October 19 - 25, 2024

CHEP 2024

Collaboration, Reinterpretation, Outreach and Education

Track 8 highlights

Lene Kristian Bryngemark (Lund), James Catmore (Oslo), **Giovanni Guerrieri (CERN)**, Jake Bennet (Mississippi) on behalf of the track 8 speakers

October 25th 2024

Our speakers

Lauren Mowberry (STFC/UKRI)

Thomas James (CERN)

Piet Nogga (Bonn)

Greg Corbett (STFC/UKRI)

Joni Pham (CDM)

Rodrigo Sierra (CERN)

Roger Jones (Lancaster)

Kevin Pedro (Fermilab)

Muhammad Imran (PK)

Kyle Knoepfel (Fermilab)

Maxim Potekhin (BNL)

Eoin J. Clerkin (FAIR)

Mindaugas Sărpis (Vilnius)

Axel Naumann (CERN)

Giovanni Guerrieri (CERN)

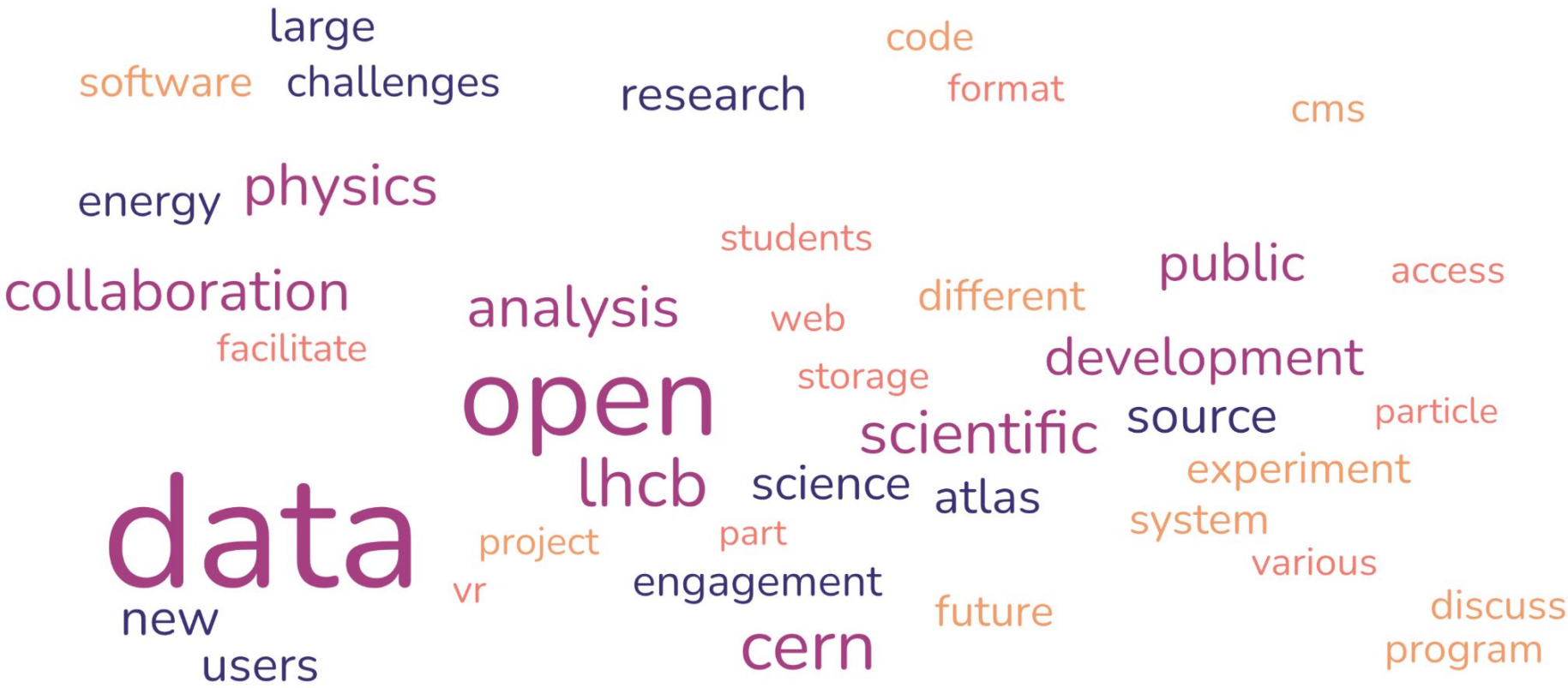
Pablo Saiz (CERN)

Gerardo Ganis (CERN)

Gordon Watts (UW/Seattle)

Kati Lassila-Perini (Helsinki)

Jim Pivarski (Princeton)


large
software
challenges
research
code
format
cms
energy
physics
students
public
access
collaboration
analysis
different
web
facilitate
development
storage
open
scientific
source
particle
lhcb
science
atlas
experiment
data
project
part
system
various
vr
engagement
future
discuss
new
cern
program
users

A tentative outline

Stealing from Axel's slides

Some of CERN's Open Science elements, from its Policy:

Open Data

Open Source

Research Integrity

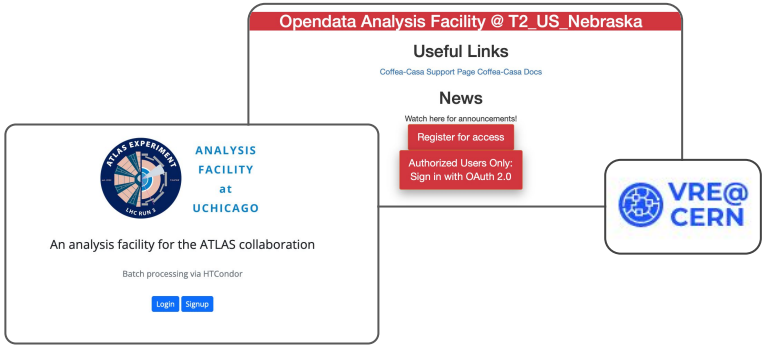
Training and Outreach

(?)

