CMS DAS in Hamburg October 13-17, 2025

Second meeting with facilitators
September 4, 2025

Maria Aldaya Martin, Juliette Alimena, Alexander Grohsjean, Matthias Schroeder

Reminders

- Agenda: https://indico.cern.ch/event/1522042/timetable/?view=standard
- More info from 1st meeting in backup

Short exercises

- Combine - limit setting (Kuan-Yu Lin, Suman Chatterjee)
- Combine - unfolding (Beatriz Lopes, Jan Kieseler)
- Machine learning (Joern Bach, Matthias Komm)
- Tracking (Brunella D'Anzi, Jan Schultz, Marco Musich)
- E/gamma (Alberto Belvedere, Ying An)
- Muons (Matteo Bonanomi)
- Jets (Andreas Hinzmann, Dennis Schwarz)
- Flavor tagging (Philipp Gadow, Uttiya Sarkar)
- Taus (Alexi Raspiareza, Andrea Cardini)
- Generators (Dominic Stafford, Mikel Mendizabal)
- REANA (Aliya Nigamova, Pallabi Das)
- Gitlab CI (Aliya Nigamova)
- Applying central corrections (Daniel Savoiu)
- Outreach (Freya Blekman)

- Please design a 2 hour exercise
- Can be repeated up to 2-3 times on Monday and Tuesday, depending on demand and available rooms

Long exercises

- 2D cross section measurement (SMP) - Patrick Connor, Ying An
- Ttbar cross section (TOP) - Evan Ranken, Dominic Stafford, Beatriz Lopes
- Higgs to 4 leptons (HIG) - Daniel Savoiu, Matteo Bonanomi
- Long-lived particle searches (EXO) - Jeremi Niedziela, Soumyaa Vashishtha
- Dark matter (EXO/NPS) - Danyer Perez Adan, Daina Leyva Pernia

- Please design an exercise that will start late on Tuesday afternoon and run through Wednesday (except some talks on Wednesday morning) and Thursday
- Participants will present their results as a group on Friday

Your to do list and questions to answer: pls fill google form

- Prepare your exercise :) Should be ready by today!
- Prepare a short description of the exercise that we can add to Indico such that the students know what to expect
- Are you looking for more facilitators?
  - We want at least 2, ideally one of them being based in Hamburg
- How much storage space do you need per participant? Any other computing requirements?
- We will recycle the pre-exercises from the last DAS at CERN (https://cmsdas-cern-2024.docs.cern.ch/). Do you have anything to add there? Do you need it all? (See next slide)
- Does your exercise depend on another exercise?
  - e.g. PAG inclusive jet depends on POG jets
- Also we request that you **register on indico** as “a facilitator or speaker”
  - https://indico.cern.ch/event/1522042/registrations/
  - No need to pay via the DESY indico
- Anything else?

2024 pre-exercises: Do you need all of this? Anything to add?

1. CMS Data Analysis School Pre-Exercises - First Set	How do you setup a CMSSW release? How do you find a dataset using the Data Aggregation Service (DAS)? What are some EDM standalone utilities and what do they do? What is MiniAOD and how do you use it?
2. CMS Data Analysis School Pre-Exercises - Second Set	How to slim a MiniAOD file? How to know the size of a MiniAOD file? How to use FWLite to analyze data and MC?
3. CMS Data Analysis School Pre-Exercises - Third Set	How do I do an analysis with so much data that I cannot run it interactively on my computer? What is CRAB? How do I use it to run an analysis on the grid? How do configuration files look like? How do I extract the luminosity of the dataset I analyzed?
4. CMS Data Analysis School Pre-Exercises - Fourth Set	How do we analyze an EDM ROOT file using an EDAnalyzer? How do we analyze an EDM ROOT file using an FWLite executable? How do we use ROOT/RooFit to fit a function to a histogram?
5. CMS Data Analysis School Pre-Exercises - Fifth Set	How do I setup git on my computer/cluster? How do I collaborate using GitHub?
6. CMS Data Analysis School Pre-Exercises - Sixth Set	What is Jupyter? What is pyROOT?
7. CMS Data Analysis School Pre-Exercises - Seventh Set	What is an image? How about a container? What is Docker/Singularity? Why is containerization useful? Ummmm…how is this different from a virtual machine?

