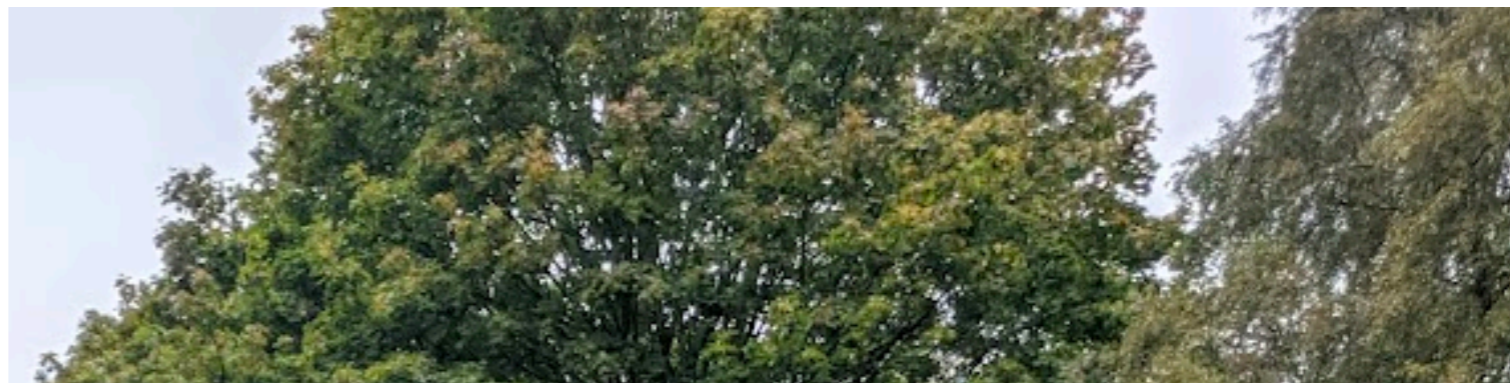




The University of Manchester



LUNDS  
UNIVERSITET



# The SMARTHEP Network

## SMARTHEP kick-off, Manchester, 22/11/22

CATERINA DOGLIONI - UNIVERSITY OF MANCHESTER & LUND UNIVERSITY

[@CATDOGLUOFM, SHE/HER](#)



# SMARTHEP

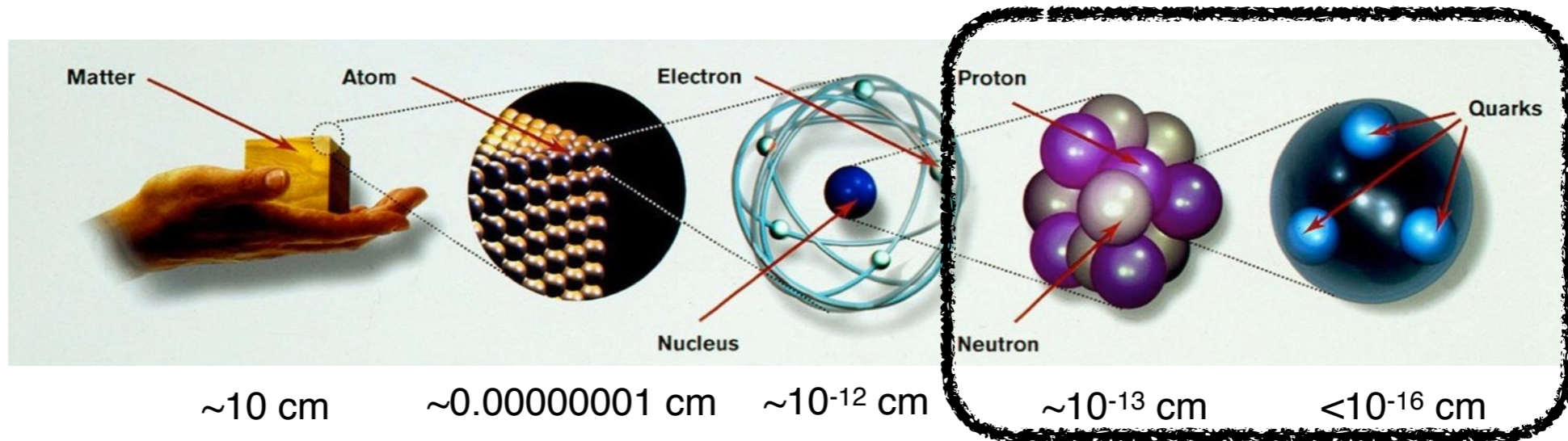
REAL-TIME ANALYSIS FOR  
SCIENCE AND INDUSTRY

[www.smarthep.org](http://www.smarthep.org)