Open Data

The dawn of analysis facilities

Sometimes our laptop is not enough



Available at the click of a button ✅

Data persistence ✅

No account restrictions ❌

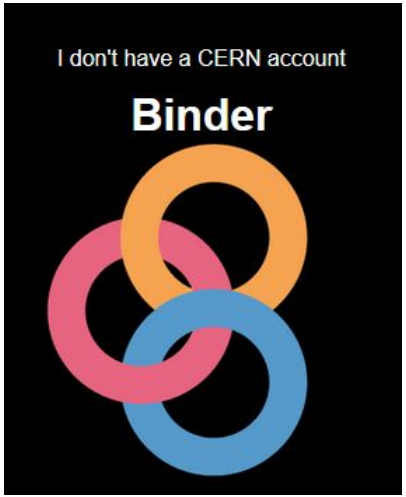
No timeout time for sessions ✅

Spawn time <1min ✅

Change easily the software stack ✅

I have a CERN account

SWAN



✅ Available at the click of a button

❌ Data persistence

✅ No account restrictions

❌ No timeout time for sessions (1 CPU-h max)

✅ Spawn time ~O(min)

❌ Change easily the software stack

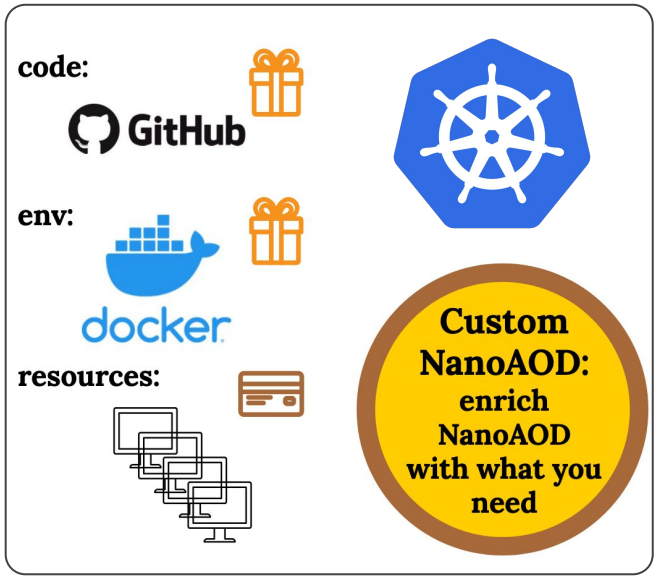
Need to re-build the underlying image

Giovanni Guerrieri (CERN)

Open Data

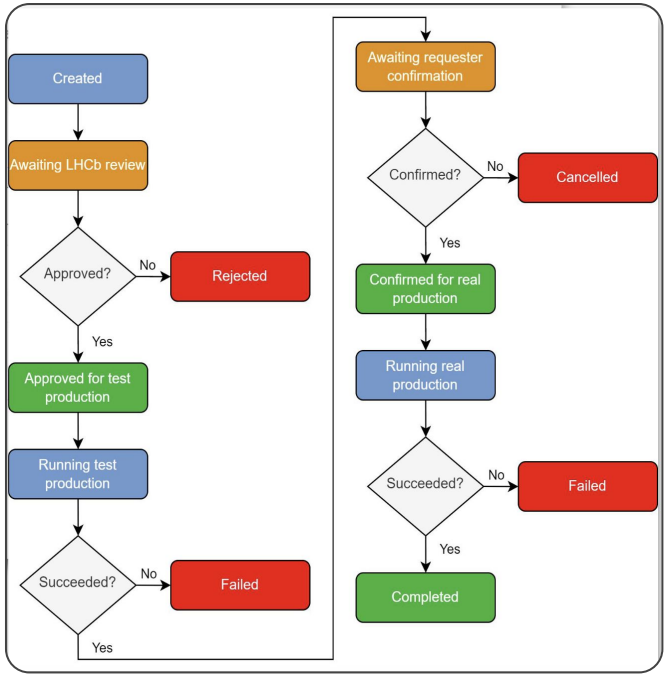
Data processing as a service

Customizing CMS NanoAOD on clouds is possible, and (reasonably) affordable



Kati Lassila-Perini (Helsinki)

The LHCb Ntupling service (coming soon - 2025)

