



Norfolk, Virginia, USA • May 8-12, 2023

CHEP  
2023

Computing in High Energy & Nuclear Physics

# Track 8 Highlights

*Collaboration, Reinterpretation,  
Outreach and Education*

*Markus Diefenthaler (Jefferson Lab), Elena Gazzarrini (CERN),  
Clemens Lange (Paul Scherrer Institut), Michel Hernandez Villanueva (DESY)*



# Topics

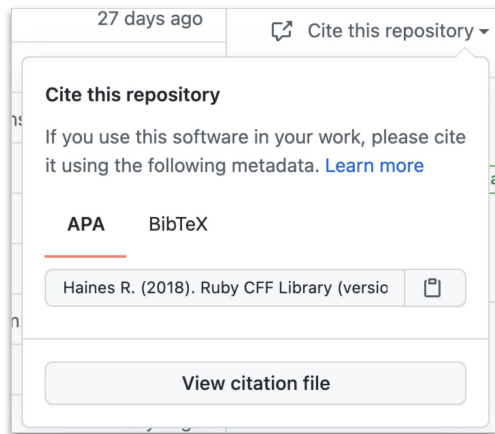
open data  
outreach  
collaborative infrastructure  
software citation  
industry  
training initiative

<https://indico.jlab.org/event/459/sessions/2042>

# Software Citation in HEP

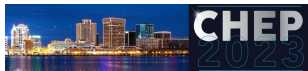
## Current State and Recommendations

- Software Citation and Recognition Workshop: **Experiments**, **SW projects** and **publishers** discussing how to handle **citation of software** and **recognition for software efforts**.
  - “An ongoing process that is not straightforward”
- Recommendations
  - Programmatic discovery of citations is important, consider using APIs
  - **Make clear how to cite in docs**
  - Adopt the Citation File Format as a common standard
  - Zenodo DOIs are common to HEP



[Feickert, Matthew \(U. of Wisconsin-Madison\) - Software Citation in HEP](#)

[How to use a CITATION.cff](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

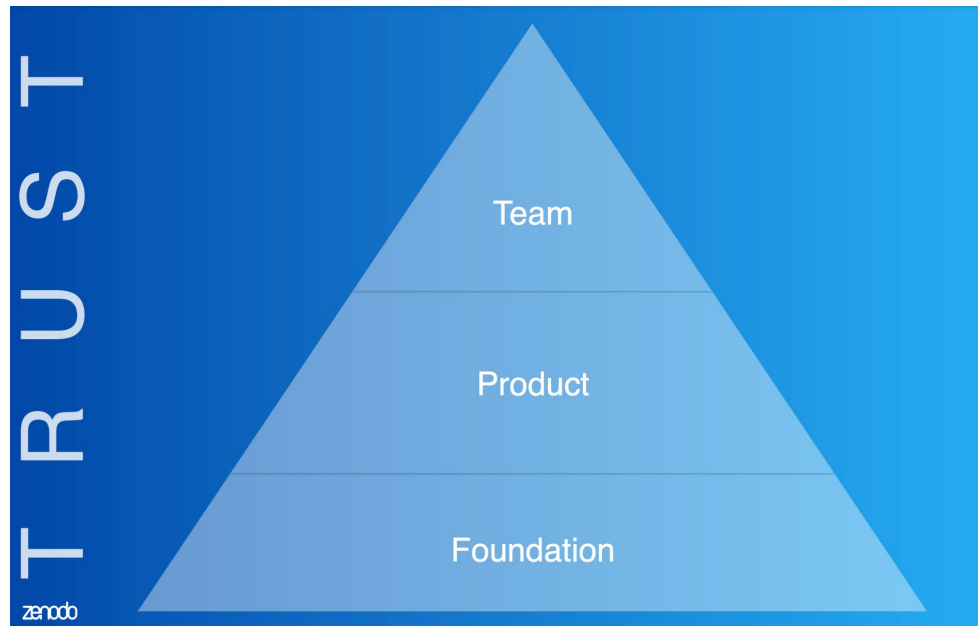


# Happy Birthday, Zenodo!



# Zenodo

300K users, 7500 organizations, 153 countries



Send them a  
birthday greeting!