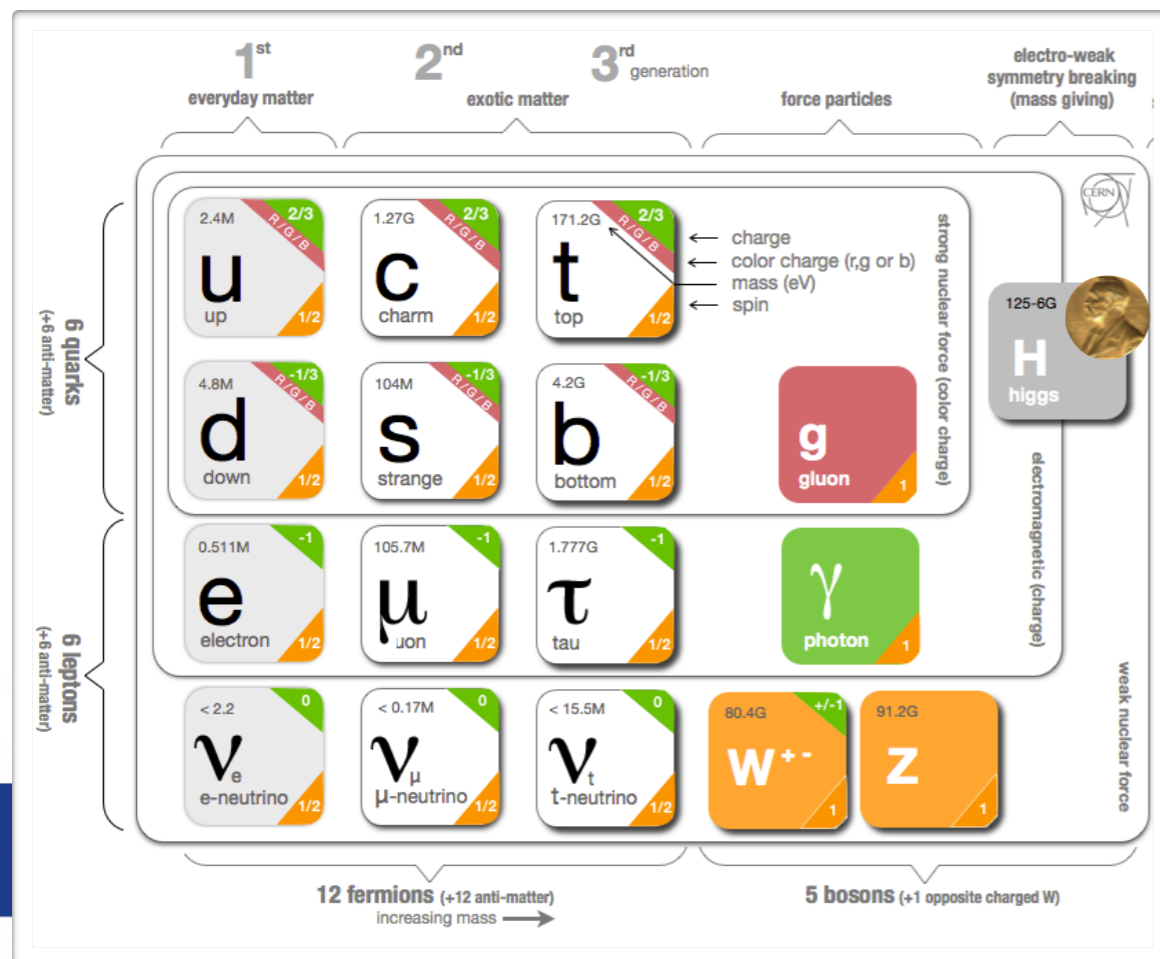
SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086



# Particle physics: a 1-slide intro



Particle physics: study of the nature and behaviour of the smallest constituents of matter



## Standard Model of Particle Physics:

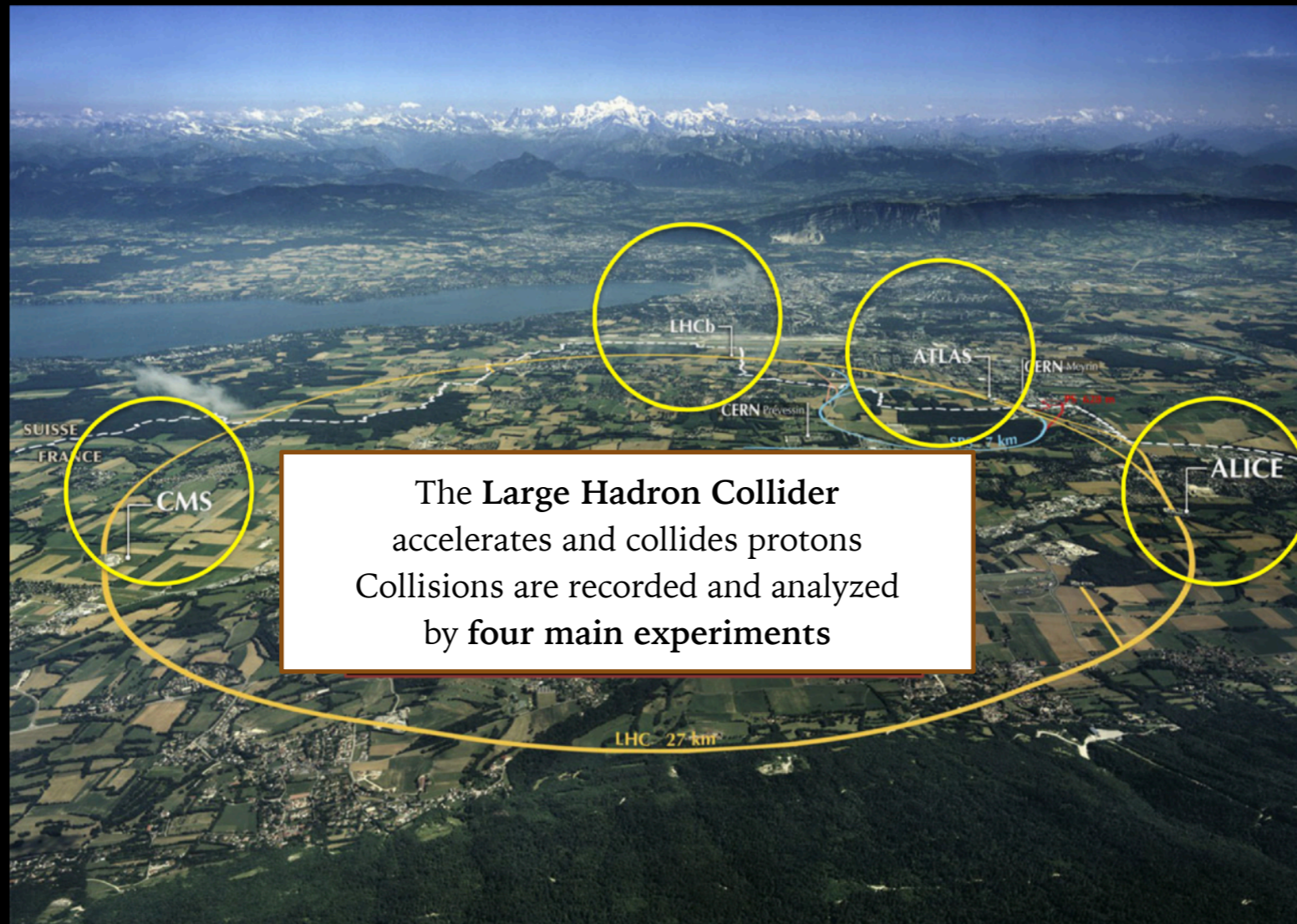
Description of all the known matter in terms of **fundamental particles and forces**

