
Metadata and bookkeeping for analysis

— Allison Hall, TJ Khoo, Chris Burr —

January 19, 2021

What is analysis metadata?

- Datasets information such file locations, included processes, cross sections,...
- Time period dependent corrections, efficiencies, fine-tuned calibrations
- Data integrated luminosity, trigger path efficiencies, prescales, data quality
- Theory cross sections, theory K-factors or higher order corrections not included in MC
- Bookkeeping information and event processing metadata
- ...anything else that is simply not in “the Event”

Some material taken from <https://indico.cern.ch/event/813208/>

Why is it worth discussing?

- Often (human) time consuming to fetch all the informations
- No current “standard” interface or payload form
 - Even within the same experiment, different pieces of meta/non-event data are served from different DBs, file based repositories, etc...
- Access pattern not necessarily uniform for different type of data
- Analysis preservation: preserving the code and the event data is just not enough, you may need external sources of non-event data
- High speed analysis (multi KHz per core) need also to optimize access to non-event data
 - Caching of payloads
 - Complex access pattern in parallel environments / GPUs

DAWG metadata discussions

Series of meetings, covering specific topics:

1. Calibrations and scale factors and cross-sections & other similar non-event data needed for analysis - **TODAY**
2. Event processing & book-keeping - Feb 2
3. How this all comes together in an analysis - Feb 16

Send us an email if you have suggestions for additional contributions!

Today's agenda

HSF DAWG: Metadata discussions

Tuesday 19 Jan 2021, 17:30 → 19:30 Europe/Zurich

Videoconference
Rooms



HSF DAWG: Metadata discussions

▶ Join



17:30 → 17:35 **Introduction**

Speakers: Allison Reinsvold Hall (Fermilab), Chris Burr (CERN), Teng Jian Khoo (Humboldt University of Berlin (DE))

🕒 5m



17:40 → 18:00 **Belle II: Global Tags**

Speaker: Paul James Laycock (Brookhaven National Laboratory (US))

🕒 20m



18:05 → 18:25 **ATLAS: Analysis releases, calibration area, cross-section metadata**

Speaker: Will Buttinger (Science and Technology Facilities Council STFC (GB))

🕒 20m



18:30 → 18:50 **CMS: Scale Factors and calibration tools**

Speaker: Loukas Gouskos (CERN)

🕒 20m

