# RECAST / REANA / CAP
# Status Update

**L Heinrich**, S Dallmeier-Tiessen[1], T Simko[1]
4th Reinterpretation Workshop

May 15th 2018

1 CERN

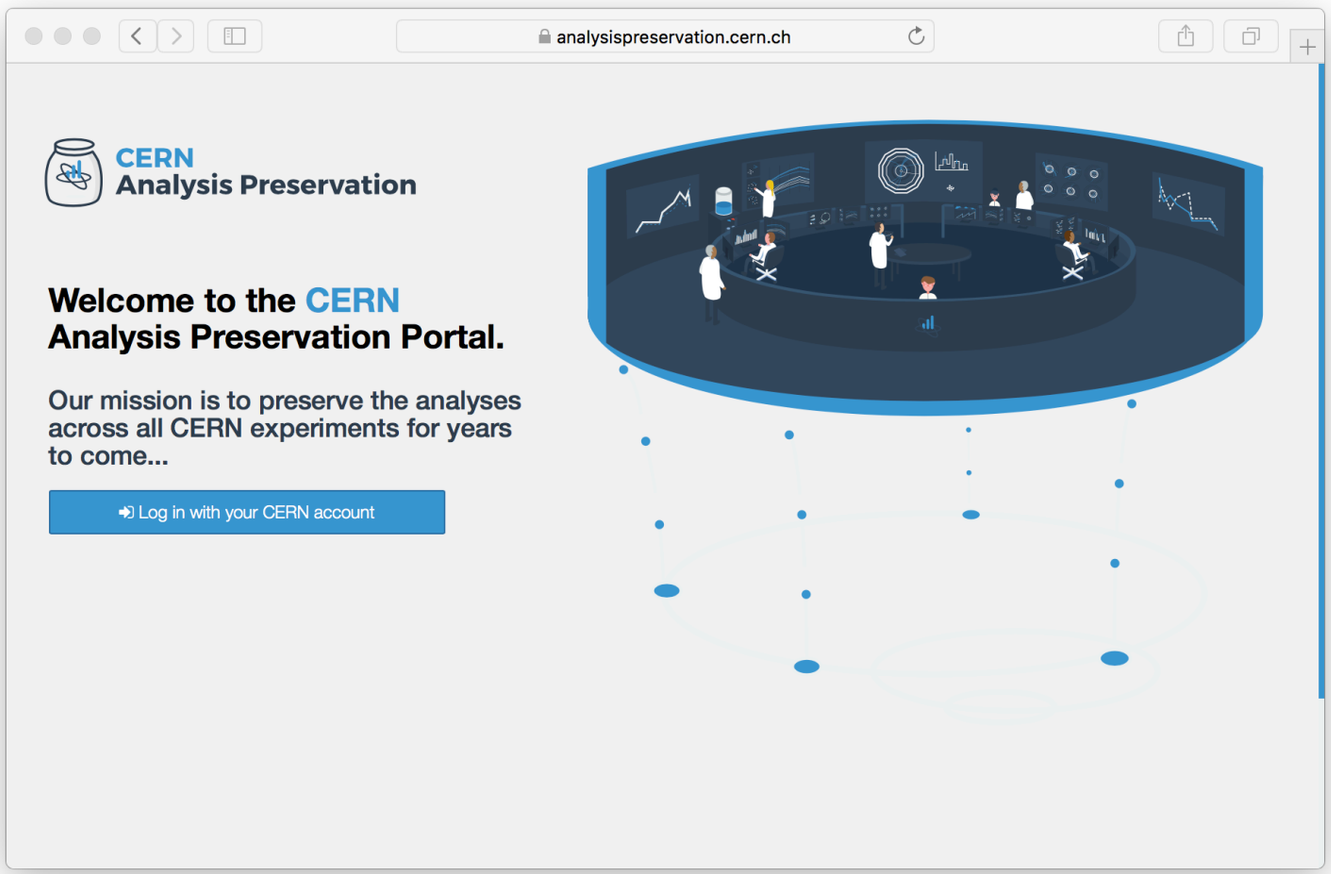
NEW YORK UNIVERSITY

/lukasheinrich

## CERN Analysis Preservation Portal

LHC-wide online portal to help experiments manage the preservation of their analysis throughout the entire analysis lifecycle

- manage meta-data: analysis team, searchable charateristics of anlaysis (final states, triggers used, etc…)
- high-value data assets that need preservation (Ntuples, final background estimates, statistical models, UFO models, etc..)
- source code preservation
- containerized software environments
- Computational Workflow Recipes

