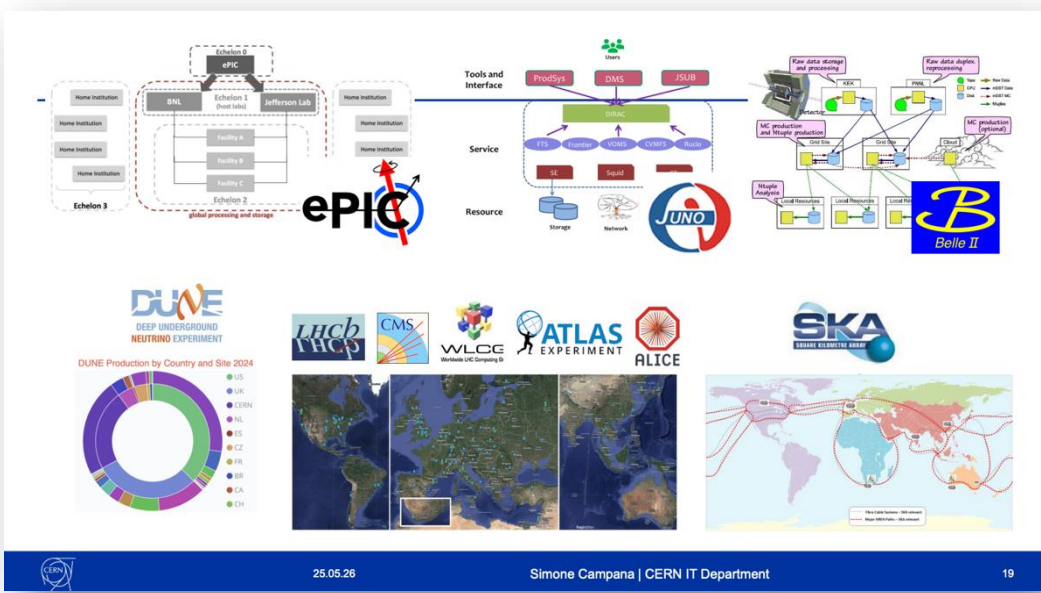
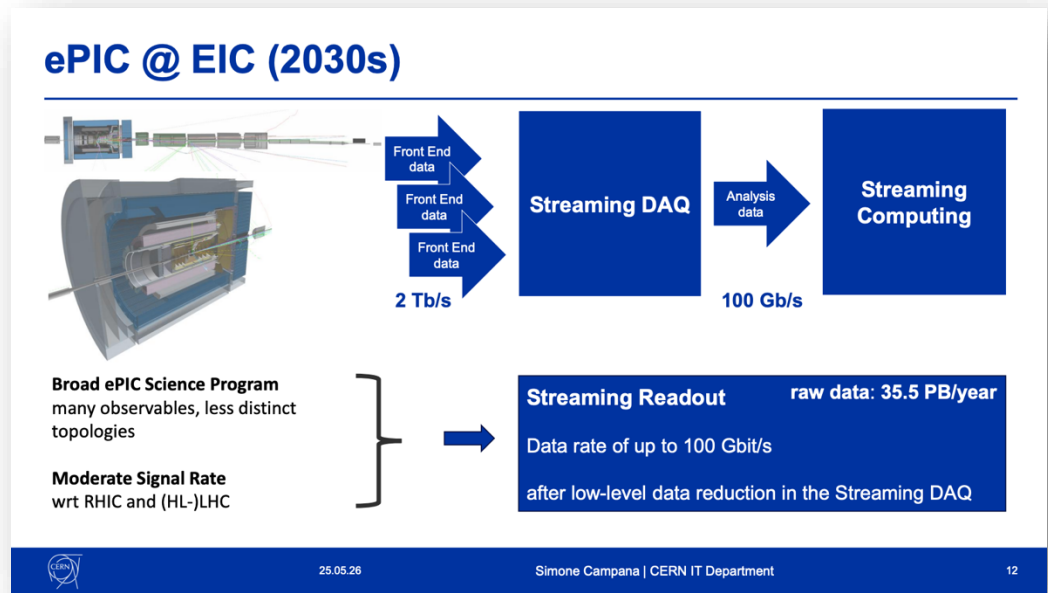


General notes from CHEP

Focus on: What are others doing? What are we missing? What are the lessons learned?

ePIC was well-represented in the opening plenary by Simone Campana: [slides](#)

