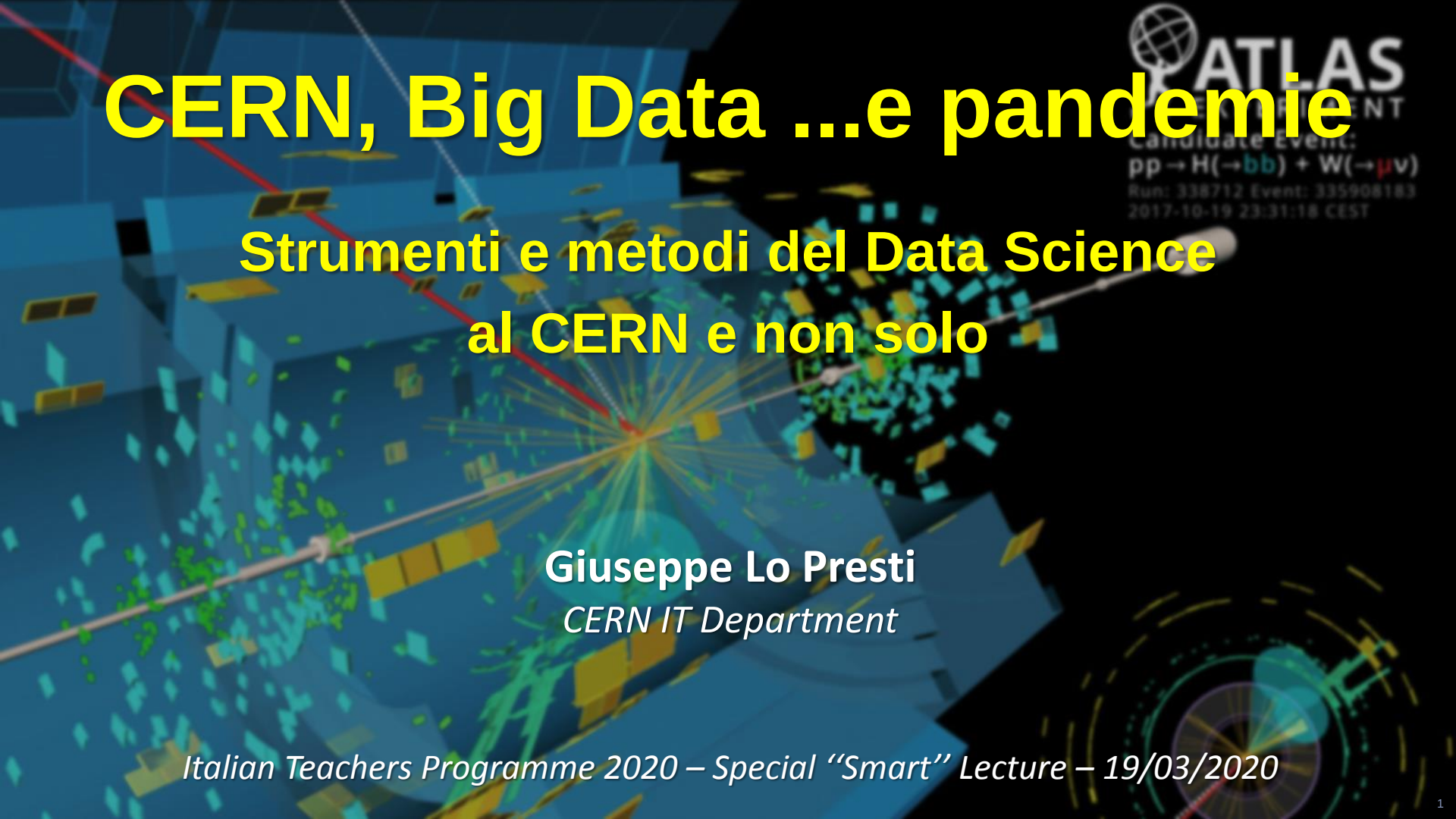


CERN, Big Data ...e pandemie

Strumenti e metodi del Data Science al CERN e non solo

Giuseppe Lo Presti
CERN IT Department

Italian Teachers Programme 2020 – Special ‘‘Smart’’ Lecture – 19/03/2020

Programma della lezione

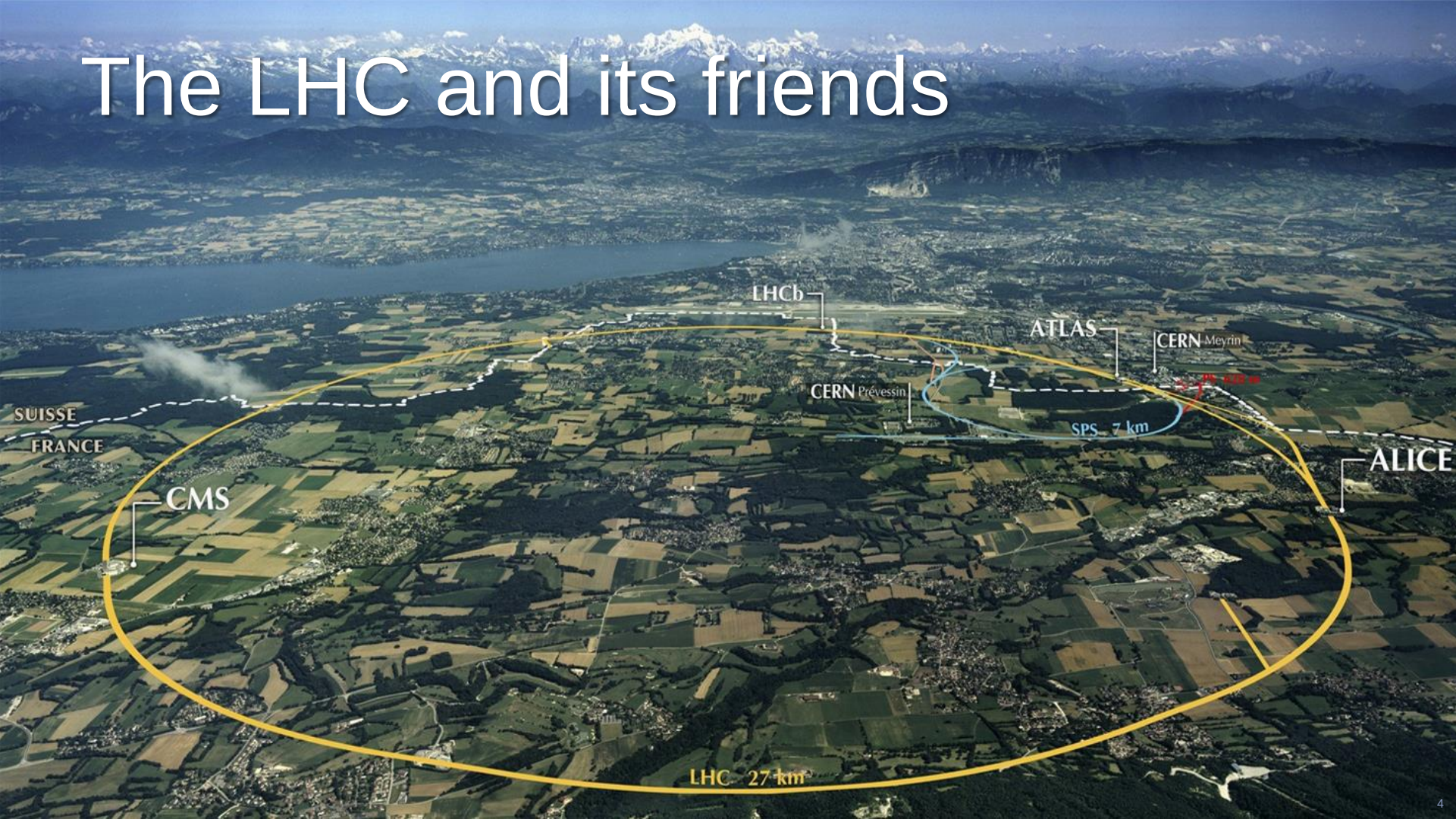
- Il contesto
  - Cos'e' il CERN, l'LHC e i *(Big)Data* di Fisica (in inglese)
- Strumenti e metodi per l'analisi dei *Big Data*
  1. Python, Jupyter notebooks e librerie per *Data Science*
  2. Esempio: analisi e modello dell'evoluzione di una pandemia
  3. Estrapolazioni e considerazioni generali di statistica
- Domande & risposte