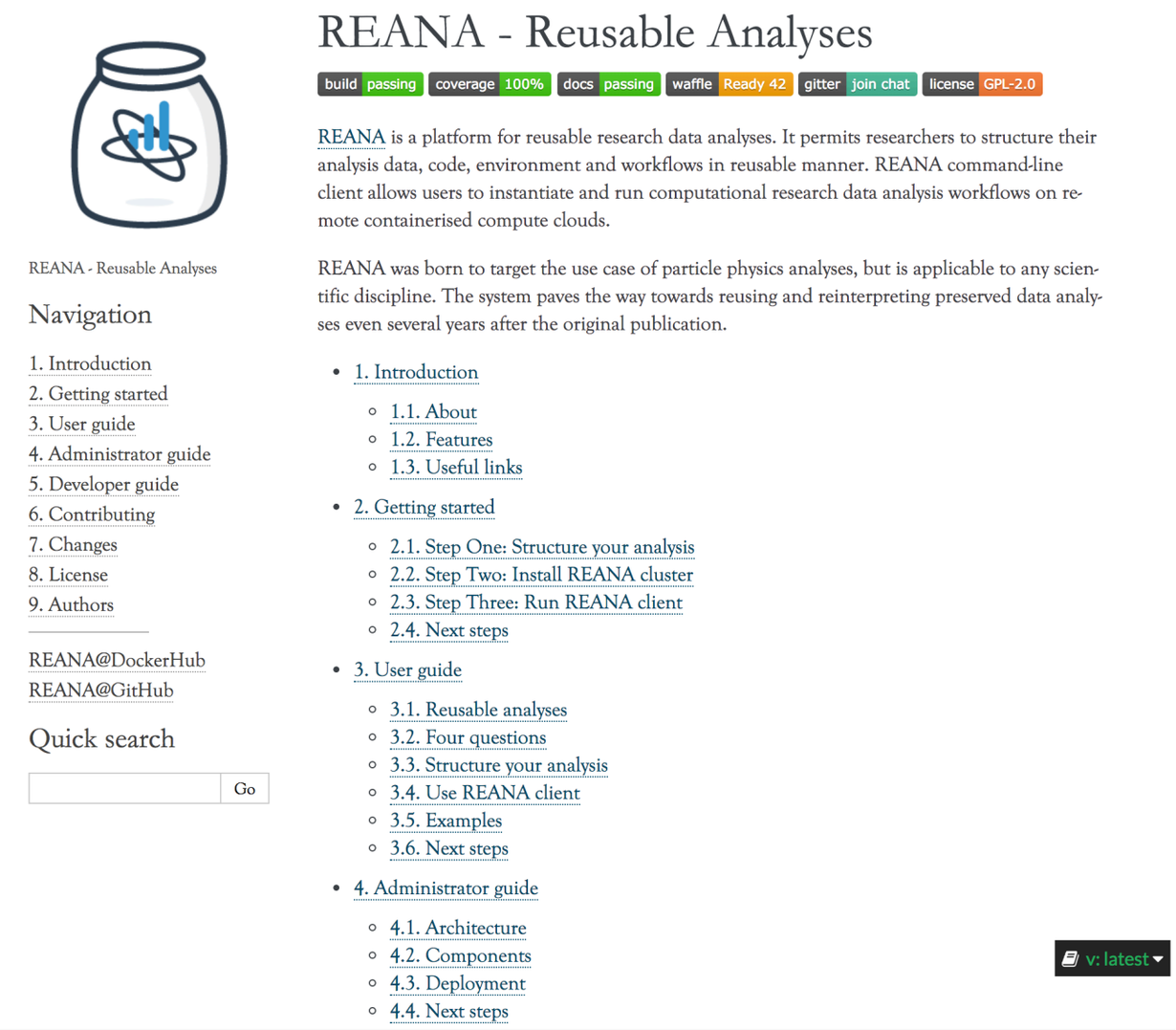


# REANA — reusable analysis platform

Cloud-backed Computational Workflow platform, both for actively developing analyses as well as re-using analyses stored in CAP

Support:

- non-interactive “batch-like” workflows
- interactive workflows steered from a CLI interface or a browser
- designed to supoprt multiple workflow engines
  - yadage
  - Common Workflow Language (CWL)
  - (nextflow? Parsl? snakemake?)



REANA - Reusable Analyses

Navigation

1. Introduction
2. Getting started
3. User guide
4. Administrator guide
5. Developer guide
6. Contributing
7. Changes
8. License
9. Authors

REANA@DockerHub
REANA@GitHub

Quick search

Go

REANA - Reusable Analyses

build passing | coverage 100% | docs passing | waffle Ready 42 | gitter join chat | license GPL-2.0

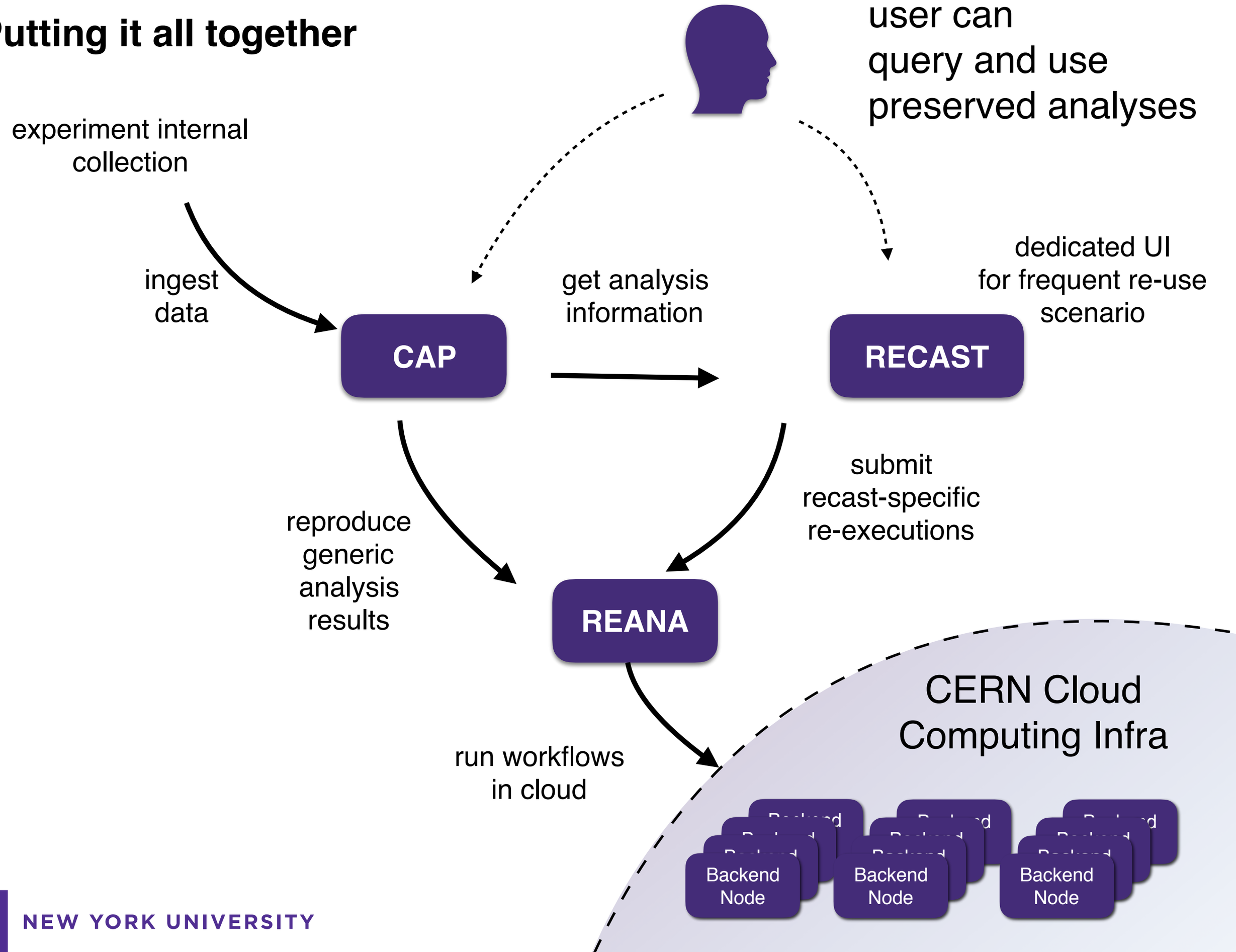
REANA is a platform for reusable research data analyses. It permits researchers to structure their analysis data, code, environment and workflows in reusable manner. REANA command-line client allows users to instantiate and run computational research data analysis workflows on remote containerised compute clouds.

