# Batch Exercise

Christoph Wissing (DESY), Ben Brueers

FH Sustainable Computing Workshop
January 2024

# Slimming & Skimming

**A very prominent use case**

**Slimming**

- Condense big data formats to more compact ones
- Keep only objects/variables that are really needed
- In this example
    - CMS AOD format is about 300kB/evt (Run 1)
    - Out will only be a few variables per event (not enough to to physics actually!)

**Skimming**

- Select on events with certain properties
- Optional: Apply also calibrations

# Repeating some Best Practices

**Plan your processing**

- One full run through the data can be resource intense
  → Get it right in the first attempt!

- Check with working group colleagues

  – Perhaps the results of the processing can be used also by others

**Test your processing**

- Understand what you are going to do!

- Do not blindly run something that you "inherited"

- Have small test runs and carefully validate your results

**Why this exercise example is actually a bad one**

- For the AOD dataset (**~300kB/evt**) there is already a nanoAOD (**a few kB/evt**) existing

- Somebody in CMS has done the processing already for you