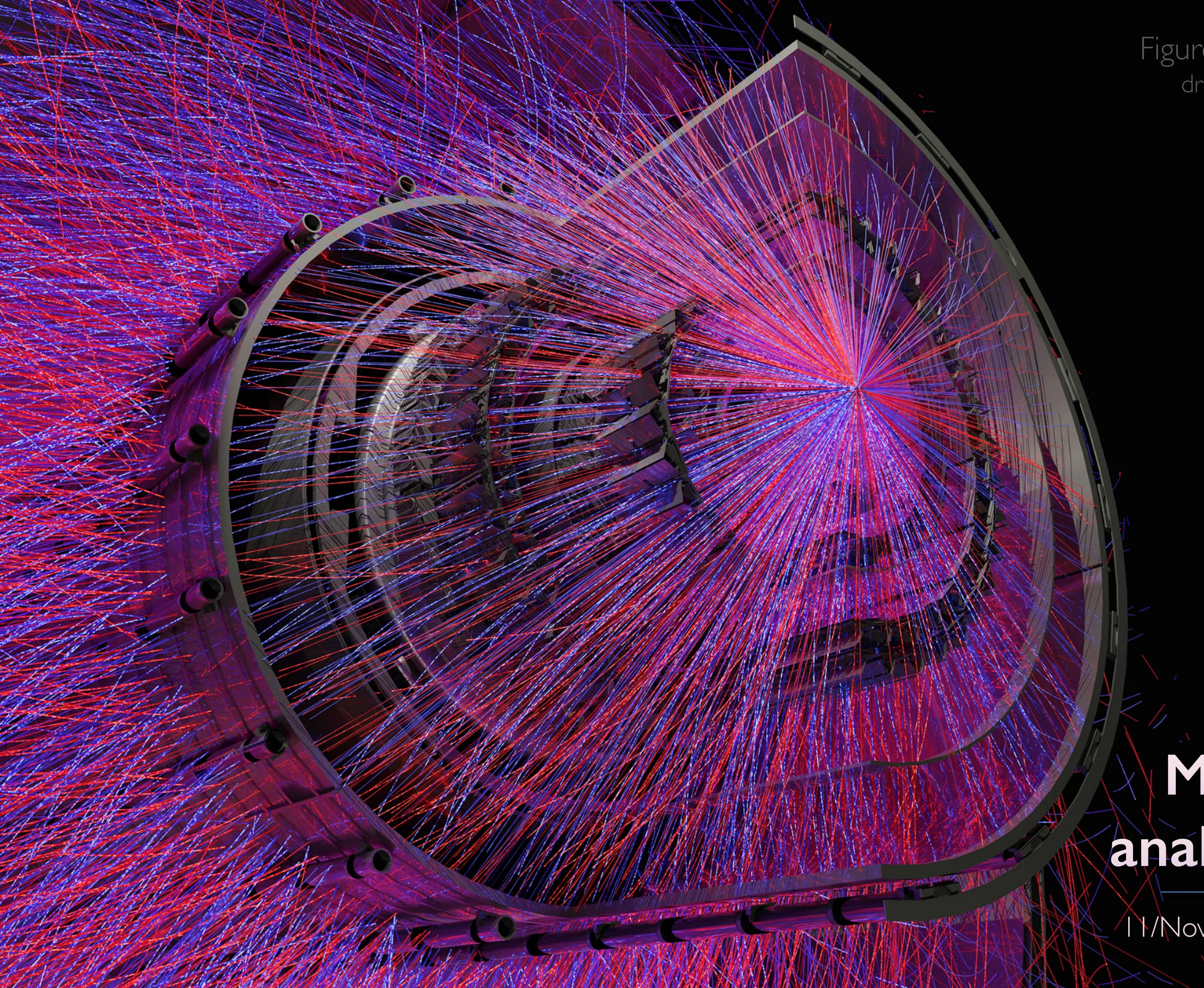


Figure: the ITS detector used in Runs 1 and 2
drawn together with a high-multiplicity Pb-Pb collision

Modern Run 2 data analysis with O2Physics

11/November/2025 - David Dobrigkeit Chinellato

The data from Runs 1 and 2: why is this interesting?

