

# AN INSTITUTE FOR SCIENTIFIC SOFTWARE

## CONNECTING APPLIED COMPUTING AND DATA INTENSIVE SCIENCES

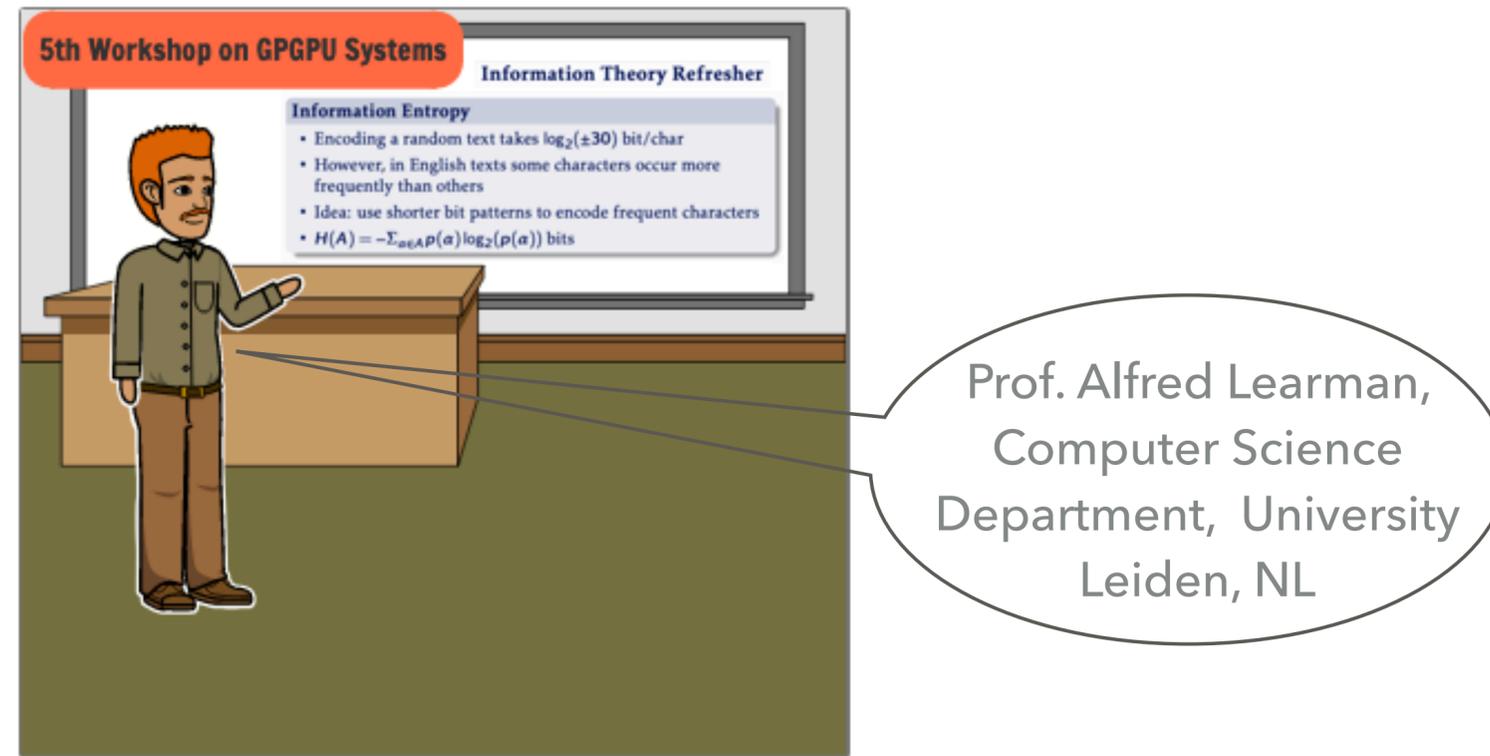
I BIRD, S CAMPANA, P MATO, S ROISER, M SCHULZ, G STEWART, A VALASSI



24th International Conference on Computing in High Energy & Nuclear Physics