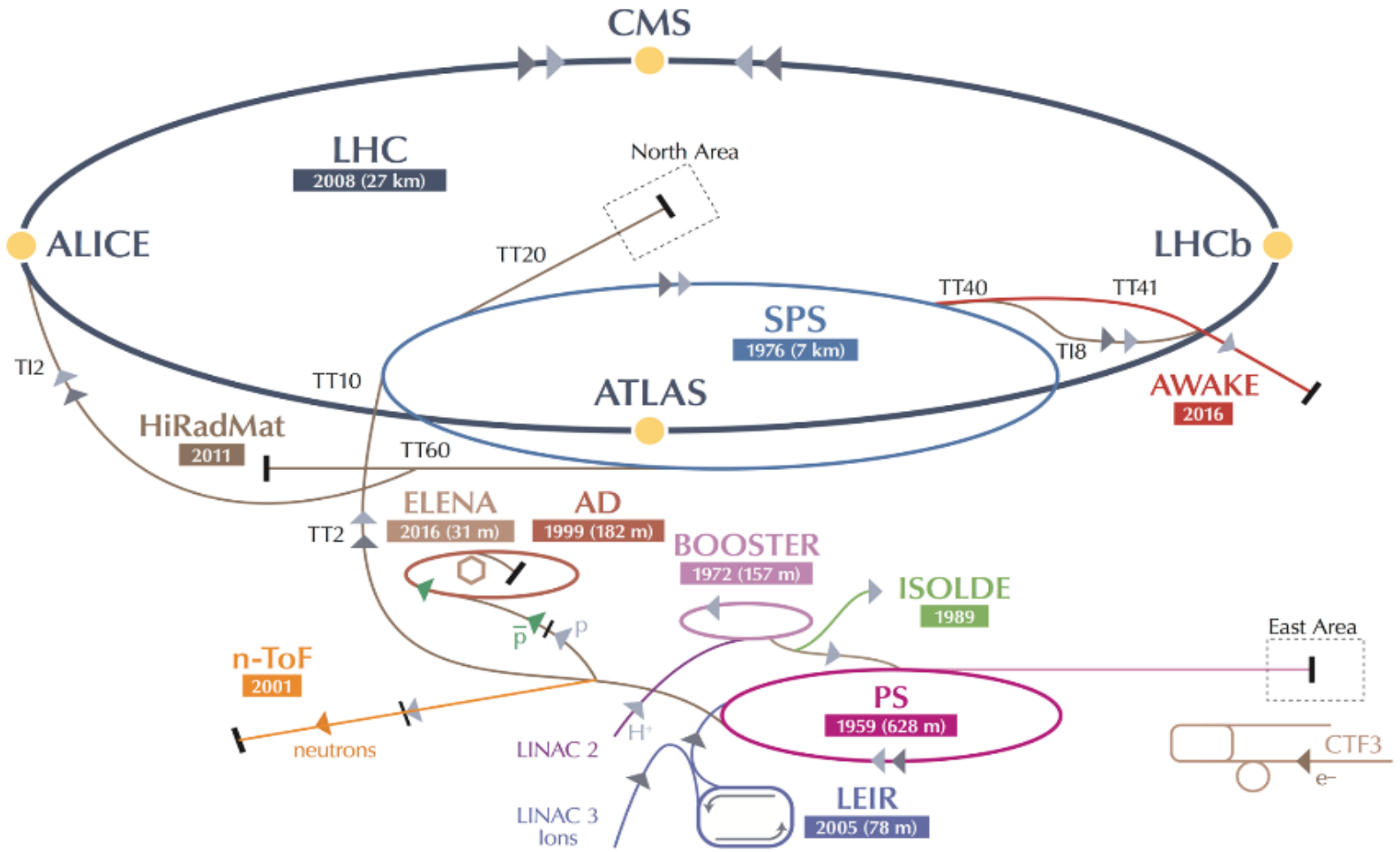
CERN in Numbers

- **1954:** foundation year, as an European Council
- **23** Member States, **8** Associates, **4** Observers, UE + UNESCO…
- **50+** Non-member States collaborate with CERN
- **2300** staff members work at CERN as personnel,

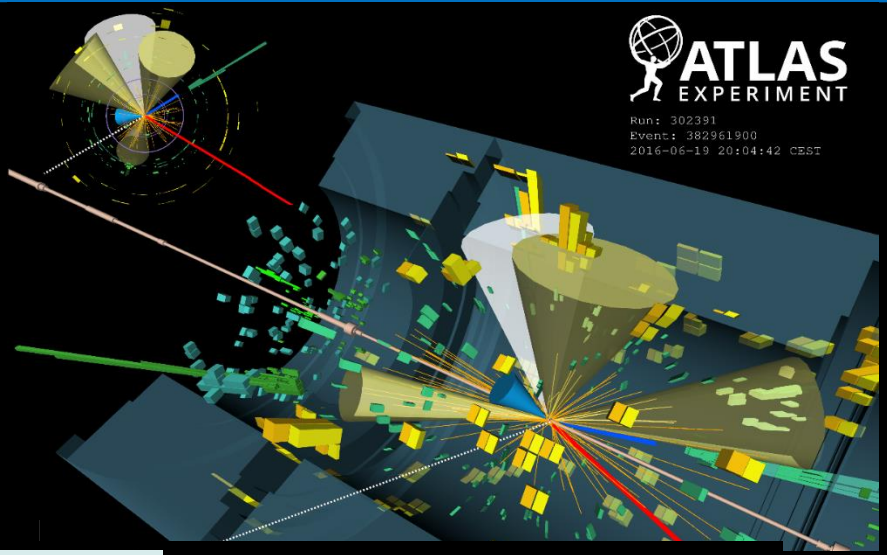
  12 000 researchers come from institutes world-wide
- **1000 MEUR** annual budget
- **5** Nobel Prizes (…**6** with Englert & Higgs)

