

ROOT 2023

Axel Naumann axel@cern.ch for the ROOT team
2023-02-08



2022

Support!

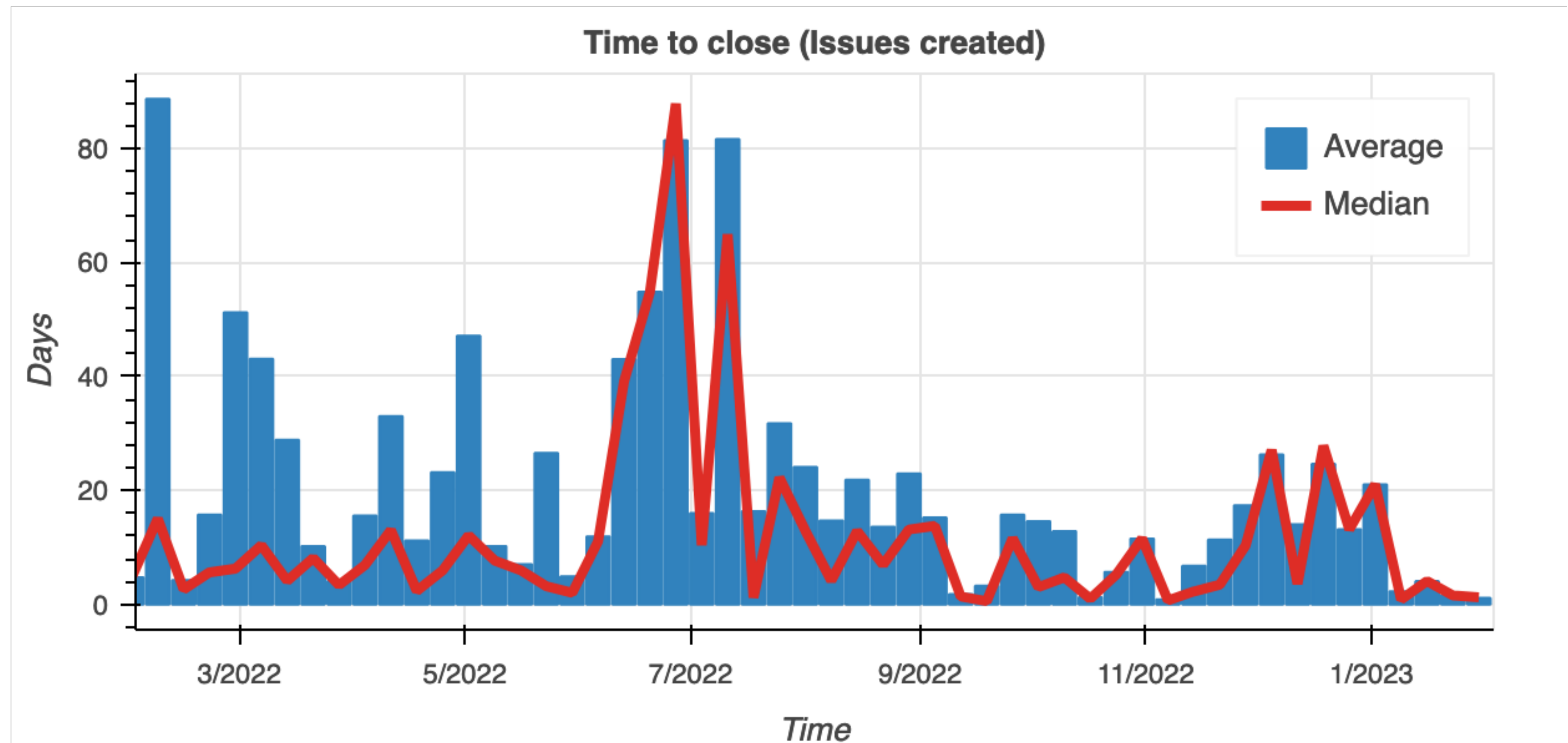
- <https://root-forum.cern.ch> CERN's #1 forum. Stats of 2022:
 - 15k posts, after 15k in 2021 and 20k in 2020
 - 1.3k new users, after 1.4k (2021) and 1.3k (2020)
 - 1st response on average after 15h, after 11h (2021), 19h (2020) [only topics with an answer are taken into account]
- We spend a considerable time here, web forum is *the* support channel
- Bidirectional: gives feedback about where our real problems are

Bugs

- 520 issues created (compared to 650 in 2021)
- 360 closed (compared to 480 in 2021)
- Current open issues: 525 in GitHub (380 in 2021), 1045 in Jira (1071 in 2021)
- This is a major concern for 2023. Not a surprise, we consciously prioritized development over bug fixes in 2022 (maybe except RooFit)

