candidates, other beneficiaries suggested other eligible candidates



**SMARTHEP**  
REAL-TIME ANALYSIS FOR SCIENCE AND INDUSTRY

**12 PhD positions in Particle Physics and Computer Science - EU project SMARTHEP**  
Geneva, Switzerland  
Full-time

**Company Description**  
The EU-funded Horizon 2020 programme has two main strategic objectives: to strengthen the scientific and technological base of European industry and to encourage its international competitiveness, while promoting research that supports EU policies.

In this context, the growing amounts of research data require new computing solutions to manage increased data storage, utilisation, and analysis capabilities.

*Synergies between Machine learning, Real-Time analysis and Hybrid architectures for efficient Event Processing and decision making (SMARTHEP) is a European Training Network (ETN). SMARTHEP will train 12 Early Stage Researchers (ESRs) across 10 host institutions across Europe to use real-time decision-making effectively leading to data-collection and analysis becoming synonymous, aided by machine learning and hybrid computing architectures.*

As part of SMARTHEP, you will join a dynamic research team that values science, technology, creativity and quality. We encourage the exploration of new ideas and innovations. You will work with a supportive international consortium which includes other enthusiastic young scientists as well as some of the world's leaders in the development of Real-Time Analysis (RTA) and key specialists from computer science and industry.

**I'm interested**

Refer a friend

Posted by  
Ingrid Haug

SHARE THIS JOB

Powered by  
**SmartRecruiters** (Data Processor)  
Privacy Policy and Terms of Use



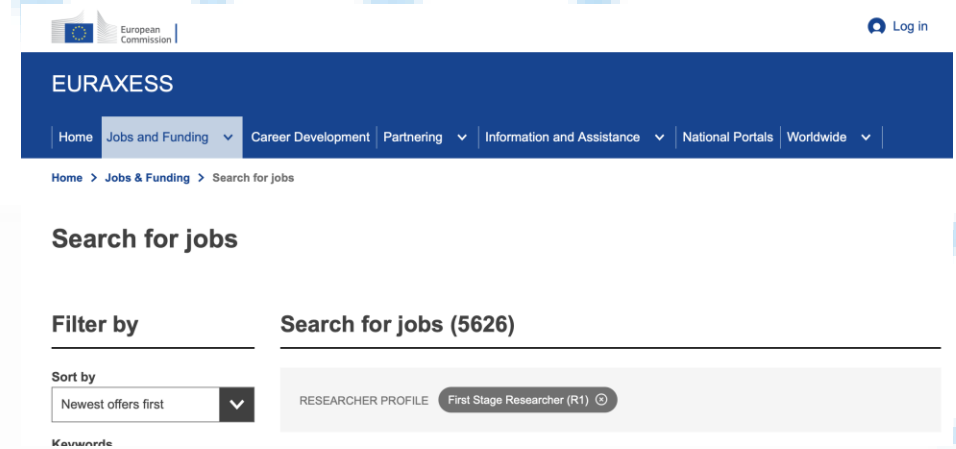
# Recruitment statistics

- Post advertised widely
  - Euraxess, external recruitment websites, social media, experiments mailing lists, beneficiary websites
- 199 applications total
  - 55 different nationalities
  - About 140 passed eligibility checks
- Selection procedure anonymized (as much as possible) for gender and age
  - Target for women ESRs: 30%, hired women ESRs: 25%
- ESRs are from India, Italy, Switzerland, Denmark, Sweden, Greece, Spain, Thailand, UK, Colombia
- Personal Career Development Plans for each student (think about whether if you would also like to do this, then ask your supervisor!)



# How do I find other networks like this?

- Since SMARTHEP, this kind of networks changed name
  - From **European Training Network to Doctoral Network**
- Calls still published on the EURAXESS website
  - <https://euraxess.ec.europa.eu/jobs/search>
- Note: PhD / postdoc positions also available on INSPIRE-HEP
  - <https://inspirehep.net/jobs?sort=mostrecent&size=25&page=1&status=open> – use search bar
- Also: ask Jon and Andy about the MCNet Collaboration!