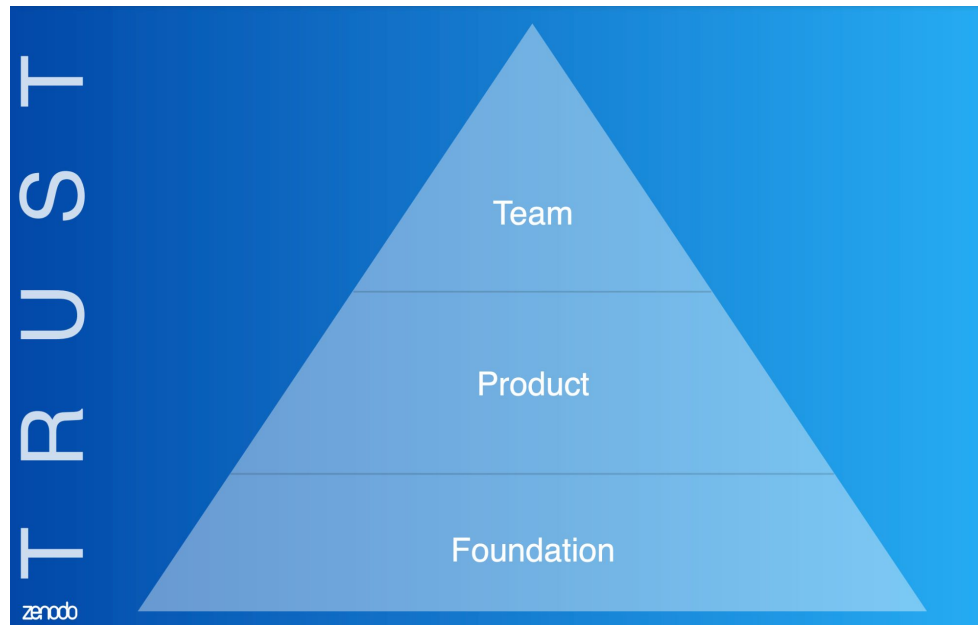


<https://zenodo.org/10years>

[Nielsen, Lars Holm \(CERN\) - 10 years of Zenodo](#)

# Zenodo

300K users, 7500 organizations, 153 countries



[Nielsen, Lars Holm \(CERN\) - 10 years of Zenodo](#)

Send them a  
birthday greeting!



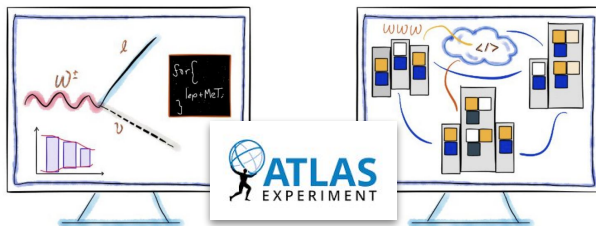
<https://zenodo.org/10years>



# Open data

The 8 TeV samples

The 13 TeV samples



Learn more about the 2016 datasets

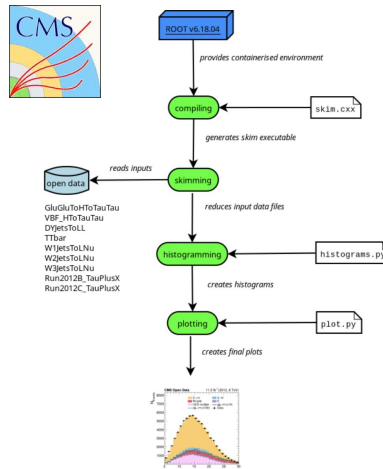
Explore the 10x more data in 2020 datasets

[Krasznahorkay, Attila \(CERN\)](#)



The Ntuple wizard!

[Fitzgerald, Dillon \(U. of Michigan\)](#)



reana

Your workflows

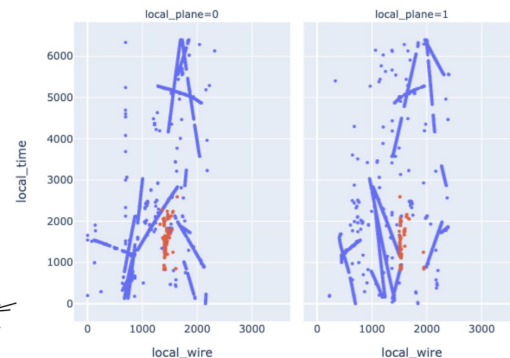
higgstotautau

Status ▼ ☐ Show deleted runs Latest first ▼

higgstotautau #1 finished in 6 min 50 sec  
Finished 11 minutes ago step 4/4

[Simko, Tibor \(CERN\)](#)

cosmic\_label plot



[Cerati, Giuseppe \(Fermilab\)](#)



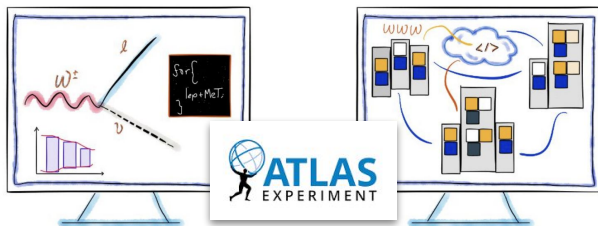
Fri 12th May, Plenary Session - Conference Highlights: Track 8



# Open data

The 8 TeV samples

The 13 TeV samples



Learn more about the 2016 datasets

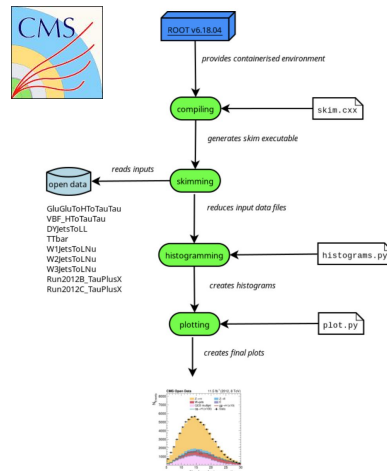
Explore the 10x more data in 2020 datasets

[Krasznahorkay, Attila \(CERN\)](#)



The Ntuple wizard!