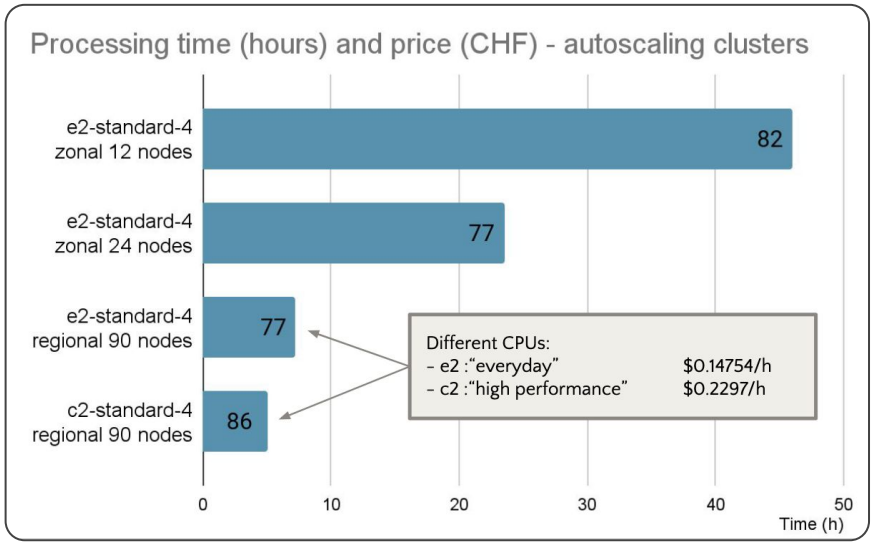


Piet Nogga (Bonn)

Open Data

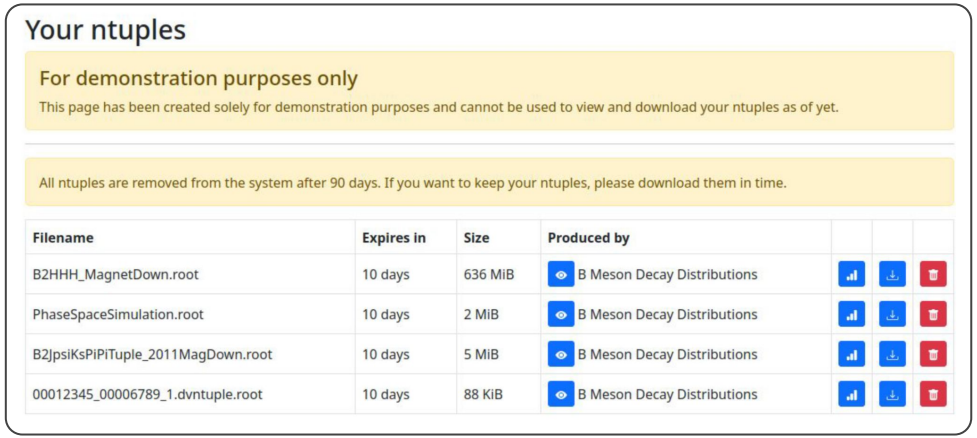
Data processing as a service

Customizing CMS NanoAOD on clouds is possible, and (reasonably) affordable



1TB of data

The LHCb Ntupling service (coming soon - 2025)



Kati Lassila-Perini (Helsinki)

Piet Nogga (Bonn)

Open Data

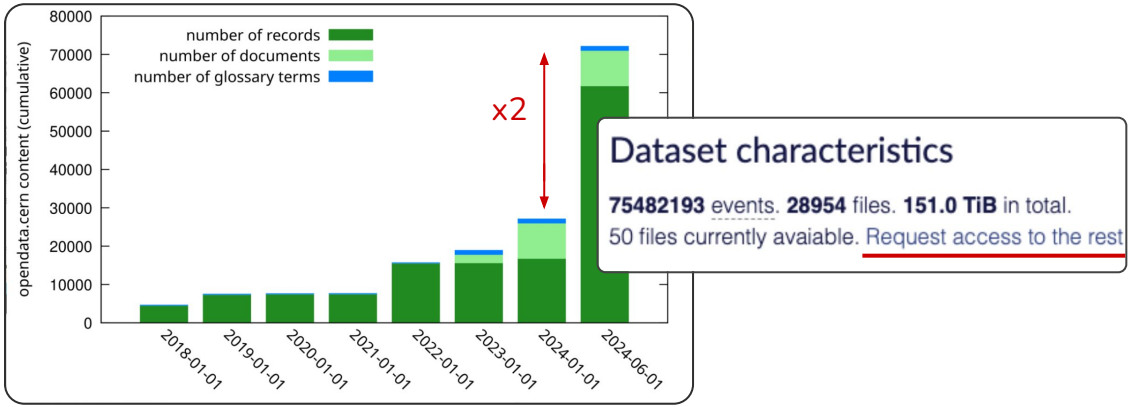
Store and preserve our data

Bringing cold storage to the CERN Open Data portal

Files might be stored on cheaper media, not immediately accessible.

Interface with FTS and CTA to archive/stage files

First round with manual operations triggered by user/curators requests



Pablo Saiz (CERN)

Data Preservation in High Energy Physics

2023 10-years report (DPHEP Collaboration)

DPO = Data Preservation Organisation

0: DPO during experiment proposal.

1: DPO during data taking.

2: DPO after data taking and during analysis-only mode.

3: DPO after the collaboration funding scheme.

4: DP Rescue organisational scheme

Taking no action == decommissioning (deleting) the data.