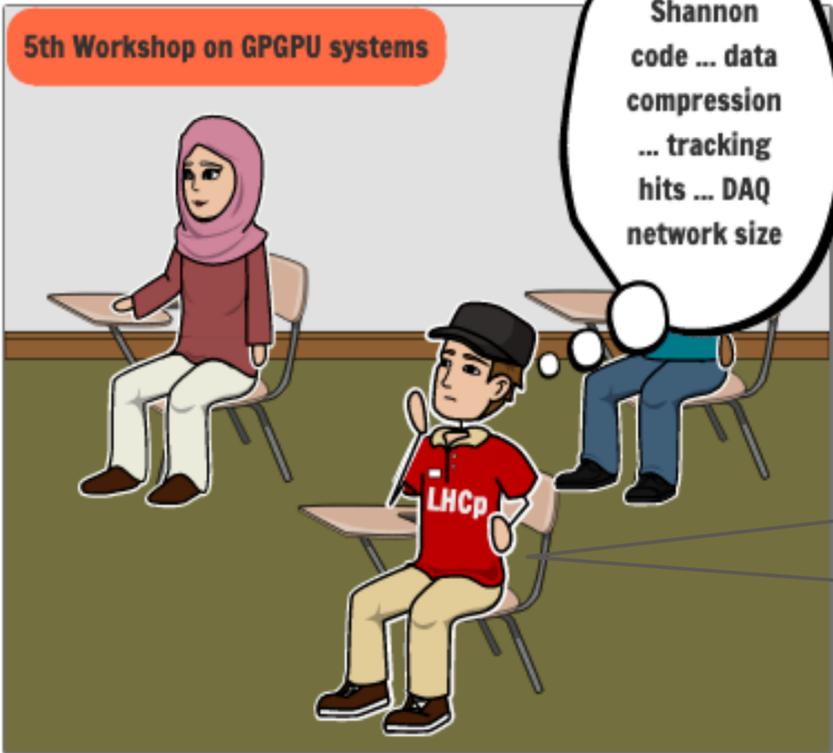
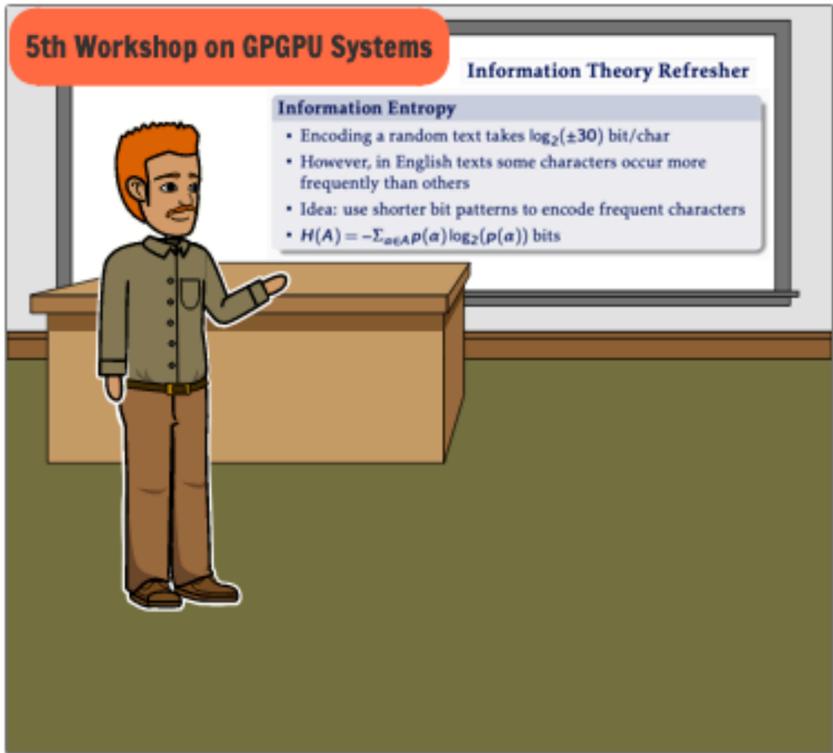
4-8 November 2019, Adelaide, Australia

# AN ALMOST TRUE STORY ON INTERACTION BETWEEN COMPUTER SCIENCE AND HEP



Names of people and institutions have been anonymised ;-)