Bugs

- 50% of issues closed after about 12 days - or summer (stable wrt 2021), all stats thanks to <https://cauldron.io/project/5676>

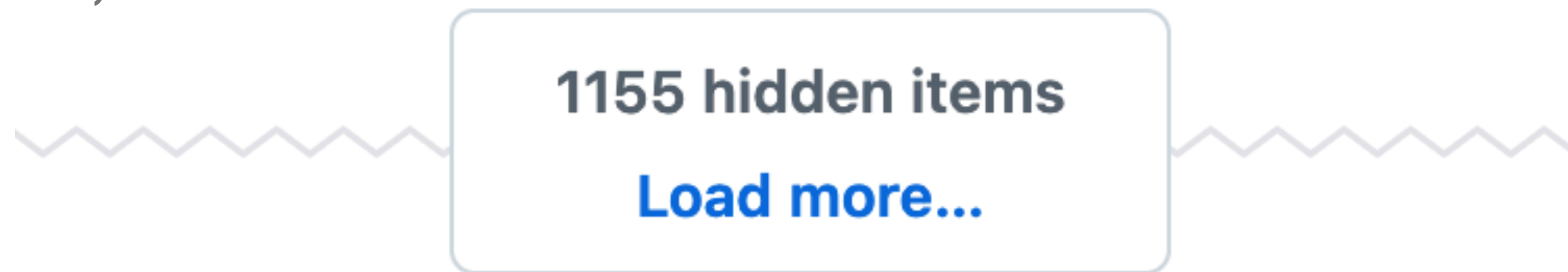


2022's Major Developments

Cling: LLVM 13

- Cling has been upgraded to LLVM / clang version 13
 - From LLVM version 9
- Major effort: lots of changes in the interfaces
 - Significant redesign of LLVM JIT (ORCv2)
- Building ROOT with C++20 is *almost* available

Cling: LLVM 13 in Practice

- [Pull request](#) opened April 1, 2022; merged Dec 9, 2022
- 154 commits, one of which is "import llvm13" 
- LLVM + HEP community effort with 10 contributors: Bertrand Bellenot (CERN), Purva Chaudhari, Simeon Ehrig (HZDR), Stefan Gränitz (freelancer), Jonas Hahnfeld (CERN), Lang Hames (Apple, maintainer of LLVM's JIT), Shahzad Muzaffar (CERN), **Vassil Vassilev** (Princeton), Jeaye Wilkerson (Electronic Arts), Axel Naumann (CERN)
- Future remedy: re-implement cling based on clang-repl, to share functionality (design and maintenance) with llvm community. Needs EP R&D Phase II.

RDataFrame / Analysis

- New binary `rootreadspeed` to measure throughput: users can find reading bottlenecks
 - Multiple files, multiple threads, selected branches of some or multiple trees
- Meta data associated with files can be passed to RDataFrame with `RDatasetSpec`: cross-section, weights, sample name ("H2bb", "MC"), and virtually anything else appear as "column"

RooFit, TMVA

- RooFit BatchMode (vectorized evaluation backend) passes all tests, tutorials, benchmarks!
- Improvements in the multi-process parallelization providing significant speed-up for state of the art models such as Higgs combination
- RooFit performance optimization for standard ATLAS and CMS workflows: new Higgs combination benchmarks
- Prototype minimization of HistFactory models with Auto-Differentiation (clad)
- Integration of SOFIE model evaluation with RDataFrame
- SOFIE can handle complex HEP models: supports several new ONNX operators