Other things needed to add to pre-exercises

- Ask participants to extend their AFS storage space on lxplus
- Reorganize the list with more basic at the start
  - Step 0: log into lxplus
  - Then ROOT, pyroot, git
  - Only then more complicated stuff like CMSSW
- Add VScode
- What else?

One other question

- We would like to offer a DESY tour on Friday at 3pm during DAS
- Are any of you available as tour guides? We probably need 2-3 guides
- Will have people sign up for the tour on Monday of DAS

Today

- Separate meetings with short and long exercise facilitators
  - To not make the meetings go too long
- Aim is for your exercises to be basically ready today
  - Take the next ~month to coordinate among ourselves and solve whatever issues come up
- Please report on the status of your exercise and give us your answers to the questions on the previous slide via the google form by September 11

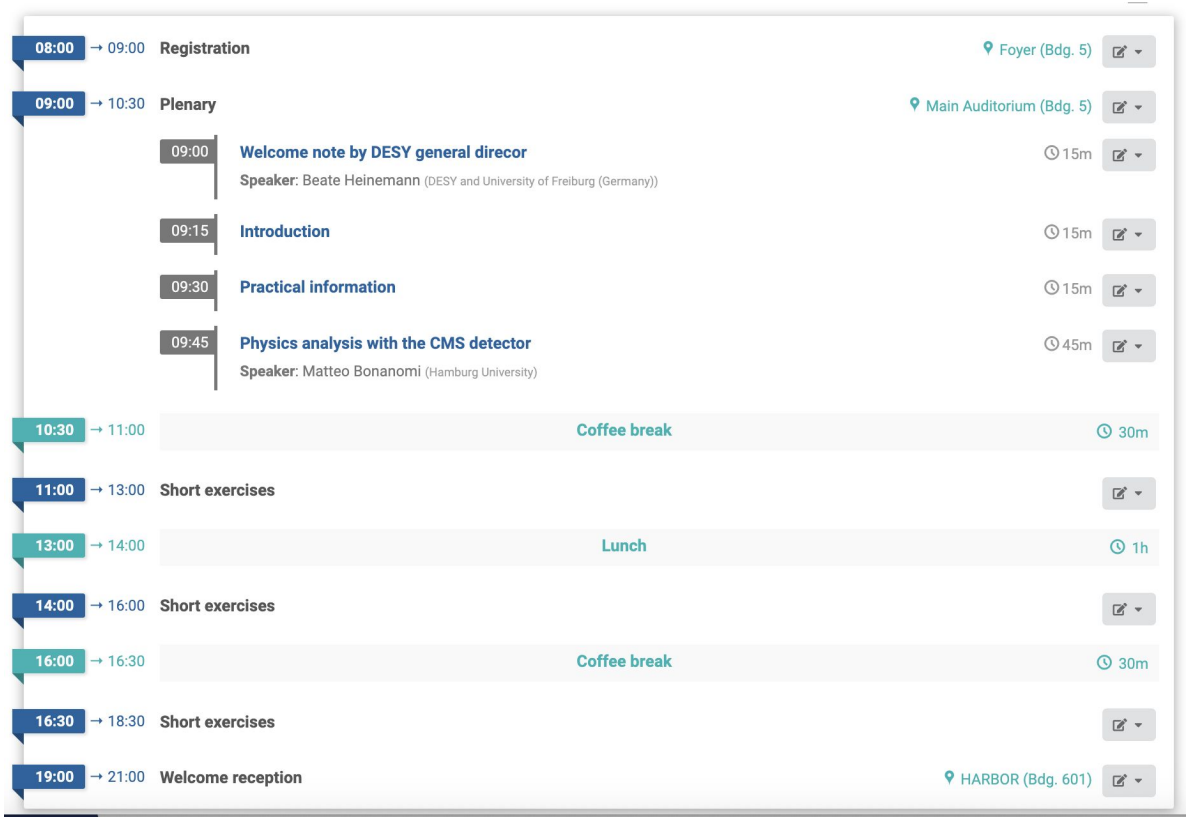
Backup

Intro

- The CMS Data Analysis School (DAS) is given several times per year to introduce newcomers to CMS on how to do an analysis
- Focus on CMS tools and practical hands-on short and long exercises
- Some talks on PAGs, POGs, DPGs, perspective from PC, etc.
- At the end of the week (Friday), the participants will present the results of their long exercise