- "ALICE 1": non-continuous readout, simple collision association
  - Plenty of high-quality data available for analysis!
  - Hundreds of papers have been written using this data!
- The previous data format was not particularly disk-efficient
  - Approximate total size of Run 1/2 data: 80 PB
- Usual analysis in the past was done via the LEGO train system
  - This still works fairly today, many years since its inception
  - This system runs on a server that's no longer maintained
- With the advent of O2/O2Physics, the question arose:
  - **Can we use O2/O2Physics for analysing this legacy data?**

Collision system	Energy(ies)
pp	5.02, 7, 13 TeV
p-Pb, Pb-p	5.02, 8.16 TeV
Xe-Xe	5.44 TeV
Pb-Pb	2.76, 5.02 TeV

The conversion from AliPhysics to O2/O2Physics data format

- The run 1 and 2 data format consisted of either ESDs (Event Summary Data) or AOD (Analysis Object Data)
- The ESD format is very complete, but very large in terms of data volume!
  - Arrays of structures approach (AOS) instead of structures of arrays
  - AliESDEvents contained AliESDTracks, AliESD[something]
  - Multiple copies of track parametrizations were stored (TPC-only, refitted, etc)

- Conversion to O2/O2Physics data format was first done in 2022 in a conversion campaign and continues taking place
  - Requires some specific care with the data model: see next slides!
  - **Converted data allows for fast analysis of legacy data with Hyperloop using O2/O2Physics analysis code!**
  - At the same time: native old format data uses **~80 PB**, but the entirety of converted data is **~800 GB today**!

- If you want to look at the old data, **please use O2/O2Physics + Hyperloop** and not AliPhysics + LEGO train
  - AliPhysics and especially LEGO trains will at some point stop working as we cannot maintain the server indefinitely
  - If we could archive the ~80 PB of old data and use only the new format, that would be fantastic!

→ **How can Run 2 data be analysed in O2Physics? Only minor differences versus Run 3!**

Notable differences: the stored data model for Run 2

this slide only lists tables that are different or differently handled!

Entity type	Run 3	Run 2	Comment
Run2BCInfos	none	aod::Run2BCInfos	Stores Run 2 event cuts, trigger information
FV0C	none	aod::FV0C	Only FV0A exists in Run 3
Tracks	aod::TracksIU	aod::Tracks	• Run 3: tracks provided at innermost update • Run 2: tracks always at PV
Run2TrackExtras	None	aod::Run2TrackExtras	• Contains ITS analogue readout signal, shared clusters • Run 3 ITS has only digital readout
On-the-fly V0s	none	aod::Run2OTFV0s	• Special type of V0s generated during tracking • Does not exist in Run 3 • present only in recent conversions
PMD data	none	aod::Pmds	Detector not present in Run 3, present in recent conversions
FMD data	none	aod::Fmds	Detector not present in Run 3, present in recent conversions

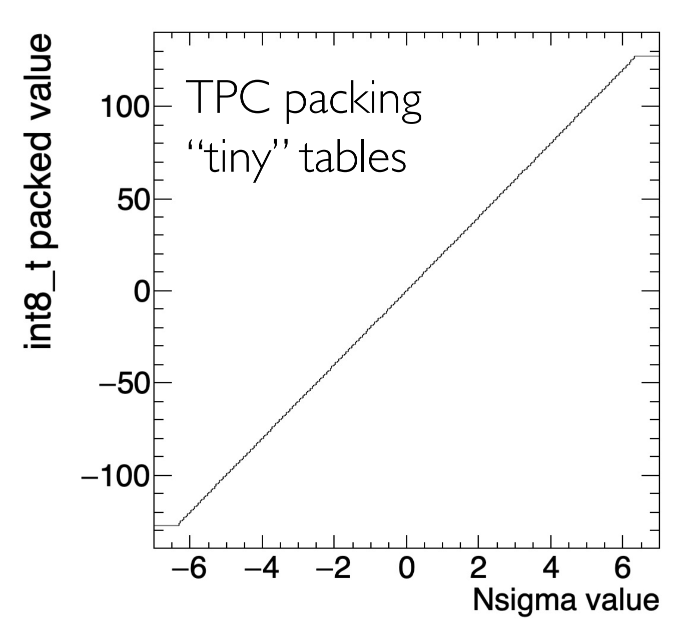
New!