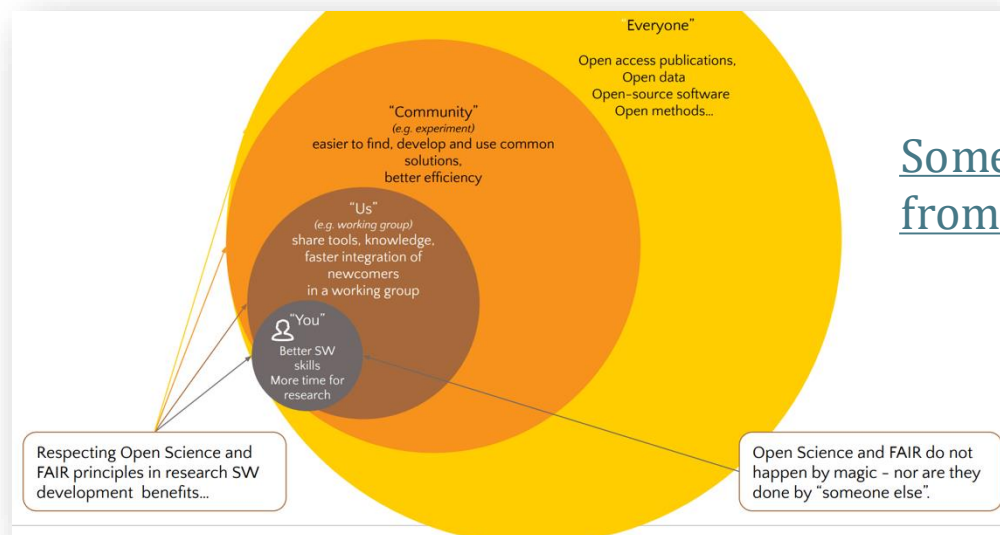
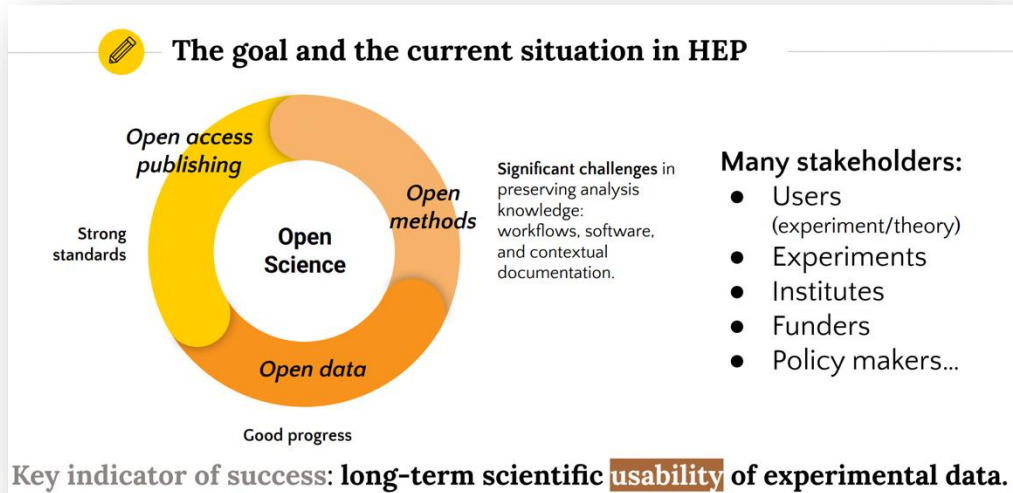


As a track convener for Track 8 (see description below), I will include notes specific from Track 8:

Track 8 - Analysis infrastructure, outreach and education

Infrastructure for interactive computing; applications and use-cases; experience with analysis facility production systems and pilots; aspects of reproducibility in interactive computing; collaboration enabling tools; reinterpretation tools; analysis preservation and reuse; data preservation for collaboration; outreach activities; open data for education and training; training initiatives; event displays.

A lot of focus on reproducible workflows and training, support of open data and findability



Some feature slides
from Caterine Doglioni

ATLAS providing fully worked out example notebooks for datasets:

[illegible]

Welcome to the Magic

What does atlasopenmagic do

- One-stop-shop for all of the ATLAS Open Data
 - Storage information (file location, size, checksum from [Rucio](#))
 - Release information (description, link to the CERN Open Data portal, etc.)
 - Physics metadata (dataset number, cross section, k factor, n Events, grid job info, etc. from [AMJ](#))
 - MC information (weights, etc. via the ATLAS Physics Modeling Group (PMG) database)
- Thoroughly documented
 - See the [Metadata tutorial](#)
 - See the [Python API documentation](#)
- Public REST APIs
 - <https://atlasopenmagic-api.app.cern.ch/docs>
 - [API specification docs](#)
- **One pip install away** (with minimal dependencies)
 - [requests, pyyaml, tqdm]

G. Guerrieri - 28th CHEP - May 2026

10

Also developed Lumi as AI support for CERN open data with relevant mcp

Some feature slides from Giovanni Guerrieri

<https://opendata.atlas.cern/docs/tutresearch/openevgentut>

REANA example

Guiding a first time User

The (growing) gallery page consists of examples created from LHCb Open Data

Gallery

Welcome to the gallery of selected 'reusable' examples that will allow you to get more familiar with the service!