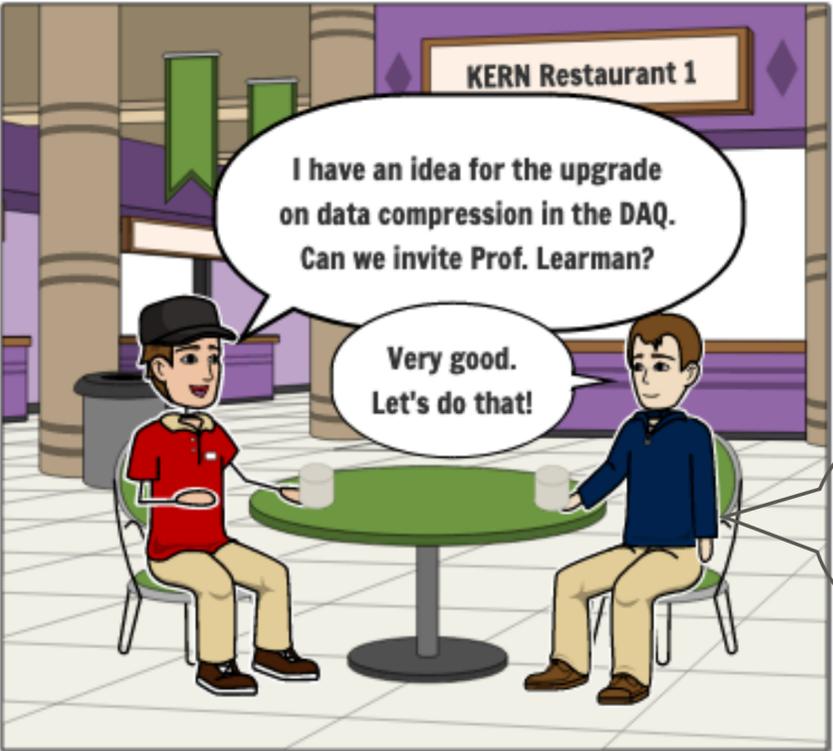
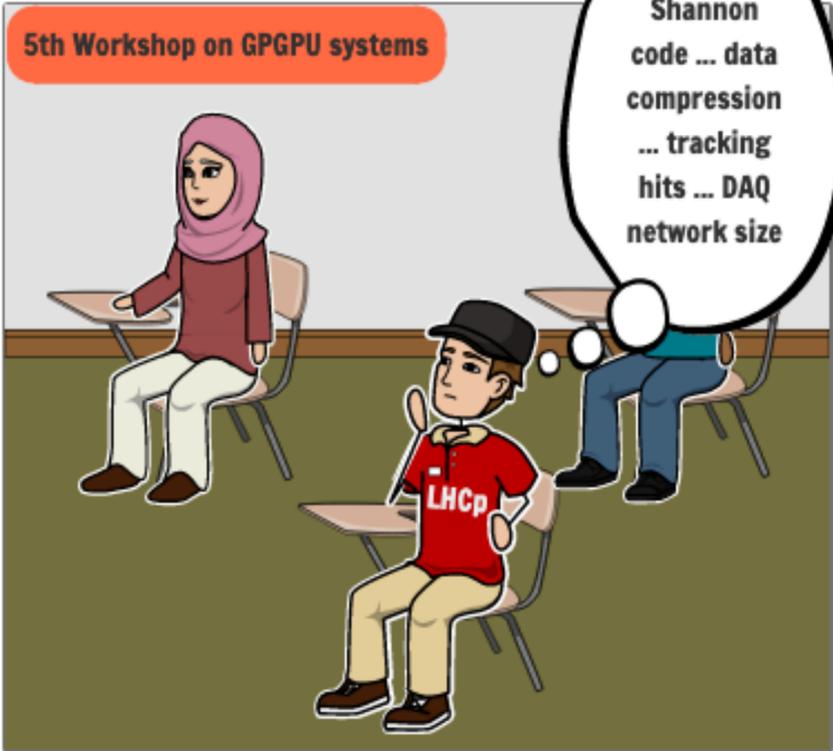
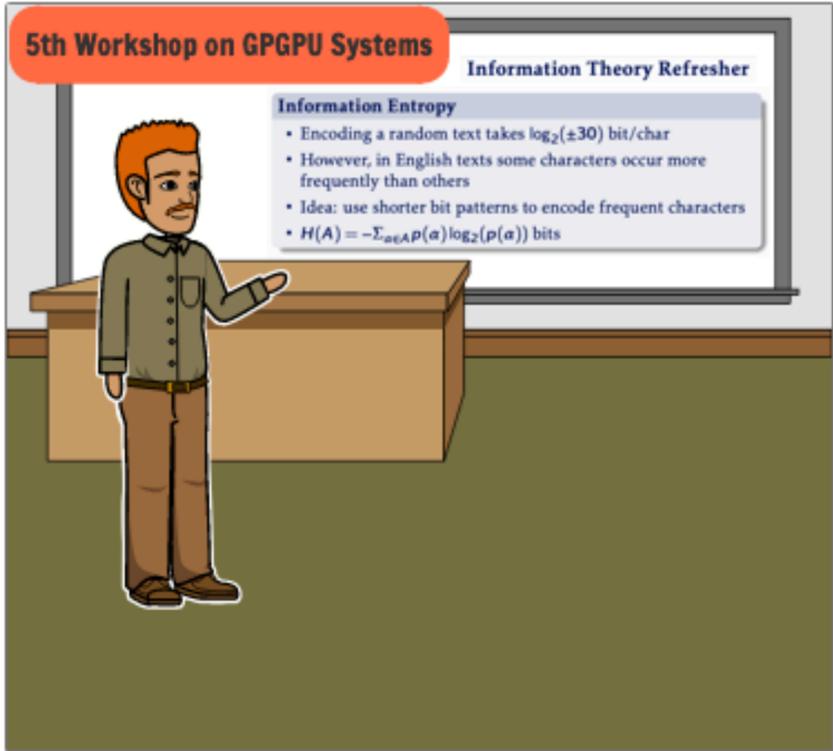
# AN ALMOST TRUE STORY ON INTERACTION BETWEEN COMPUTER SCIENCE AND HEP