- N.B.: Some tables are Run 3 exclusive and not mentioned above! They include: MFT tables, ambiguous track tables, strangeness tracking tables, track QA table

Ensuring reproducibility: some extra stored information

- In addition to standard tables, and given the small size of the Run 1 and 2 data, in recent conversions:
  - **A few tables traditionally produced by helper tasks have been stored to converted AO2D**
  - This ensures that we **don't need to convert a decade's worth of calibration databases!**
  - This has the side effect that a few heavy core service tasks can be totally skipped

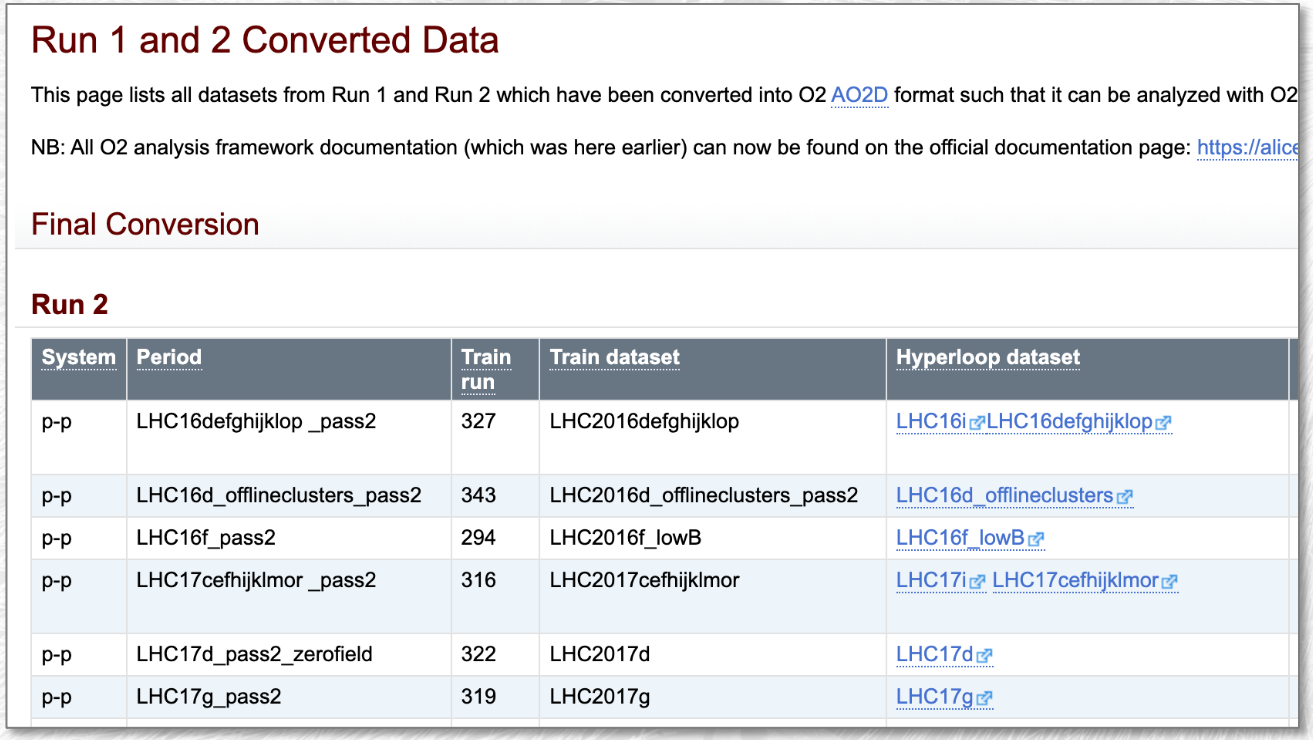
- Tables stored to disk:
  - **TPC N_{sigma} tables for all species**: dispenses with the TPC PID task altogether
    - N_{sigma} encoded into dynamic range from -6.3 to +6.3 with 0.05 precision
    - De-encoder needs to be run to produce "Full" tables
    - Provided for $e^{\pm}$, muons, pions, kaons, protons, deuterons, tritons, he-3 and alpha
  - **Centrality tables:** dispenses the use of the multcenttable task altogether
    - Provided: V0M, V0A, CL0, CL1, reference multiplicity in $|\eta| < 0.5$ and $|\eta| < 0.8$
  - **HF 2-prong, 3-prong, D*, cascade index tables:** dispenses HF vertexing
    - Stores indices meant for building with HF tools (similarly to V0s and Cascades)