Gerardo Ganis (CERN)

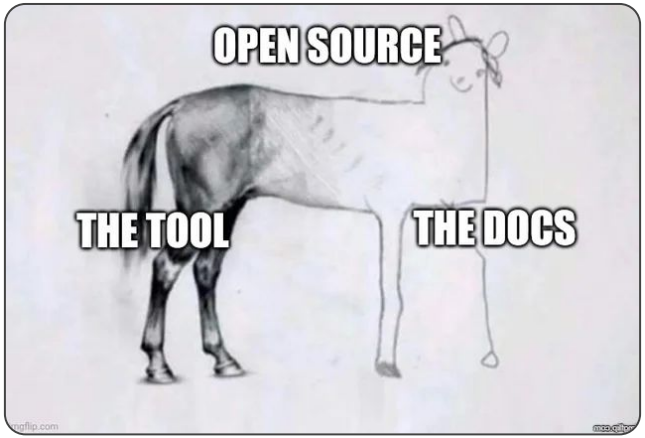
Open Source

Changing the (HENP) software world, pragmatically

CERN Open Source Program Office (OSPO)

OSPO ≠ open-source evangelists (but they are, really)

- CERN-wide official recommendations about procedures and best practices
- Guidance on how to run open-source projects, hardware-specific aspects
- Providing open-source hardware and software catalogs
- Center of expertise, providing KPIs and story telling



Axel Naumann (CERN)

Open Science

The CMB collaboration successfully tackling Open Science best practices

The Compressed Baryonic Matter experiment

Ensuring that Open Science policy aids advancement to the collaboration's physics goals

Validation and verification tests on an open data policy

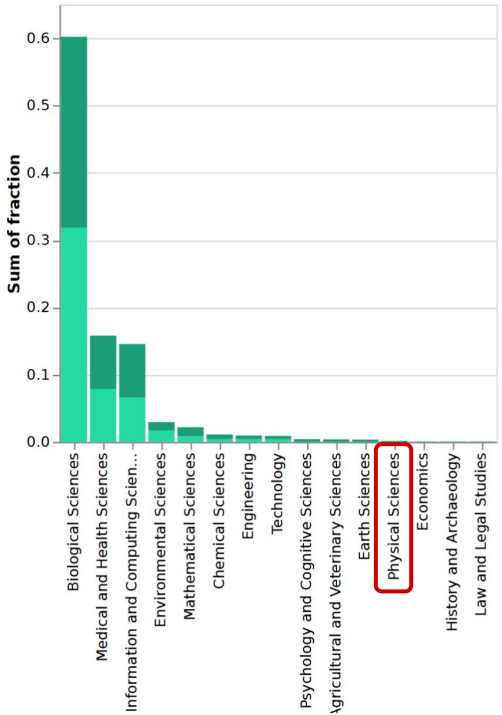
Providing an open and licensed software suite

Eoin J. Clerkin (FAIR)

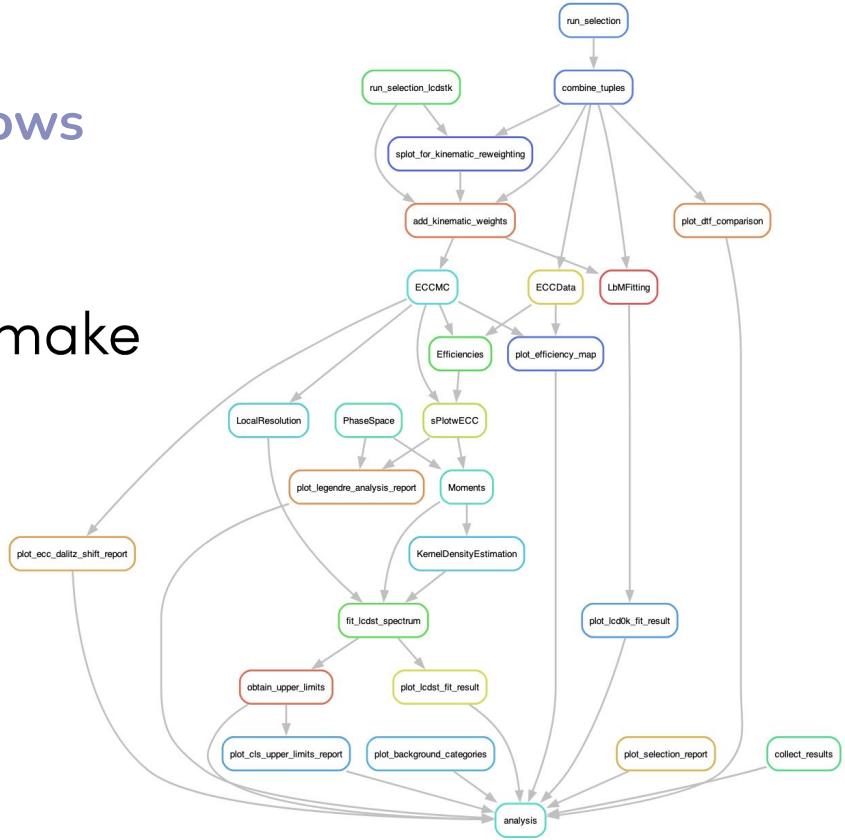
Facility for **A**ntiProton and **I**on **R**esearch goes **F**indable **A**ccessible **I**nteroperable **R**eusable