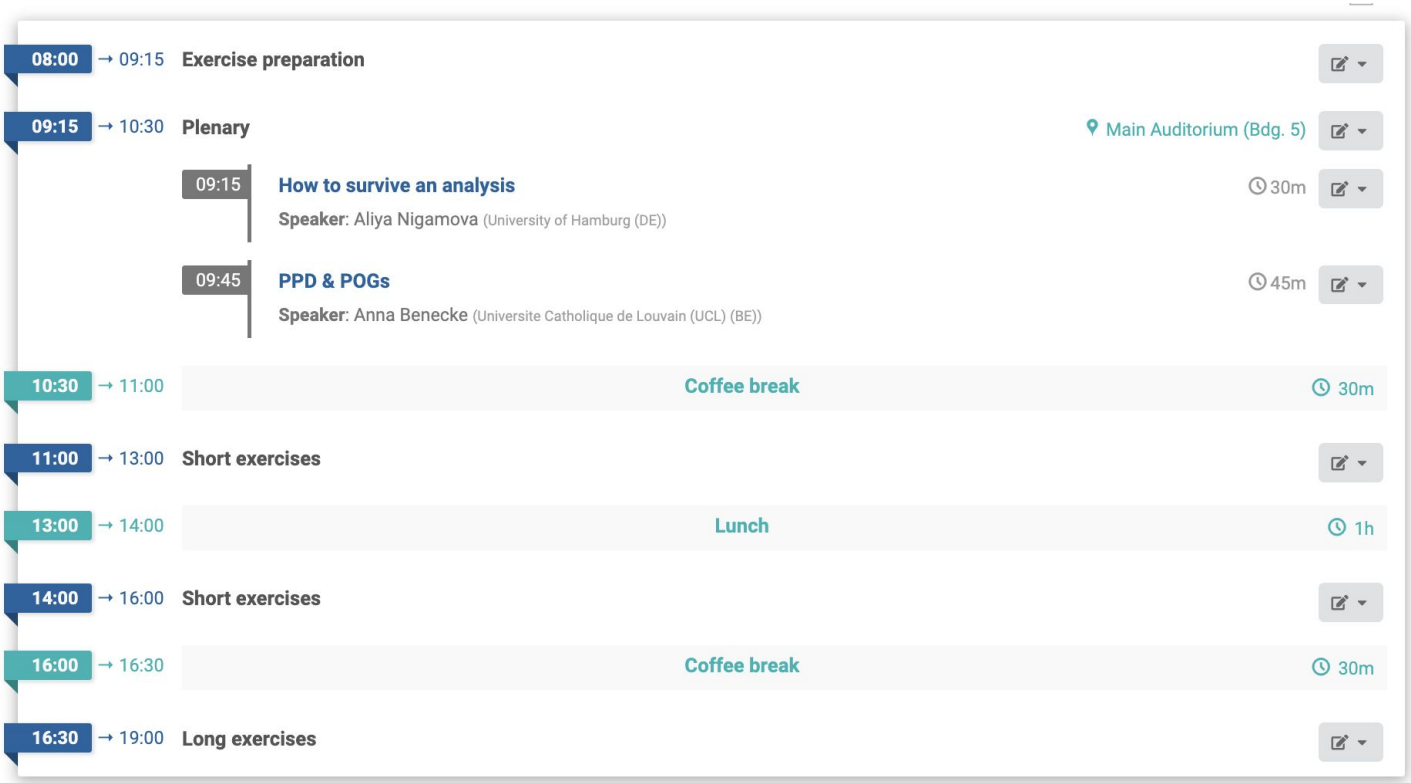
Preliminary agenda: Monday October 13

https://indico.cern.ch/event/1522042/timetable