Available converted data in various formats

A large conversion campaign was carried out in 2022

- Converted dataset list can be found here
- This should be good for most purposes but does not include latest developments



Run 1 and 2 Converted Data

This page lists all datasets from Run 1 and Run 2 which have been converted into O2 AO2D format such that it can be analyzed with O2

NB: All O2 analysis framework documentation (which was here earlier) can now be found on the official documentation page: https://alice

Final Conversion

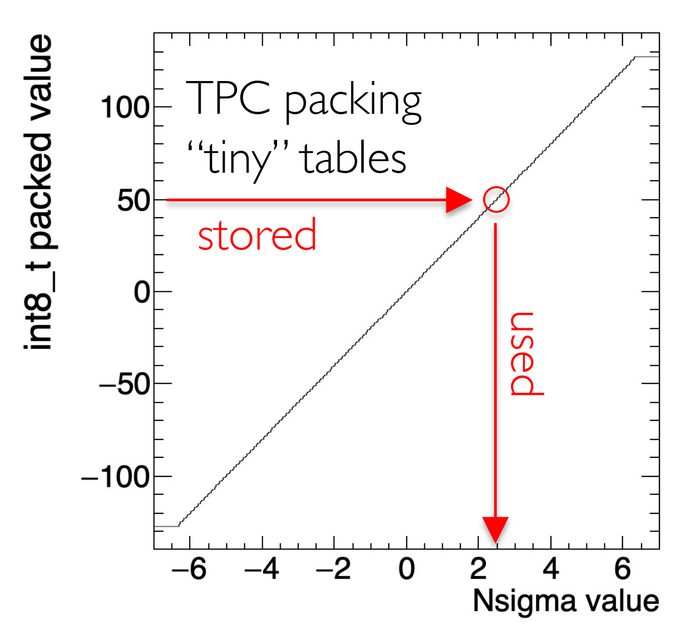
Run 2

System	Period	Train run	Train dataset	Hyperloop dataset
p-p	LHC16defghijklop _pass2	327	LHC2016defghijklop	LHC16i LHC16defghijklop
p-p	LHC16d_offlineclusters_pass2	343	LHC2016d_offlineclusters_pass2	LHC16d_offlineclusters
p-p	LHC16f_pass2	294	LHC2016f_lowB	LHC16f_lowB
p-p	LHC17cefhijklmor _pass2	316	LHC2017cefhijklmor	LHC17i LHC17cefhijklmor
p-p	LHC17d_pass2_zerofield	322	LHC2017d	LHC17d
p-p	LHC17g_pass2	319	LHC2017g	LHC17g

- If one of the recent features is desired, then a reconversion can be performed on demand
- **The LHC15o dataset and a few selected pp datasets** have already been reconverted with the latest features
  - See an example exercise later!
- **A larger reconversion campaign to replace the existing data is in progress** pending cross-checks from EMCal, PMD and FMD

