

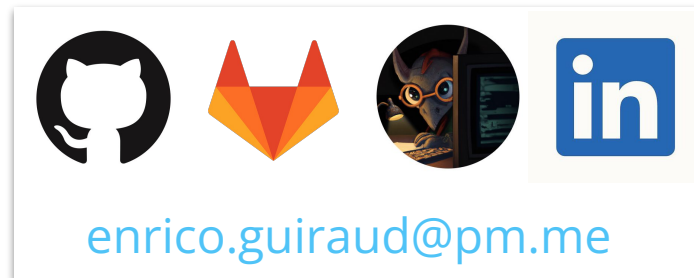
# Final fellowship report

Enrico Guiraud  
SFT meeting, 27/2/2023

# About me



- Physics student at Uni Milan (2010-2016)
- Openlab summer student @ ROOT (2015)
- Master's thesis in unsupervised learning @ Uni Oldenburg (2016)
- Gentner scholarship for doctoral @ ROOT & Uni Oldenburg (2016-2019)
- **Senior fellowship @ ROOT (2020-2023)**
- Collaboration w/ Princeton on HEP analysis tools (½ time, 2023)





What this fellowship  
was about



# What this fellowship was about

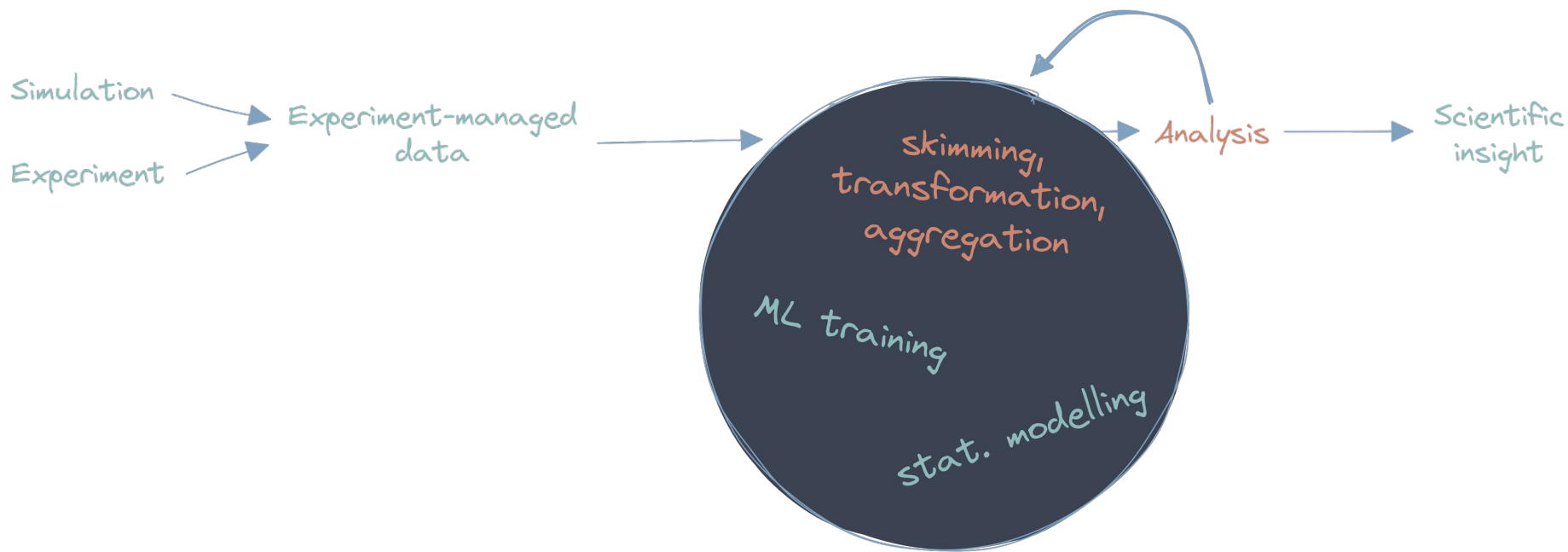
In short, **making ROOT users happier** in the context of HEP data analysis.