[Fitzgerald, Dillon \(U. of Michigan\)](#)



**Common challenge:  
publish data in formats  
other than ROOT.**



reana

Your workflows

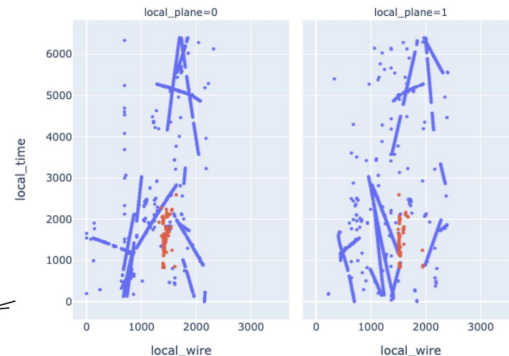
higgstotatau

Status ▼ ☐ Show deleted runs Latest first ▼

higgstotatau #1 finished in 6 min 50 sec  
Finished 11 minutes ago step 4/4

[Simko, Tibor \(CERN\)](#)

cosmic\_label plot



[Cerati, Giuseppe \(Fermilab\)](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Preservation

Analysis preservation in LHCb

Environment and initial workflow preserved →



**Metadata is key:** datasets can be identified by WG / analysis / tags.

Productions / [SL](#) / rds\_hadronic

Grouped tags: ☐ config ☒ datatype ☒ eventtype ☒ polarity ☐ sign

Drag to sort: config datatype eventtype polarity

rds_hadronic 24			
mc 20		lhcb 4	
2012 20		2012 4	
23903000 4	13563002 4	13266069 2	90000000 4
magdown 2	magdown 2	magdown 1 magup 1	magdown 2

[Couturier, Benjamin \(CERN\) - Facilitating the preservation of LHCb analyses](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Preservation

Analysis preservation in LHCb

Environment and initial workflow preserved →



**Metadata is key:** datasets can be identified by WG / analysis / tags.

Productions / **SL** / rds\_hadronic

Grouped tags: ☐ config ☒ datatype ☒ eventtype ☒ polarity ☐ sign

Drag to sort: config datatype eventtype polarity

rds_hadronic 24			
mc 20		lhcb 4	
2012 20		2012 4	
23903000 4	13563002 4	13266069 2	90000000 4
magdown 2	magdown 2	magdown 1 magup 1	magdown 2

[Couturier, Benjamin \(CERN\) - Facilitating the preservation of LHCb analyses](#)

BaBar experience with preservation

**“Data and analyses preservation can be done, but planning early for it can help a lot”.**

[Accessing the BaBar dataset via the BaBar Associates open-access program](#)



To contact the BaBar Management Team: [\[Click here\]](#)

To invite BaBar speakers to present results at physics conferences: [\[Click here\]](#)

**BaBar Management Team**

Michael Roney (Spokes) No of records displayed are: 28

	DOC #	BAIS Title
Fabio Anulli (Physics C)	<a href="#">NOTE-1572</a> ...	Corrections to EMC Timing
Tina Cartaro and Marcu	<a href="#">NOTE-1571</a> ...	The Care and Feeding of BaBar
	<a href="#">NOTE-1570</a> ...	EMC radFET radiation monitoring system
Janis McKenna (Spokes)	<a href="#">NOTE-1569</a> ...	Performance test and comparison of two storage management solutions using BaBar software framework at the Italian Tier-A
William Garv (Publicat	<a href="#">NOTE-1568</a> ...	Performance test and comparison of two storage management solutions using BaBar software framework at the Italian Tier-A

[Ebert, Marcus \(U. of Victoria\) - BaBar's Experience with Data Preservation](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Preservation

Analysis preservation in LHCb

Environment and initial workflow preserved →



**Metadata is key:** datasets can be identified by WG / analysis / tags.

Productions / 54 / rds\_hadronic

Grouped tags: ☐ config ☒ datatype ☒ eventtype ☒ polarity ☐ sign

Drag to sort: config datatype eventtype polarity

rds_hadronic 24			
mc 20			
2012 20			
23903000 4	13563002 4	13266069 2	lhcb 4
magdown 2	magdown 2	magdown 1 magup 1	2012 4

[Couturier, Benjamin \(CERN\) - Facilitating the preservation of LHCb analyses](#)

BaBar experience with preservation

**“Data and analyses preservation can be done, but planning early for it can help a lot”.**

[Accessing the BaBar dataset via the BaBar Associates open-access program](#)

To contact the BaBar Management Team: [\[Click here\]](#)

To invite BaBar speakers to present results at physics conferences: [\[Click here\]](#)