REANA was born to target the use case of particle physics analyses, but is applicable to any scientific discipline. The system paves the way towards reusing and reinterpreting preserved data analyses even several years after the original publication.

- 1. Introduction
  - 1.1. About
  - 1.2. Features
  - 1.3. Useful links
- 2. Getting started
  - 2.1. Step One: Structure your analysis
  - 2.2. Step Two: Install REANA cluster
  - 2.3. Step Three: Run REANA client
  - 2.4. Next steps
- 3. User guide
  - 3.1. Reusable analyses
  - 3.2. Four questions
  - 3.3. Structure your analysis
  - 3.4. Use REANA client
  - 3.5. Examples
  - 3.6. Next steps
- 4. Administrator guide
  - 4.1. Architecture
  - 4.2. Components
  - 4.3. Deployment
  - 4.4. Next steps

v: latest

# Putting it all together



# Running analyses on the Cloud

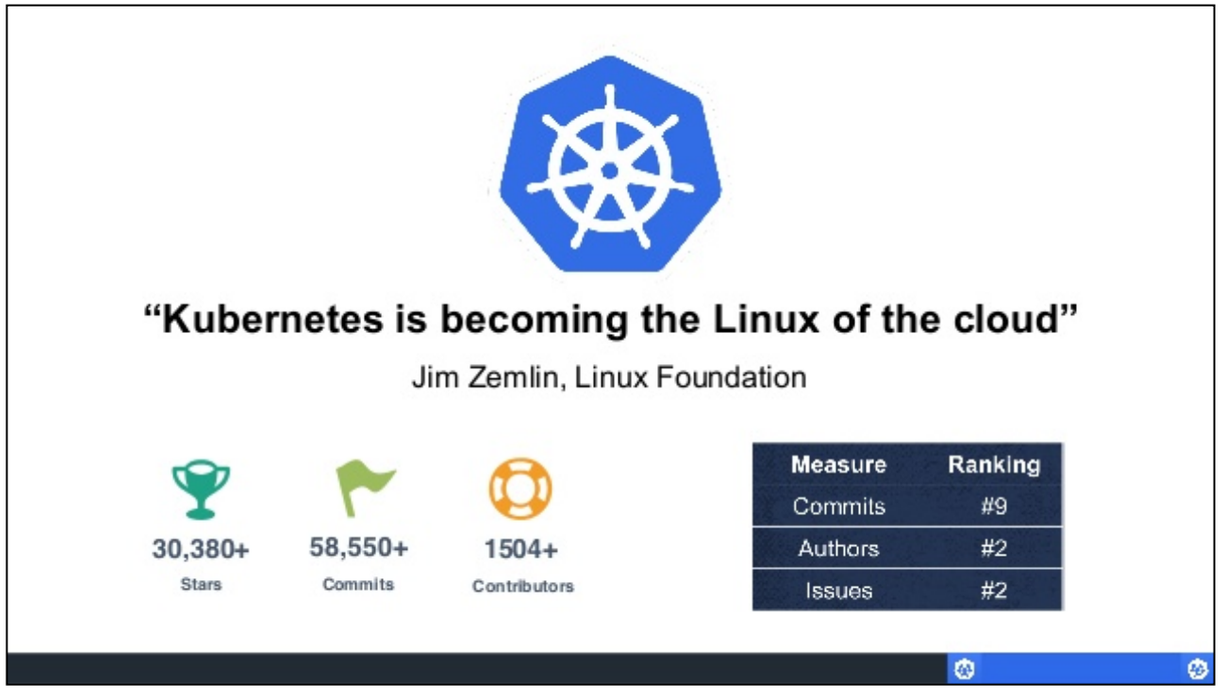
full-fidelity analyses need larger scale than e.g. CheckMate, Rivet, etc..

**Computational Workflow**
**+ Container-based infrastructure**
**are a winning combination**

*(already widely used in bio-informatics and genomics, but very applicable in HEP — backend of REANA/RECAST)*

**Linux containers are ideal packaging tool to archive software environments and distribute them across clouds**

Experiments starting to adopt them as part of their analysis workflow




R Rocha (CERN IT) presenting core REANA component (distributed yadage) during a KubeCon 2018 keynote