Raul, LHCp physicist and trigger expert

Names of people and institutions have been anonymised

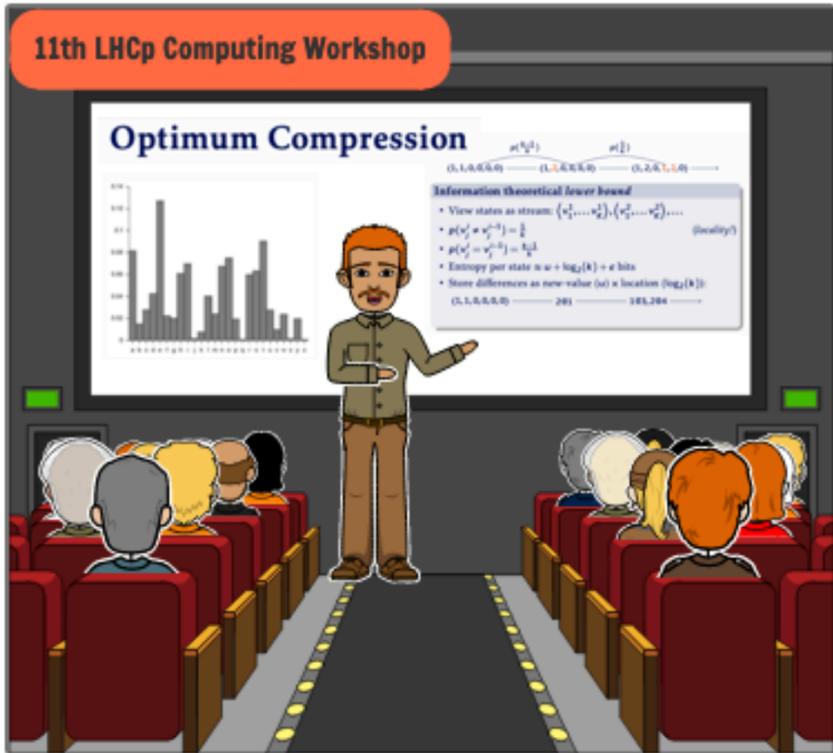
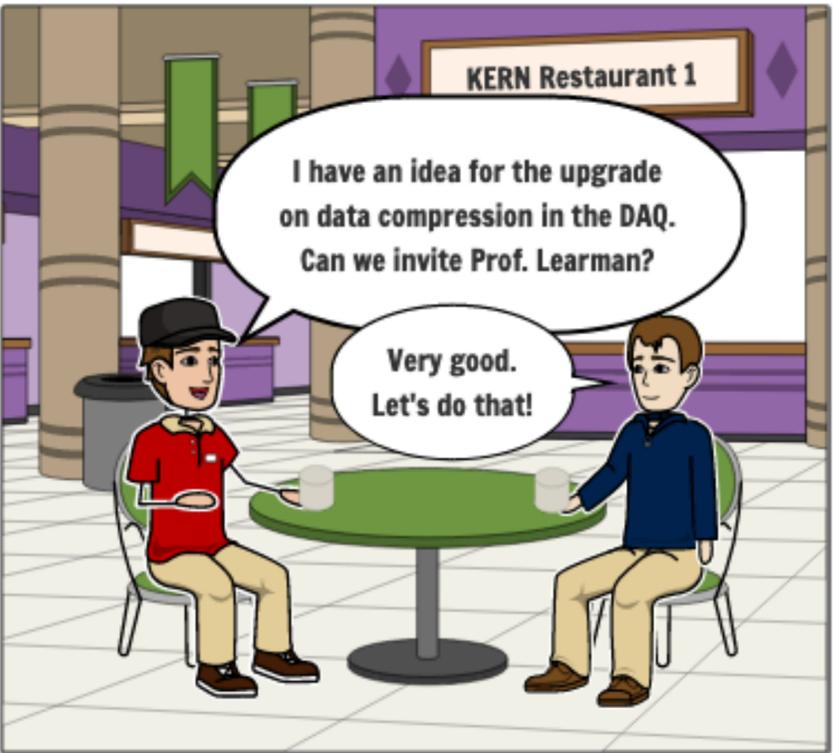