BaBar Management T

Michael Roney (Spokes)

No of records displayed are: 28

Fabio Anulli (Physics C

Tina Cartaro and Marcu

Janis McKenna (Spokes

William Garv (Publicat

DOC #

BAIS Title

NOTE-1572

NOTE-1571

NOTE-1570

NOTE-1569

NOTE-1568

Corrections to EMC Timing

The Care and Feeding of BaBar

EMC radFET radiation monitoring system

Performance test and comparison of two storage management solutions using BaBar software framework at the Italian Tier-A

Performance test and comparison of two storage management solutions using BaBar software framework at the Italian Tier-A



[Ebert, Marcus \(U. of Victoria\) - BaBar's Experience with Data Preservation](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

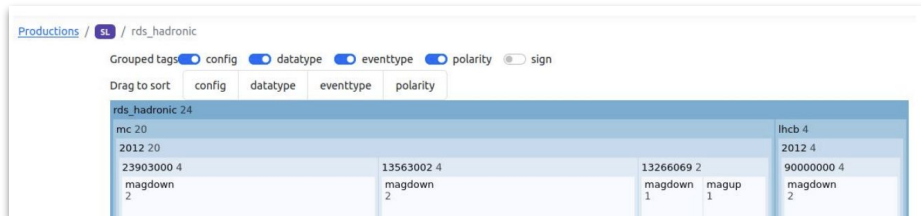
# Preservation

Analysis preservation in LHCb

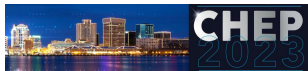
Environment and initial workflow preserved →



**Metadata is key:** datasets can be identified by WG / analysis / tags.



[Couturier, Benjamin \(CERN\) - Facilitating the preservation of LHCb analyses](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

BaBar experience with preservation

**“Data and analyses preservation can be done, but planning early for it can help a lot”.**

[Accessing the BaBar dataset via the BaBar Associates open-access program](#)

To contact the BaBar Management Team: [\[Click here\]](#)

To invite BaBar speakers to present results at physics conferences: [\[Click here\]](#)

**BaBar Management T**

Michael Roney (Spokes) No of records displayed are: 28

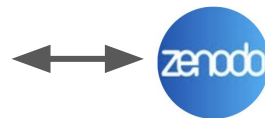
DOC #	BAIS Title
NOTE-1572 ...	Corrections to EMC Timing
NOTE-1571 ...	The Care and Feeding of BaBar
NOTE-1570 ...	EMC radFET radiation monitoring system
NOTE-1569 ...	Performance test and comparison of two storage management solutions using BaBar software framework at the Italian Tier-A
NOTE-1568 ...	Performance test and comparison of two storage management solutions using BaBar software framework at the Italian Tier-A



[Ebert, Marcus \(U. of Victoria\) - BaBar’s Experience with Data Preservation](#)

Universal FeynRules Output (UFO) Models

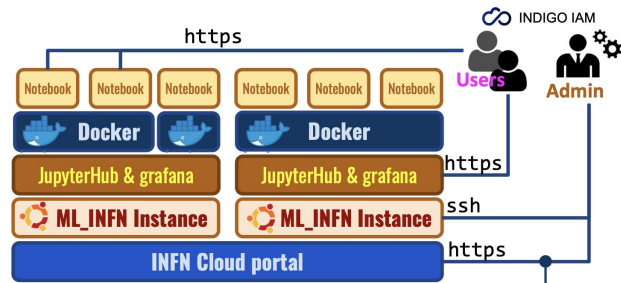
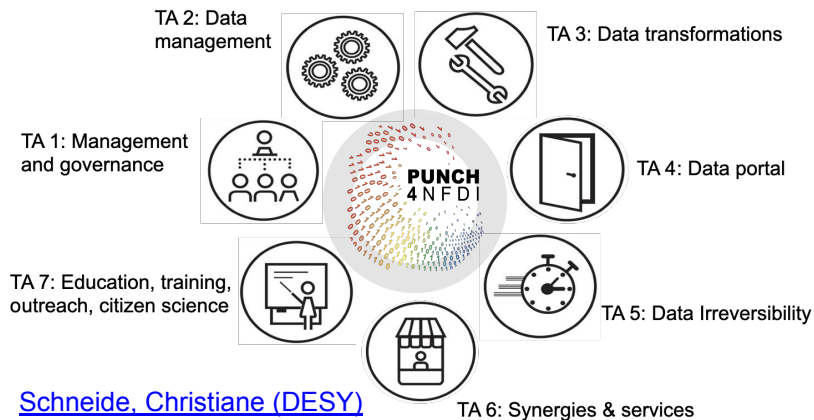
Model-independent files	Model-dependent files
__init__.py	particles.py
object_library.py	coupling_orders.py
function_library.py	parameters.py
write_param_card.py	vertices.py
	couplings.py
	lorentz.py



[Roy, Avik \(U. of Illinois\) - FAIR principles for Digital Objects in HEP](#)