Run 2-specific core service wagons on Hyperloop

- o2-analysis-run2-tiny-to-full-pid
  - The TPC PID Nsigmas that are stored are in the "Tiny" format, which encodes information in int8_t variables
  - Unpacking this is necessary for analysis tasks that require tables in the pidTPCFull[species] format
  - Extremely fast and lightweight (no database access) but still necessary
- o2-analysis-track-dca-cov-filler-run2
  - In Run 2, Tracks are provided already in their position closest to the PV
  - Fully dispenses the propagation of tracks to PV…
  - …but TrackDCAs and corresponding covariance matrices are not stored!
  - This track calculates them and populates the usual tables: this is a fast operation
- o2-analysis-propagationservice-run2 (n.b.: different executable!)
  - Creates standard V0, Cascade information (does not create Tracks)
- o2-analysis-event-selection-service-run2 (n.b.: different executable!)
  - Creates timestamps / event selection criteria
  - Old kINT7 selection: use **collision.sel7() && collision.alias_bit(kINT7)**
  - Old kMB selection (recommended only in Run 1 pp!): special instructions, see backup
  - More info in documentation page



Technical: some cases require different executables for now but will eventually be unified into a single executable pending the capability to customize workflows with defineDataProcessing directives and metadata

An exercise for homework: Run 2 analysis example

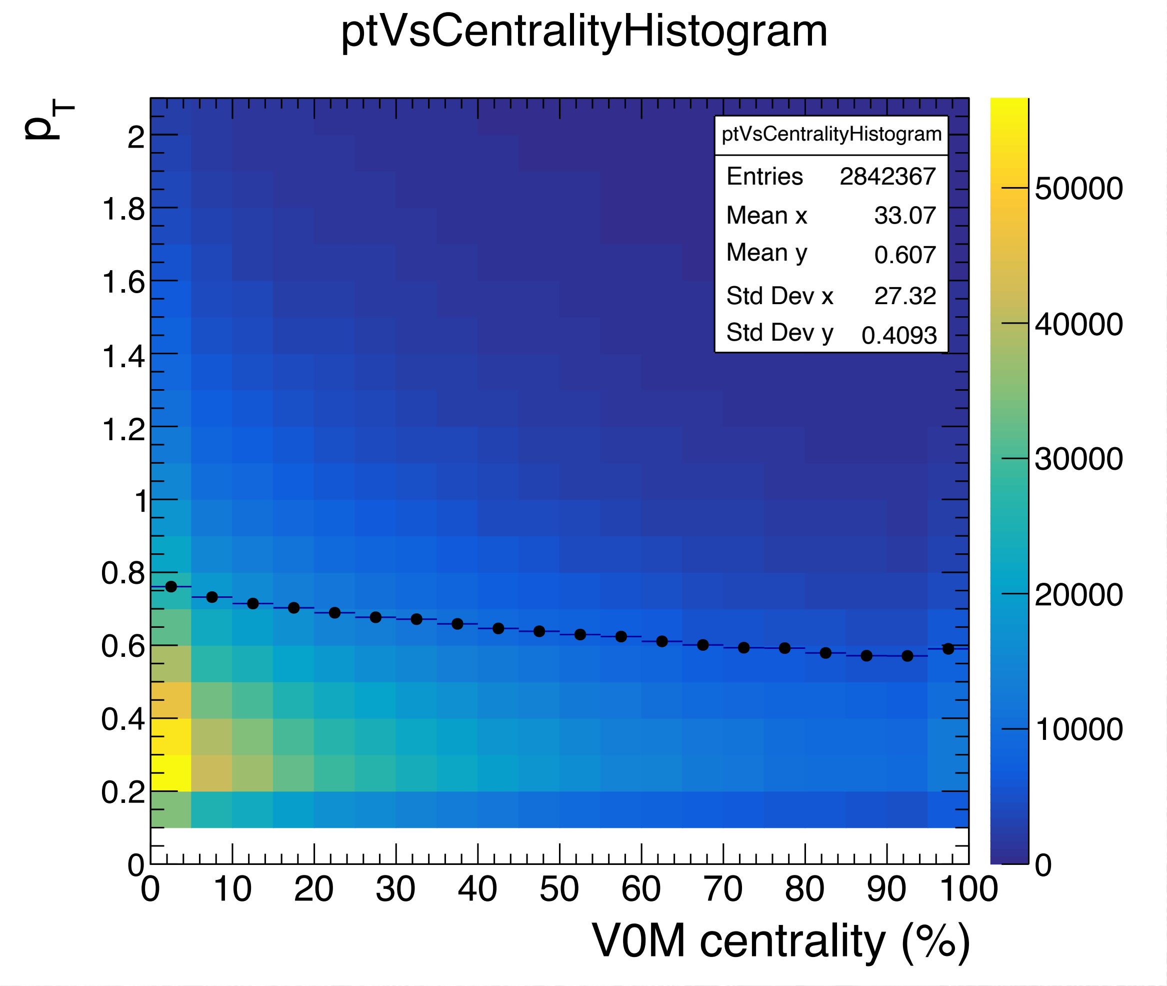
- The sample: a single AO2D from LHC16 pp 13 TeV converted data
  - Download here: dropbox, cernbox
- **Try to play around with this file!** The example task at the end of the first hands-on will work wonderfully, since it subscribes to `Collisions`, `Tracks`, `TracksExtra` and `TracksDCA`. You will only need **two helpers**: the tracks-dca-cov-filler-run2 to fill `TracksDCA` and event-selection-service
- You can then seamlessly add a subscription to `CentRun2V0Ms` [1], for instance, in a join with Collisions, and you will be able to access centrality percentiles via collision.centRun2V0M() without adding any centrality helper!
- This centrality was pre-calculated with the tried-and-true centrality framework and just passed along the conversion!
- **As an exercise, can you calculate the raw <p_T> for all charged particles versus V0M centrality percentile?**
- You can also play with track selection criteria to stabilize the result
  - Note that the Run 2 Tracks table will include "tracklets": short two-point tracks created in the SPD
  - This is already removed with our number of crossed rows cut from the example task!