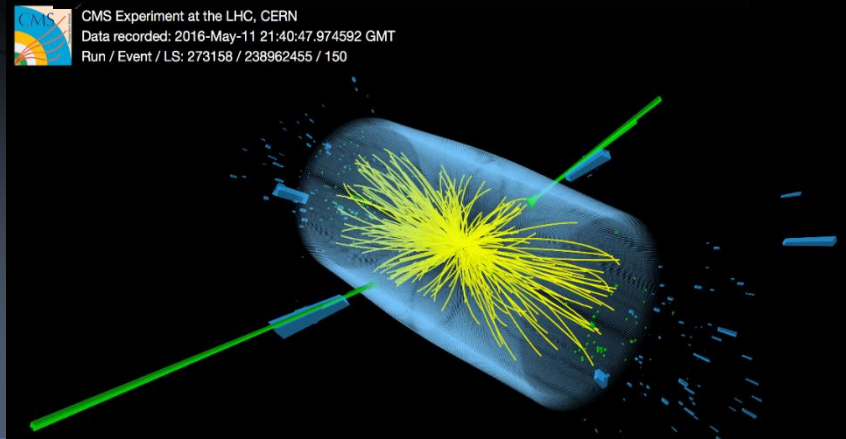
The LHC and its friends

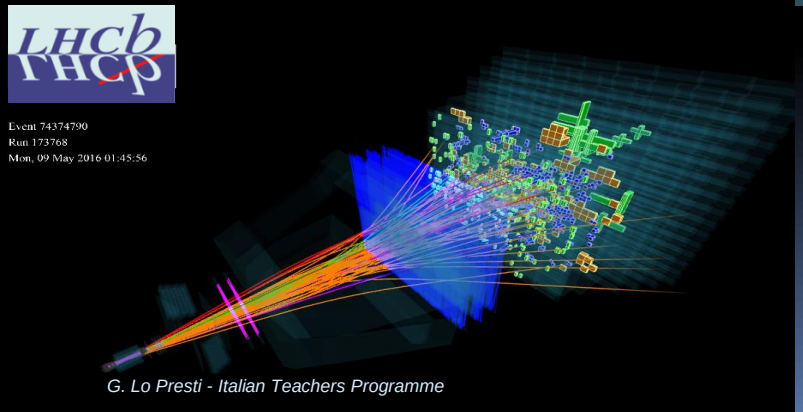



CMS
LHC
2008 (27 km)
North Area
ALICE
LHCb
TT20
TT40
TT41
SPS
1976 (7 km)
TI8
TI2
TT10
ATLAS
AWAKE
2016
HiRadMat
2011
TT60
ELENA
2016 (31 m)
AD
1999 (182 m)
TT2
BOOSTER
1972 (157 m)
ISOLDE
1989
East Area
n-ToF
2001
PS
1959 (628 m)
neutrons
LINAC 2
LINAC 3
Ions
LEIR
2005 (78 m)
CTF3