relatively new **discovery: Higgs boson**

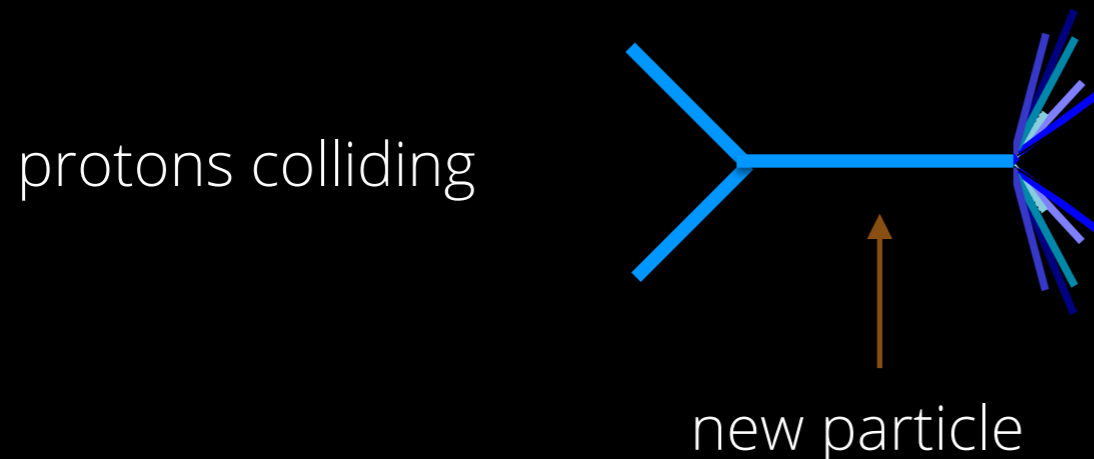
There are also unexplained observations in our universe...like **Dark Matter**



# Scientific tool: the Large Hadron Collider at CERN



Proton-proton collisions at the LHC can e.g. produce new particles



collision outcome, visible in the experiments

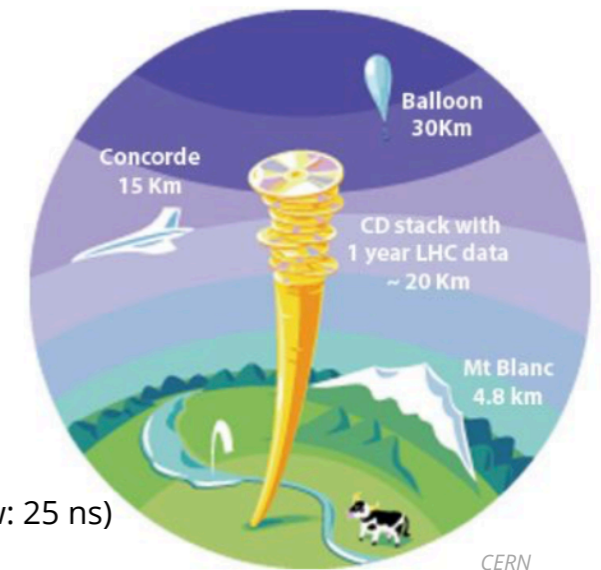
# Video: LHC collisions and data selection

CERN-MOVIE-2013-041-001



# A “Big Science” problem to solve: *too much data*

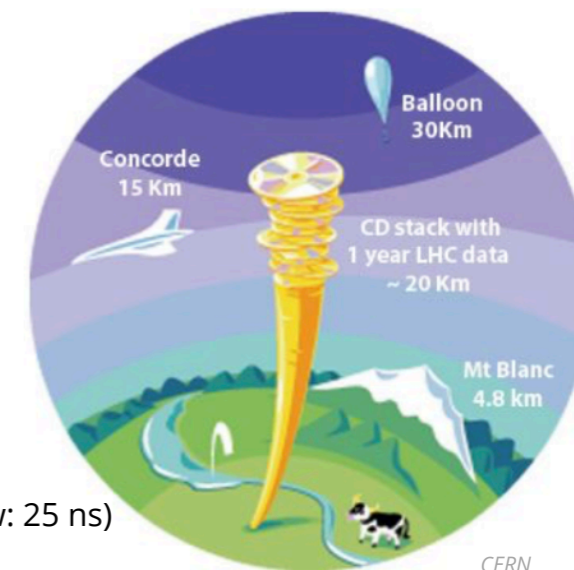
- The **new physics signals** we are looking for are **rare**  
→ need enormous amount of collisions to produce them
- Their **backgrounds** often look the same and are **much larger**
- **Problem:** recording all LHC data takes 400000 PB/year [Ref]
  - up to 30 million proton-proton collisions/second (MHz)
  - ~ 1-1.5 MB/data per collision event, including raw data
- Future hadron colliders plan to collide beams up to every 5 ns (now: 25 ns)
  - and Moore’s law / storage costs don’t scale as fast as that yet



after selection of “interesting” data

# A “Big Science” problem to solve: *too much data*

- The **new physics signals** we are looking for are **rare**  
→ need enormous amount 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 30 million proton-proton collisions/second (MHz)
  - ~ 1-1.5 MB/data per collision event, including raw data
- Future hadron colliders plan to collide beams up to every 5 ns (now: 25 ns)
  - and Moore’s law / storage costs don’t scale as fast as that yet