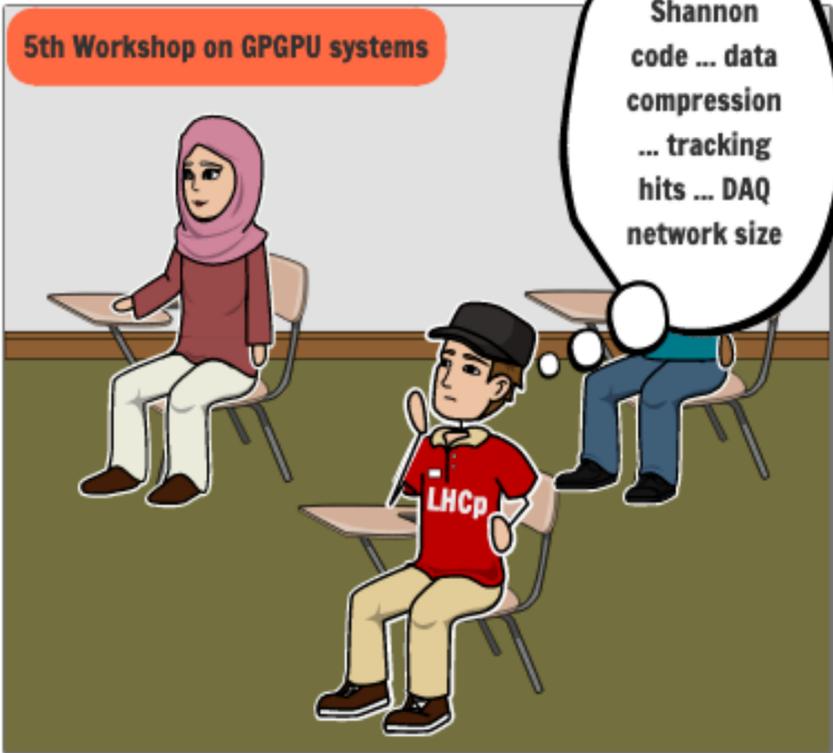
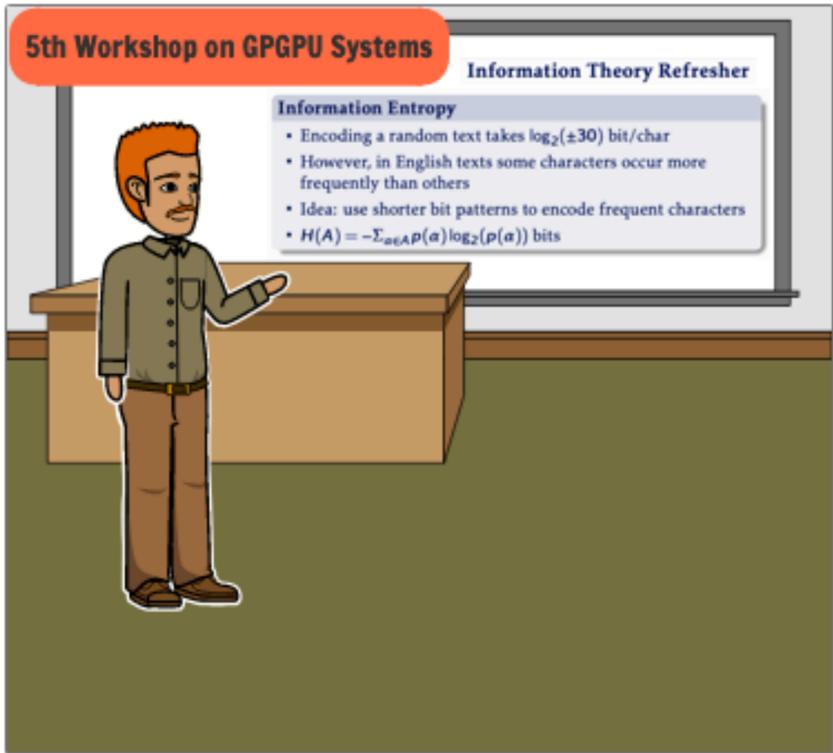
# AN ALMOST TRUE STORY ON INTERACTION BETWEEN COMPUTER SCIENCE AND HEP