Research integrity

Or how to manage our physics workflows



snakemake

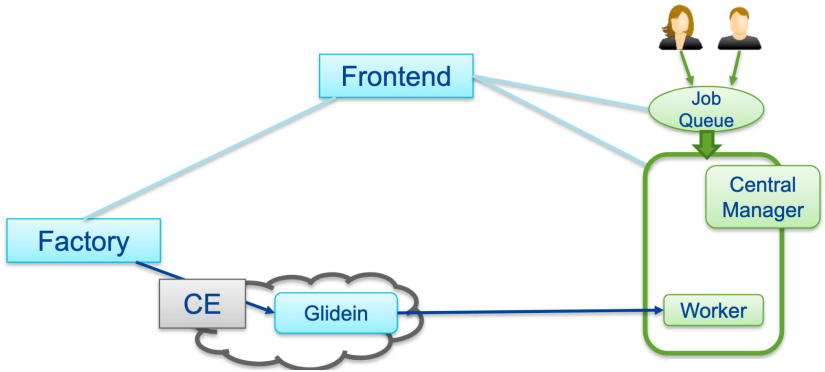


Mindaugas Sărpis (Vilnius)

Training and Outreach

Building the next generation of scientists

Learn how to use GlideinWMS with containers here



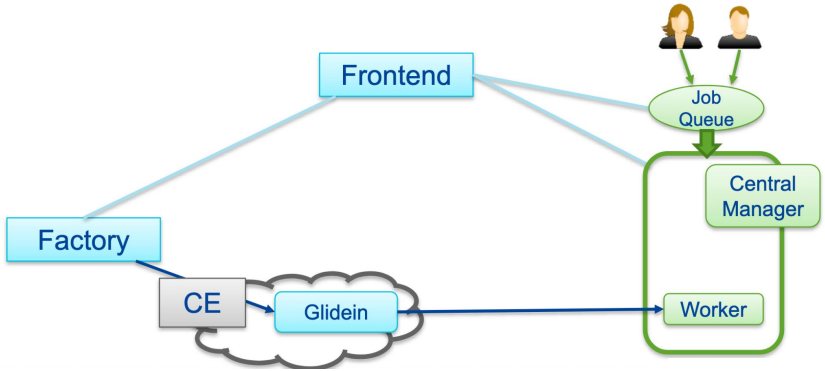

```
mkdir ws-test; cd ws-test
TEST_DIR=$(pwd)
git clone https://github.com/glideinWMS/containers.git
cd containers/workspaces
mkdir "$TEST_DIR"/gwms #Optional
GWMS_PATH="$TEST_DIR"/gwms/ podman-compose up -d
```

Kevin Pedro (Fermilab)

Training and Outreach

Building the next generation of scientists

Learn how to use GlideinWMS with containers here

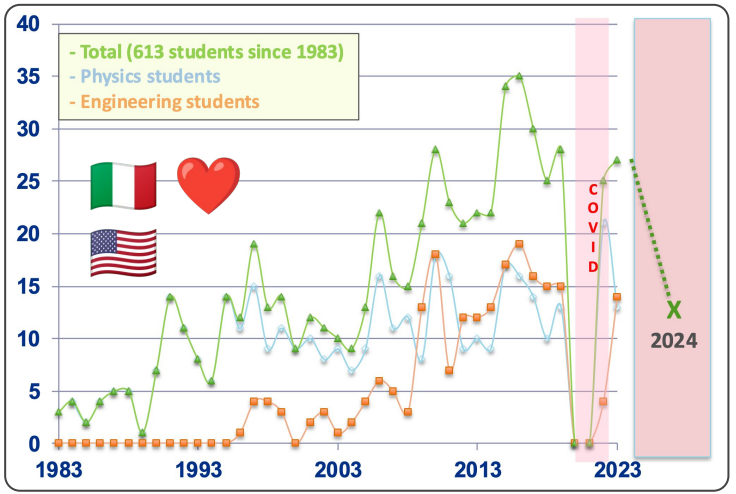



```
mkdir ws-test; cd ws-test
TEST_DIR=$(pwd)
git clone https://github.com/glideinWMS/containers.git
cd containers/workspaces
mkdir "$TEST_DIR"/gwms #Optional
GWMS_PATH="$TEST_DIR"/gwms/ podman-compose up -d
```

Kevin Pedro (Fermilab)

Are you Italian? Are your students italian?

Try the Italian summer student program (not only) at Fermilab!



Kyle Knoepfel (Fermilab)

Training and Outreach

Discovering experiments

The ATLAS Virtual visit program

Joni Pham (CDM)

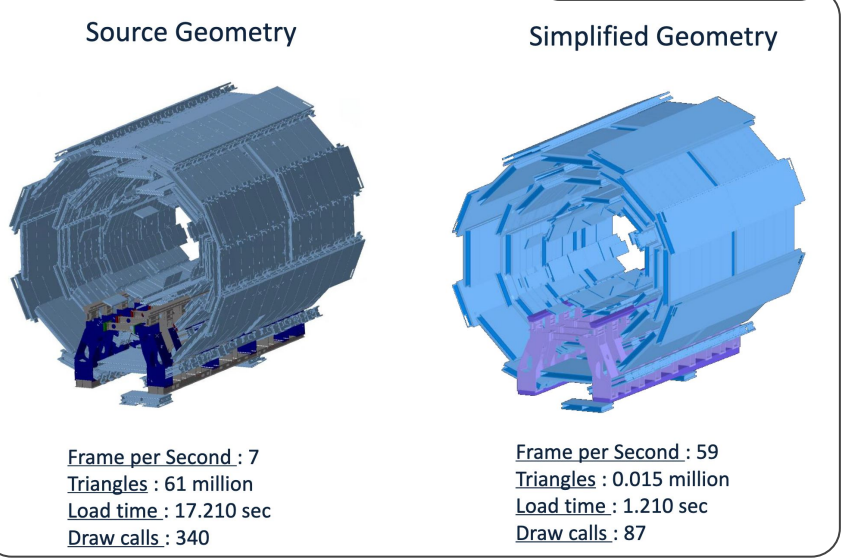
Training and Outreach

Discovering experiments

The ATLAS VR Application

The ATLAS Virtual visit program

Joni Pham (CDM)