/?view=standard



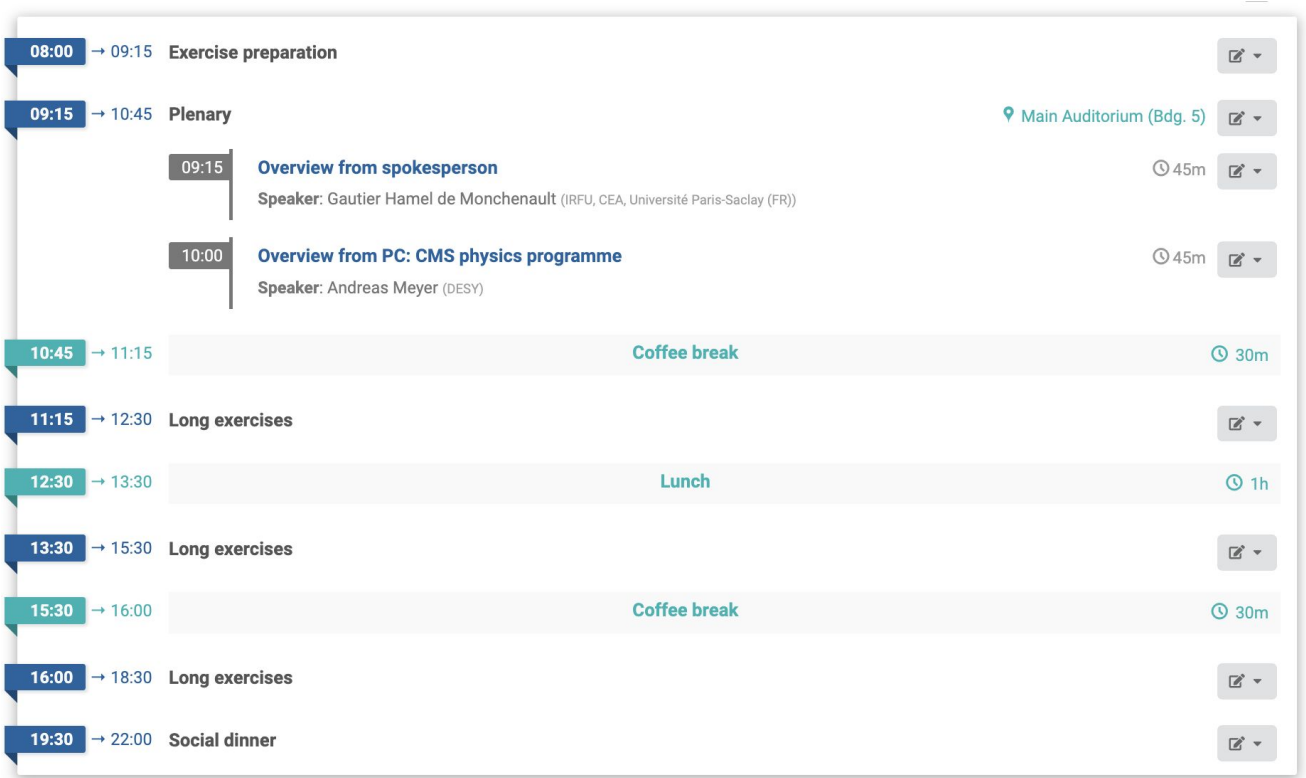
Preliminary agenda: Tuesday October 14

https://indico.cern.ch/