[1] Note that you will need to add the corresponding data model include:

```
#include "Common/DataModel/Centrality.h"
```

Result of the exercise

Feel free to play around with it further!



- Still other items missing: event selection, efficiency
  - meant solely as a demonstration!
  - Event selection: simply check sel7()
  - Efficiency: requires similar developments as in yesterday's second hands-on and corresponding application

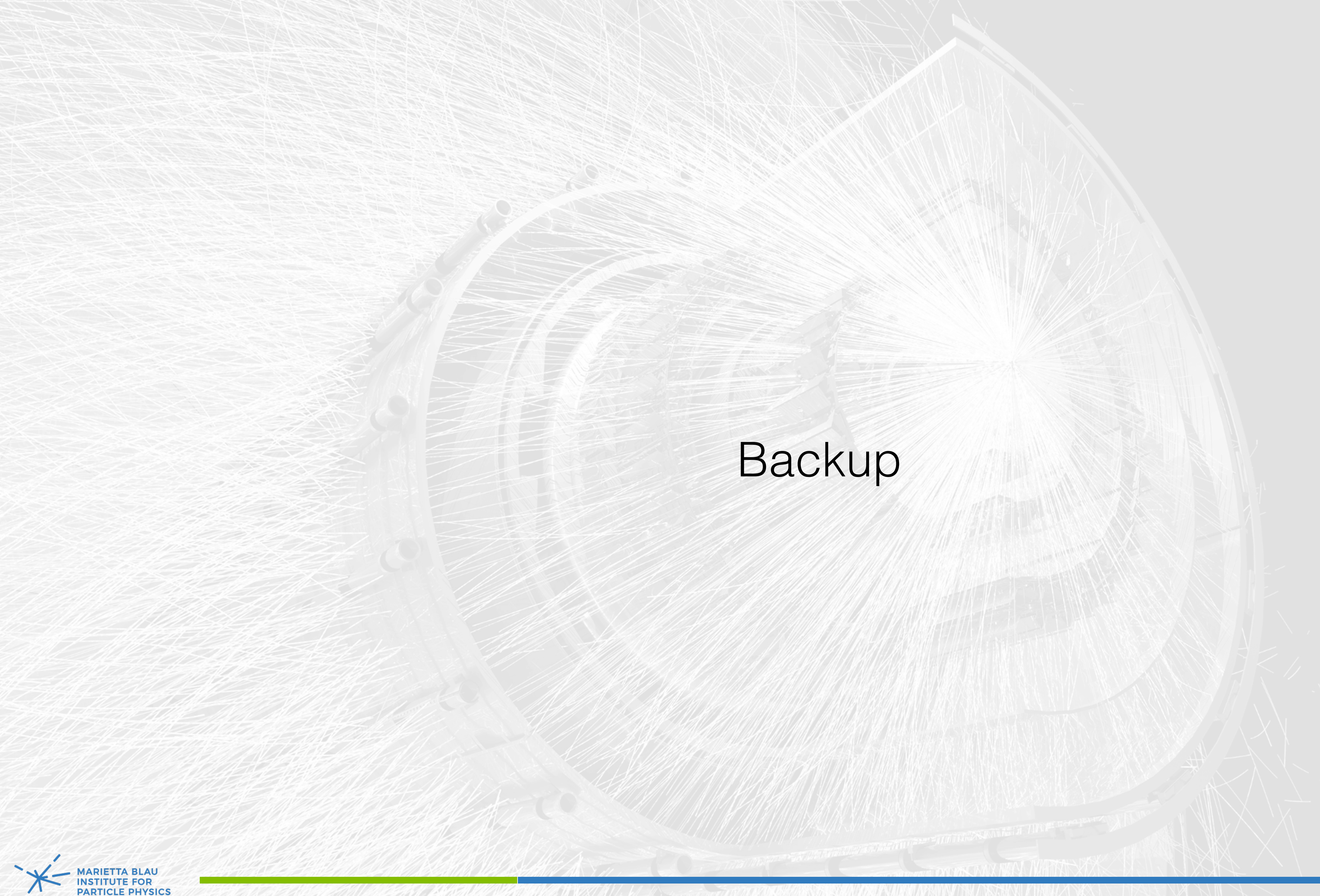
Summary

- **ALICE Run 1 and 2 data is a very good target for analysis!**
  - **High-quality data that is well understood, including pp, p-Pb/Pb-p, Xe-Xe and Pb-Pb**

- But: we should avoid using AliPhysics and LEGO trains and should rather be using O2Physics and Hyperloop
  - It's much faster and coding is ultimately much more convenient
  - The legacy system is already becoming increasingly more difficult to maintain

- Final version of converted data with latest features remais a challenge to organize
  - Will be carried out as soon as input on features added since first campaign have been fully validated
  - If you want to get on this activity, please reach out!

Thank you!



Backup

Event selection in converted data

https://aliceo2group.github.io/analysis-framework/docs/analysis-tools/EventSelection.html#basic-usage-in-user-tasks

```
// kINT7 equivalent – everything that’s not pp Run 1
bool selected = collision.sel7() && collision.alias_bit(kINT7);

// kMB equivalent – ONLY for pp Run 1 data
// first check: sel1
bool sel1 = collision.selection_bit(kIsINT1);
      sel1 = sel1 && collision.selection_bit(kNoBGV0A);
      sel1 = sel1 && collision.selection_bit(kNoBGV0C);
      sel1 = sel1 && collision.selection_bit(kNoTPCLaserWarmUp);
      sel1 = sel1 && collision.selection_bit(kNoTPCHVdip);
bool selected = sel1 && collision.alias_bit(kINT7); // <- recommended selection
```