## Experience inside of ATLAS

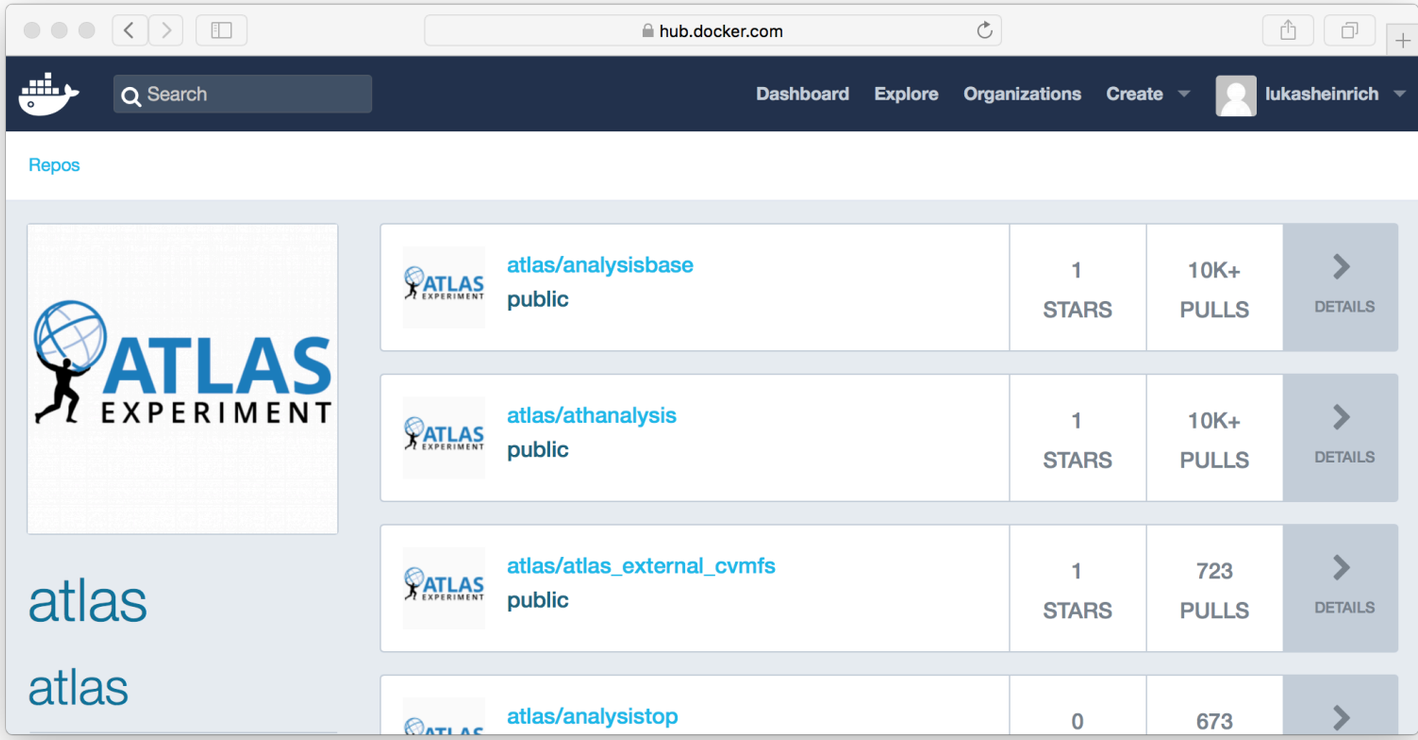
ATLAS has been heavily investigating using containers as a part of the analysis model / analysis preservation strategy

- officially supported base software environments onto which user analysis code can be added + preserved as a single package
- preserving analyses (e.g. event selection) becomes a matter of a few-line text file
- is integrated in official documentation / recommendation on how to setup new ATLAS analysis code (e.g. for upcoming full Run-2 analyses)

**Dockerfile** 190 Bytes

```
FROM atlas/athanalysis:21.2.23
ADD . /xampp/XAMPPmonoH
WORKDIR /xampp/build
RUN source ~/release_setup.sh &&  \
    sudo chown -R atlas /xampp && \
    cmake ../XAMPPmonoH && \
    make -j1
```

