

# CMS OpenData

CMS experiment has pioneered open data for the LHC

Published data & software

- Derived data (AOD and miniAOD format) – No RAW data
- Collision event data and corresponding MC
- Run1 almost entirely, fraction of Run2 data, more data to be added
- Total volume: several PB
- CMS software in form of image for virtual machines (container in progress)
- Detailed web pages with descriptions

Infrastructure

- Public (EOS) storage instance at CERN

**Several papers in peer reviewed journals (mainly by theorists)**

Other LHC ramping up their open data efforts

# OpenData at DESY

Mirroring the CERN (LHC) open data is certainly not of interest

Opportunities by combined analyses of existing (or coming) open data

- Common analyses of open data from different experiments (even beyond collider experiments)
- Results or combined data sets might be re-published
- Some ideas being discussed in PUNCH4NFDI (and elsewhere)

# Technical Aspects

## Access

- Streaming via XROOTD protocol is very common in HEP applications
- Read access often without restrictions – writing (of course) not
- For authenticated access OIDC tokens will be used (perhaps X509 during a transition period)

## Prototyping within PUNCH4NFDI

- PUNCH VO enabled on the DESY dCache instance
  - Very simple authorization scheme
  - VO managed via Helmholtz-AAI (based on UNITY)
- Token based authentication is rather challenging to implement
- Metadata management
  - Possible solution being looked at (or planned to) in PUNCH4NFDI