Finding and downloading data from the Bookkeeping

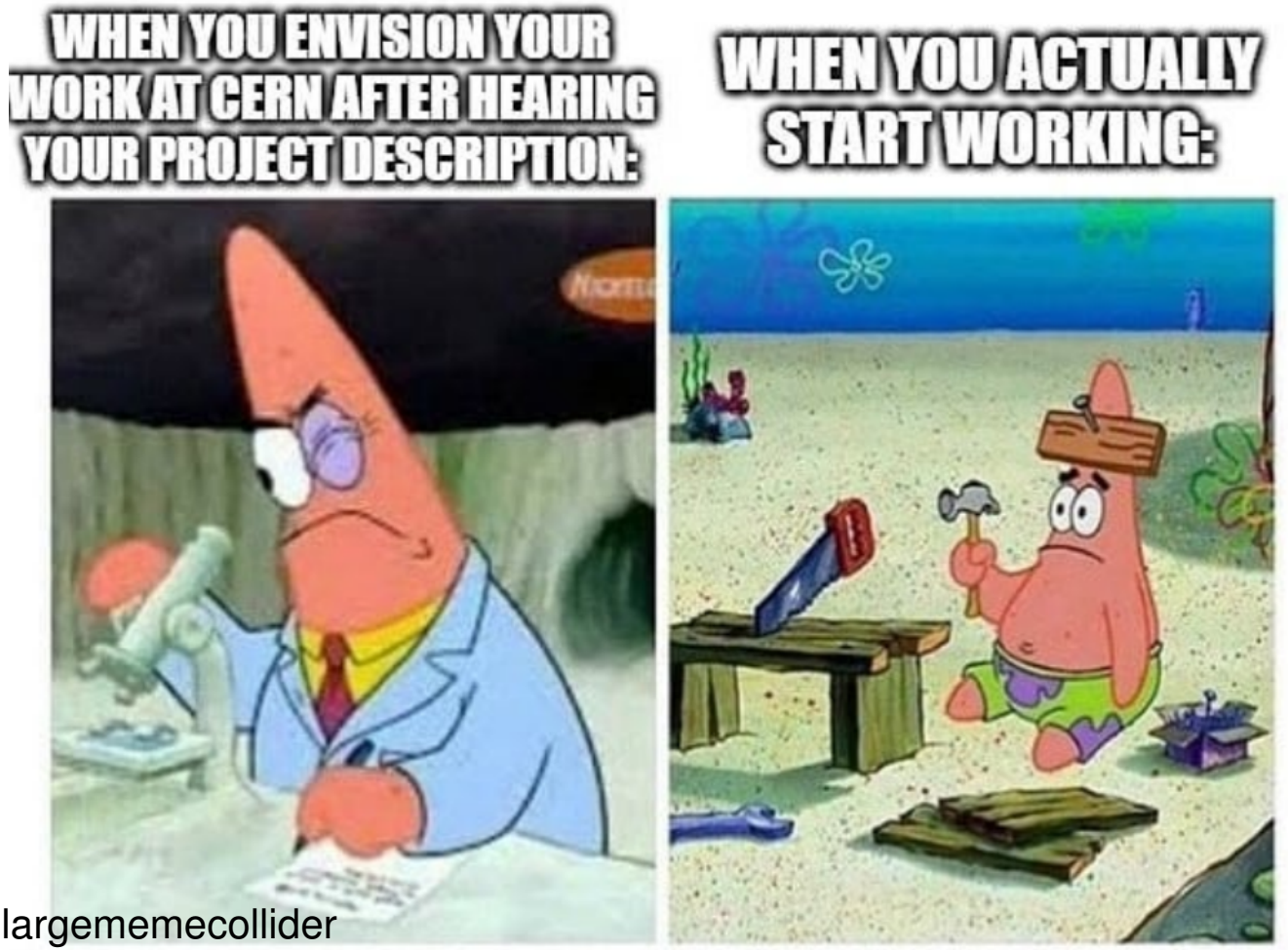
Renata Kopečná

Physikalisches Institut, Heidelberg University

November 4, 2020

Goals of this lecture

- Find your favorite dataset in the bookkeeping
- Find your favorite decay in the bookkeeping
- Download DST file from a grid



Meme credit: @largememecollider

Certificate settings

- In case you have trouble accessing the Bookkeeping, let us know!