LHC & future hadron collider experiments need to select “interesting” events (=trigger) in real-time (milli/microseconds)

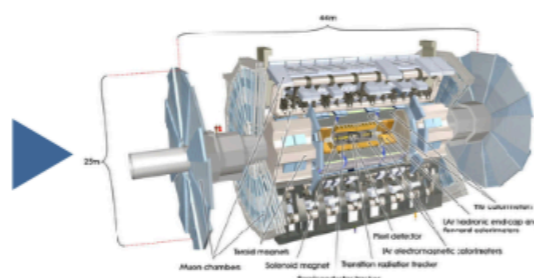
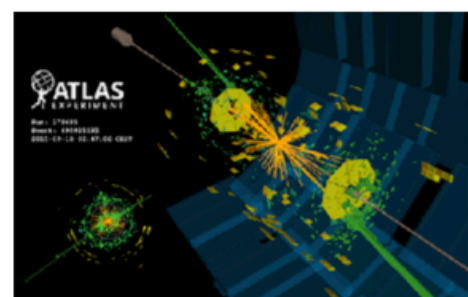
after selection of “interesting” data

Collisions at ~30 MHz  
(~1 MB of info each)

Hardware trigger  
outputs ~100 kHz

Software trigger  
outputs ~1 kHz

Online ← → Offline

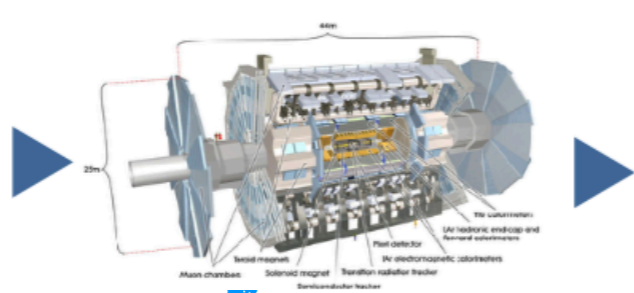
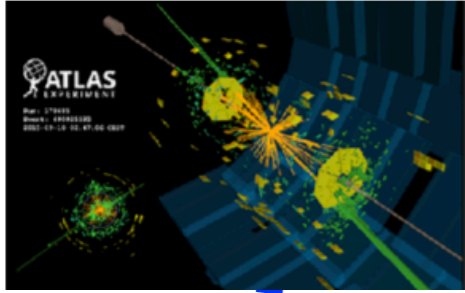


Event selection  
(trigger)

Object  
reconstruction  
and calibration

Data analysis





Online ← | → Offline

Event selection (trigger)

Object reconstruction and calibration

Data analysis

Data I like

...maybe discoveries?

*Data I would have used to like if I had been a data hipster*

Data you like

What we can record in full after the trigger

Data ~nobody likes

Data produced by the LHC

(multiplied by large number)



...so are we missing something?

*What we can trigger on and record*



# Solution: do analysis in real-time

Traditional data analysis is asynchronous:

First record and store data,  
then reconstruct/analyze it



Real-time data analysis

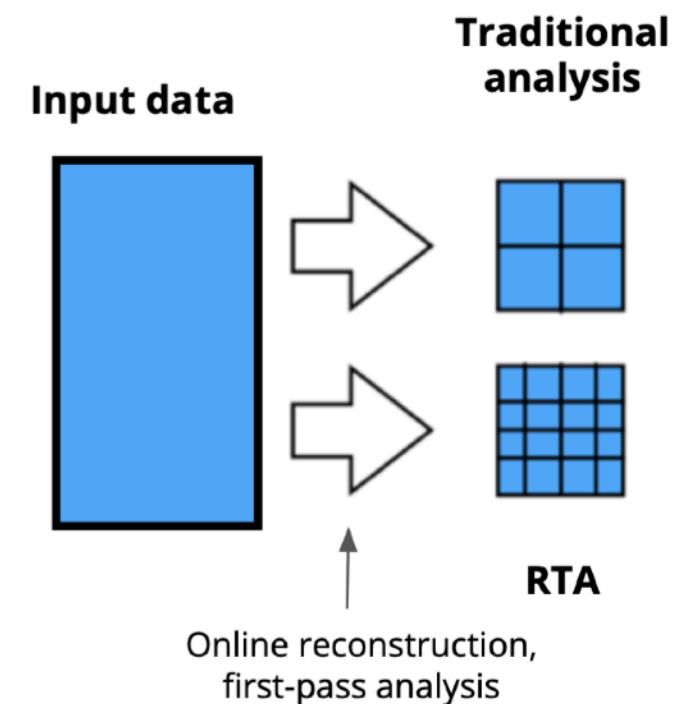
Reconstruct/analyse data as soon as it is taken, so that:  
only (smaller) final-state information needs stored

ALICE: [online reconstruction \(O2\)](#) ATLAS: [Trigger Level Analysis](#) (2 ERC StG, 1 ERC CoG)

CMS: [Data Scouting](#) (1 ERC CoG), LHCb: [Turbo stream](#) (2 ERC CoG)