# What this fellowship was about

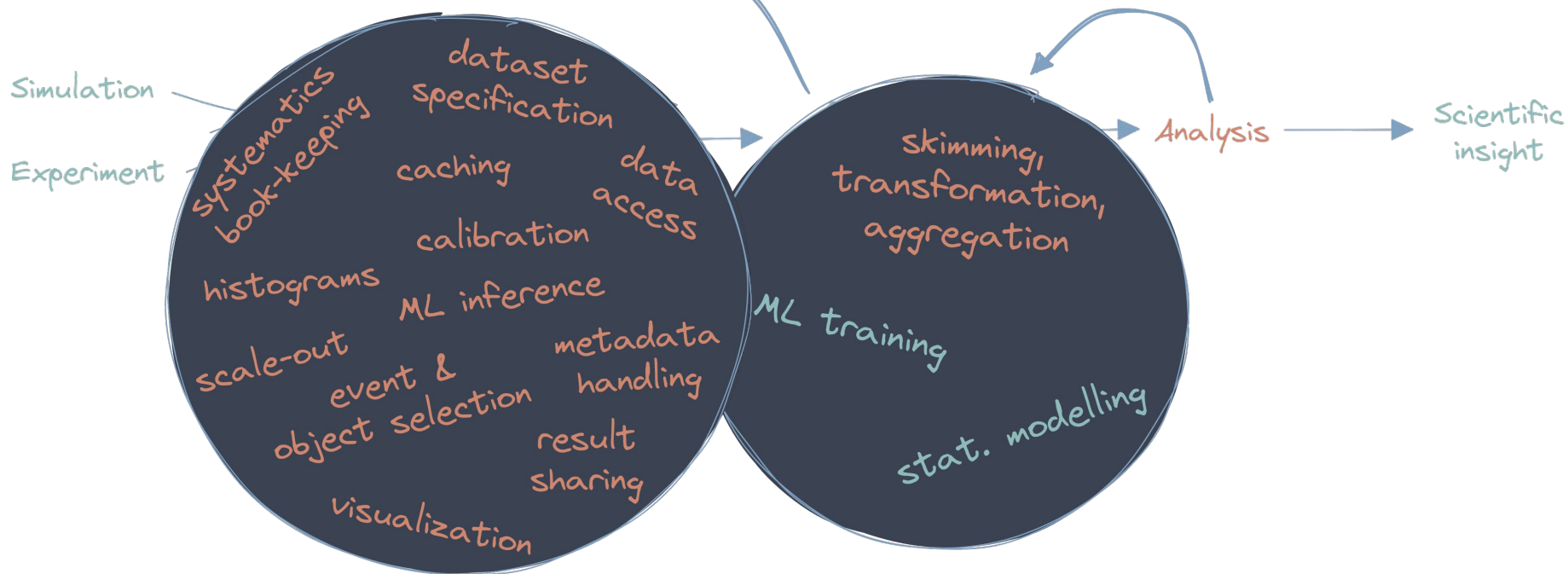
In short, **making ROOT users happier** in the context of HEP data analysis.





# What this fellowship was about

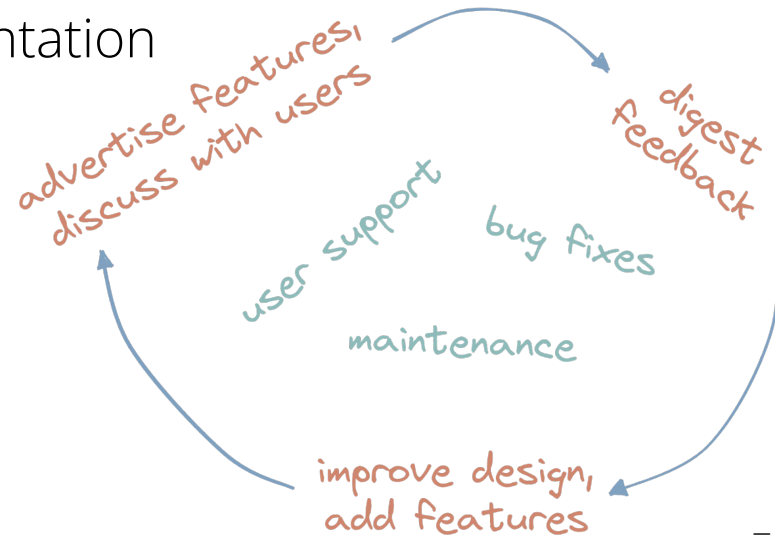
In short, **making ROOT users happier** in the context of HEP data analysis.