# Collaborative Infrastructure



National Czech Programme

eosc

<b>AUGER</b>	<b>AUGER2</b>	<b>CTA-N</b>
1622659 images	80352 images	1161783 images
First Night: 20140101	First Night: 20220401	First Night: 20181017
Latest Data: 20230502	Latest Data: 20230502	Latest Data: 20230502

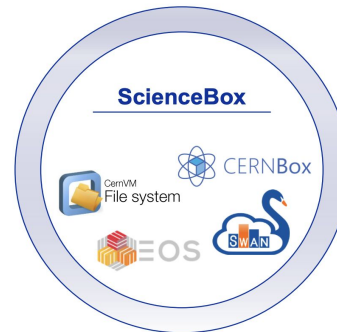
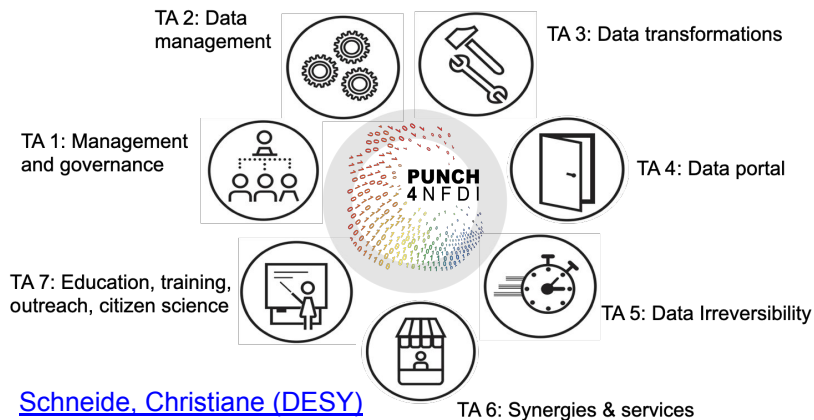
**Giommi, Luca (INFN Bologna)**

**Chudoba, Jiri (FZU)**



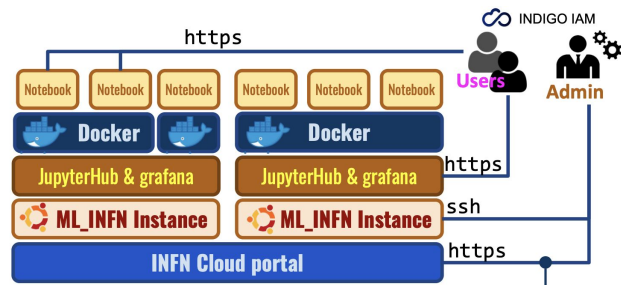
Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Collaborative Infrastructure



“A self-contained and easy-to-use package”.

[Bocchi, Enrico \(CERN\)](#)



National Czech Programme

eosc

<b>AUGER</b>	<b>AUGER2</b>	<b>CTA-N</b>
1622659 images	80352 images	1161783 images
First Night: 20140101	First Night: 20220401	First Night: 20181017
Latest Data: 20230502	Latest Data: 20230502	Latest Data: 20230502

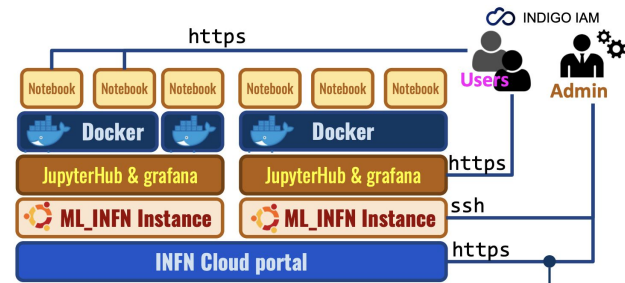
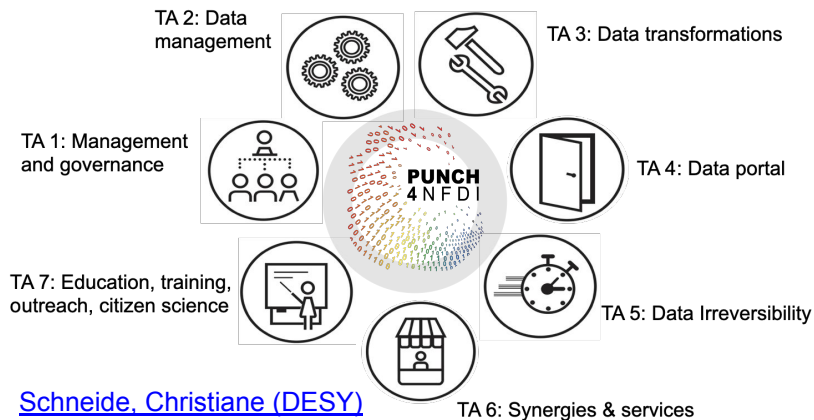
[Giommi, Luca \(INFN Bologna\)](#)

[Chudoba, Jiri \(FZU\)](#)