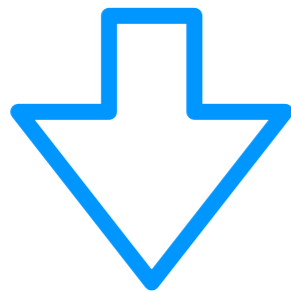
time-to-insight is reduced (e.g. with outlier detection)

*Beyond-state-of-the-art tools necessary:  
machine learning, hybrid computing architectures*



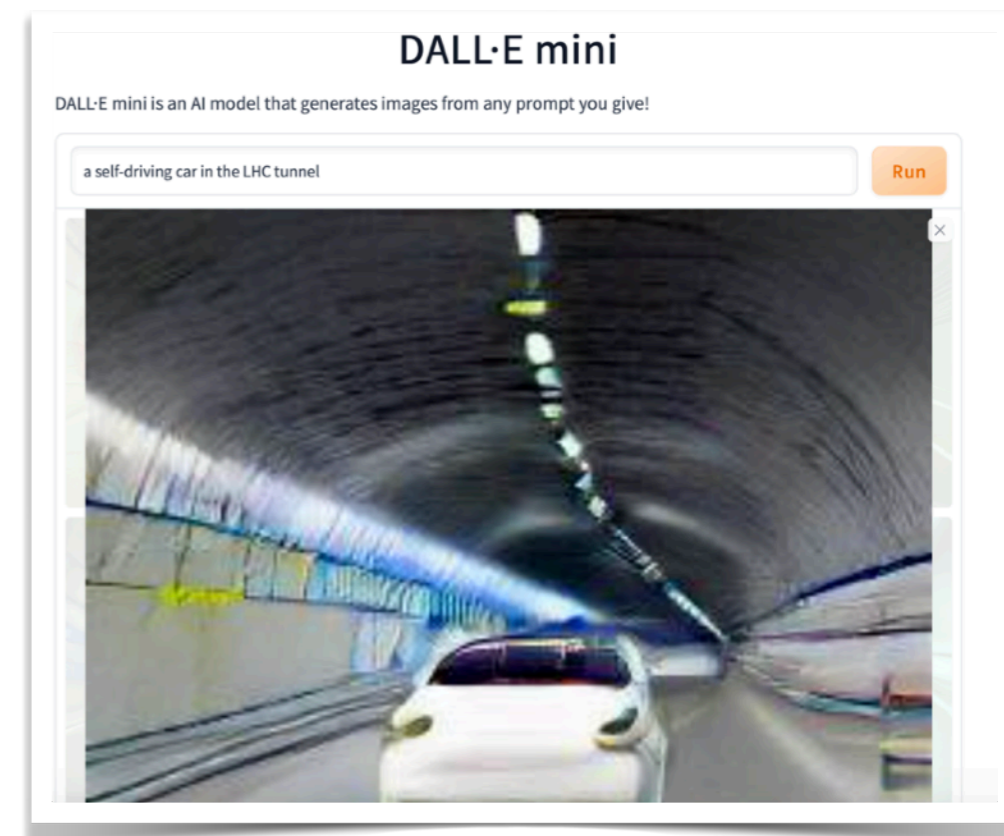
# The need for real-time analysis isn't unique to HEP

- Data is nowadays very **cheap to record**
- Best use of the information in data if analysis happens right away  
⇒ **time-to-insight** is the key metric



**Real-time analysis is a solution for industry as well:**

- Monitoring of financial transactions
- Car fleets/public transport/traffic monitoring
- Self-driving cars
- Predictive maintenance of machinery



# Our industrial beneficiaries & partners



IBM  
Research




verizon  
connect

## Mutual benefit:

- to advance real-time analysis further, research needs tools and infrastructure beyond the state-of-the art in hybrid computing architectures & machine learning
- industry often pioneers these, then can be tested on LHC data



# SMARTHEP goals

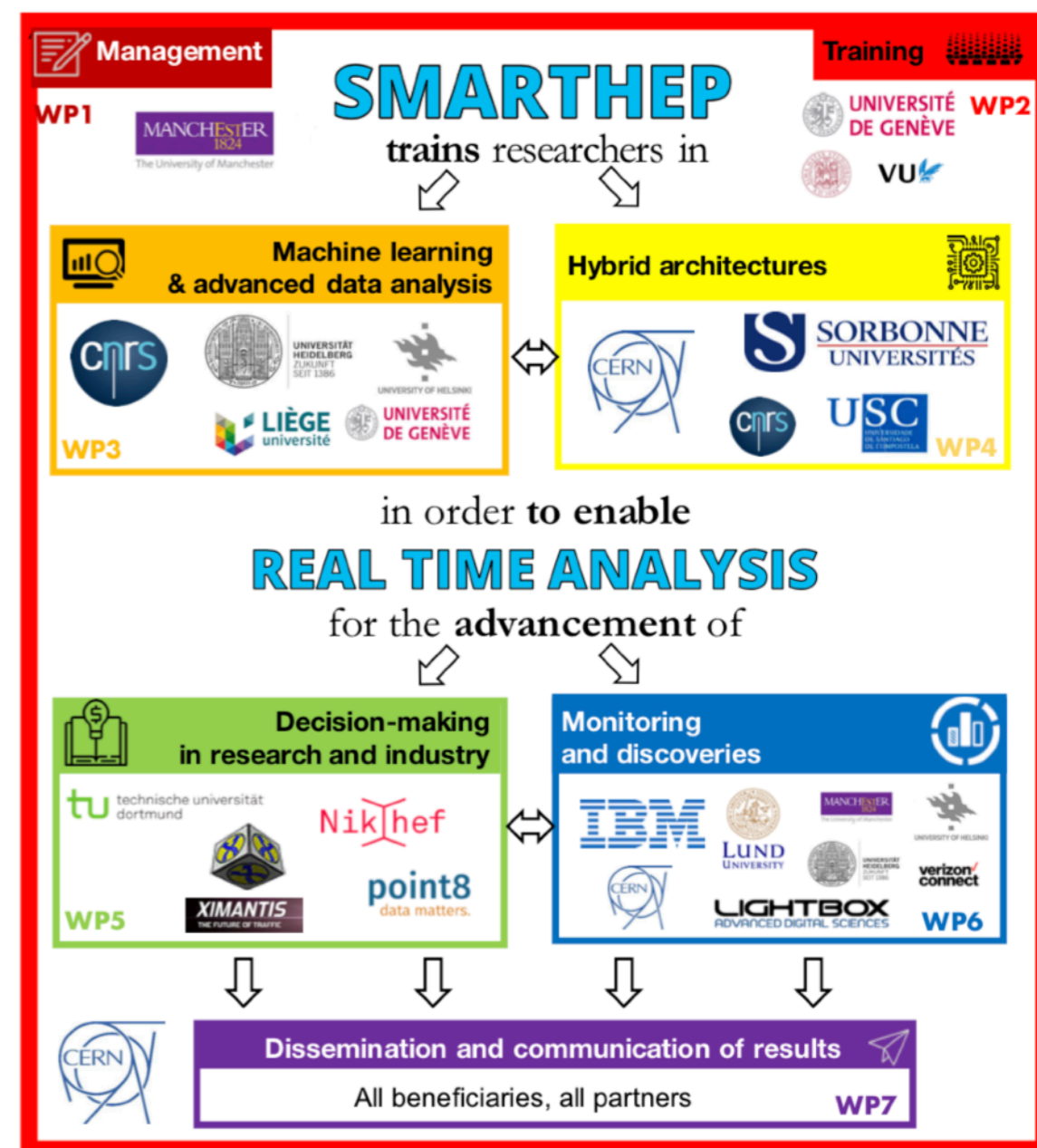
- **Training** 12 PhD students from LHC experiment, computer science and industry
- **Advancing real-time analysis and decision-making in high energy physics and industry, using machine learning and hybrid computing architectures**