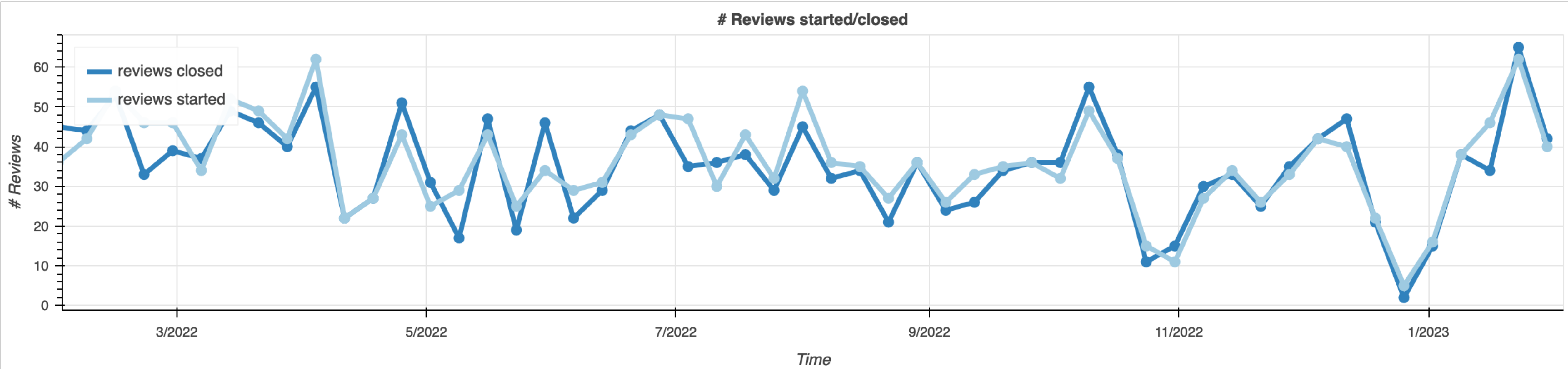
Heavy Lifting

- RNTuple: many, many new features towards required feature set
 - Probing design sanity and feature set: experiments' AODs, object store, etc
- RDataFrame: performance improvements, feature parity for distributed RDataFrame
 - Development of bulk processing: significant speed-up; to be released in 2023

2022 Dev Statistics

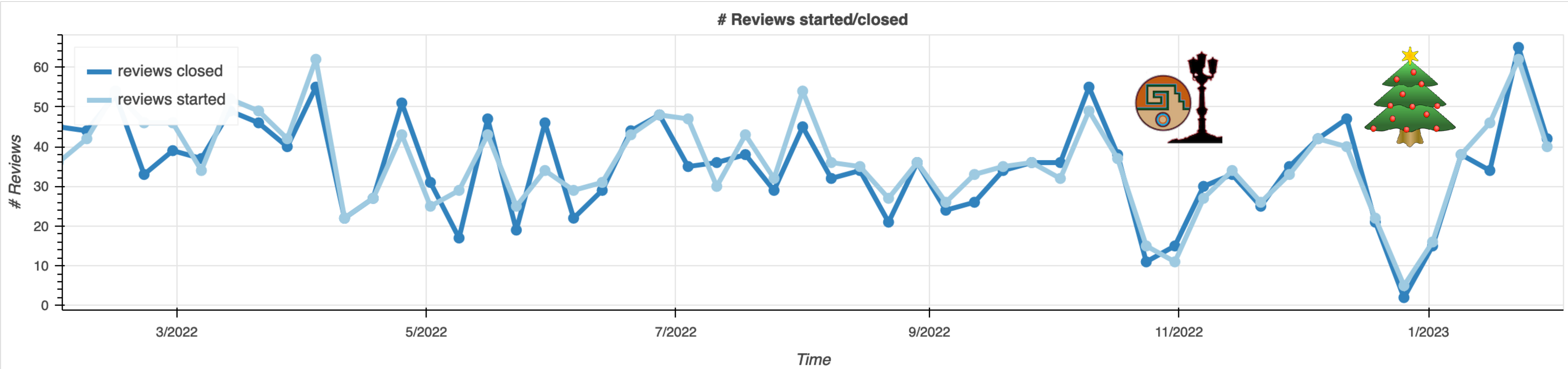
Code Change = Pull Requests

- About 1900 PRs over 2022 (stable wrt 2021), PRs per week:



