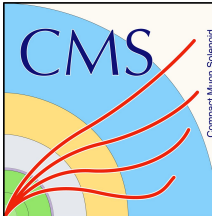


# CMS perspectives on Open Science

Dietrich Liko (ÖAW), Clemens Lange (CERN)  
for the CMS Collaboration

20<sup>th</sup> September 2021  
Open Science Strategy Working Group meeting



# Mission



- The CMS Collaboration strives to perform world-class research to uncover what the universe is made of and how it works
- Our unique research unites people from all over the world to push the frontiers of science and technology, for the benefit of all
  - 4000 particle physicists, engineers, computer scientists, technicians and students from around 200 institutes and universities from more than 40 countries
- The scientific results are shared openly with the world

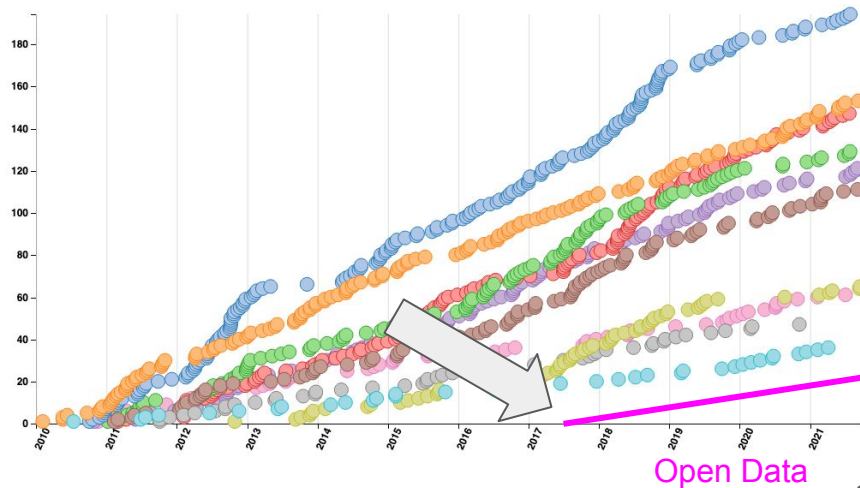


Fully aligned with CERN's mission

# Activities and projects: open data



- CMS is leading the high-energy physics open data effort
  - > 2 PB of research-level data currently dwarf the efforts of all other experiments
- Two well-attended open data for theorists workshops ([2020](#), [2021](#)) have been instrumental to making the open data more accessible
  - They helped us ensure our open data are usable by non-CMS physicists, also exercising our own workflows
  - By now, the number of results based on CMS open data has reached a level as if CMS had an additional physics analysis working group
- Open data released on CERN Open Data Portal with additional code/analysis examples (also on GitHub)



# Activities and projects: results dissemination



- All our publications are open access
  - We have recently released a [search tool](#) to make details of our analyses easier findable
- A large fraction of our results are also made available as HEPData records
  - [HEPData lib](#) tool to significantly simplify getting data into HEPData developed and maintained by CMS physicists, now de-facto standard tool
- CMS has two Physics Communication Officers who work on making CMS results more accessible, e.g. via [physics briefings](#)
- Some results are provided with additional information such as simplified likelihoods, ROOT ntuples etc. for easier reinterpretation

CERN Accelerating science

**CMS Public Results** Visit us: [CMS Public Website](#) | [CMS Publications](#) | Contact us: [CMS Publications Committee](#)

Select [Papers](#) [arXiv](#) [INSPIRE](#) Show 10 entries Previous 1 2 3 4 5 ... 105 Next

Only show results that match any all of the selected categories [Clear all filters](#)

**Filters**

- Collision system
- Accelerator parameters
- Physics theme
- Working group
- Final state
- Final state signature
- Interpretation
- SM analysis characteristics
- Further search categorisation
- Heavy ions: further categorisation

Code	Title	Status/Link	Date
B2Q-19-002	Search for heavy resonances decaying to WW, WZ, or WH boson pairs in the lepton plus merged jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to Phys. Rev. D</a>	September 13, 2021
SMP-20-010	Study of quark and gluon jet substructure in Z-jet and dijet events from pp collisions	<a href="#">Submitted to J. High Energy Phys.</a>	September 8, 2021
HIN-19-011	Observation of $B_s^0$ mesons and measurement of the $B_s^0/B^+ \pi^+$ yield ratio in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV	<a href="#">Submitted to Phys. Lett. B</a>	September 4, 2021
TOP-20-002	Observation of $tW$ production in the single-lepton channel in pp collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to J. High Energy Phys.</a>	September 3, 2021
TOP-19-009	Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to J. High Energy Phys.</a>	August 24, 2021
TOP-20-001	Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton-jets events from proton-proton collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to Phys. Rev. D</a>	August 5, 2021
TOP-21-001	Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to J. High Energy Phys.</a>	July 29, 2021
EXO-20-004	Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to J. High Energy Phys.</a>	July 27, 2021
SUS-20-003	Search for chargino-neutralino $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ decays into $WZ$ and $W\tilde{\chi}_1^0$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to Phys. Lett. B</a>	July 26, 2021
HIG-20-015	Measurement of the inclusive and differential Higgs boson production cross sections in the decay mode to a pair of $\tau$ leptons in pp collisions at $\sqrt{s} = 13$ TeV	<a href="#">Submitted to Phys. Rev. Lett.</a>	July 23, 2021

Showing 1 to 10 of 1,046 entries (filtered from 1,807 total entries) Previous 1 2 3 4 5 ... 105 Next

© Copyright CERN 2021 for the benefit of the CMS Collaboration To the top

# Activities and projects: analysis preservation



- We are trying to make use of CERN-provided tools for analysis preservation
- CERN Analysis Preservation Portal (CAP) currently actively used for:
  - Internal statistics committee questionnaire
  - Internal machine-learning group questionnaire
  - Physics analysis categorisation/tagging (for public results search)
- REANA:
  - Encouraging analysts to use the platform, no significant effort though
  - Use of analysis-specific software containers has little adoption in CMS as no CMS-wide common high-level analysis tools exist
- Note: both projects seem understaffed and under-equipped affecting adoption
- We are now putting more efforts on analysis preservation internally
  - Physics analysis (tools) task force recently put in place

# Activities and projects: OD in education and outreach



- CMS Open Data sets are widely used in education
  - in physics courses at universities
  - in Bachelor's theses or student assignments
  - at high-school level (simplified format derived from the research-level data)
  - Master classes and workshops
- They also provide very important inputs to various benchmarking studies (e.g. for ROOT developments and general HEP software developments such as <https://arxiv.org/abs/2104.12615>)

# Current/Anticipated challenges

- Open data storage space limited, but data set is growing continuously
  - CERN OD policy and its implementation plan has been a real relief for our efforts, which have previously done on pilot resources
- Open science costs actual money and cannot be based entirely on good will
  - Computing power
  - Storage space
  - Personnel (library, developers, physicists)
  - Long-term support and (knowledge) preservation
- More training required
  - Big interest in analysis preservation workshops --- need to build up more expertise

# The future of open science



- Funding agencies have required statements in line with open science for years now
  - Little activity/requirements beyond open access publications though
- Expecting more due diligence in the future
- Open science tooling will need to grow with the demands
- CERN could lead the way (beyond time-limited national funding efforts)



# Further important issues or policy matters for consideration



- Open science != citizen science
  - Research-level science needs far more more expertise and resources
  - Important to enable reinterpretations
- Open data analyses are real research projects
  - Threshold to get started, but new results have been published with CMS Open Data
- What will be provided by CERN and what will collaborations and other institutions have to provide?