For each example, you can view the production/reproducible configuration together with the produced results that you can directly download or study.

If you like a particular example, you could use it as a model, clone its configuration and amend to parameters in order to start creating your own reusable!

B⁻K⁺ meson mass example

This example studies $B^{\pm} \rightarrow K^{\mp} \pi^{+}$ decay demonstrating a B⁻ meson mass peak.

Each example is accompanied by:

- Minimal [REANA reproducible analysis examples](#)
- Records on the [Open Data portal](#) (including DOIs & citation templates)
- Detailed [documentation](#) on how to analyze the data

See [Mindaugas'](#) Talk later today on fully reproducible LHCb Analyses

You can **clone** the example and **build your own request!**

[B⁻ → J/ψ(→ μ⁺μ⁻)K⁺]CC Ntuples 4530

LHCb collaboration

Cite as: LHCb collaboration (2020). [B⁻ → J/ψ(→ μ⁺μ⁻)K⁺]CC Ntuples 4530. CERN Open Data Portal. DOI:10.7554/CDPDATA.LHCb/B-TOJPSI

Data recorded in 2016 and published in 2020

Clone
Download

Usage

The following resources illustrate how to work with these reunits:

- if LHCb's Open Data Guide explains how the `$for(reunit, $(ls | grep ^[A-Za-z0-9_]+\.json$)) do { echo "reunit: $reunit"; }` isolates the desired signal and shows how you can analyse the reunits with ROOT or python.
- if `reana-demo-lhc-run2-hqdm` repository contains a runnable containerized workflow for this analysis.

Slide from Piet Nogga

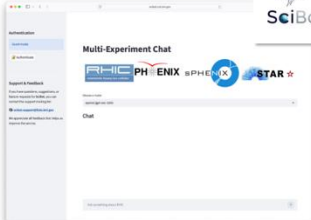
RHIC datacards, data preservation and workflows

slides from Eric Lancon

Pillar 2: SciBot

Knowledge accessibility & transfer

- **RAG** over 25 years of notes/slides/code/Indico
- LLM-agnostic via **MCP**
- Collaboration aware **access control**
- **AI agent** under development for workflows



sciobot.csi.bnl.gov

Key limitation
SciBot surfaces what is documented
It cannot recover undocumented expertise — active participation from the experiments is essential

Brookhaven
National Laboratory

E. Sannoni, RHIC DAP, CHEP 2028

Use of SciBot

Eric Lancon

Datacards

A structured description that makes a dataset searchable and citable

```
run_period: Run14
software: SL22d
collision_system: Au+Au
energy: 200 GeV
```

Why they matter

- Machine-readable dataset descriptions
- Uniform metadata across experiments
- Built for long-term discoverability

DataCard

- DOI
- Run period
- Software versions
- Configuration
- Calibration
- Trigger configuration
- Provenance
- Pointers
- How to use
- ...

AI Integration

- Indexed by SciBot
- Natural-language searchable
- Compatible with emerging national AI initiatives
- Generated from internal documentation

Datacard makes a dataset searchable, citable, and machine-readable

Brookhaven National Laboratory

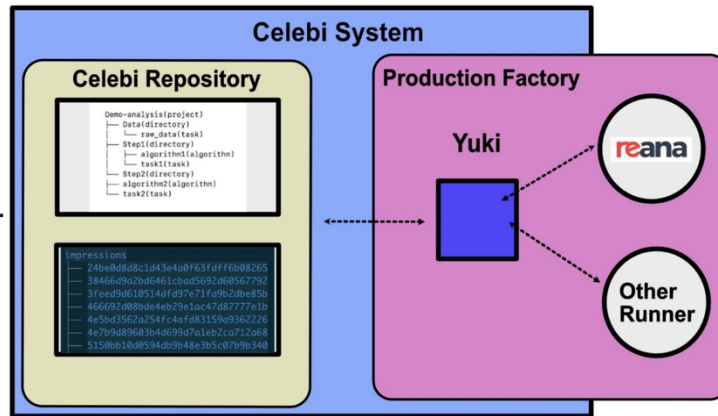
E. Lancon, RHIC DAP, CHEP 2026

12

Increasing use of automated workflows for reproducibility

Overview of the celebi system

- **Celebi repository**
 - Workflow system.
 - Versioning system (impressions).
- **Yuki (DITE) middleware.**
- **Runners (external)**
 - REANA.
 - Local runner.
 - Other runners.