Etienne, LHCp computing project leader

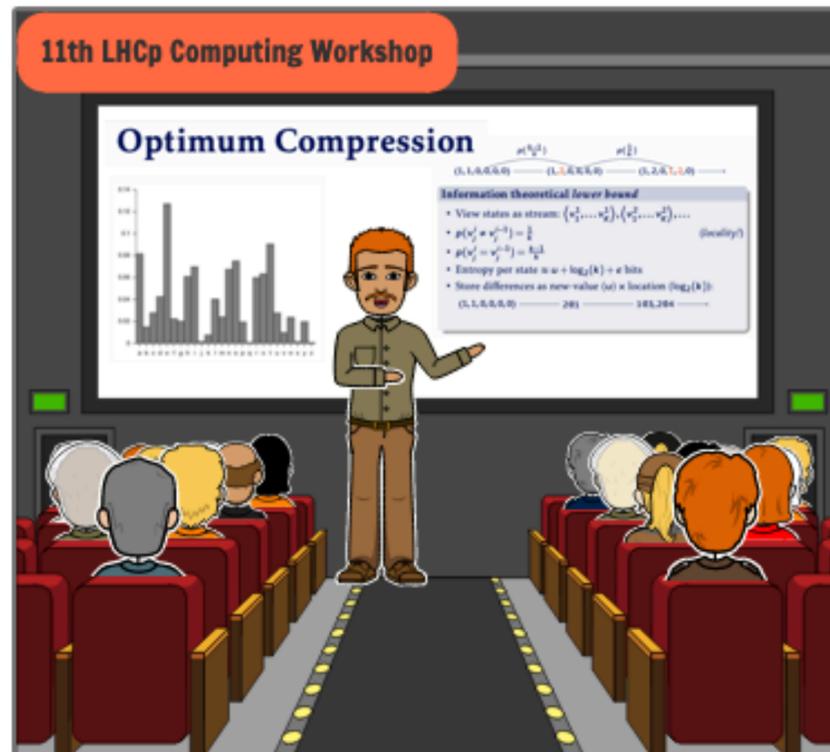
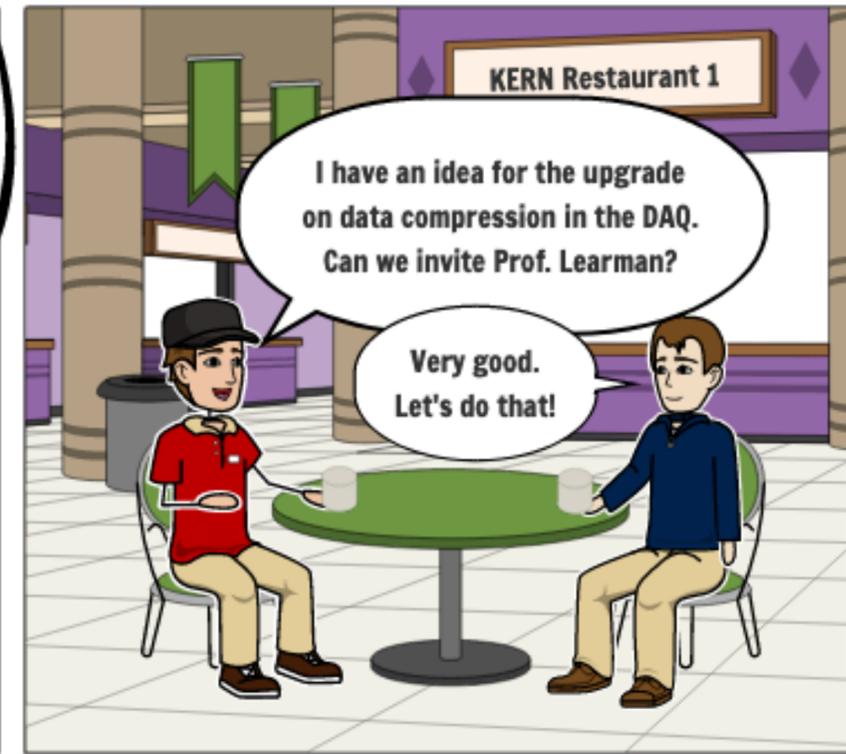
