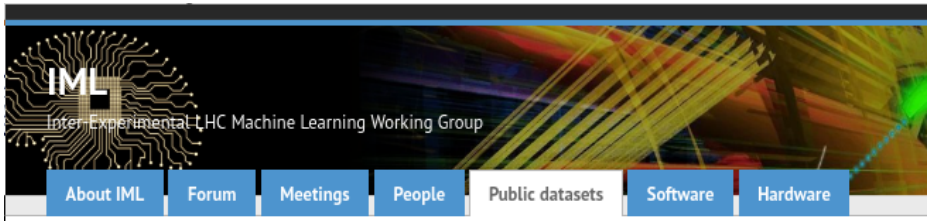
IML meeting – NEWS

Lorenzo Moneta (CERN) Markus Stoye (Imperial College, DSI)
Paul Seyfert (CERN) Rüdiger Haake (Yale University)
Steven Schramm (Université de Genève)

HEP Software Foundation – Community White Paper

- organised by HSF
- writing sessions e.g. at the first IML workshop in 2017
- now on arXiv
- Updates are planned
  - Editors are preparing a new version
  - Community feedback will be solicited on the next draft
  - More details and open discussion to follow

public datasets



Public datasets

This page collects public datsets that are used for machine learning studies at the LHC. The resources come from voluntary contributions of authors from papers as well as challenges. The text on the linked description pages is not in the responsibility of the IML. We want to emphasis that public datasets are typically done with simplified simulation of the real detectors and with much smaller samples than available to the collaborations. Best results in simplified simulation with limited number of samples do not automatically suggest an optimal strategy for application in real experiments, however currently there is no other option for comparisons as most of the collaborations' data is not public.

Simplified datasets for benchmarking:

- Top tagging without heavy flavour & pileup: Data and details of arXiv:1707.08966
- Jet substructure: Data from arXiv:16107.08633 at UC Irvines page MLPhysics
- Flavour tagging without pileup: Data from arXiv:1603.09349 at UC Irvines page MLPhysics

Datasets for developing simulation:

- Data for jet images from LAGAN
- Data for 3D jet images from CaloGan
- Electromagrentic jet images

Realistic datasets from the CMS experiment:

- CMS open data (non trivial data format (CMS software knowledge of advantage), limited in size and older samples 2011)

LHC Kaggle challenges:

- TrackML particle tracking challenge (starting May 2018)

- public ML datasets linked on the IML website
- IML now also a community on zenodo
- files from the workshop challenge available on zenodo (in addition to eos)