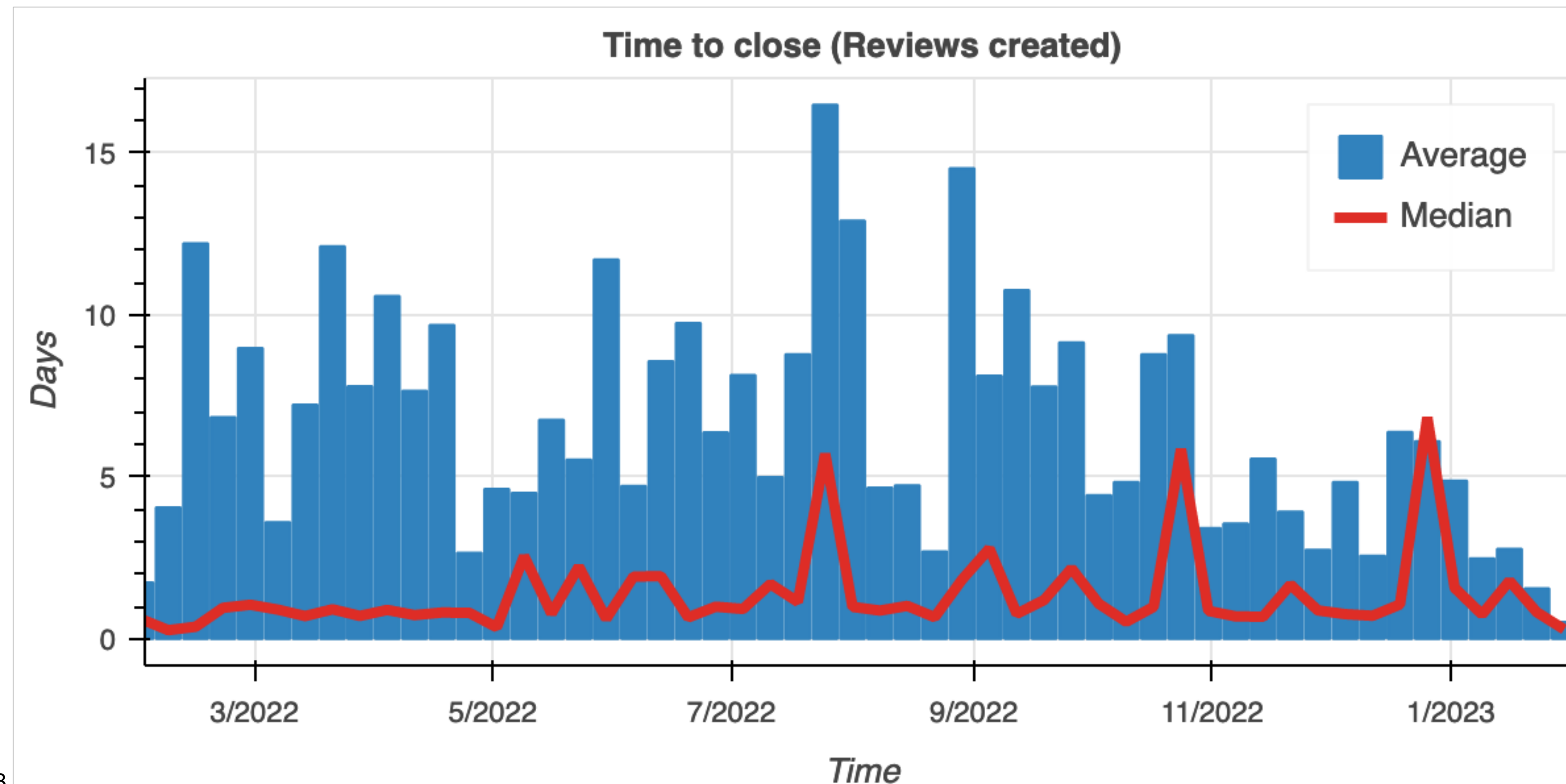
Code Change = Pull Requests

- About 1900 PRs over 2022 (stable wrt 2021), PRs per week:



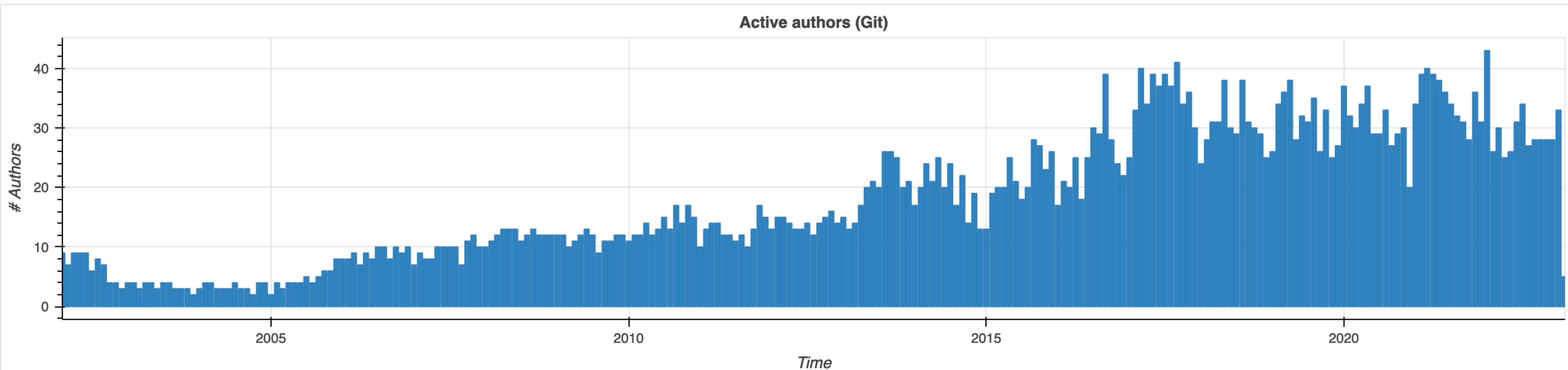
Are PRs working?