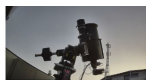


Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Collaborative Infrastructure

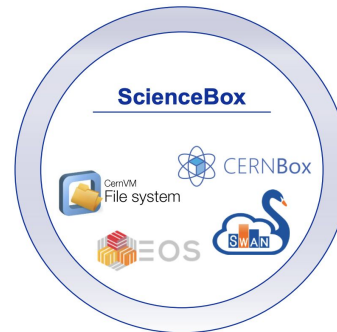


National Czech Programme

eosc

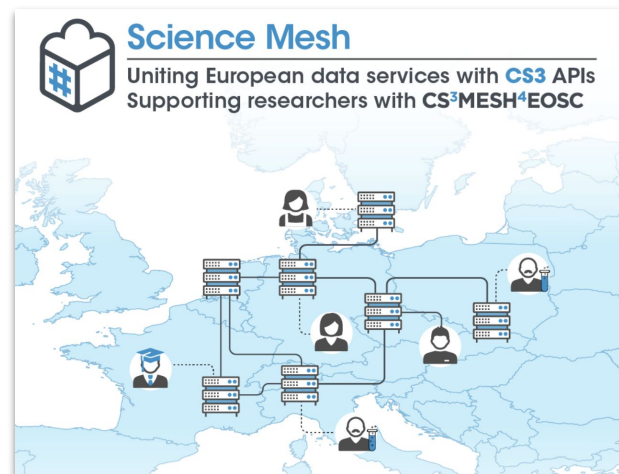
		
<b>AUGER</b>	<b>AUGER2</b>	<b>CTA-N</b>
1622659 images	80352 images	1161783 images
First Night: 20140101	First Night: 20220401	First Night: 20181017
Latest Data: 20230502	Latest Data: 20230502	Latest Data: 20230502

**Chudoba, Jiri (FZU)**



“A self-contained and easy-to-use package”.

**Bocchi, Enrico (CERN)**



**Gonzalez, Hugo (CERN)**



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Training

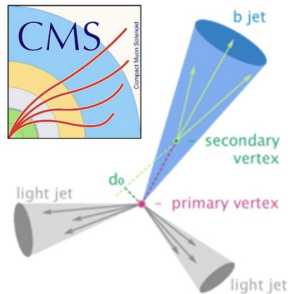
The ML\_INFN lecture Program



Jet b-tagging at CMS

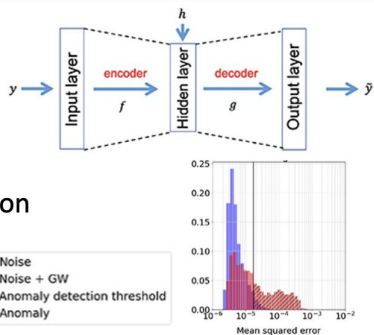
Recurrent Neural Networks with LSTM

Hands-on and exercises



Gravitational Waves with Virgo

Autoencoders, anomaly detection and compression



[Giommi, Luca \(INFN Bologna\)](#)

FPGAs in the Cloud for students!

Develop



- VMs hosted on the INFN Cloud infrastructure;
- Python environment with ML libraries to develop Neural Networks;
- Command line interface with Jupyter notebooks support;
- HLS4ML and Vivado Design Suite to produce FPGA firmware;
- Available during and after the workshop.

Deploy



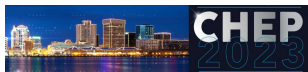
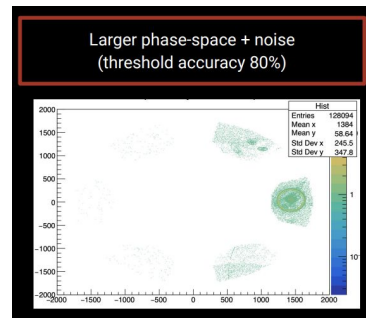
- VMs hosted by Amazon Web Services (see next slide);
- All set-up with drivers and libraries to program the included FPGA;
- Vitis-AI Docker container;
- Available during the workshop and after if requested.



[Lorusso, Marco \(University of Bologna\)](#)



[Fanelli, Cristiano \(W&M\)](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Training and Onboarding Initiatives in HEP

How do experiments teach analysis software?

Virtual

Hybrid

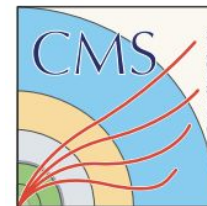
In person



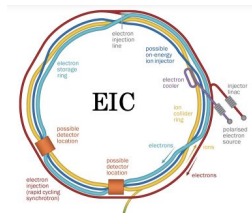
Online book



Starter Kit



Data Analysis Schools



Online tutorials



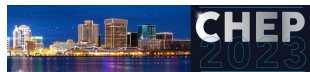
Software tutorials



Synchronous tutorials  
“Carpentries-style”

**“Software is different, but challenges are common”**

[Reinsvold Hall, Allison \(US Naval Academy\)](#)



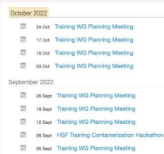
Fri 12th May, Plenary Session - Conference Highlights: Track 8



# Building a Global HEP Training Community

We can cover more ground together!