Some samples of what we see…

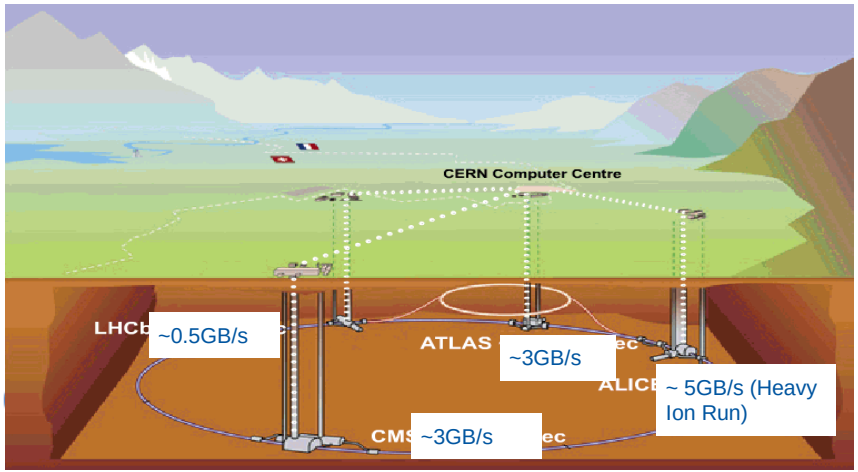




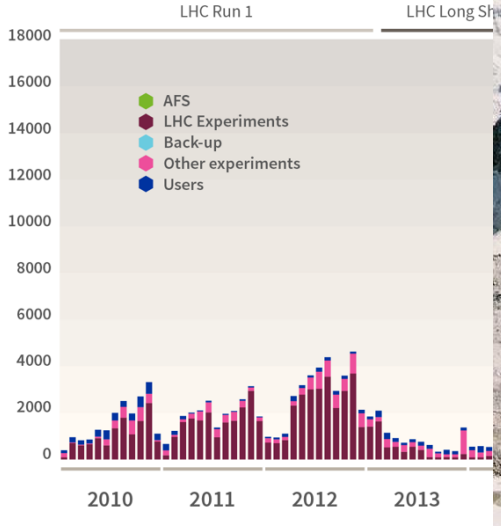


Data Processing

- Experiments send over 10 PB of data per
  - 115 PB from all experiments in 2018
- The LHC data is aggregated at the CERN be stored, processed, and distributed



It



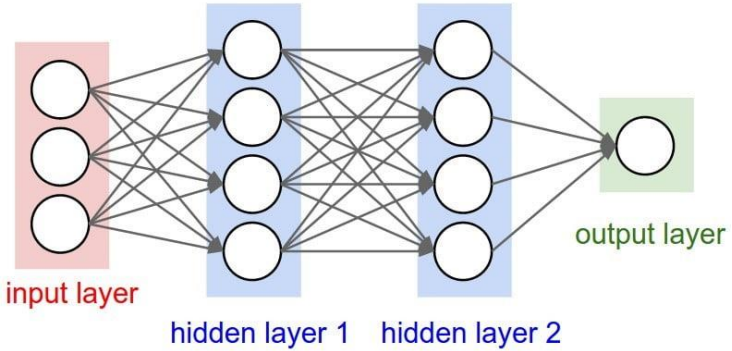
Big Data ?

- *Big data* is a field that treats of ways to analyse […] or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software (*Wikipedia*)
  - **Moving target** by definition!
- From structured data, relational DBs, centralized processing…
- …To ***unstructured*** data and decentralized (i.e. parallel and loosely-coupled) processing, more adapted to the Cloud
  - E.g. trend analysis, pattern recognition, image segmentation, natural language interpretation/translation, …

Big Data and **Data Science**

- **The Power of Data:** Neural Networks are well known since the 1990s, but it's only now with **very large** and **easily accessible** data sets that they become effective!

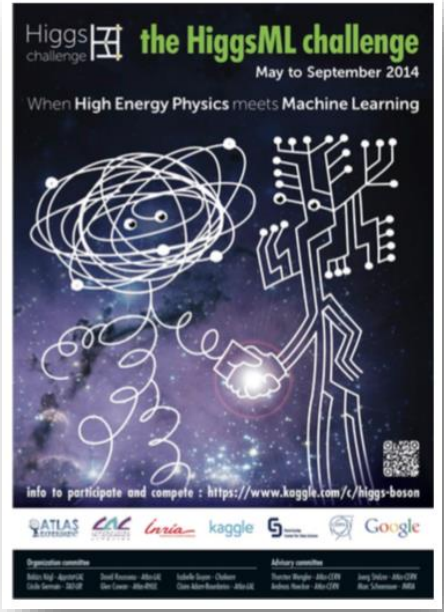
  - Lots of software frameworks for *Deep Machine Learning* with NNs coming up



Big Data at CERN

- Experiments have long used Machine Learning (once called Multi-Variate Analysis) techniques

- In recent years, *Challenges* (competitions) organized among students and young researchers to motivate ML research
  - BUT: in 2018, best results obtained without ML!
  - Nevertheless, the push from industry makes ML approaches attractive



Big Data and Future Experiments

- **High Luminosity LHC**: an upgrade of all LHC Experiments that increases data rates by more than 10x
- **New Big Science experiments** coming up:
  - **LIGO-Virgo** Gravitational Astronomy
  - Square Kilometer Array (**SKA**)
  - Deep Underground Neutrino Experiment (**DUNE**)

- Time for further research, opportunity for new **synergies**
  - Increasing role of Machine Learning techniques
    - E.g. LIGO: real-time Gravitational Waves signal detection

End of the introduction
Questions before the interactive part?

Accélérateur de science