Machine Learning related sessions at summer conferences

- ICHEP
- CHEP
- BOOST
- QCHS

upcoming events

2018-09-10 – 2018-09-13 ROOT Users' Workshop in Sarajevo

2018-09-17 – 2018-09-18 AI at CERN and SKA

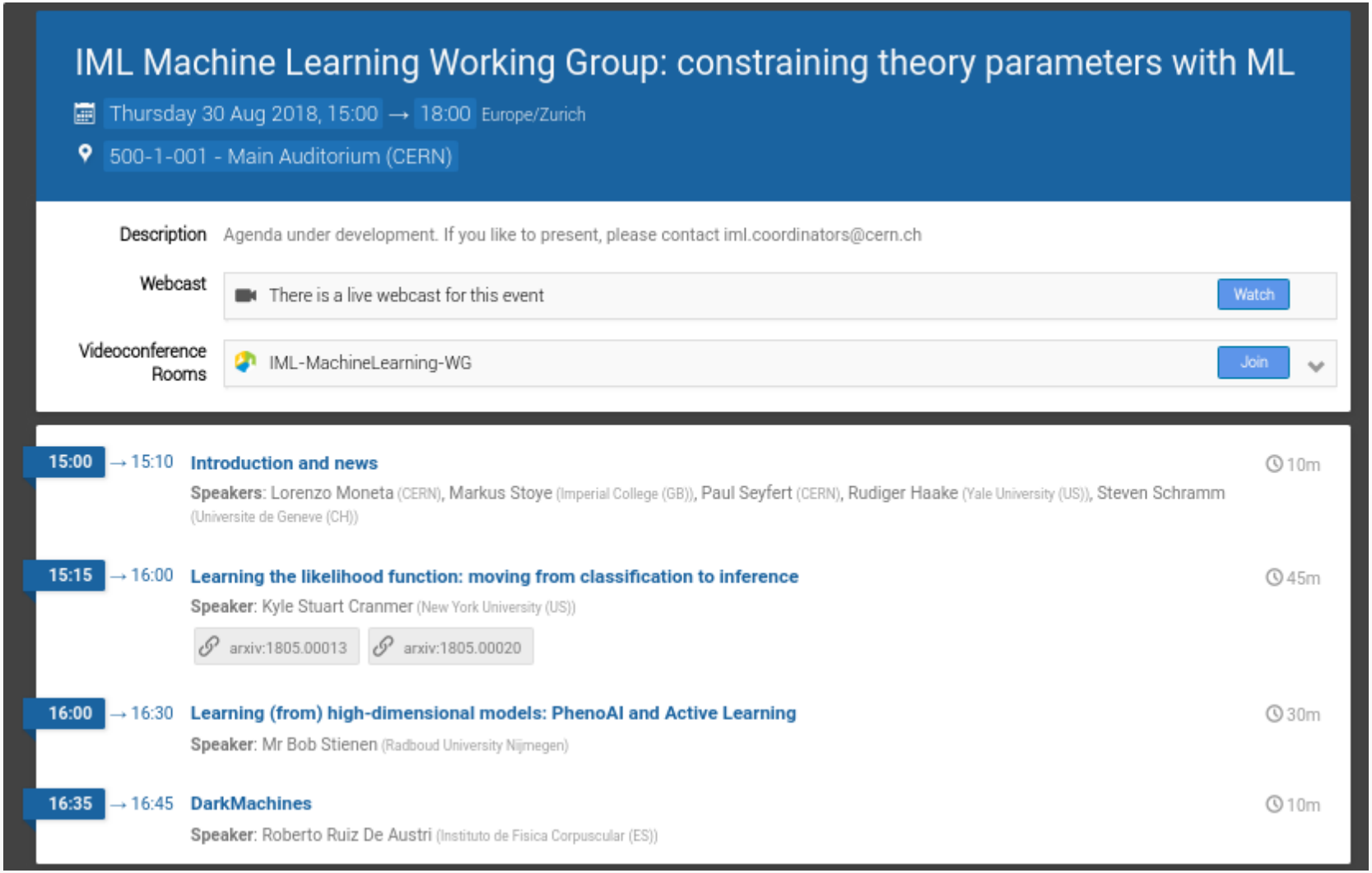
2018-09-17 – 2018-09-21 INSIGHTS Workshop on Statistics and Machine Learning

2018-09-19 joint Outreach Workshop together with the aMVA4NewPhysics Training Network

2018-09-21 CERN Alumni -- Moving Out of Academia to Big Data

2018-10-12 next IML meeting (Salle Anderson 40/S2-A01)

today's meeting



IML Machine Learning Working Group: constraining theory parameters with ML

Thursday 30 Aug 2018, 15:00 → 18:00 Europe/Zurich

500-1-001 - Main Auditorium (CERN)

Description Agenda under development. If you like to present, please contact iml.coordinators@cern.ch

Webcast There is a live webcast for this event — Watch

Videoconference Rooms IML-MachineLearning-WG — Join

15:00 → 15:10 **Introduction and news** — 10m

Speakers: Lorenzo Moneta (CERN), Markus Stoye (Imperial College (GB)), Paul Seyfert (CERN), Rudiger Haake (Yale University (US)), Steven Schramm (Universite de Geneve (CH))

15:15 → 16:00 **Learning the likelihood function: moving from classification to inference** — 45m

Speaker: Kyle Stuart Cranmer (New York University (US))

