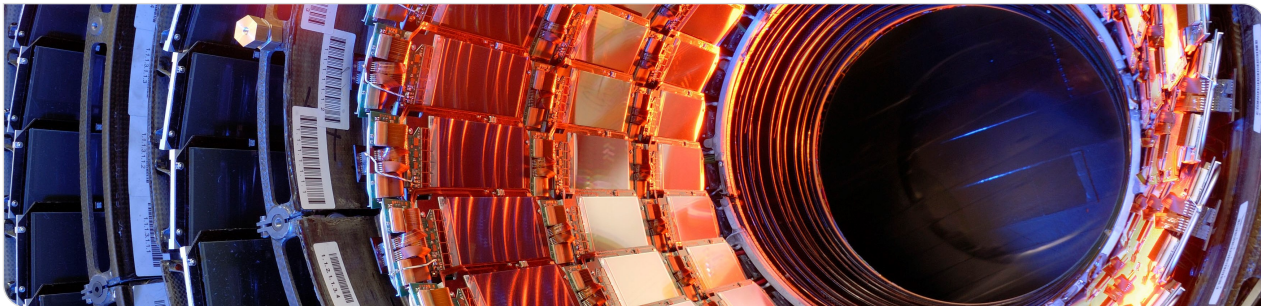# User Samples at Global CMS Storage
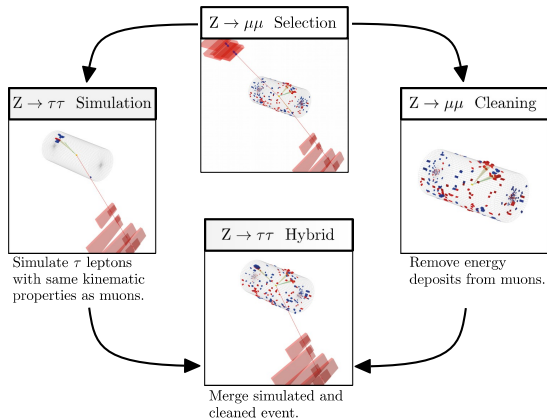
Example: Run 2 Tau Embedded Samples

S. Brommer, **A. Gottmann**, M. Klute, G. Quast, R. Schmieder, C. Winter, R. Wolf | June 23, 2022

Introduction

- Discussed with CMS about moving user samples to global CMS area
- Made 2 presentations on Tau Embedding in PPD (and corresponding Run 2 sample):
  - April 14, 2022
  - June 2, 2022
- Obtained Green Light from O&C Co-coordinator and PPD
  → No objections for small datasets, which are relevant for CMS as collaboration
- Potential work to be invested by us, if required

$Z \to \mu\mu$ Selection

$Z \to \tau\tau$ Simulation

$Z \to \mu\mu$ Cleaning

Simulate τ leptons with same kinematic properties as muons.

Remove energy deposits from muons.

$Z \to \tau\tau$ Hybrid

Merge simulated and cleaned event.

from JINST 14 (2019) 06, P06032

Current Status

- Details on current situation given in CMS Talk and GGUS Ticket. Summary:
  - All Run 2 Tau Embedded Samples were/are produced at opportunistic resources available at KIT
  - Currently stored on limited Tier 3 dCache disk storage at KIT and DESY (user area)
    → This storage is dedicated to German CMS (dCMS) users for their analyses
  - This dCache disk storage is **not** suitable for long-term storage and data preservation

Our Conclusions from a KIT Management Meeting

- Need to ensure that Tau Embedded Samples are secured and available for
  - future analyses,
  - and reproduction of existing results.
- This can be accomplished by moving them to global area known to Rucio and DBS
  - Allows storage on tape and replication to disk and other sites