## Shared supervision:

- mixed academia/industry supervision
- secondments at universities / labs / industries

## Shared tools and competences:

- hardware (FPGA, GPU) & software
- **machine learning**



MANCHESTER  
1824

The University of Manchester



LUNDS  
UNIVERSITET

M. Pierini - ML @ LHC



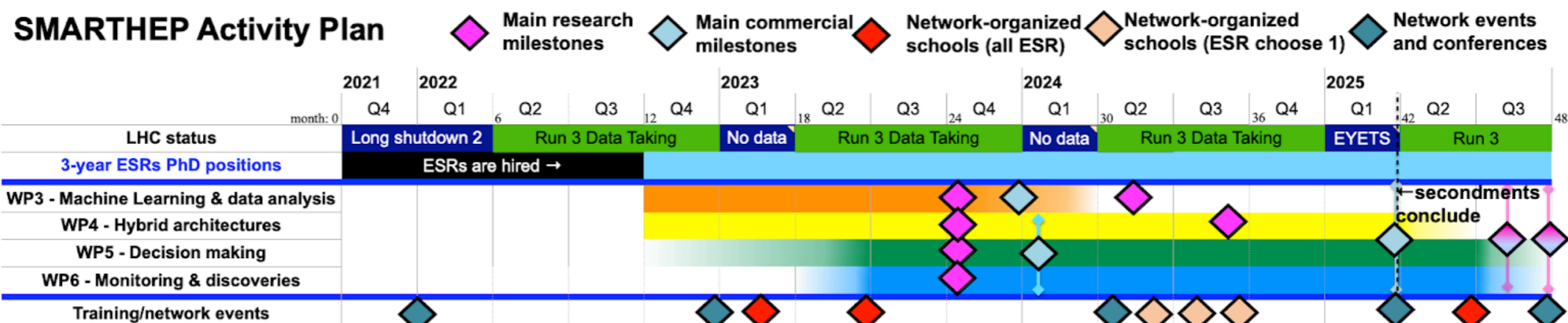


Thanks for your attention!