event/1522042/timetable/?view=standard



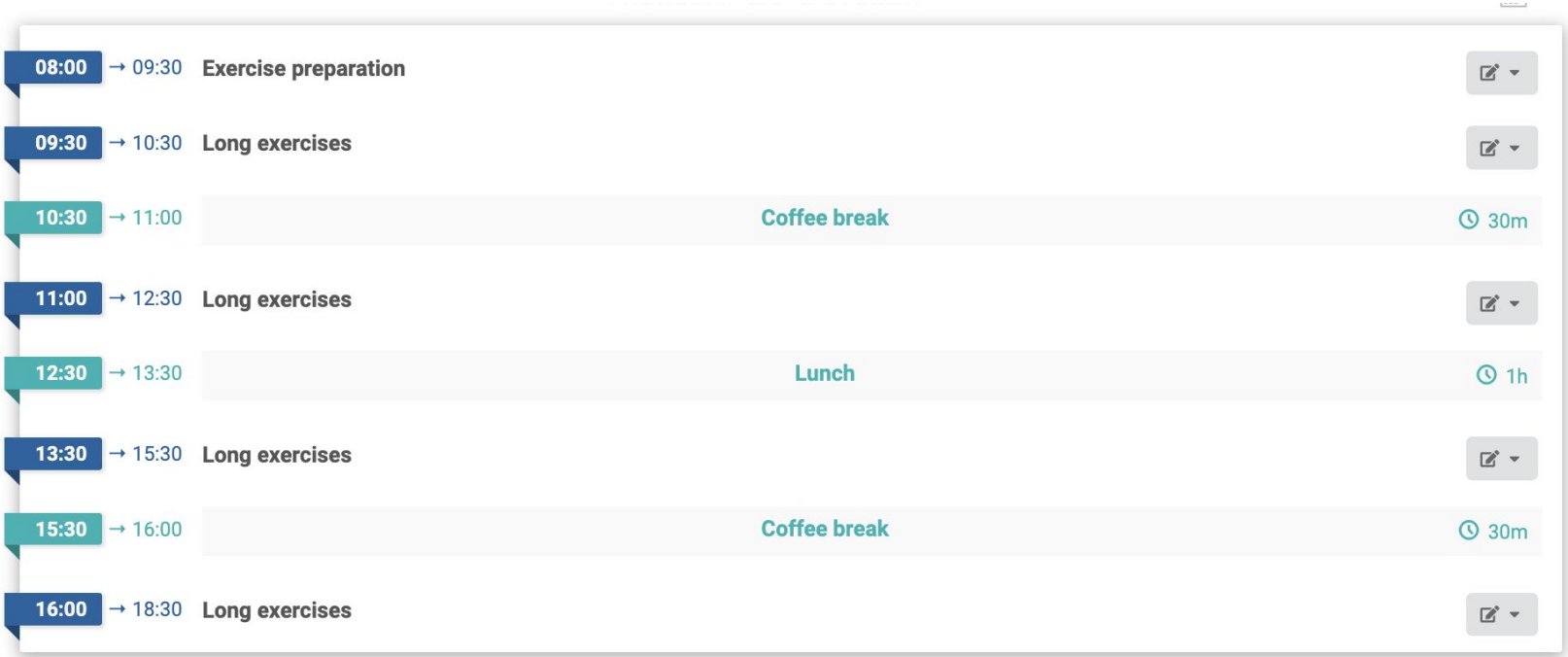
Preliminary agenda: Wednesday October 15