- Team invests in high PR review throughput: <1 day, limited by continuous integration system (to be addressed in 2023)



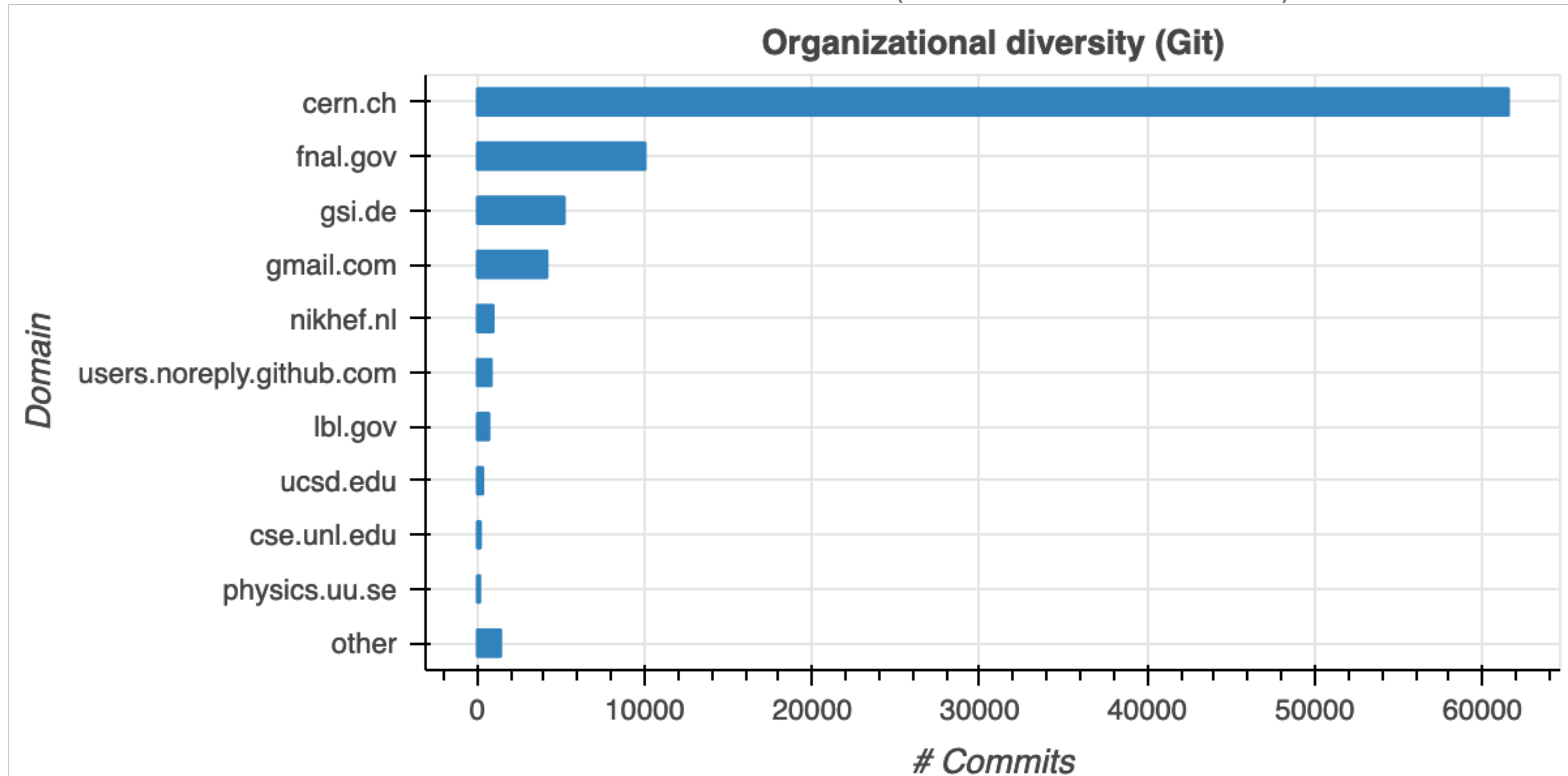
Contributors

- Consistently high number of contributors / month



Contributors

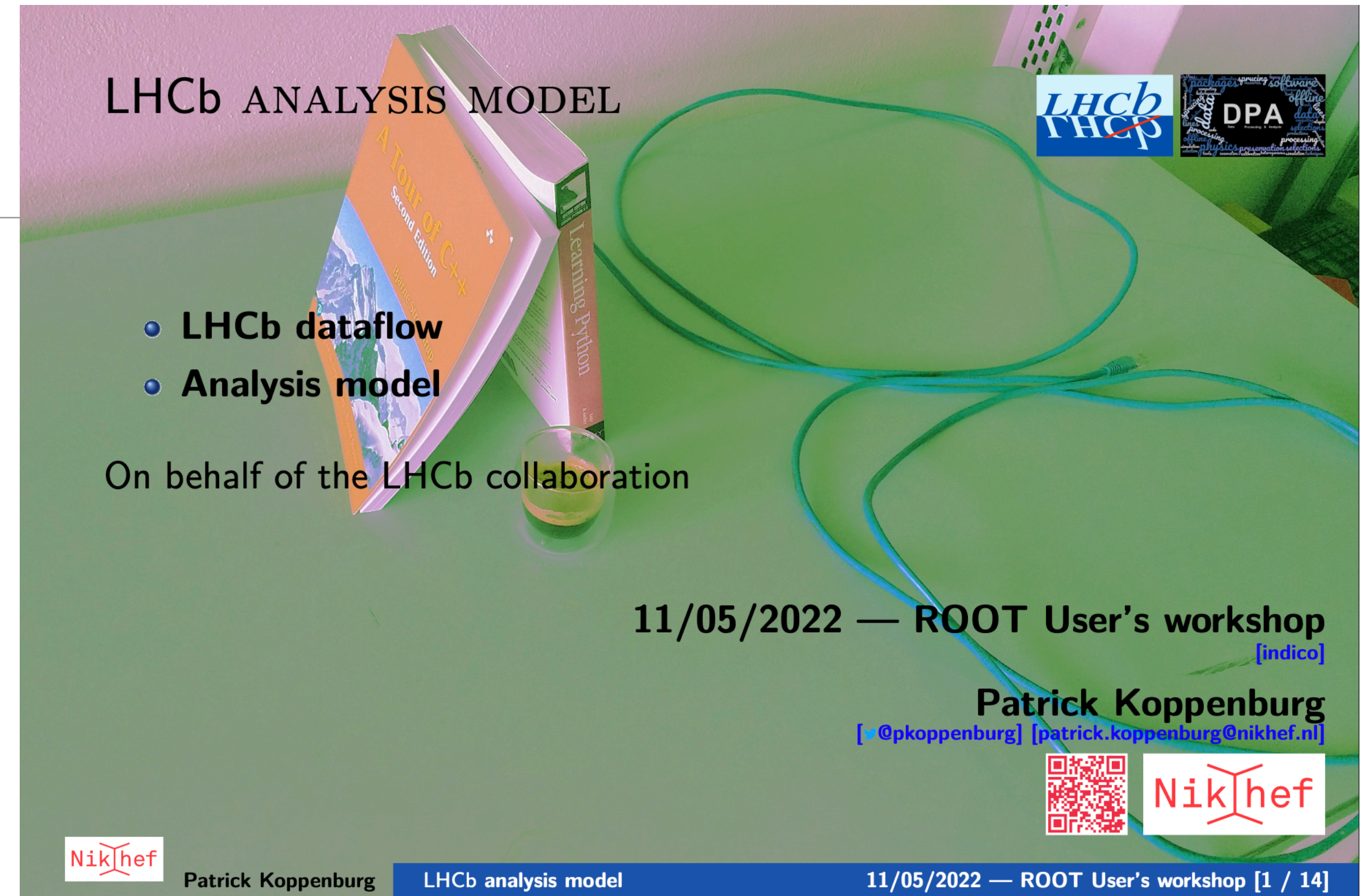
- Reasonable distribution of contributor institutes (email addresses)





Communication

ROOT Workshop

- "At Fermilab": thank you!
- Fantastic overview presentations by experiments on their analysis model, LHC and non-LHC!
- Followed by Train the Trainer event
- 250 registered participants, out of which 150 joined
- Summary



LHCb ANALYSIS MODEL



 


- LHCb dataflow
- Analysis model

On behalf of the LHCb collaboration

11/05/2022 — ROOT User's workshop [\[indico\]](#)