CHEP2026

25-May-2026

4

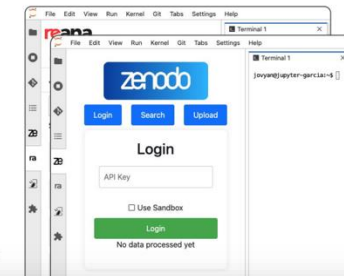
[Use of Celebi, slide from Mingrui Zhao](#)

Jupyter workflows as the standard

More Jupyter magic

Supporting the end-to-end user analysis

- **REANA Jupyter extension**
 - To browse, create, and run workflows
- **Zenodo Jupyter extension**
 - To browse the Zenodo Catalog, and be able to pull / push to the repository without getting out of the session

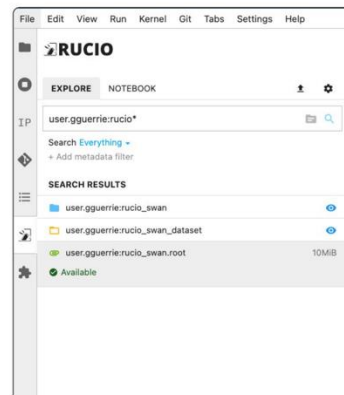


[Presentation link](#)

The Rucio Jupyterlab extension

What does it do?

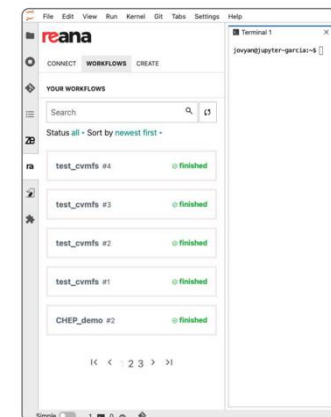
- Browse Rucio data from the Lab sidebar
- Replicate data with just **one click**
- **Resolves file path automatically**
- **Inject** path to notebook as a variable
- Supports three methods of authentication
 - Username & Password
 - X.509 User Certificate (or Proxy)
 - OIDC tokens (with limitations)
- Supports two modes of operation:
 - **Replica mode:** uses network-attached storage as a Rucio Storage Element (RSE), utilizes Rucio's file transfer capability.
 - **Download mode:** downloads data directly to the user's directory



More Jupyter magic

Supporting the end-to-end user analysis

- **REANA Jupyter extension**
 - To browse, create, and run workflows



Live student progress when teaching in a Jupyter notebook: <https://cadence-dash.com/>



cadence

Live student **progress** for Jupyter teaching.

Drop two lines into any notebook and watch the class solve in **real time** — per-checkpoint solve rates, common wrong answers, who needs help.

IN THE MOMENT

Improved communication

- Pace your lessons perfectly
- See who's stuck in real time
- Showcase solutions without pushing files
- Monitor student understanding and quick-patch gaps
- Optional auto-hints when they're stuck
- Built-in solutions after N attempts, no second notebook

OVER TIME

Improving teaching

- Measure your lesson effectiveness
- Measure how hard your sessions really are
- Surface the common wrong answers
- Compare solutions across students
- Spot AI use, copy-paste, convergent approaches

[Check out slides for Liv Helen Vage](#)

1



Cadence

live dashboard for Jupyter

cadence-dash.com

2



ML resources

a study companion for ML

livaage.github.io/ml_resources

3



hep-ml-templates

a modular ML pipeline

[github.com/livaage/hep-](https://github.com/livaage/hep-ml-templates)

HEP training, HSF training, and ROOT training all using Jupyter notebooks for teaching!