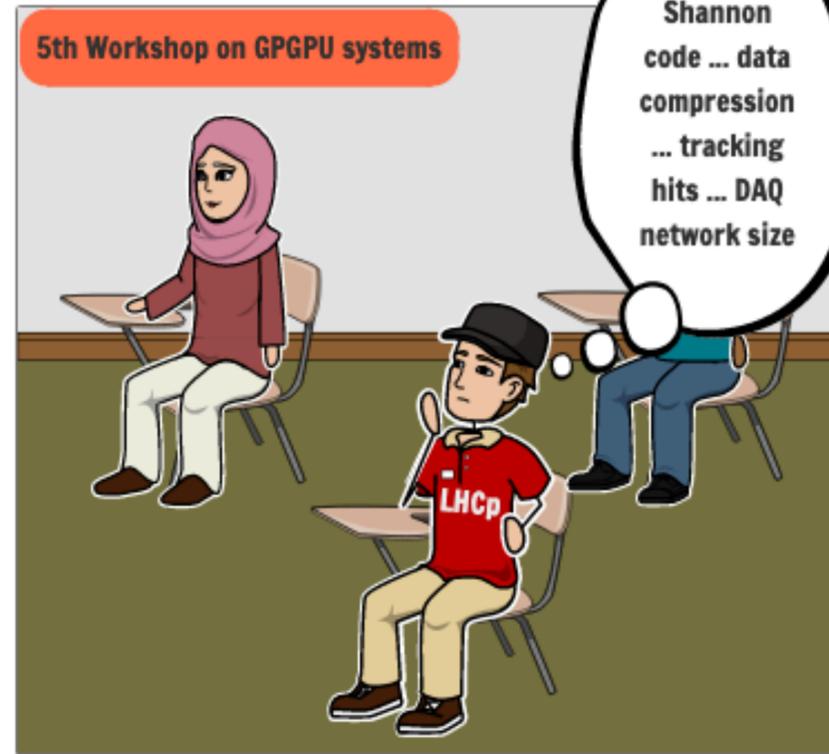
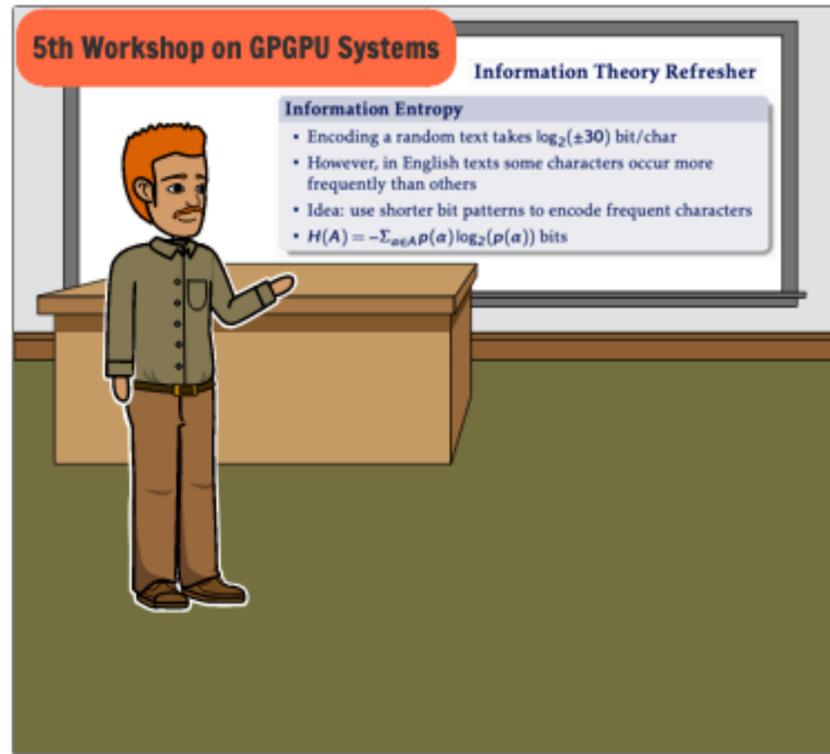
Names of people and institutions have been anonymised