Technical Aspects

- Prepared Run 2 UL Tau Embedded Samples to test & perform the migration:
  - 2017 data-taking: around 8 TB in total, file sizes 1-3 GB, modified MiniAOD format
  - 2018 data-taking: around 11 TB in total, file sizes 1-3 GB, modified MiniAOD format
  - File paths contained in these lists match CMS file naming scheme
- Following technical steps are suggested:
  - Move all files and directories (can be done by KIT admins, if needed)
    from `/pnfs/gridka.de/cms/disk-only/store/user/sbrommer/ul_embedding/large_miniAOD/`
    to `/pnfs/gridka.de/cms/store/results/tau_embedding/`
    → Allows to trigger transfers to KIT tape system
  - Publish Tau Embedded Samples under `prod/global` in DBS with names matching file paths, e.g.:
    `/EmbeddingRun2018A/ElMuFinalState-inputDoubleMu_106X_ULegacy_miniAOD-v1/MINIAOD`
    → We have a script for publishing under `prod/phys03`, but no permissions for `prod/global`
  - Introduce these samples to Rucio database → no expertise and no permissions to do that
- More to come: Run 2016 UL in production; produced pre-UL Run 2 datasets in preparation
  → In total **around 60 TB** of data to be moved to global space

Discussion with CompOps and PPD

Technical constraints to be taken into account:

- Samples need to be in the global space (e.g. under `/store/results`) for the following reasons:
    - to be able to run subsequent steps (e.g. NanoAOD) with CMS system,
    - and to be able to replicate to tape systems and disk storage including also other sites.
    - $\rightarrow$ There is no policy for user-produced samples to do that
- Currently, there is no CMS procedure available to elevate samples from user area to global area
    - StoreResults.py procedure from PhEDEx times deprecated
- Even if we are able to copy files to KIT tape system ourselves, it is unclear yet, how to make them known to Rucio and DBS

Conclusions from the Discussion

- This a non-standard request and requires several custom operations
- Work to be invested by us on voluntary basis (no EPR forseen for that)

More General Perspective

- At KIT, we have discussed the migration of user-produced datasets to global area in general, in case these are useful for the entire collaboration.
- We expect in future, that HPC centers independent from HEP will become much more important.
- Some of these centers may allow access only to users coming from a region in a country Example: NEMO Cluster in Freiburg
  $\rightarrow$ These HPC centers can not be included into CMS infrastructure open to all CMS users.
- However, they are able to provide a large amount of resources to a **local CMS community**
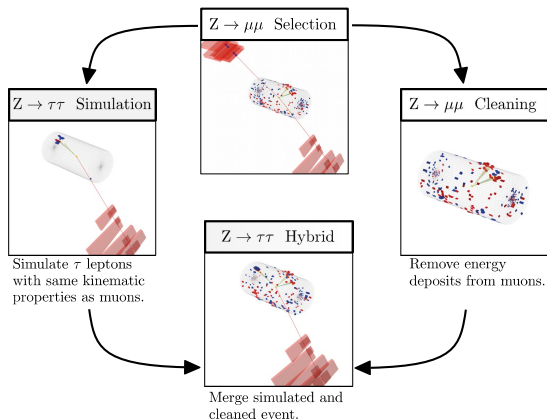
Our Conclusions

- In this context, it would be good to have a procedure to move opportunistically produced samples from user area to global area, if needed
  $\rightarrow$ Despite the fact, that this will only happen occasionally
- Otherwise, we miss a large fraction of opportunistic resources available to the collaboration
- Suggestion from O&C: create a special VOMS group to integrate HPC into CMS infrastructure

Backup

Introduction (Embedding-specific)

- Previous PPD presentation on Tau Embedding focused on the workflow itself:
  - Explained technical setup and concept.
  - Gave historical perspective and considered current use-cases.
  - Discussed boundary conditions at Tau POG and KIT, and future plans.
- Current Status for Tau Embedding WF:
  - Adaptions to CMSSW to test vanilla Tau Embedding already merged.
  - Test workflow submitted to CMS system.
  - Ongoing effort to move latest developments to official CMSSW.
- **Topic for today**: Discussion about already produced Run 2 Tau Embedded Samples

$Z \to \mu\mu$ Selection

$Z \to \tau\tau$ Simulation
Simulate τ leptons with same kinematic properties as muons.

$Z \to \mu\mu$ Cleaning
Remove energy deposits from muons.

$Z \to \tau\tau$ Hybrid
Merge simulated and cleaned event.

from JINST 14 (2019) 06, P06032