https://indico.cern.ch/event/1522042/timetable/?view=standard



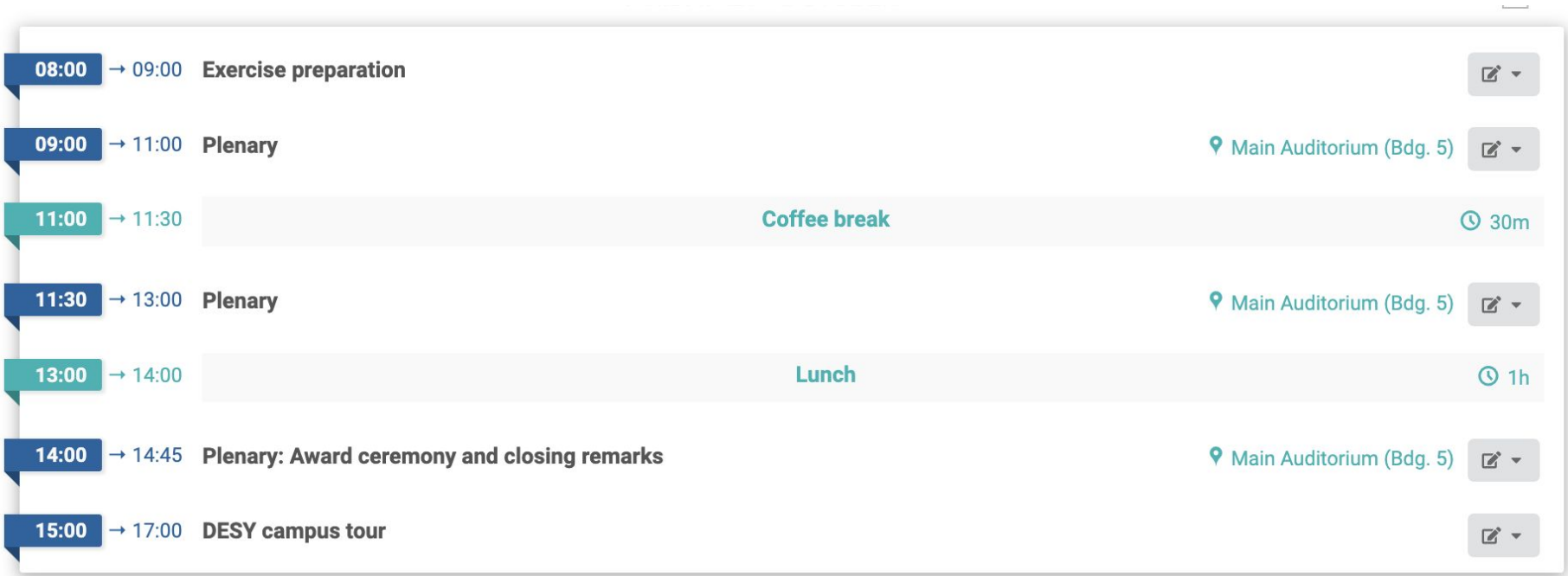
Preliminary agenda: Thursday October 16

https://indico.cern.ch/event/1522042/timetable/?view=standard



Preliminary agenda: Friday October 17

https://indico.cern.ch/event/1522042/timetable/?view=standard



In preparing your exercises

- Should include a very practical approach from the user point of view, rather than a too theoretical approach
- Hands-on
- See previous DASes for examples

Computing resources

- We will use **lxplus** for everything
  - Easiest to interface with previous versions of the exercises
  - Best for future maintenance of code and samples, and better transference to students' future use
- From O&C:
  - Don't expect any major issues running analysis code, as long as the job isn't too long (limit probably around 4 hours)
  - Main concern is the HOME directory, which has a 10 GB quota
    - Checking out code and running it should be fine, but users should avoid producing large outputs there.
  - One option is /tmp, but note that it's machine-specific, you'll need to log in to the same machine each time.
    - Also, if /tmp usage exceeds 90%, the automatic cleanup mechanism will start. More details are here: https://lxplusdoc.web.cern.ch/evolution/fair/tmp/
  - Alternatives:
    - Running on EOS, if the user has EOS space available
    - Requesting additional AFS space (should avoid this, requesting for 50 participants for tutorial does not fit the strategy we have at the moment)
- Avoid relying on the Grid and CRAB
- **tl;dr: Could you estimate how much storage space you'll need per participant?**
- **Any other computing requirements on your side?**