# The approach

- improve UX of ROOT's analysis interfaces
- optimize performance of (parallel) data processing
- advertise new features
- provide tutorials and documentation
- user support and bug fixes
- continuous "market research"





What this fellowship was about  
(concretely)





# Technical work

- code owner of **ROOT RDataFrame** and related libraries, tools ([887 merged PRs](#), [133 closed GitHub issues](#), [85 closed JIRA tickets](#))



# Technical work

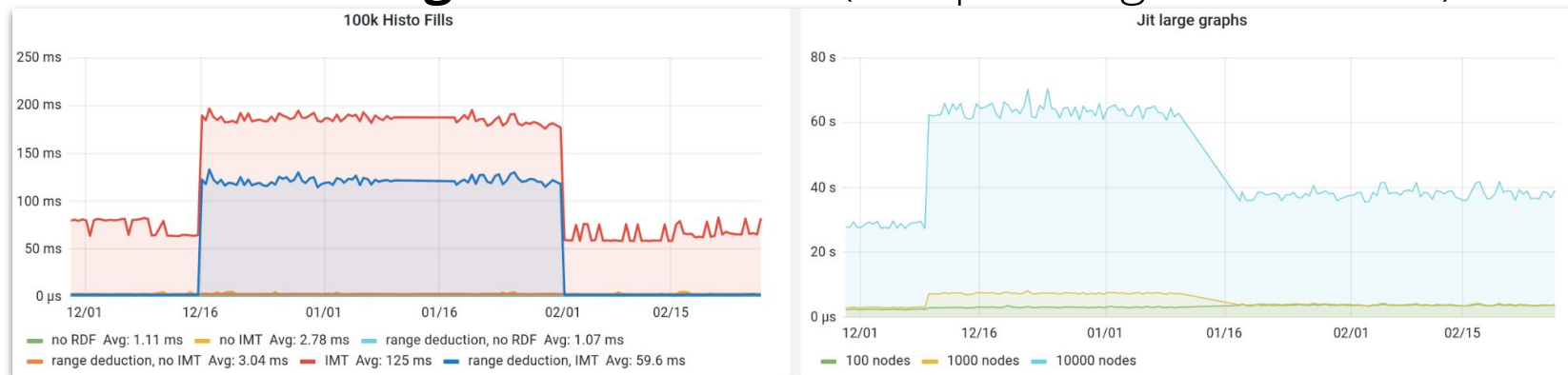
- code owner of **ROOT RDataFrame** and related libraries, tools ([887 merged PRs](#), [133 closed GitHub issues](#), [85 closed JIRA tickets](#))
- CI setup for [nightly builds of ROOT conda packages](#)

```
$ REPO=https://root.cern/download/conda-nightly/latest  
$ conda create -n root-nightly -c $REPO -c conda-forge root-nightly
```

- responsible for ROOT's [official Docker images](#)

```
$ docker run -it rootproject/root  
$ docker run -it rootproject/root:6.24.06-centos7
```

- code owner of **ROOT RDataFrame** and related libraries, tools ([887 merged PRs](#), [133 closed GitHub issues](#), [85 closed JIRA tickets](#))
- CI setup for [nightly builds of ROOT conda packages](#)
- responsible for ROOT's [official Docker images](#)
- integration of real analysis applications into our **continuous benchmarking infrastructure** (and perf. regression fixes!)