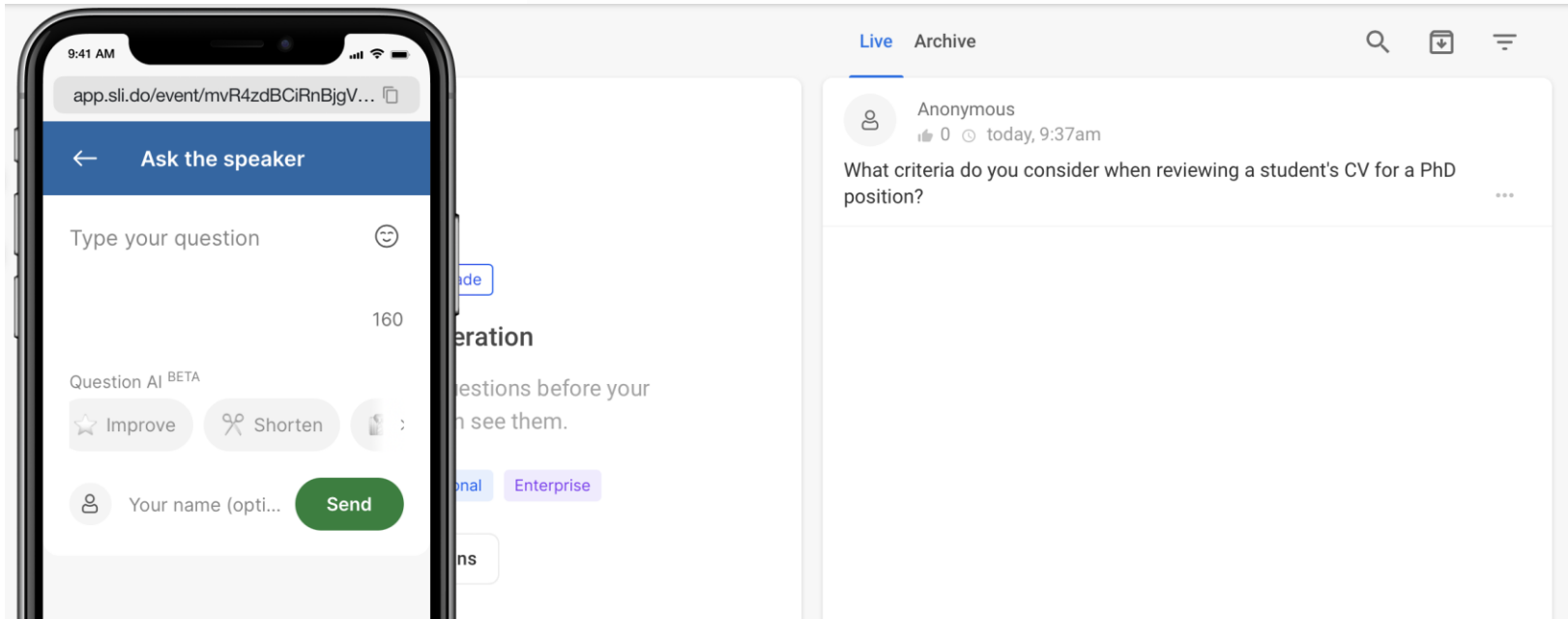
*Unsolicited advice for job seeking (more in panel discussion): whatever job you end up looking for/doing, be yourself and have fun*



# Could the lecturers give me some advice on...*anything you'd like*

Raise your hand, or add a question to slido.com → event code 9876543210

Or click this link: <https://app.sli.do/event/mvR4zdBCiRnBjgVtgh9a78>



Or use this QR code:



# SMARTHEP

REAL-TIME ANALYSIS FOR  
SCIENCE AND INDUSTRY



## Backup slides



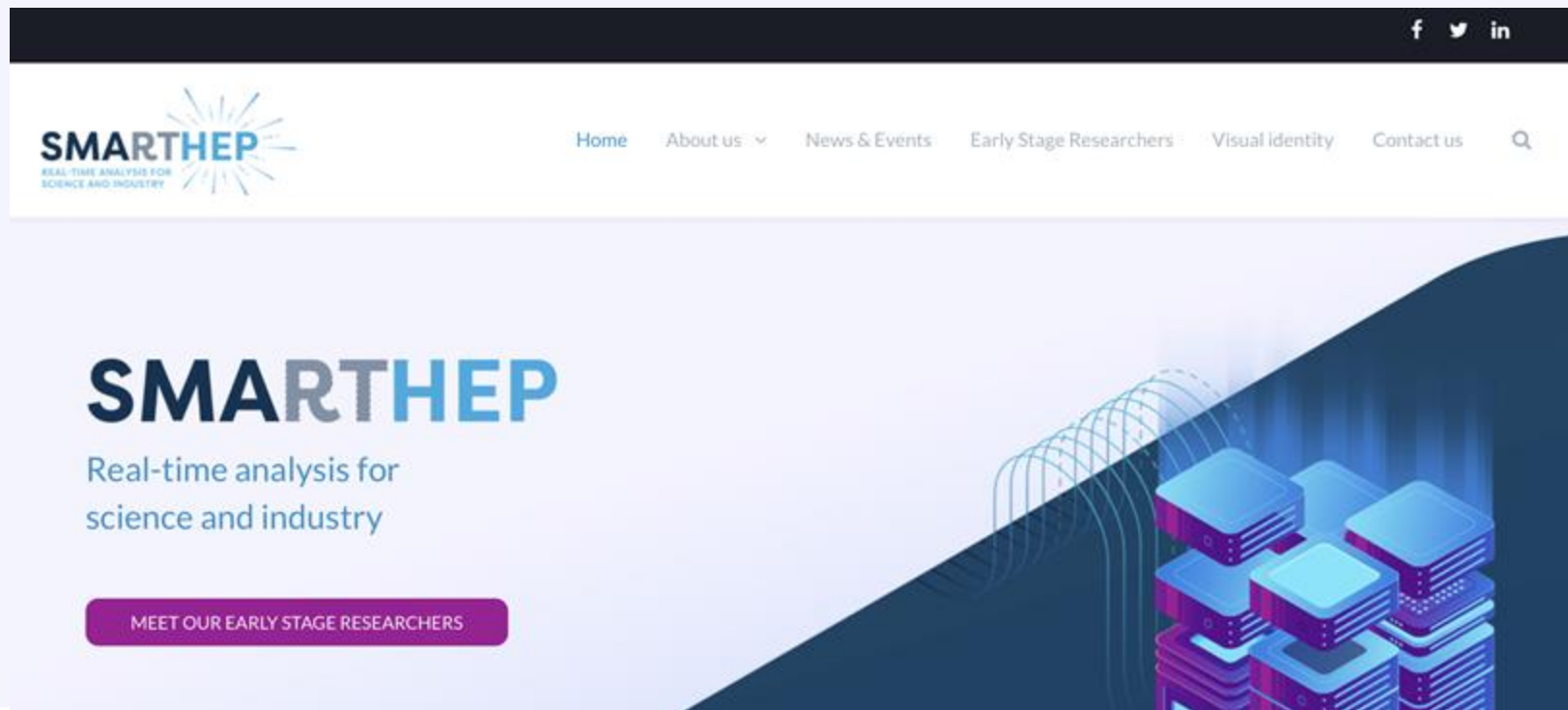


# The SMARTHEP website

<https://smartheep.org>

Developed in collaboration with Nectar UK  
(who also provided logo and marketing strategy)

**Went online in January 2022, main source of information for recruitment**



SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086



# Open and FAIR data

Findable, Accessible, Interoperable, and Reusable digital assets.

**With SMARTHEP, we intend to be part of the effort to improve data *FAIR-ness* and *reproducibility & sustainability* of software**

SMARTHEP has a **Data Management Plan**

- Data from LHC experiments follows the CERN Open Data policy (FAIR)
- Individual discussions involving Ethics Officer for industry beneficiaries and partners

Software discussions ongoing in high energy physics within the **HEP software Foundation** (see <https://hepsoftwarefoundation.org>)

# Kick-off meeting: Code Of Conduct

- Discussed and agreed upon a code of conduct
  - Based on CoC from University of Helsinki / Kumpula Campus
  - Complementing the European Code of Conduct for Research Integrity and the MSCA Researchers Rights and Obligations
  - Reflecting and pledging conduct in terms of:
    - Truth and knowledge
    - Autonomy
    - Creativity
    - Critical Mind
    - Edification
    - Well-being

## **SMARTHEP Network Code of Conduct**

### **Truth and Knowledge**

- We are guided in our actions by our core values of truth and knowledge, autonomy, creativity, critical mind, edification and wellbeing.
- We take as a starting point for our research, teaching, learning and other activities the pursuit of truth and new knowledge.
- We respect and value difference.
- We are open to new ideas and approaches.
- We structure our efforts so that others can get involved, and continue or extend our work.
- We do not deceive others, whether by unintentional omission or by deliberate act.
- We respect the privacy of others, and the confidentiality of information, documents and data.
- We do not commit plagiarism, or misinterpret or falsify data.

### **Autonomy**

- We recognise that our behaviour may reflect upon the reputation of the SMARTHEP Network.
- We respect the limited human, financial and material resources available to the Network community.
- We advise and guide each other where appropriate.
- We exercise adequate supervision when in a position of authority, or when delegating tasks, avoiding excessive workloads.
- We do not abuse our authority, position or power to obtain special treatment or undue influence for ourselves or others.
- We are familiar with, and follow, all relevant rules and regulations.
- We strive to avoid conflicts of interest, whether real or perceived, and disclose them otherwise.

### **Creativity**

- We are open to new ideas and approaches.
- We value all areas of academic endeavour equally highly.
- We keep up-to-date with developments that affect our work, studies or research.
- We apply our learning, skills and professional experience constructively for the benefit of all.
- We share any knowledge that could benefit each other in our work or studies.
- We adopt alternative approaches in order to generate new thoughts and concepts.
- We give credit to others for their contributions.


# Potentially useful material


- **UniGe First Collider Physics and Machine Learning school** [[agenda](#)]
  - Particle physics (theory: T. Sjöstrand, experiment: A. Sfyrla) + ML with hands-on component (M. Pierini)
  - Seminars (+ apero) on multimessenger astronomy & CERN experimental programme, slides available
- Presentation course by Alice Ohlson from the 2023 Yearly Meeting [[link](#)]
- More resources on C++/Git:
  - Step by step course on C++: <https://www.learncpp.com> (suggestion: use an adblocker)
  - Step by step course on Git: <https://swcarpentry.github.io/git-novice/>


**SMARTHEP**  
REAL-TIME ANALYSIS FOR  
SCIENCE AND INDUSTRY

First Network School on  
**Collider Physics and  
Machine Learning**

10-13 January 2023  
University of Geneva  
[https://indico.cern.ch/e/SMARTHEP\\_SchoolOne](https://indico.cern.ch/e/SMARTHEP_SchoolOne)

 UNIVERSITÉ  
DE GENÈVE  
FACULTÉ DES SCIENCES





We acknowledge funding from the European Union Horizon 2020 research and innovation programme call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