Roger Jones (Lancaster)

Training and Outreach

Discovering STEM in the UK

UK Research and Innovation

Science and Technology Facilities Council

REMOTE³

Students work to program LEGO Mars Rovers that would operate 1.1km underground.

[Lauren Mowberry (STFC/UKRI)]

The STFC Open Weeks project

-Engage and involve under-served communities with STFC's work

-Enable participants to feel that "science and technology are for people like me"

-STFC Open Weeks run every 4 years. Open Week 2027?

[Greg Corbett (STFC/UKRI)]

Trainin

Discovering

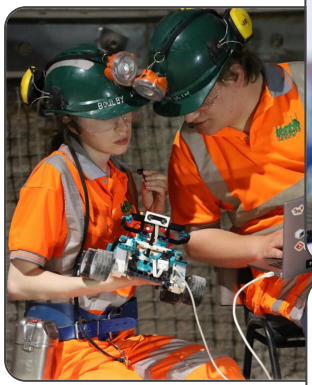
Science and Technology Facilities Council

n Weeks project

olve under-served
n STFC's work

nts to feel that
nology are for

eks run every 4 years.
7?



"I loved working with all your colleagues but in my opinion, Will was my favourite but you're all fantastic"
[A Will's fan]

Lauren Mowberry (STFC/UKRI)

Greg Corbett (STFC/UKRI)

Fueling innovation through collaboration

The CERN Openlab industry-science journey

Summer Student Programme

Provides undergraduate and master's level students with an opportunity to work on one of the R&D projects for nine weeks under experts' supervision

This year there was a record of more than 6600 applicants!

CURRENT INDUSTRY AND RESEARCH MEMBERS

SIEMENS · micron · E4 COMPUTER ENGINEERING · ORACLE · cerabyte · UNIVERSITÀ DEGLI STUDI DI TRIESTE · SIMONS FOUNDATION · INFN · intel

PRE-AGREEMENT STAGE

PURESTORAGE · PASQAL · NVIDIA · Johnson&Johnson

In discussions and negotiations with ~10 more

OBJECTIVES

High-level: Accelerating Computing for Science

Pioneering sustainable and emerging computing and storage solutions

Harnessing heterogeneous computing and AI for a greener future

Fostering synergies and technology transfers between industry and sciences

Sustainable Infrastructures

- Heterogeneous computing platforms and infrastructures
- Computer architectures and software engineering
- Storage and data management
- Artificial intelligence algorithms, platforms and applications
- Applications for society and environment

Emerging Technologies

- New materials for long term digital storage
- Digital twins
- Quantum computing and networks

Thomas James (CERN)

AI

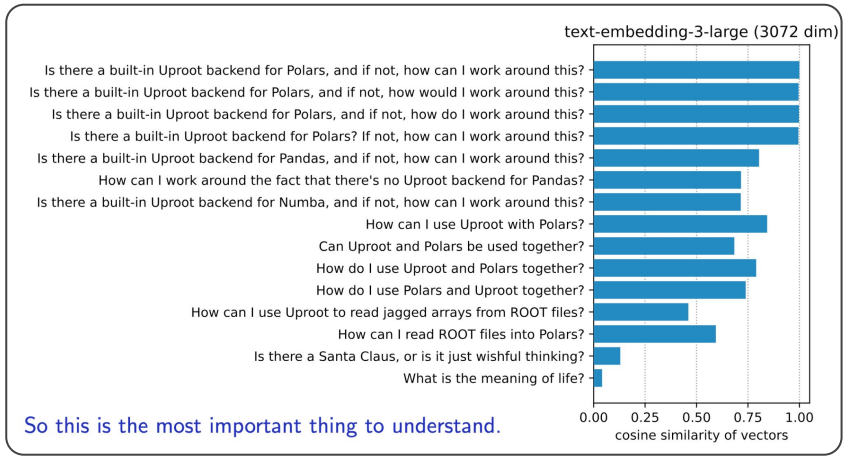
Harness LLMs to help us

A careful approach

https://hep-help.org

GitHub discussion format, with LLM-chatbot to help.