Patrick Koppenburg
[\[@pkoppenburg \]](#) [\[patrick.koppenburg@nikhef.nl\]](mailto:patrick.koppenburg@nikhef.nl)

 Patrick Koppenburg LHCb analysis model 11/05/2022 — ROOT User's workshop [1 / 14]

ACAT

- Many contributions:
 - ROOT Machine Learning Ecosystem for Data Analysis
 - RNTuple: Towards First-Class Support for HPC data centers
 - RDataFrame: a flexible and scalable analysis experience
 - Precision Cascade: A novel algorithm for multi-precision extreme compression
 - Automatic differentiation of binned likelihoods with RooFit and Clad
 - Efficient and Accurate Automatic Python Bindings with Cppyy and Cling
 - Plenary: Simpler, faster and bigger: HEP analysis in the LHC Run 3 era
- Essential for communication between team and community

Presentations, Working Groups

- Several ICHEP presentations
- Several presentations with experiments' physics groups
- Participated at IRIS-HEP+HSF analysis ecosystem workshop, [summary](#)
- Member of CERN's Open Science Policy working group
- Member of CERN's team to bootstrap an Open Source Program Office

Trainings

- Contributions to CMS Data Analysis school
- Software carpentries
- C++ course
- Recording of ROOT training videos, to be published

2023

ROOT Team Situation in 2023

- Several key contributors left (or reduced their presence)
 - Hopeful to capture some of them back, soon!
- Significant influx of young contributors, thanks to EP R&D, experiments, and EU project SYCLOPS: seven students! Three fellows!
- Will have plenty of innovation, though bandwidth limited by experienced developers' guidance
- 2023: focus on bug fixes, open PRs, cleanup, rather than new major developments

Plan of Work

- Result of discussions within the team, with experiments, with "godparents"
- Priorities and goals ~~can~~ shift
 - Will communicate with experiments / users if that happens
- Priorities: **super high**, medium high, fairly high

I/O

- **Address scaling issues with MT-writing to TBufferFile**
- **TBufferFile > 1GB**
- Schema evolution improvements
- Beta release of lossy compression + Incorporation in RNTuple
- Support for STL Collection of std::array

RNTuple

- **Bulk I/O API and RDataFrame connection**
- **Late schema extension**
- Support for merging and chains
- Unaligned friends
- Support for lossy compression (low-precision floats)
- S3 Backend
- Implementation of meta-data API
- Prototype schema evolution based on existing I/O customization support

RDataFrame

- **Bulk processing, also with RNTuple readers**
- **Default values for missing columns**
- **Distributed support for RDatasetSpec**
- Prototype CUDA kernels in RDF
- Varied snapshot
- Live histograms (streaming results as they come from the mappers)

RooFit (1/2)

- **Execute LHCb benchmark fits fully on the GPU (result for CHEP 2023)**
- **Engine for C++ code generation from RooFit model (as an input for AD)**
- **Finalize redesigned minimizer interfaces for better support of automatic differentiation (clad)**
- **Further consolidate JSON standard, joint target with ATLAS: publish joint ttW and ttH multilepton full Run-2 analysis**
- Support for likelihood parallelization with new test statistics and improve scheduling of gradient parallelization

RooFit (2/2)

- Consolidate new test statistics classes: deduplicate common code, unify interfaces, enable vectorized/gpu + multiprocessing fits, etc.
- Stabilize and improve the code, and speeding up the HistFactory
- Pythonize RooWorkspace factory language (a factory based on the JSON dictionaries would fit very well into Python and get people used to the JSON IO format, GSoC project if there is a good candidate)
- Create more ROOT benchmarks that compare RooFit also with other fitting tools to get a better overview on fitting tools in HEP

Math

- **Apply several improvements in Minuit2 (e.g. Fumili algorithm)**
- **Complete Pythonizations of Histogram and Graph classes**
- Benchmark Minuit2 against optimizers of scipy and eventually integrate some of those algorithms in ROOT

TMVA

- **Batch generator integrated with RDataFrame to train Machine Learning models**
- **Add support in SOFIE for inference of GNN**
- **Consolidate SOFIE adding support for missing ONNX operators according to user requests**
- Make SOFIE inter-operable with HLS4ML
- Consolidate RBDT (fast BDT inference)

Visualization / UI

- **REve - RenderCore using GPU, window manager, drop ROOT::Experimental namespace**

User landonth
at Unity Forum

The bad news: WebGL is dead

The good news: WebGPU will soon take its place and provides a great foundation for graphics and GPU compute on the web.