## Weekly meetings

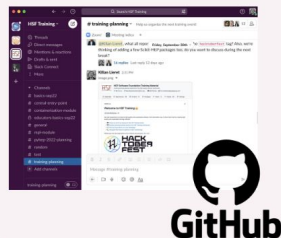


October 2022	
01 Oct	Training WS Planning Meeting
08 Oct	Training WS Planning Meeting
15 Oct	Training WS Planning Meeting
22 Oct	Training WS Planning Meeting
September 2022	
01 Sep	Training WS Planning Meeting
08 Sep	Training WS Planning Meeting
15 Sep	Training WS Planning Meeting
22 Sep	Training WS Planning Meeting
29 Sep	HSF Training Containerization Hackathon
06 Oct	Training WS Planning Meeting

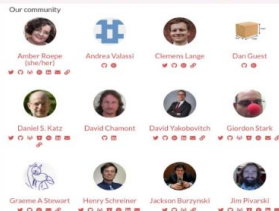
## Monthly Hackathons



## Platforms



## Community pages



## How-to guides



## Software Development and Deployment

### Version controlling with git

Track code changes, undo mistakes, collaborate. This module is a must.

Start learning now!

Contribute!

### Advanced git

Learn to work with branches and more with this interactive webpage.

Start learning now!

Contribute!

### CI/CD (gitlab)

Continuous integration and deployment with gitlab.

Start learning now!

Watch the videos!

Contribute!

### CI/CD (github)

Continuous integration and deployment with github actions.

Start learning now!

Watch the videos!

Contribute!

### Docker

Introduction to the docker container image system.

Start learning now!

Watch the videos!

Contribute!

### Singularity

Introduction to containerization with Singularity/Apptainer.

\* Status: Beta testing

Start learning now!

Watch the videos!

Contribute!

### Unit testing

Unit testing in python.

\* Status: Beta testing

Start learning now!

Contribute!

### Level up your python

Advanced bits of python (testing, debugging, logging, and more)

Start learning now!

Contribute!

[Lieret, Kilian \(Princeton University\)](#)

Fri 12th May, Plenary Session - Conference Highlights: Track 8



# Academia – Industry



## Main affiliation



## HEP secondment



## Industry secondment



Automated decision-making for fraud detection, collaboration detecting anomalies/patterns to classify HEP observations.



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386



Real-time dark photon candidate selection at LHCb, collaboration processing vehicle video and sensor data in real-time.



LUNDS  
UNIVERSITET

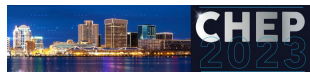


ximantis



Online calibration of ALICE time-projection chamber, collaboration with Ximantis monitoring/forecasting traffic information in real-time.

[Gooding, Jamie \(Dortmund\) - The SMARTHEP European Training Network](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Academia – Industry



## Main affiliation



## HEP secondment



## Industry secondment



Automated decision-making for fraud detection, collaboration detecting anomalies/patterns to classify HEP observations.



Real-time dark photon candidate selection at LHCb, collaboration processing vehicle video and sensor data in real-time.



Online calibration of ALICE time-projection chamber, collaboration with Ximantis monitoring/forecasting traffic information in real-time.

[Gooding, Jamie \(Dortmund\) - The SMARTHEP European Training Network](#)



## Key Goals

Digital Competencies

Networking

Transfer & Communication

ErUM-internal

*How do next-gen analyses environments look like?*

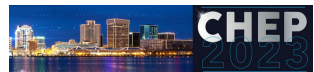
*How do experiments handle (big) data?*

ErUM ↔ Industry

*How to do sustainable computing?*

*How is AI used in Industry?*

[Fackeldey, Peter \(RWTH Aachen\) - ErUM-Data-Hub](#)

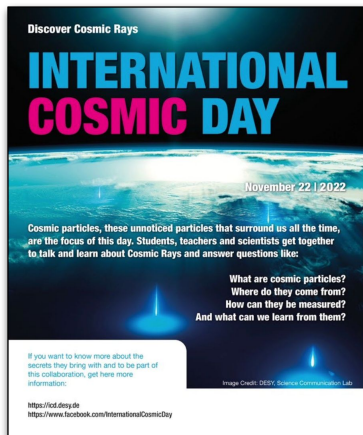


# Outreach

Reaching society, decision makers, and future scientists.

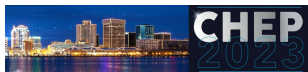


## CERN's Beamline 4 Schools



**“It’s not AI/ML, there are no Qubits, or cutting edge technology  
Except that too many people --- it is!”**

[Yacoob, Sahal \(University of Cape Town / CERN\) - IPPOG](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8

# Outreach

Reaching society, decision makers, and future scientists.