preserving a event selection

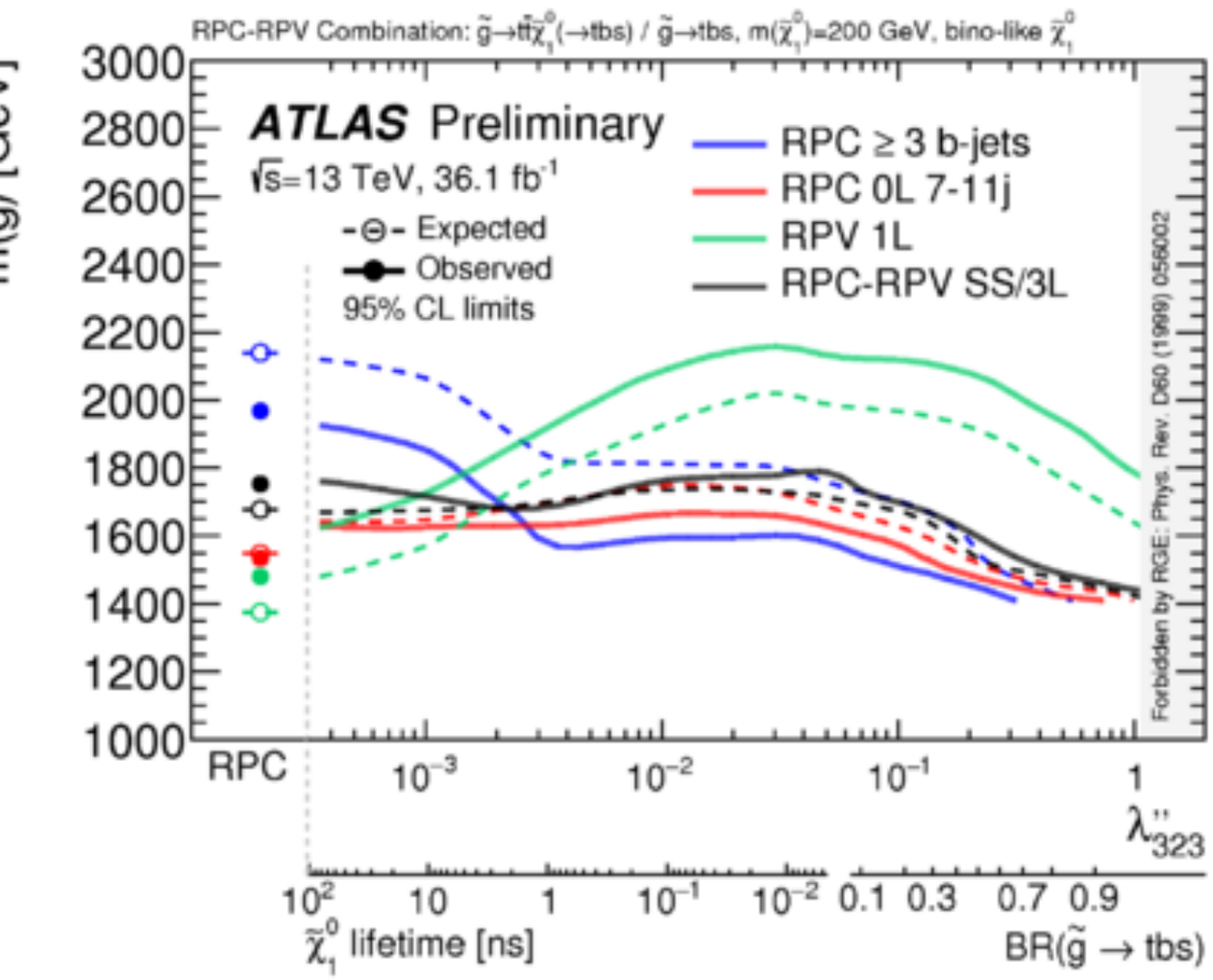


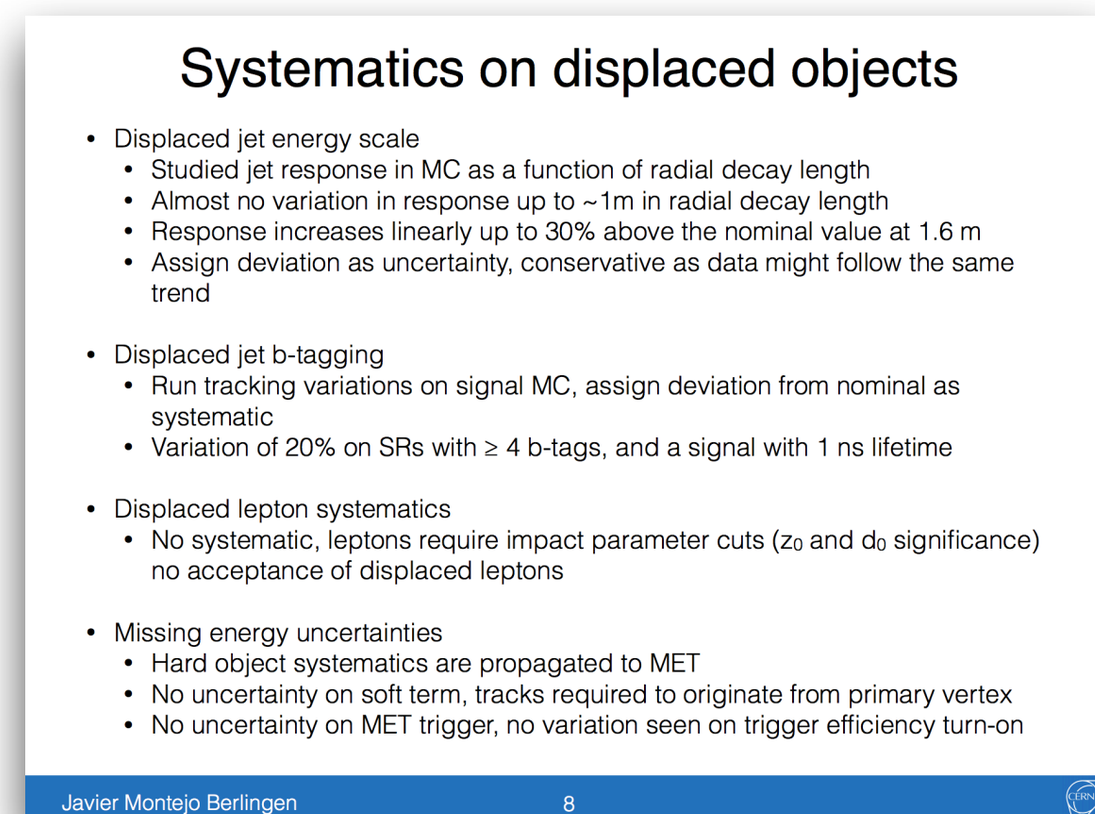
## Experience inside of ATLAS

Successfully tested: Re-using analyses with minor modifications

- preserved analyses are not “black boxes”
- preservation format suitable such that modifications to analyses can be made at time of re-use

RPV-RPC project: Adding new systematics for displace objects



### Systematics on displaced objects

- Displaced jet energy scale
  - Studied jet response in MC as a function of radial decay length
  - Almost no variation in response up to ~1m in radial decay length
  - Response increases linearly up to 30% above the nominal value at 1.6 m
  - Assign deviation as uncertainty, conservative as data might follow the same trend
- Displaced jet b-tagging
  - Run tracking variations on signal MC, assign deviation from nominal as systematic
  - Variation of 20% on SRs with ≥ 4 b-tags, and a signal with 1 ns lifetime
- Displaced lepton systematics
  - No systematic, leptons require impact parameter cuts ($z_0$ and $d_0$ significance) no acceptance of displaced leptons
- Missing energy uncertainties
  - Hard object systematics are propagated to MET
  - No uncertainty on soft term, tracks required to originate from primary vertex
  - No uncertainty on MET trigger, no variation seen on trigger efficiency turn-on

Javier Montejo Berlingen 8

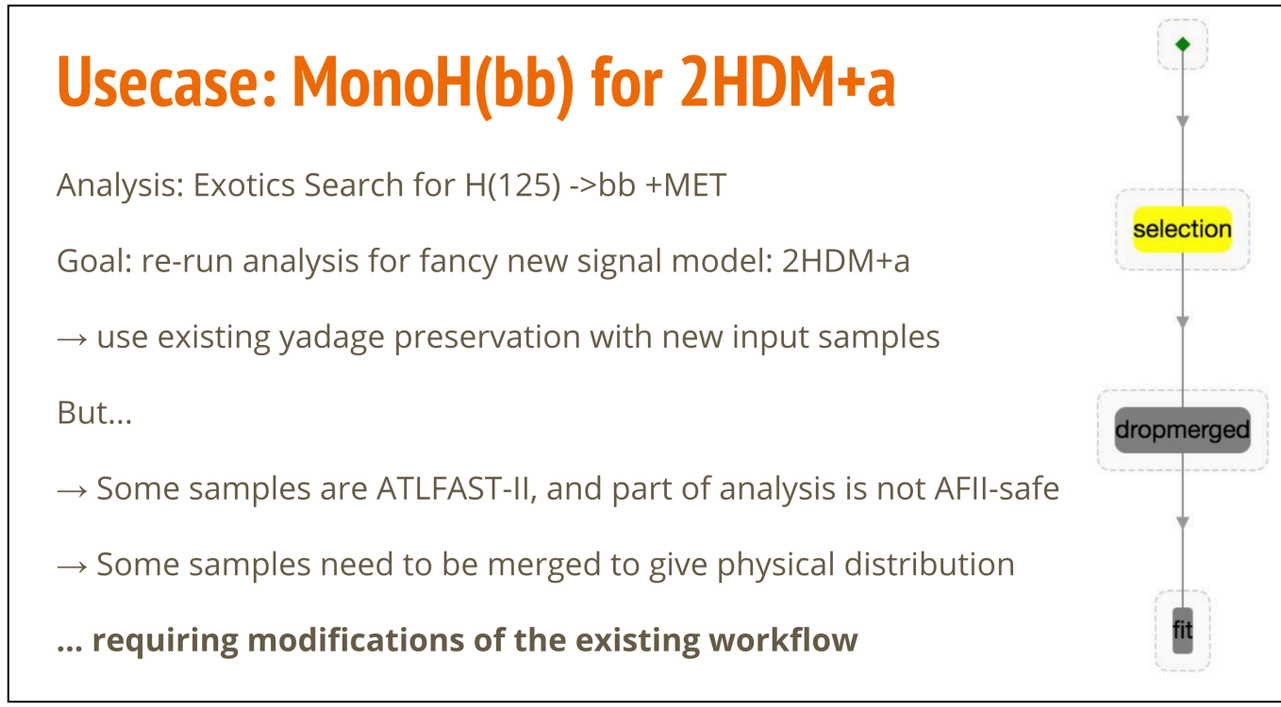
J. Berlingen

# Experience inside of ATLAS

Reinterpreting mono-H searches (work in progess)