- **TWebCanvas - use by default as with TBrowser**
- Optimize object Paint() methods - avoid gPad usage as much as possible
- Support RWebWindow in JupyterLab - make it fully interactive

Interpreters

- PyROOT
 - Seamlessly translate PyROOT-bound C++ code via Numba
- Cling
 - **RISC-V support for cling**
 - **cling: O2 for non-interactive ROOT on Linux, Mac**
 - Reduce dependency on precise version of SDK headers, esp. for macOS

Builds, Binaries

- **CI rewrite including PRs**
- **Add .deb package generation with CPack**
- Prototype CMake superbuilds
- pip install ROOT

Conclusion

Why?

ROOT Forum



Hello everyone,

first off, great job with RDataFrame. It is such a great tool to have. I have one minor question: Is there a

LHCC Minutes,
CERN-LHCC-2022-014

The **LHCC supports** the continuous efforts put into the modernisation of the infrastructure, along with the HL-LHC Software and Computing review recommendations, and **is glad** about the continuous improvements obtained by the experiments, such as a 10% gain in the CMS raw size, by using a new compression algorithm provided by the ROOT software.

O. Lantwin,
ROOT Workshop

- Coupling cling/pyROOT with jupyter creates a very powerful interactive environment for prototyping, teaching and exploratory analysis
- RDataFrame alone already has profoundly influenced how I and others think about their analyses, looking forward to the rest of ROOT 7

And sometimes...

Build a better product, and release a complete product when it is ready.

The worst part of doing physics is using ROOT.

IT DOESN'T HAVE TO BE THAT WAY.

User who shall remain unnamed

And sometimes...

~~Build a better product and release a complete product when it is ready.~~

Thank you for your time listening to me today, Axel!

I am doing to look into RDataFrames now....

Best,

User who shall remain unnamed



for your help in 2022!

Thank you

- Our users and the experiments do a terrific job at providing early feedback
- Thank you for your patches, bug reports, discussions, ideas, reproducers
 - The community invests a lot
 - We do our best to convert that into a better ROOT for physicists, again in 2023

Plan of Work 2022 vs Reality

Legend:

 Not started

 Ongoing

 Finished

	Priority (1=highest)	
Core I/O	Address scaling issues with MT-writing to TBufferFile	1
	TBufferFile > 1GB	1
	Schema evolution improvements	2
	Incorporation of lossy compression	2
RNTuple	Add (preliminary) support for schema evolution and I/O customization	1
	Finalize support for DAOS: backend improvements and data mover	1
	Add hadd support / fast merging	2
	Implement backfilling	1
	Finalize support for chains and (unaligned) friends	2
	Disk-to-disk converter	2
	Add bulk I/O API	2
RDataFrame	Bulk processing of RNTuple data	1
	Make nested parallelism safe (ROOT-10269)	1
	TMVA+RDF: Finalize fast inference from ONNX format for DNN+CNN, including RDF adapter	2
	RNTuple+RDF integration: fix RVec usage	2
distRDF	More pythonic usage, less C++ code in strings	2
	Allow default values for missing branches	2
	Assess need for C++ DistRDF	1
	Reduce initialization time in client due to file metadata queries	1
	JIT only once	2
	Better execution logs from distributed execution	2
	Decide on API for user code distribution	2
RootFit	Systematic variations	1
	Prototype usage of automatic differentiation	1
	Consolidate work on GPU support	1
	Roll out parallel gradient likelihood and parallel Hessian computation	1
	Further optimize HistFactory implementation for speed	2
	Stabilize RooWorkspace to JSON conversion tools	3
	More benchmarks with recent experiment workflows	1
Further improvements for PyROOT	2	
Math	In Minuit2, add support for external computation of Hessian	1
	Make Minuit2 default minimizer for ROOT fitting and RootFit	3
	Improve Pythonization of Math libraries: direct Numpy interface to histograms and graphs	2
	RHist: revisit class layout, RHistView (range), tutorials	1
	Consolidate SOFIE	1
TMVA	Finalize RReader interface and make it default for TMVA inference	2
	Pythonic interfaces, e.g. direct conversion between numpy and RTensor	2
	REve: support of adoption of further experiments	2
Visualization	REve: support of CMS prototype	1
Interpreters		
	PyROOT: Feasibility study on Numba understanding cppy	2
	PyROOT: Move cppy to use more cling interfaces instead of ROOT meta	3
	cling: LLVM upgrade	1
Build	cling: Address unloading issues	3
	CI rewrite including PRs	2
	Windows 64 bit including roottest	2
	Windows cxxmodules	3
	Prototype CMake superbuilds	3
	Add .deb package generation with CPack	1