# User support, teaching

- **user support** on the ROOT forum
- curation of [root.cern/install](https://root.cern/install)
- **mentoring** for the [HSF C++ course](#), [CMS data analysis school](#), [Software Carpentries](#)
- recording of **ROOT video tutorials**

## FORUM STATS

**1.9k** days visited    **18d** read time    **6h** recent read time

**7.0k** topics viewed    **38.8k** posts read    **931** ❤️ given

**627** ❤️ received    **10** topics created    **4.0k** posts created

**566** ✓ solutions

RDataFrame in a nutshell: from data to aggregations

th	pt	pz	eta	Define
[1,2,3]				[2,3]
[4,5,6]				[4,6]
[7,8,9]				



# Advertisement, market research

- chairing ROOT's [Performance, Programming model and Parallelism meetings](#)
- talks at [ICHEP](#), [EPS-HEP](#), [ACAT](#), [Analysis Ecosystem Workshop](#), **CMS/ATLAS/LHCb meetings**. Chairing for vCHEP '21.
- ROOT presence at **CMS Analysis Tools Task Force**, Analysis Facility @ CERN Working Group
- consulting for different projects benchmarking ROOT against Python tools, Julia libraries, [cloud query engines](#)
- **close collaborations with analysts** from the University of Pisa, [KIT](#), INFN, and CERN (mainly CMS, ATLAS, FCC)



It's been a busy three years!



Spotlight on a couple of  
interesting developments



# Handling metadata in RDF

1. **attach metadata** to samples in the dataset specification

```
"samples": {  
  "files": ["a.root", "b.root"],  
  "trees": ["Events"],  
  "metadata": { "weight": 0.5 }  
}
```

2. **access metadata** for the current sample via the RDF API

```
df.DefinePerSample("weight",  
  [](unsigned int, const RSampleInfo &id) {  
    return id.GetD("weight");  
  });
```

See also  
[Ivan's final report](#)





# A new way to handle systematics

```
1  ROOT.EnableImplicitMT() # enable multi-threading
2  h_nominal = (
3      RDataFrame('Events', 'root://eos.server/data/*.root')
4      .Vary('Muon_pt', 'RVec<RVecF>{0.9*Muon_pt, 1.1*Muon_pt}', ['down', 'up'])
5      .Filter('nMuon == 2 && Muon_charge[0] != Muon_charge[1]')
6      .Define('mass', 'InvariantMass(Muon_pt, Muon_eta, Muon_phi, Muon_mass)')
7      .Histo1D('mass')
8  )
9  # dictionary with keys 'nominal', 'Muon_pt:down', 'Muon_pt:up'
10 h_dict = ROOT.RDF.VariationsFor(h_nominal)
```

**Variations automatically propagate** to selections, derived quantities and results.

**Multi-thread** and **distributed** execution **just works**.

Only needed quantities are re-computed, all in a **single event loop**.



# A new CLI tool: rootreadspeed

“why is my analysis so slow?”

```
$ rootreadspeed --files root://eospublic.cern.ch//eos/opendata/cms/derived-data/AOD2NanoAOD0
utreachTool/Run2012BC_DoubleMuParked_Muons.root --trees Events --branches nMuon Muon_charge
Muon_pt Muon_eta Muon_phi Muon_mass --threads 32
Total number of tasks: 75
Thread pool size:          32
Real time to setup MT run: 0.466447 s
CPU time to setup MT run:  0.45 s
Real time:                 9.47371 s
CPU time:                  27.01 s
Uncompressed data read:    3232610772 bytes
Compressed data read:      2243313339 bytes
Uncompressed throughput:   325.412 MB/s
                           10.1691 MB/s/thread for 32 threads
Compressed throughput:     225.824 MB/s
                           7.057 MB/s/thread for 32 threads

CPU Efficiency:            8.90953%
Reading data is likely I/O bound.
For details run with the --help command.
```

More info in the [README](#)



# Conclusions



# Thank you, everyone

Too many people to thank to fit here!

The ROOT team, many people in SFT, at CERN,  
and in the rest of the world.

Special thanks to **Axel, Enric** and **Vincenzo** without whom  
these three years wouldn't have been what they have  
been.



# Thank you, everyone

Too many people to thank to fit here!

The ROOT team, many people in SFT, at CERN,  
and in the rest of the world.

Special thanks to **Axel**, **Enric** and **Vincenzo** without whom  
these three years wouldn't have been what they have  
been.

As for me...



*I'm going on an adventure!*