- needed to drop certain signal regions, as simulation was not reliable there
- re-use the prepackage code, but introduce additional steps in the workflow to handle details

**Experiment-internal preservation of analysis allows us to carefully re-use specific pieces of an analysis and adapt where necesary.**




→ hard to do outside of the experiments

L Henkelmann

## Experience inside of ATLAS

Within ATLAS, we are building infrastructure to allow to easily re-run analyses*,  in which they were not originally involved.

```
$> yad submit workflow.yml \
   gitlab-cern:prieck/monoHbb16_preservation:yadage \
  -o fit/limits.txt -o fit/plots \
  -f pars.yml
```

specify which analysis

specify desired outputs

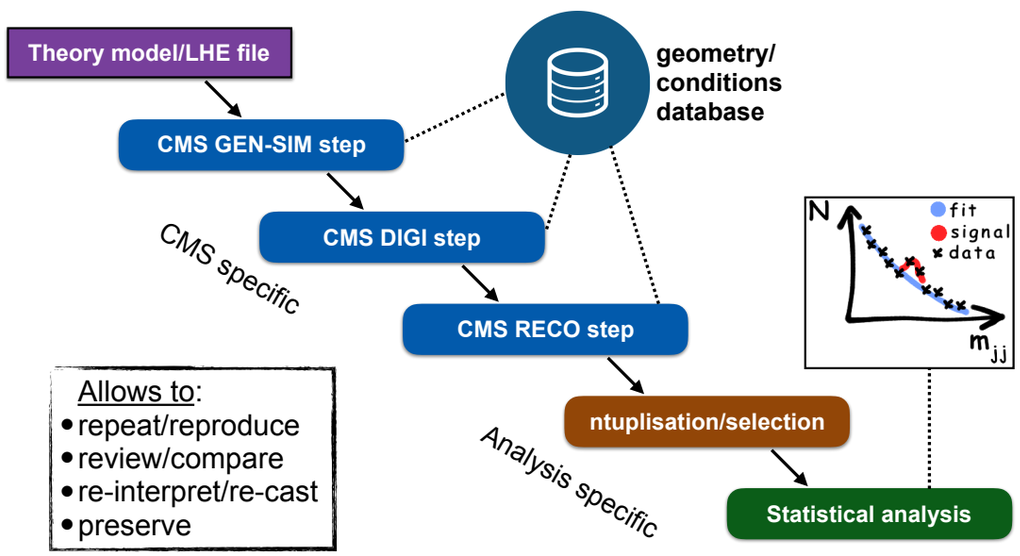
specify new inputs (new signal files, cross-sections)



* when preserved using latest recommendations

# Developments in CMS (Clemens Lange, CERN)

Investigating similar strategies as ATLAS: e.g. preserve analyses using experiment-provided software images. Use Kubernetes-based workflow engines to steer the analysis. Looking to integrate with REANA




PUBLIC REPOSITORY

clelange/cmssw ☆

Last pushed: 21 hours ago

Repo Info | Tags

| Tag Name | Compressed Size | Last Updated |
|---|---|---|
| 7_1_32_patch1 | 7 GB | 21 hours ago |
| 9_4_4 | 6 GB | 2 months ago |

## Preservation options

> Docker containers:

- fully self-contained images (15-17 GB) + 1 GB of conditions database per CMSSW release
- light-weight image mounting CVMFS (via FUSE) inside container (~700 MB) — slow
- light-weight image mapping CVMFS (via GraphFS) as volume (~700 MB) — fast (not yet in repository)
- repository: https://github.com/clelange/cmssw-docker
- these containers are already being used for analysis development by individual analysts

> Kubernetes:

- possibility to pass kerberos token and VOMS proxy to Kubernetes jobs — fully authenticated
- example runs data set query and determines cross section for a given simulation sample (requires CMS VOMS membership)
- repository: https://github.com/clelange/cmssw-k8s
- currently exploring different workflow tools for chaining

## Integrating alterantive analysis implementations

REANA / RECAST are purposfully agnostic to how the analysis is implemented (→needed to support full analyses)

Makes it easy to integrated other analysis catalogues. Currently have

1. SimpleAnalysis (ATLAS-internal, truth-level implementations)

2. Rivet (using the official Rivet Docker Images)

3. CheckMate (current using custom image, but helping CM authors to automate release of image together with HepForge → GitHub transition)
4. MadAnalysis: not yet :( but interested to make this work

## REANA Integration / Status

Much of ATLAS-internal / RECAST-internal infrastructure will be offloaded to REANA.

- REANA 0.2.0 released in April 2018
  - support Common Workflow Language (CWL) in addition to yadage
  - improved command lines tools, monitoring
- Work in progress
  - more workflow engines, for easy on-boarding of analyzers
  - support more container runtimes (singularity, etc…)
  - seamless integration into CAP
  - *lots* of documentation. It's a new paradigm in HEP, so need good docs to make user's feel at home.