- Getting the certificate:
  `https://twiki.cern.ch/twiki/bin/view/LHCb/FAQ/Certificate`
- Exporting certificate to Firefox:
  `https://ca.cern.ch/ca/Help/?kbid=040111`
- Exporting certificate on Mac:
  `https://ca.cern.ch/ca/Help/?kbid=060111`
- The manual is also in the Pre-workshop checklist

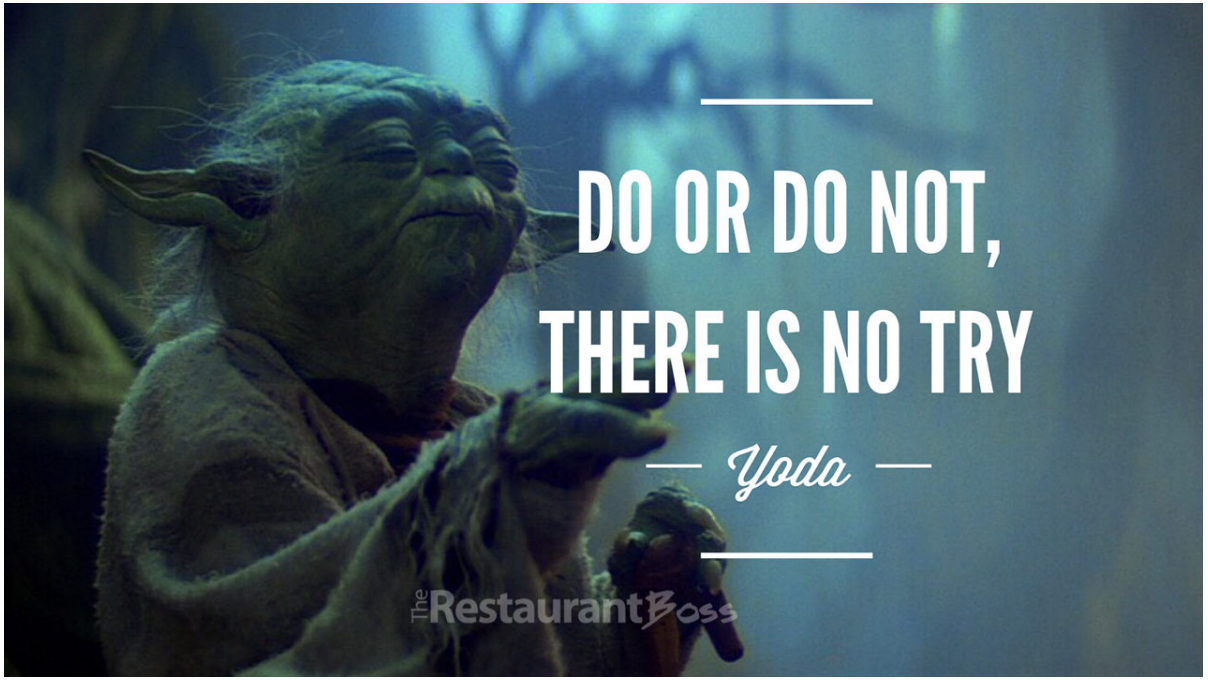
Navigating through Bookkeeping

1. Open `https://lhcb-portal-dirac.cern.ch/`
   - Documentation: here
2. Open Bookkeeping Browser by clicking *once*
3. Lookup according to event type
4. Looking for $D^{*+} \to D^0 (\to K^- K^+)\pi^+$
   - Event type 27163002 (Dst_D0pi,KK=DecProdCut) Internal note, Twiki
   - MagDown, Sim09c, Trig0x6138160F, Reco16, Turbo
   - Flagged
   - Or use the address bar at bottom right!

   `evt+std://MC/2016/27163002/Beam6500GeV-2016-MagDown-Nu1.6-25ns-Pythia8/`

Flagged vs filtered sample

- Filtered
  - All events not passing trigger and stripping are thrown away
- Flagged
  - Events are flagged accordingly if they do or do not pass the trigger and stripping
  - Useful in MC to study the rejected events



Saving the Bookkeeping addresses

0. Right bar contains information about your dataset
1. `Save as...`
   - The full data sample is divided into several LFS files, each typically has several hundreds of MB
   - One can select to save only the address of *n* files in given range
2. Save as .py somewhere you can find it
   - Make sure the file name has .py at the end
   - You can also just change its name later after you download it

- In case you cannot access the file, here is a link:
  ALLSTREAMS.DST.py

Move the file to lxplus

- Open a lxplus session
- Do `lhcb-proxy-init` (uploads proxy to DIRAC)
- I highly recommend moving to your `/work` directory and creating a new folder there:

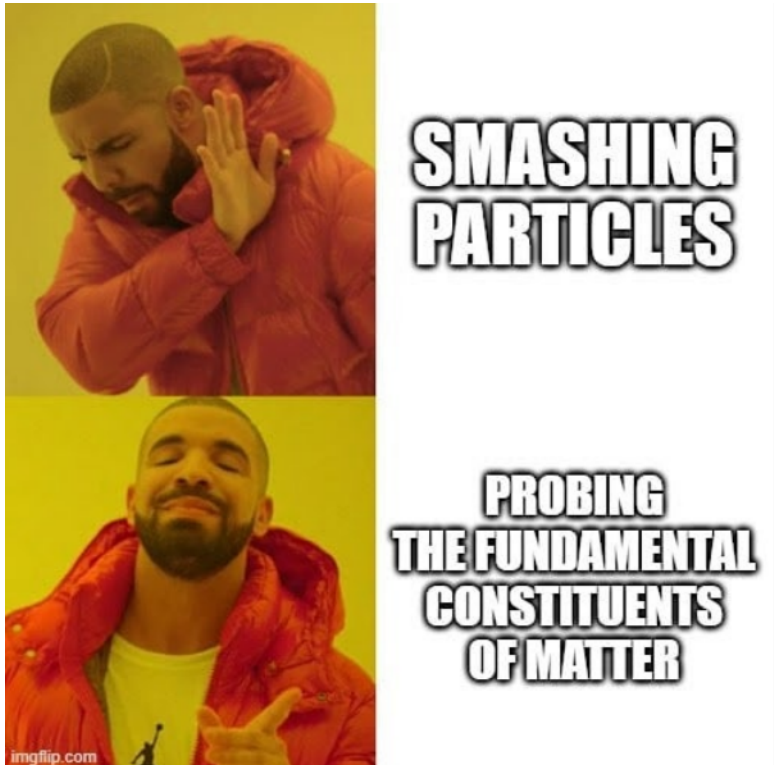
```
cd /afs/cern.ch/work/u/username/
mkdir Starterkit
```

  - We will be downloading files from the grid which can be large
  - In case you don't have enough space and you haven't done so yet, request more space in your work dir here: `https://resources.web.cern.ch/resources/Help/?kbid=067040`
- Copy the download `.py` file onto lxplus

```
scp MC_2016_..._ALLSTREAMS.DST.py
user@lxplus.cern.ch:/afs/cern.ch/work/u/username/Starterkit
```

Play around

- Search for your favorite decay
- The default is sorting by Simulation Condition
  - Try searching for 27163002 (Dst_D0pi,KK=DecProdCut) using this sorting
  - ! The path in the address bar changed!
    `sim+std://MC/2016/Beam6500GeV-2016-MagDown-Nu1.6-25ns-Pythia8/Sim09c/.../27163002/ALLSTREAMS.DST`