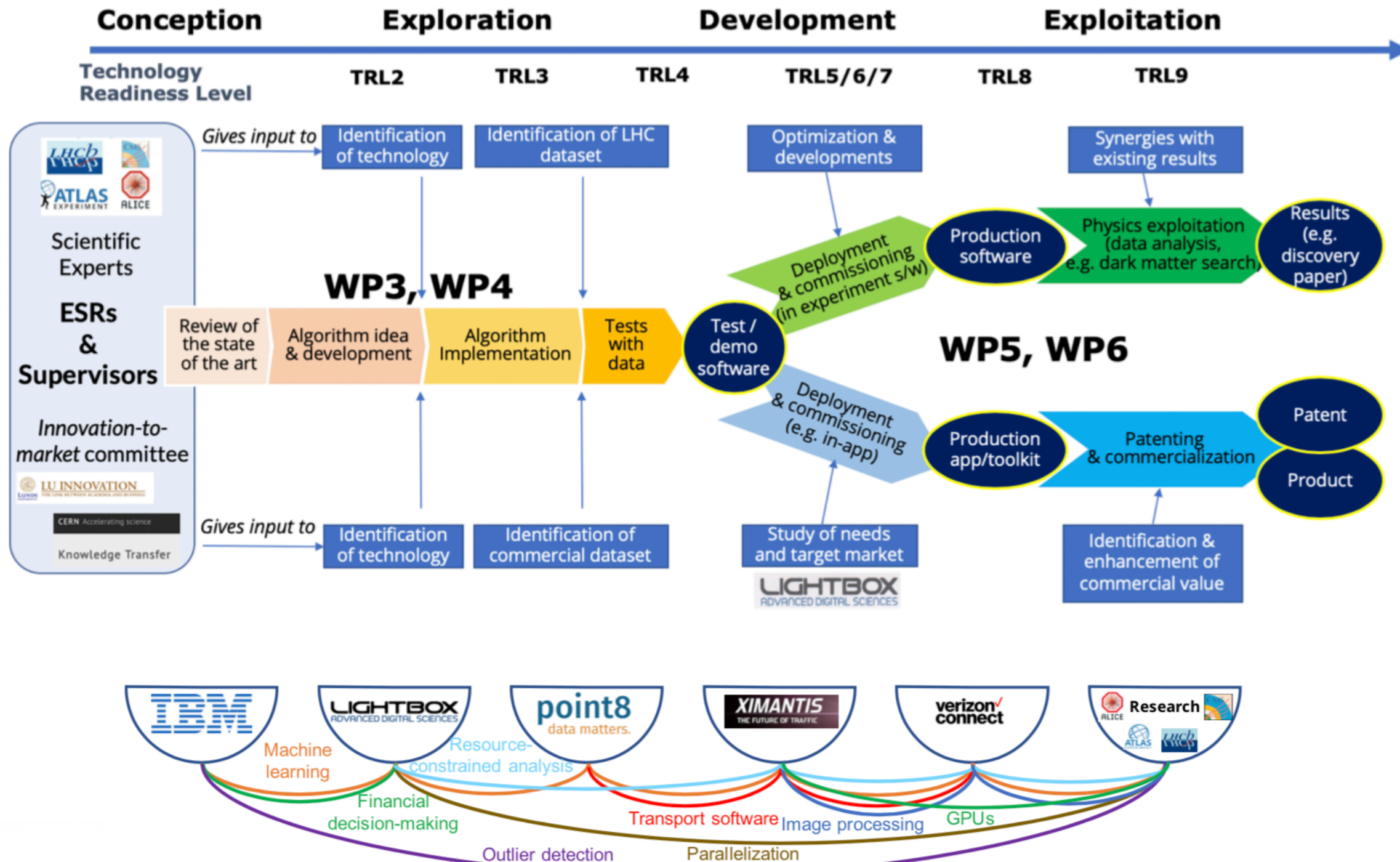
# SMARTHEP status & plans

- **2022:**
  - Hiring of the Early Stage Researchers
  - Kick-off meeting and check-in meeting (University of Manchester): [this meeting](#)
- **2023:**
  - yearly meeting, with more transferrable skills training
  - Deliverables: short whitepapers written by ESRs describing state-of-the-art for RTA, ML, hybrid architecture
  - Schools and training:
    - physics and machine learning school (University of Geneva), for network participants

**Following years:** more schools / activities, including data challenges, outreach, participation/ presentation at conferences



# Collaborating & learning from industry





# The PhD positions (Early Stage Researchers)

PhD student position with **CMS** at University of Helsinki [Link](#)

ESR1: Machine learning and Real Time Analysis for **Higgs boson measurements and fleet safety**

PhD student position with **IBM** at Sorbonne University [Link](#)

ESR2: Real-time rule induction in **fraud detection and HEP**

PhD student position with **ATLAS** at University of Geneva [Link](#)

ESR3: Real time analysis strategies for **reconstruction, exotic physics, and market analysis**

PhD student position with **ATLAS** at CERN and at the University of Geneva [Link](#)

ESR4: Efficient RTA in **ATLAS and finance** using multithreading

PhD student position with **LHCb** at Sorbonne University [Link](#)

ESR5: RTA on heterogeneous architectures for **LHC and self-driving cars**

PhD student position with **LHCb** at VU Amsterdam [Link](#)

ESR6: Optimization of RTA resources for LHCb **LFV search and manufacturing**

PhD student position with **LHCb** at TU Dortmund [Link](#)

ESR7: Machine Learning for Real Time Analysis of **Lepton Flavour Violation** in neutral meson decays and **traffic predictions**

PhD student position with **LHCb** at TU Dortmund [Link](#)

ESR8: Real Time Analysis for global event triggering in **LHCb and manufacturing**

PhD student position with **LHCb** at University of Heidelberg [Link](#)

ESR9: Real Time analysis for Dark Photons search in **LHCb and smart vehicles**

PhD student position with **ALICE** at Lund University [Link](#)

ESR10: Real Time calibration of **ALICE Time Projection Chamber and ML traffic predictions**

PhD student position **Verizon Connect** in Florence and University of Bologna [Link](#)

ESR11: Real Time Analysis through computer vision on **dashcams and triggers** in HEP

PhD student position with **ATLAS** at University of Manchester [Link](#)

ESR12: Accelerated **anomaly detection**

My jobs (1)

**12 PhD positions in Particle Physics and Computer Science - EU project SMARTHEP**

Geneva, Switzerland • Jan 19, 2022 • MSCA-ITN-SMARTHEP-2022-1

● FILLED • Not Published

- 199 candidates applied, of which approx. 150 eligible
- selection procedure was anonymised (as much as possible) for gender and age
- hired PhD students are from India, Italy, Denmark, Sweden, Greece, Spain, Thailand, UK
- PhD programs starting in Autumn 2022



# How to collaborate

We want this network to be as **inclusive** as possible  
(within funding & logistics constraints)

Would you like to connect to our activities,  
or participate in the planned network training events?  
⇒ universities, industries and organizations can still become  
*SMARTHEP partners*

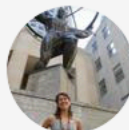
The procedure requires network approval && signature of addendum to existing Consortium Agreement, where intellectual property is discussed

Partner students can be affiliated to SMARTHEP as *trainees*,  
with a possibility to participate to events remotely or in person if logistics allows  
Lecturers will be reimbursed for their participation to events

Let's chat if you'd like to **collaborate** or have ideas  
on how to expand the network!