# REANA Examples:

reanahub / reana-demo-lhcb-d2pimumu

Unwatch 10 | Star 0 | Fork 1

Code | Issues 0 | Pull requests 0 | Projects 0 | Wiki | Insights

REANA demo example of LHCb D->pimumu analysis

6 commits | 1 branch | 0 releases | 1 contributor

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

atrisovic smaller image ··· Latest commit 7991e34 30 minutes ago

| | | |
|---|---|---|
| code | add description | an hour ago |
| data | first commit | 23 hours ago |
| docs | smaller image | 30 minutes ago |
| README.md | Update README | 33 m |
| reana.yaml | first commit | 23 |

README.md

## REANA example

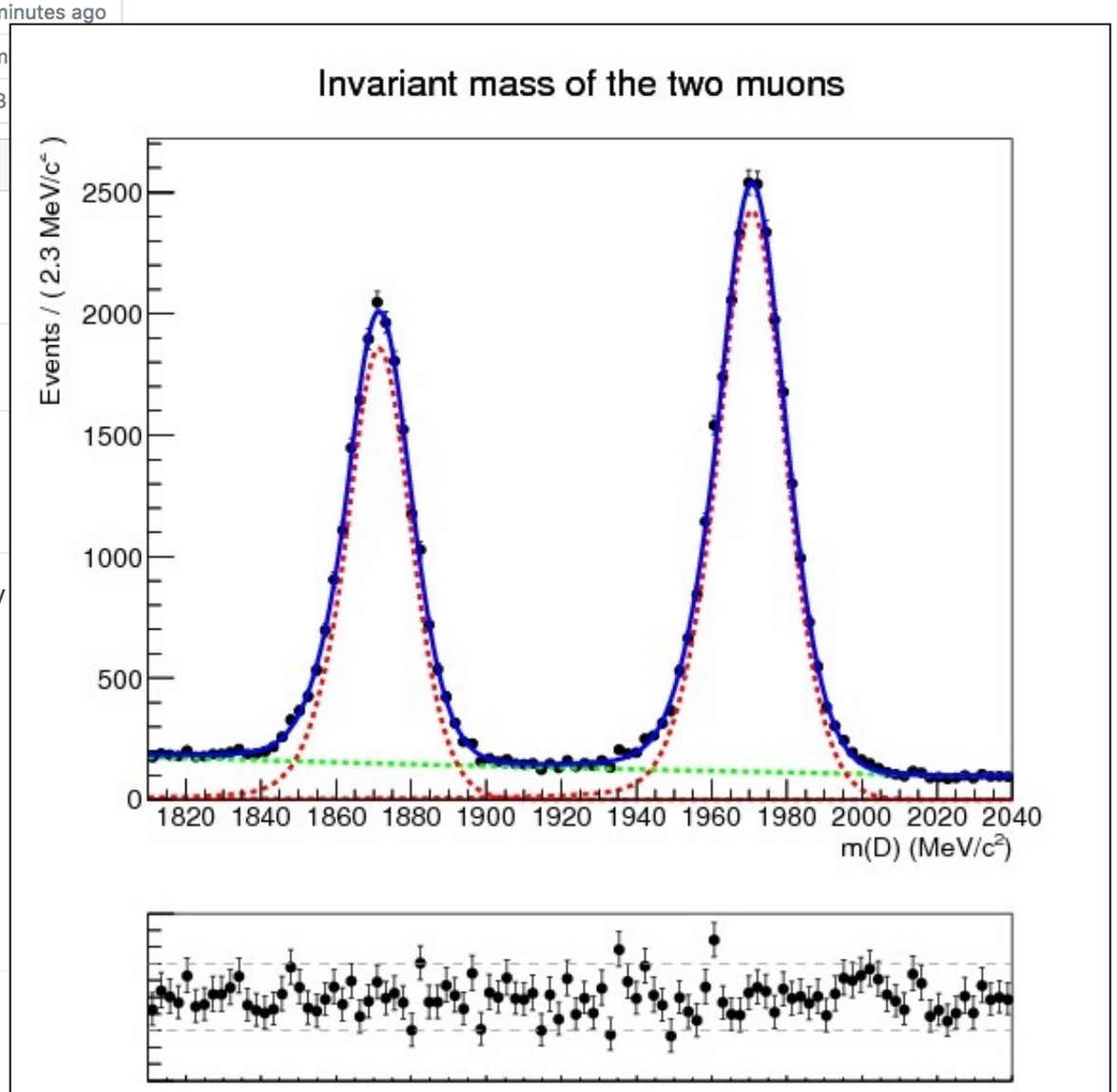
### About

This example is based on analysis-case-study.

### Mass fit results

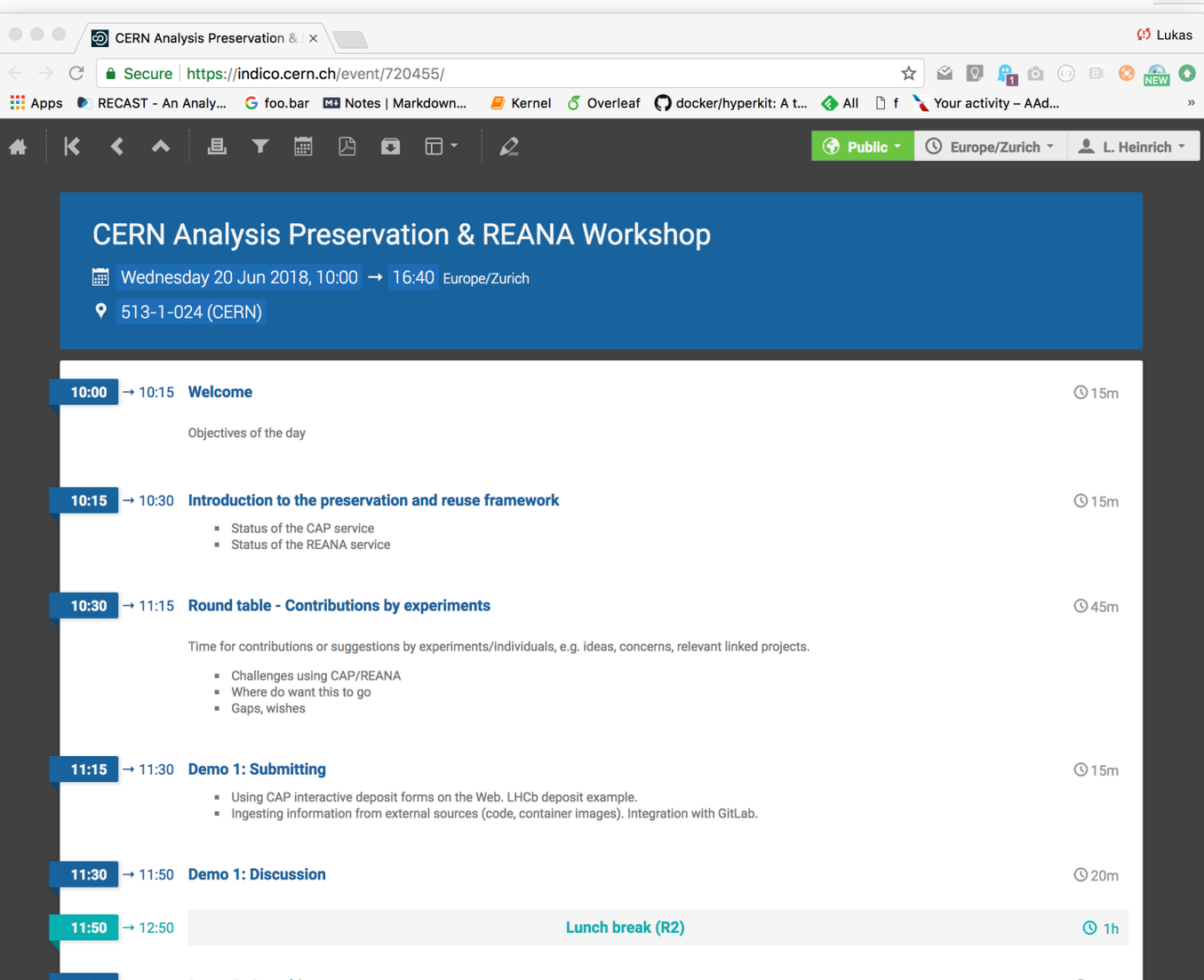
The result of this analysis are the following plots in various dimuon mass ranges. We studied the three body decay high dimuon and low dimuon mass range, and we did not observe any signal.