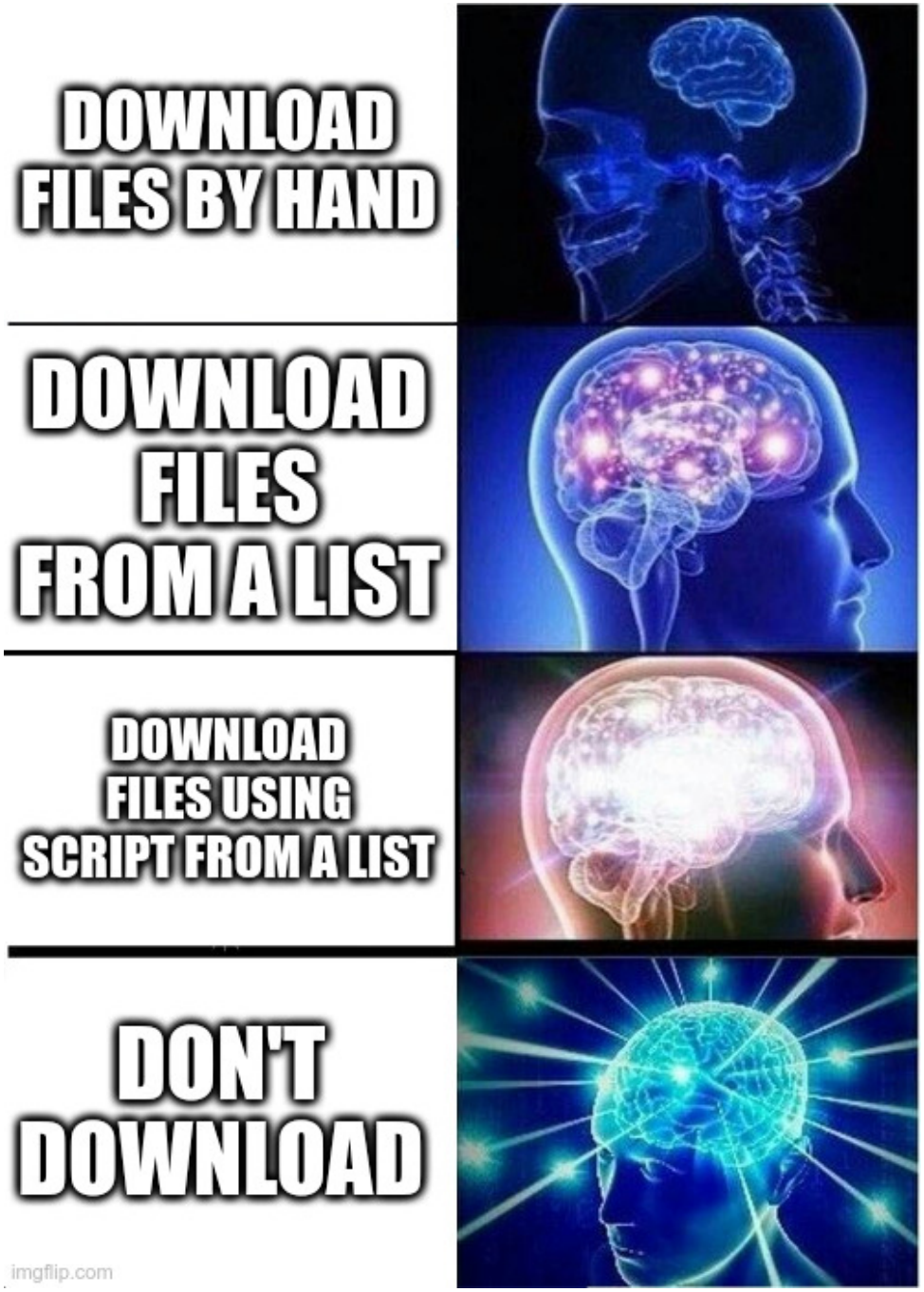


Download a file from the grid

1. Open the file and see what is inside

`LFN:/lhcb/MC/2016/ALLSTREAMS.DST/00070793/0000/00070793_00000001_7.AllStreams.dst`

- LFN = Logical File Name
- LFN:/exp/data or MC/year/data type/production ID/ file number.YourStream.M/L/DST
  - From production ID, one can obtain request ID
  - In BKK click on the left on Data
  - Select Transformation Monitor, put request ID
  - From there, you can see details of the production by right-clicking on the production, such as show request, which opens a request manager with the simulation steps (right click and `View`)


DOWNLOAD FILES BY HAND
DOWNLOAD FILES FROM A LIST
DOWNLOAD FILES USING SCRIPT FROM A LIST
DON'T DOWNLOAD
imgflip.com

Download a file from the grid

- To download the file, standard procedure is to do
  `lb-dirac dirac-dms-get-file LFN:/lhcb/MC/2016/ALLSTREAMS.DST/00070793/0000/00070793_00000001_7.AllStreams.dst`
- There is a lot of us, it would take super long, use
  `xrdcp root://eosuser.cern.ch//eos/user/l/lhcbsk/data-sets/00070793_00000001_7.AllStreams.dst ./`
- This downloads the DST directly to the location you're at

Download a file from the grid, smarter

- In case you need to download more files, instead of

```
lb-dirac dirac-dms-get-file
LFN:/lhcb/MC/2016/ALLSTREAMS.DST/00070793/0000/
00070793_00000001_7.AllStreams.dst
```

  you can do (not now!)

```
lb-dirac dirac-dms-get-file
--File=MC_2016_27163002_Beam6500GeV2016MagDownNu
1.625nsPythia8_Sim09c_Trig0x6138160F_Reco16_Turbo03
_Stripping28r1NoPrescalingFlagged_ALLSTREAMS.DST.py
```

- This downloads all the files listed in your .py we just downloaded from the bookkeeping
- You can also write a python script to download only *n* files from the .py (see Starterkit)

Download from the grid, even smarter

- In case you need play around with the LFN paths, instead of downloading the .py manually, you can list the available LFNs using

```
dirac-bookkeeping-get-files
```

```
 lb-dirac dirac-bookkeeping-get-files
--BKQuery=/MC/2016/
Beam6500GeV-2016-MagDown-Nu1.6-25ns-Pythia8/
Sim09c/Trig0x6138160F/Reco16/Turbo03/
Stripping28r1NoPrescalingFlagged/27163002/
ALLSTREAMS.DST
```

Not downloading a file from the grid

- For this exercise, we need to download the files actually. This is often not the case, *e.g.* when you just need to test your DaVinci script.
- Create a xml catalog! (Also not now!) `lb-dirac dirac-bookkeeping-genXMLCatalog -Options=MC_2016_27163002_..._ALLSTREAMS.DST.py -Catalog=myCatalog.xml`
- Okay, but this requires to search for the correct file in BKK, which is a lot of clicking unless you already know the path...
- If you know your decay ID, simply do
  `lb-dirac dirac-bookkeeping-decays-path 27163002`
  - This also shows you the simulation conditions!

That's that from me!

- Now you will play around with the .DST file you just downloaded
- If you have any more questions, mattermost/email me anytime!

BACKUP