Without a good embedding, we won't find the right documents



Jim Pivarski (Princeton)

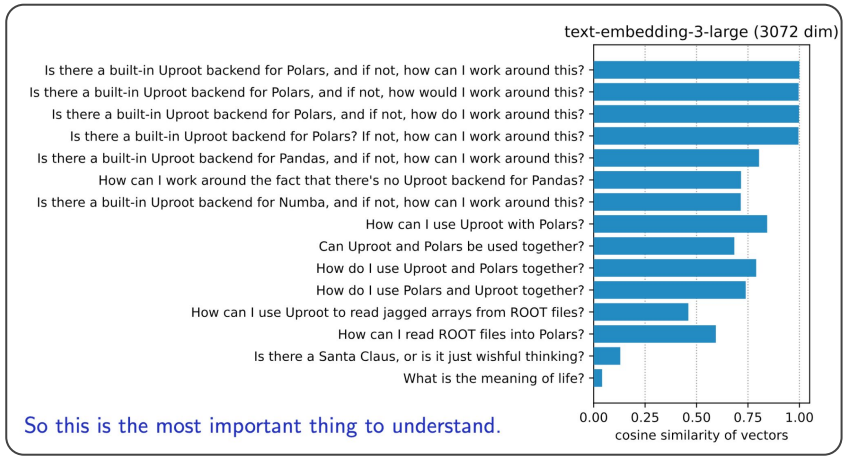
Harness LLMs to help us

A careful approach

https://hep-help.org

GitHub discussion format, with LLM-chatbot to help.

Without a good embedding, we won't find the right documents



Jim Pivarski (Princeton)

Navigating conference abstracts

Open-Source configuration and running is painful

The large models really do make a big difference for this sort of task

```
# Raw prompt for the LLM
abstract_ranking_prompt = """Help me judge the following conference presentation as interesting or
not by summarizing the abstract and ranking it according to topics I'm interested in or not.
"""

interested_topics = [
    "Hidden Sector Physics",
    "Long Lived Particles (Exotics or RPV SUSY)",
    "Analysis techniques and methods and frameworks, columnar analysis, particularly those based around python or "
    "ROOT's DataFrame (RDF)",
    "Machine Learning and AI for particle physics",
    "The ServiceX tool",
    "Distributed computing for analysis (e.g. Dask, Spark, etc)",
    "Data Preservation and FAIR principles",
    "Differentiable Programming",
]

not_interested_topics = [
    "Quantum Computing",
    "Lattice Gauge Theory",
    "Neutrino Physics",
]
```

Gordon Watts (UW/Seattle)

Harness LLMs to help us

A careful approach

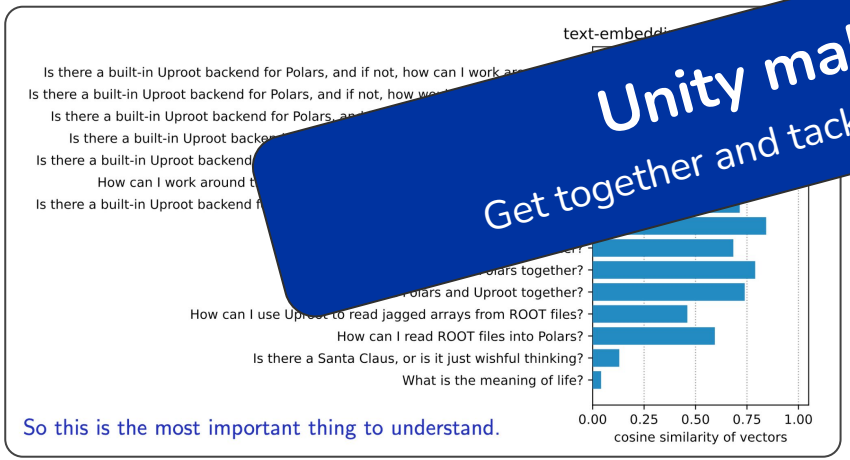
https://hep-help.org

GitHub discussion format, with LLM-chatbot to help.

Without a good embedding, we won't find the right documents

Unity makes strength

Get together and tackle common LLM challenges!



Jim Pivarski (Princeton)

Navigating conference abstracts

Open-Source config… g is painful

The t… ifference

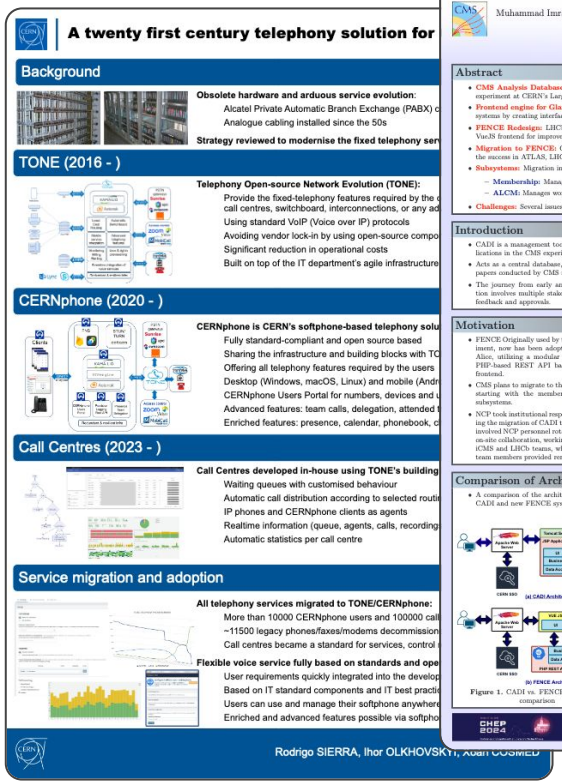
```
…ollowing conference presentation as interesting or
…nking it according to topics I'm interested in or not.

…ted_topics = [
    "Hidden Sector Physics",
    "Long Lived Particles (Exotics or RPV SUSY)",
    "Analysis techniques and methods and frameworks, columnar analysis, particularly those based around python or "
    "ROOT's DataFrame (RDF)",
    "Machine Learning and AI for particle physics",
    "The ServiceX tool",
    "Distributed computing for analysis (e.g. Dask, Spark, etc)",
    "Data Preservation and FAIR principles",
    "Differentiable Programming",
]

not_interested_topics = [
    "Quantum Computing",
    "Lattice Gauge Theory",
    "Neutrino Physics",
]
```