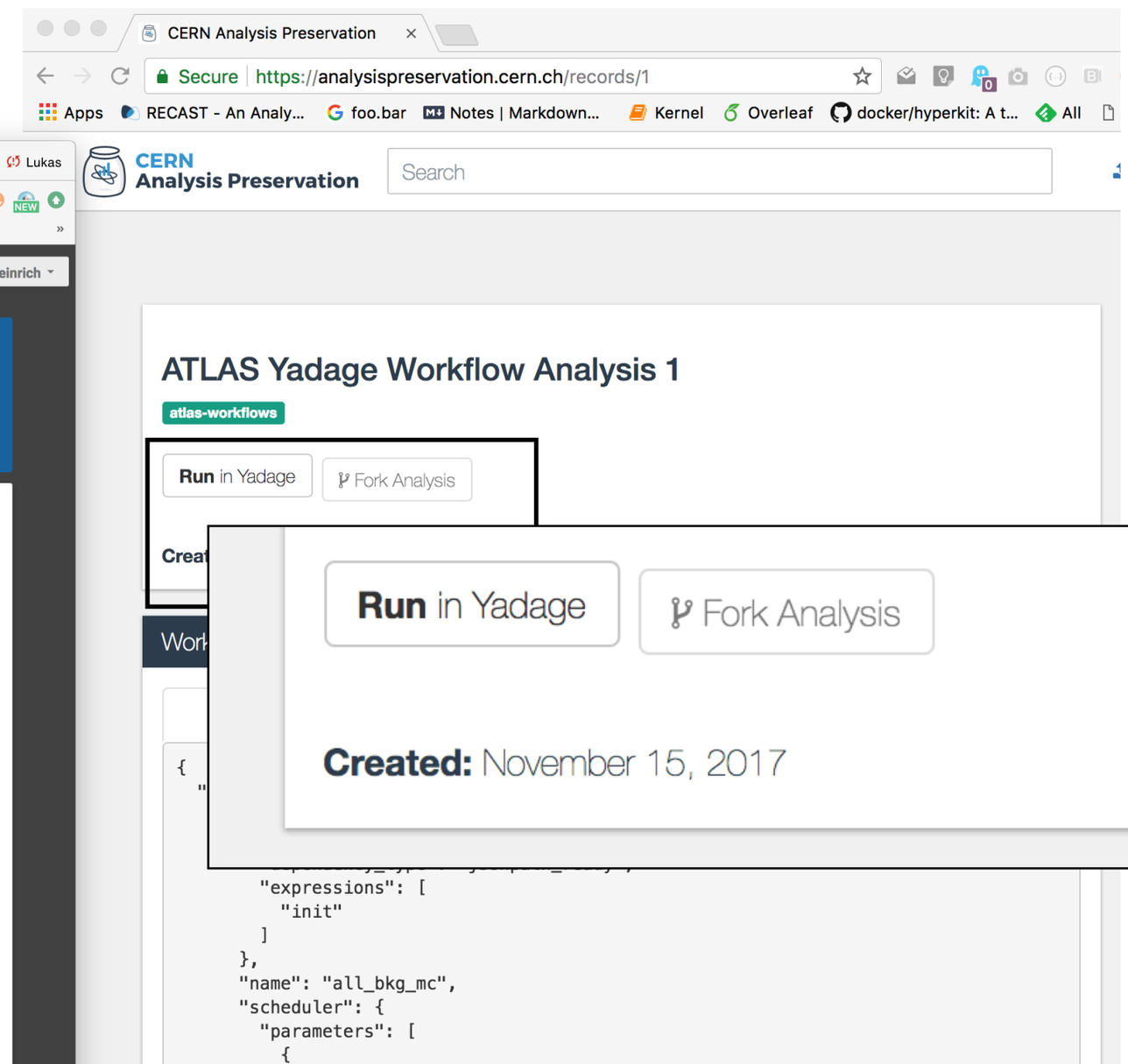
| Dimuon resonances | Dimuon mass range (MeV) | Plot |
|---|---|---|
| Three body decay (low dimuon) | 250 - 525 | `low_dimuon_signal.pdf` |
| $\eta$ | 525 - 565 | `eta.pdf` |
| $\rho, \omega$ | 565 - 850 | `rho_omega.pdf` |
| $\phi$ | 850 - 1250 | `phi.pdf` |
| Three body (high dimuon) | 1250 - 2000 | `high_dimuon_signal.pdf` |



# CAP Status

- CAP/REANA workshop on June 20 (please join if you are part of the experiments!)
- Lots of development leading up to workhop
  - new UI, powerful search, many data ingestion pipes from expts
  - integrating with REANA
  - code, s/w, data archiving



**CERN Analysis Preservation**

http://analysispreservation.cern.ch
http://github.com/cernanalysispreservation
analysis-preservation-support@cern.ch

**REANA**
http://reanahub.io
http://github.com/reanahub
reanahub
info@reanahub.io

**RECAST**
http://recast.cern.ch