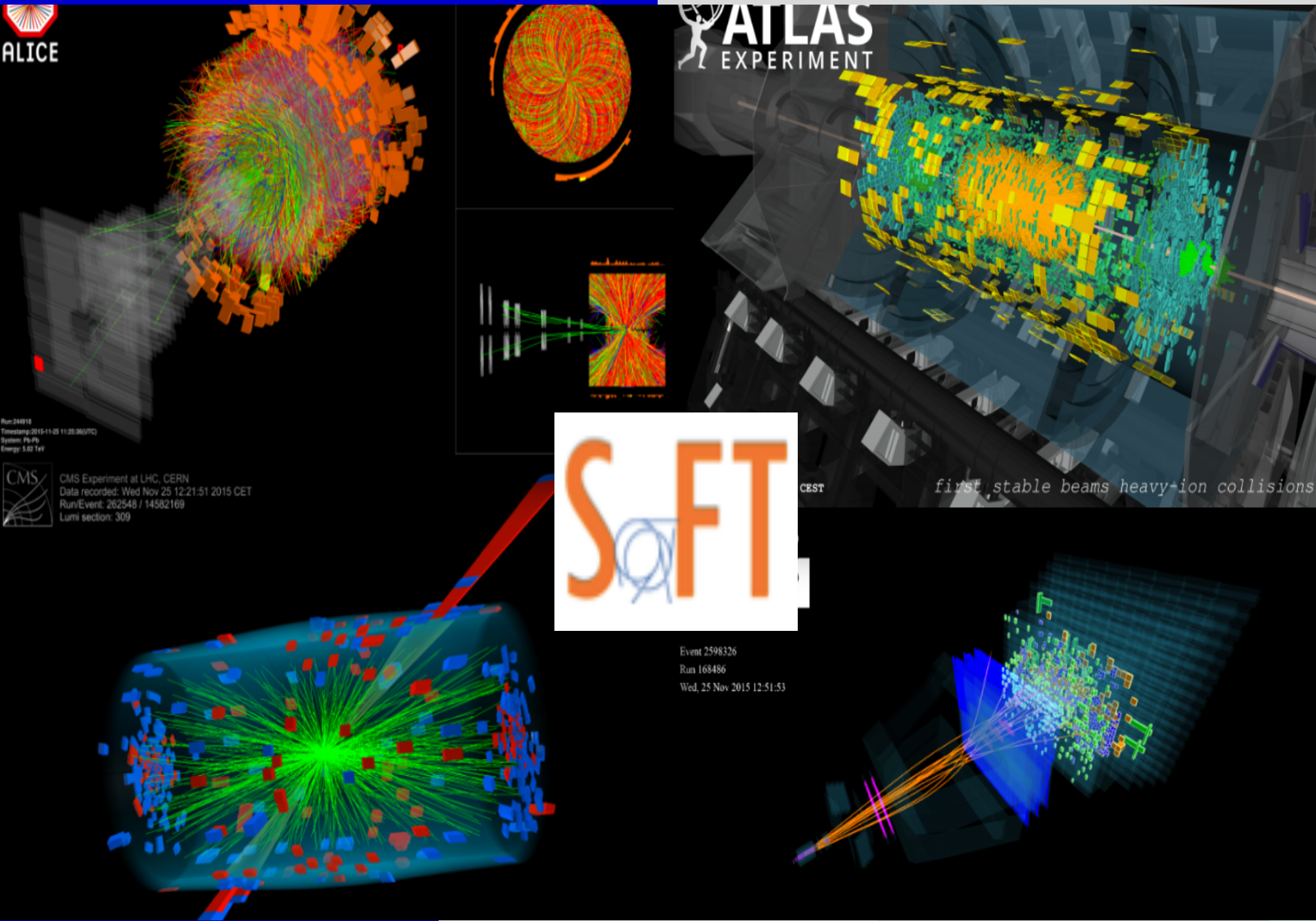
arxiv:1805.00013 arxiv:1805.00020

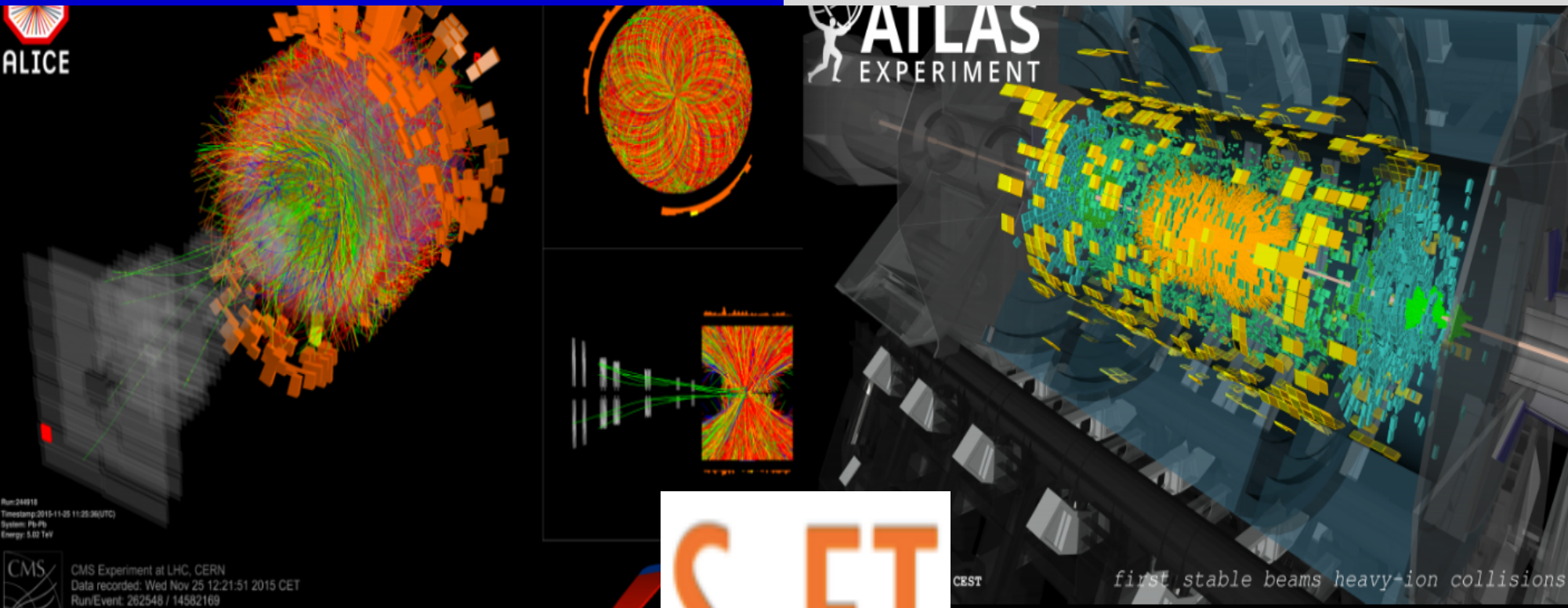
16:00 → 16:30 **Learning (from) high-dimensional models: PhenoAI and Active Learning** — 30m

Speaker: Mr Bob Stienen (Radboud University Nijmegen)

16:35 → 16:45 **DarkMachines** — 10m

Speaker: Roberto Ruiz De Austri (Instituto de Fisica Corpuscular (ES))


ALICE
ATLAS
EXPERIMENT
CMS Experiment at LHC, CERN
Data recorded: Wed Nov 25 12:21:51 2015 CET
Run/Event: 262548 / 14582169
Lumi section: 309
first stable beams heavy-ion collisions
SoFT
Event 2598326
Run 168486
Wed, 25 Nov 2015 12:51:53



slides (excl. cern logo) will appear on

https://gitlab.cern.ch/pseyfert/slides-imlnews-2018-08-30