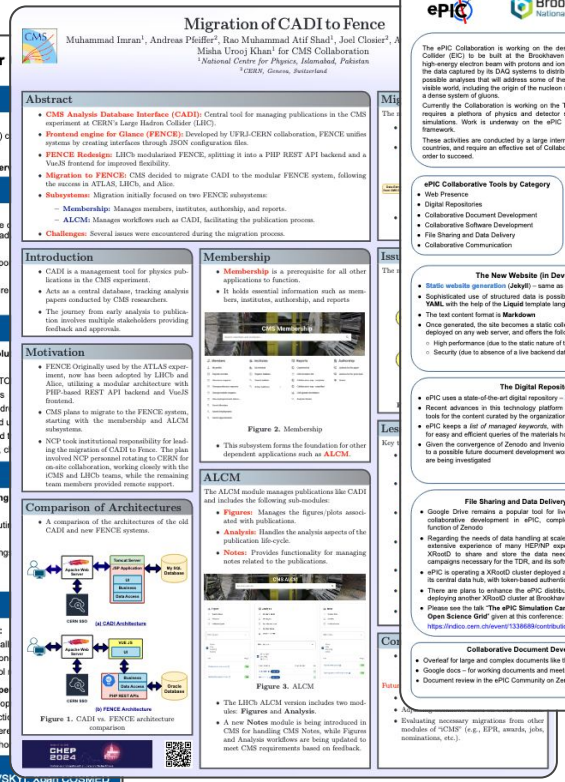
Gordon Watts (UW/Seattle)

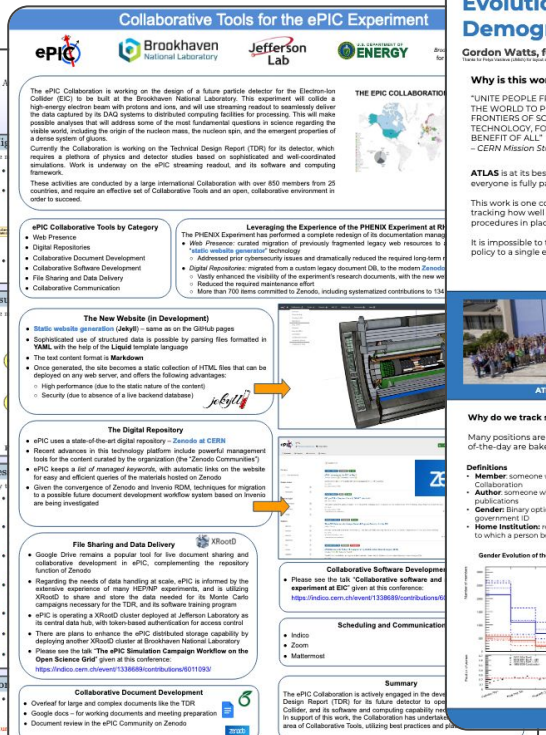
Summary

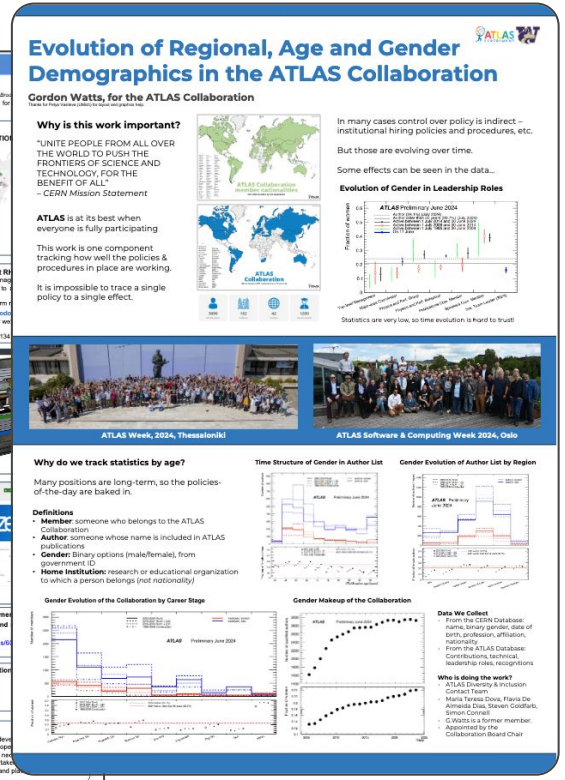
- Open Data is looking good and is gaining momentum
  - Education, training, research
- Enabling analysis and data preservation is critical (always has been)
  - Sustainable workflows, data preservation best practices
- Collaborating to accelerate physics research!
  - Enable researchers to do better science with less computing challenges
- Investing in the next gen of HEP is important
  - Analysis facilities, LLM helpers
- Building the next gen of scientists
  - Outreach, training, and public engagement

Posters









[Rodrigo Sierra (CERN)](#) [Muhammad Imran (PK)](#) [Maxim Potekhin (BNL)](#) [Gordon Watts (UW/Seattle)](#)

Thanks

To our speakers, to our organisers, and to everyone!