# Extra: Unsolicited advice on ITNs (and grants?)

 **Caterina Doglioni** @CatDogLund · Jan 8

Research collaboration persistence (memes) paying off...with @particleist @xmpierinix @MartinoBorsato @ASfyrlla @skvarkki & many others. More news on the collaboration/physics involved soon!

**How it started (2016, 2018, 2019)**

European Commission  
Your EU proposal 861250 - SMARTHEP; rejection

**Europa / Funding & Tenders Portal notification**

Dear Coordinator,

We regret to inform you that your above proposal has been rejected. The Rejection Letter is available.

**How it's going (2020)**

European Commission  
Your EU proposal 956086 - SMARTHEP; start of grant preparation

**Europa / Funding & Tenders Portal notification**

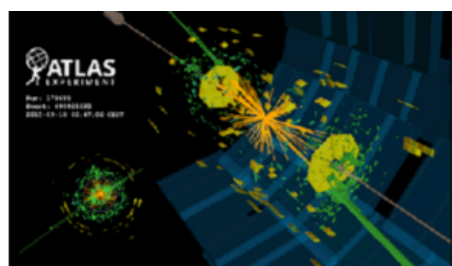
Dear Coordinator,

7 5 49

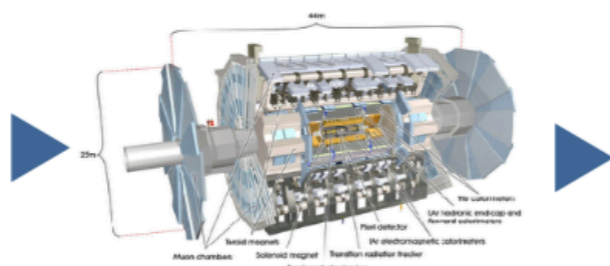
- We persevered, in spite of the “outcome fluctuations” (not a linear slope in grade from referees, even though we thought the proposal improved with time)
- We followed ITN writing courses, multiple times
- Most of all, we understood there is a component of randomness in all grants where there are many excellent projects, so we didn't get discouraged!



# Where are the limitations to record (more) data?



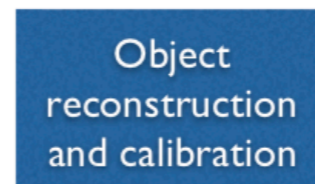
**Detector** readout to hardware trigger



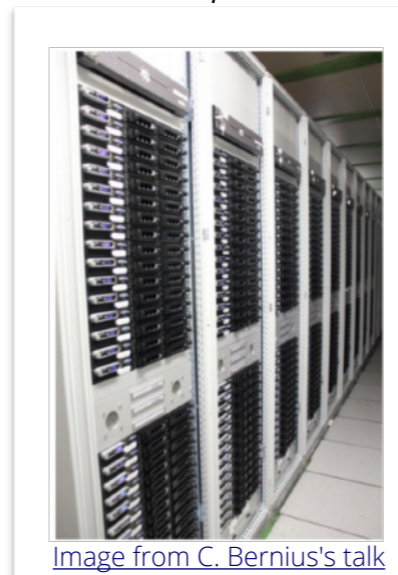
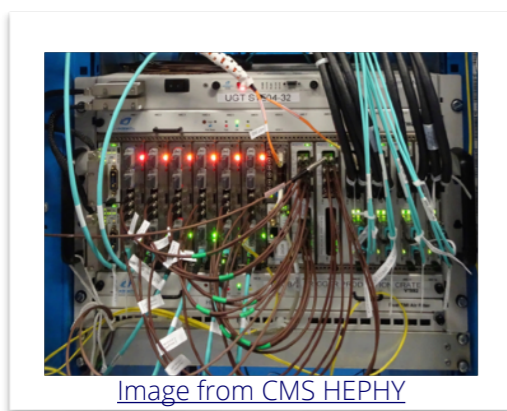
**L1** hardware trigger



**CPU** for processing events  
*large but limited processing power*



**Disk/tape** to store events



With the other SMARTHEP participants from the main LHC experiments, we want to overcome these limitations so we can make the most of the upcoming data



SMARTHEP first kick-off, 2017  
Supported by Grace & Filip Sandblom Fund



# Where to discuss trigger&reconstruction in HEP & beyond

## HSF = High Energy Physics (HEP) Software Foundation

- Forum for physicists with interest in software for HEP
  - ...and beyond: contacts and shared meetings with nuclear physics, accelerator, DM experiments
- Latest [whitepaper](#) on common software and techniques
  - Initial whitepapers helped shape [IRIS-HEP](#) US/NSF effort
- Working groups including [trigger & reconstruction](#)
  - Trigger & reconstruction [plans](#) for 2021 include discussions of ML for hardware triggers, heterogeneous architectures



[Website](#)  
[Discussion Forums](#)

Many HSF participants also in cross-field [SIDIS](#)  
**Software Institute for Data-Intensive Sciences**



STFC-funded effort co-led by UofM: **SWiFt-HEP**  
Reconstruction & trigger Work Package



# Joint / related activities



- REALTIME interdisciplinary Advanced Study Group, 2019-2020
  - Pufendorf IAS is a great source of interdisciplinary ideas/connections!
  - REALTIME had members from engineering, IT, maths, astrophysics, law, social sciences
  - concluded their preliminary discussions, unfortunately dampened by corona
- **Much more** than real-time analysis: HELIOS Lund-Hamburg research school funded by Helmholtz on **intelligent instrumentation in physics**
  - similar “knowledge environments” of university + labs (ESS, European xFEL, DESY + CERN...)
  - will collaborate with SMARTHEP