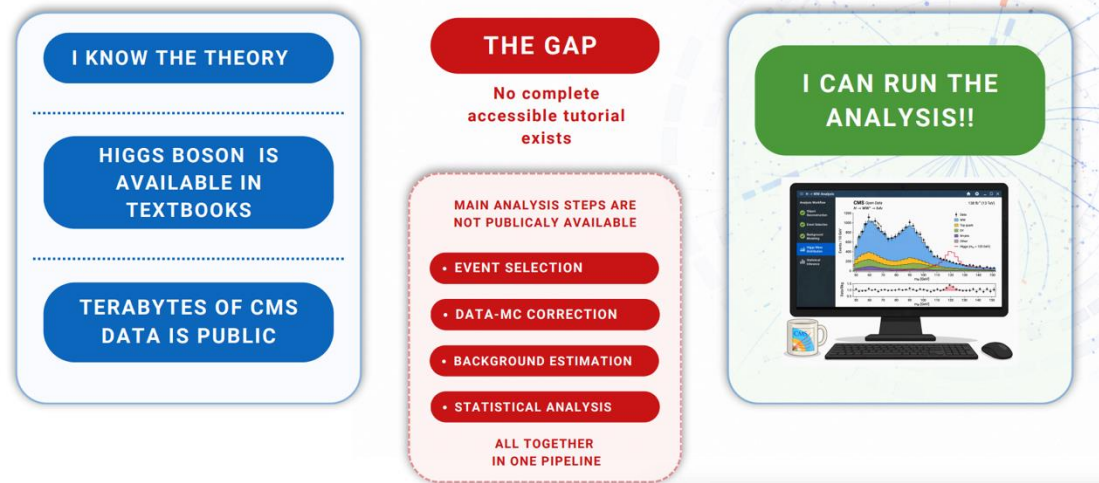
How to add content?

- **Case 1: you have one or two training resources**
-> login using CERN SSO and manually register a material/event and enter as much information as possible to make it discoverable, e.g., title, URL, description, keywords, etc.,
- **Case 2: you have or know a webpage with lots of training resources** (e.g., HSF, indico)
-> contact us at contact.heptraining@cern.ch
- **Case 3: you are a project, institution, experiment or university having multiple training resources to share**
-> contact us at contact.heptraining@cern.ch to be featured as a Content Provider.

Can register course content with HEP training catalog ([HEP talk slides](#))

Full student analysis of open data with reproducible workflow ([CMS](#)):
- Using only open data, reproduce published result workflow, [example is here](#)

MOTIVATION—TO BRIDGE THIS GAP



Hosted BinderHub for tutorials:

[Slides from Fengping Hu](#)

Summary

A robust, scalable training platform for the HEP community

01

Hosted, not just launched

Uniform, reproducible, browser-based environments — zero install for participants.

02

Built for workshop scale

GPU, guaranteed QoS, pre-pulling, and multi-cluster federation out to NRP and beyond.

03

Operated as a platform

CI/CD, Flux GitOps, and a 5-step remote-cluster recipe keep it maintainable.

Acknowledgments

IRIS-HEP (NSF OAC-1836650) · National Research Platform · OSG / PATH · HSF community · Princeton, UNL, UChicago

TRY IT

binderhub.ssl-hep.org

CONTACT

ssl-team@iris-hep.org

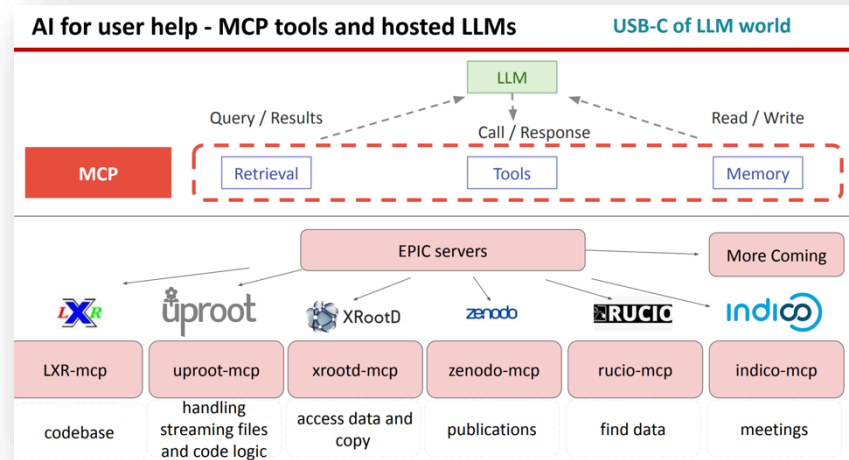
CODE

maniaclab/binderhub



So how's ePIC doing?

Nice talk from Sasha with demo!



[Link to slides](#)

- Use of AI to support documentation
- Moving towards reproducible workflows with JupyterHub
- Use of datacards, and already setting up “data” preservation

"With great software comes great responsibility
— to teach the people who'll use it."

Food for thought:

- Would new students/collaborators be interested in creating a workflow to reproduce TDR/early science plots?
- Database for reproducible workflows, linked to “published” plots, tutorials

“Every hour we save a collaborator is an hour returned to physics.” (also featured in Track summary)