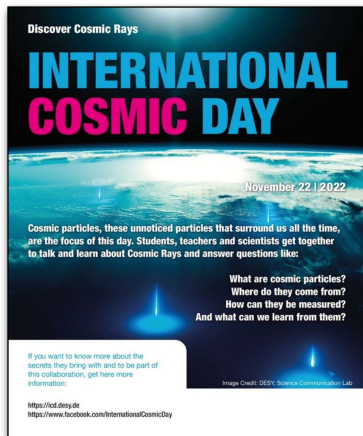


## CERN's Beamline 4 Schools

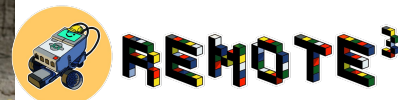


**“It’s not AI/ML, there are no Qubits, or cutting edge technology  
Except that too many people --- it is!”**

[Yacoob. Sahal \(University of Cape Town / CERN\) - IPPOG](#)

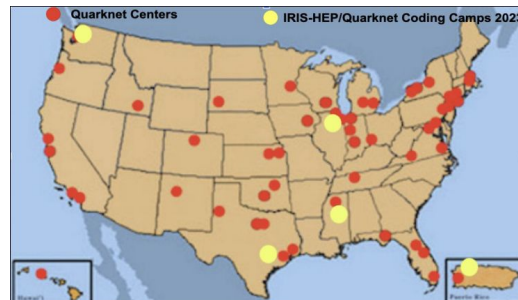


**“Get the best requires to reach everyone”**

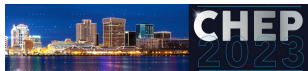


**Remote sensing by Remote schools  
in Remote environments**

[Collier, Ian \(UKRI-STFC\) - strategy and evaluation in public engagement](#)



[Cordero, Danelix \(CROEM\) - Software Training Outreach In HEP](#)



Fri 12th May, Plenary Session - Conference Highlights: Track 8



# Virtual Visits



[DeMuth, David \(Valley City State University\) - vDUNE](#)

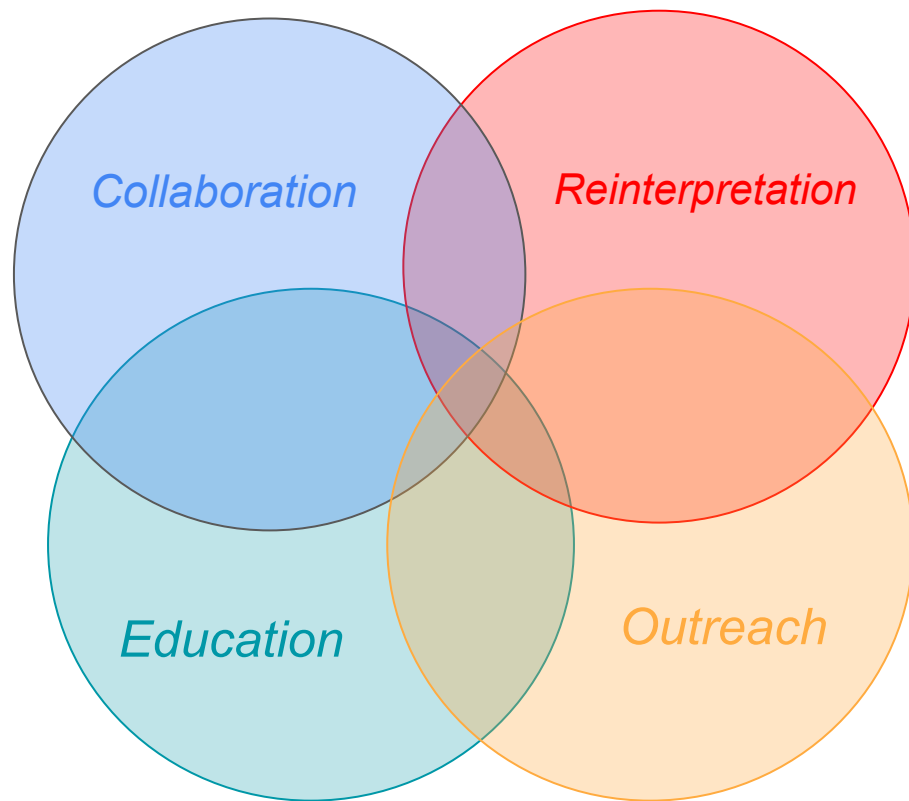


[Le Boulicaut, Elise \(Duke University\) - ATLAS Virtual Visits](#)

**“In six more years, I could be doing that!” – A student.**

# Investing in the current and next generation of researchers

- Enable analysis preservation, sustainable workflows.
- Help to accelerate physics research!
- Investing in the next gen of HEP is important!



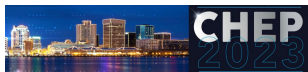
# Thank you!

*Michel*

*Clemens*

*Elena*

*Markus*



Fri 12th May, Plenary Session - Conference Highlights: Track 8