# AN ALMOST TRUE STORY ON INTERACTION BETWEEN COMPUTER SCIENCE AND HEP



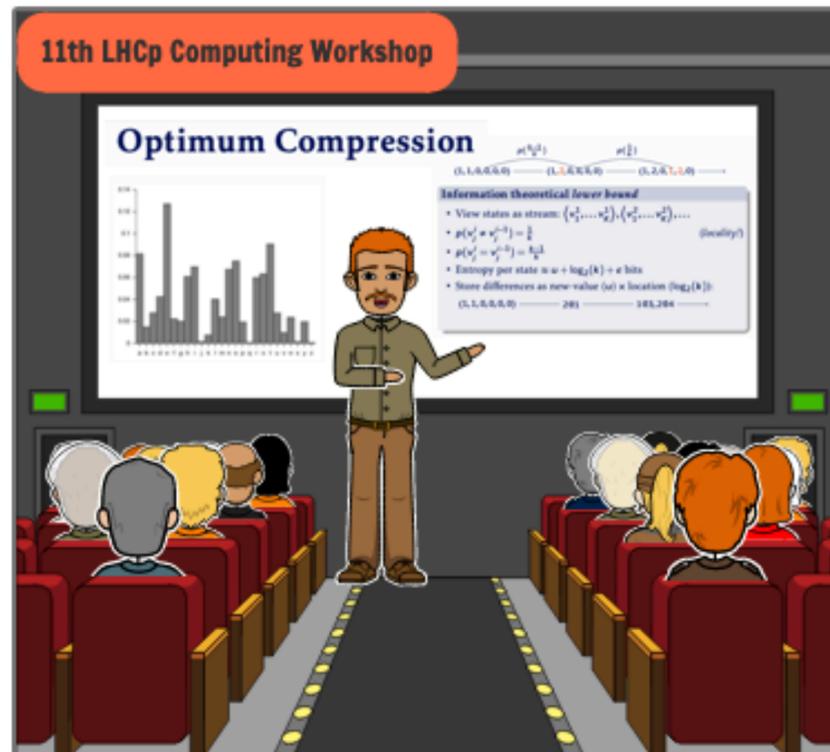
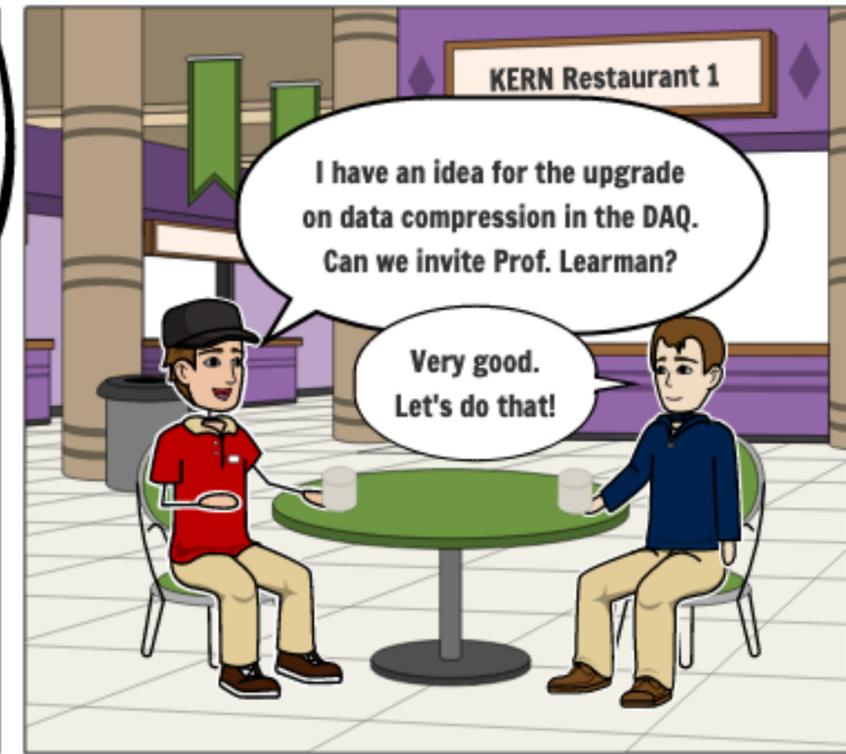
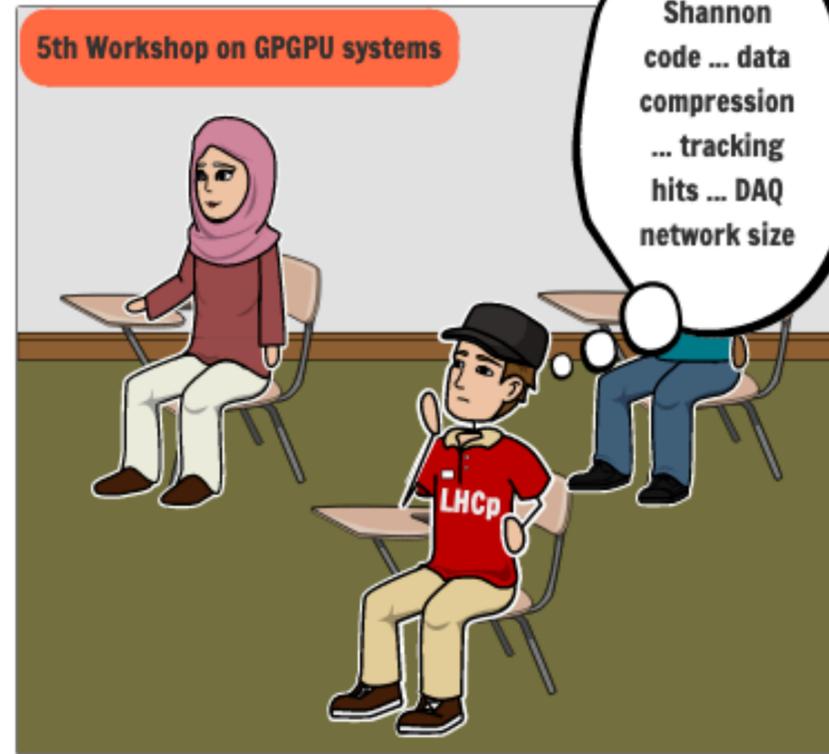
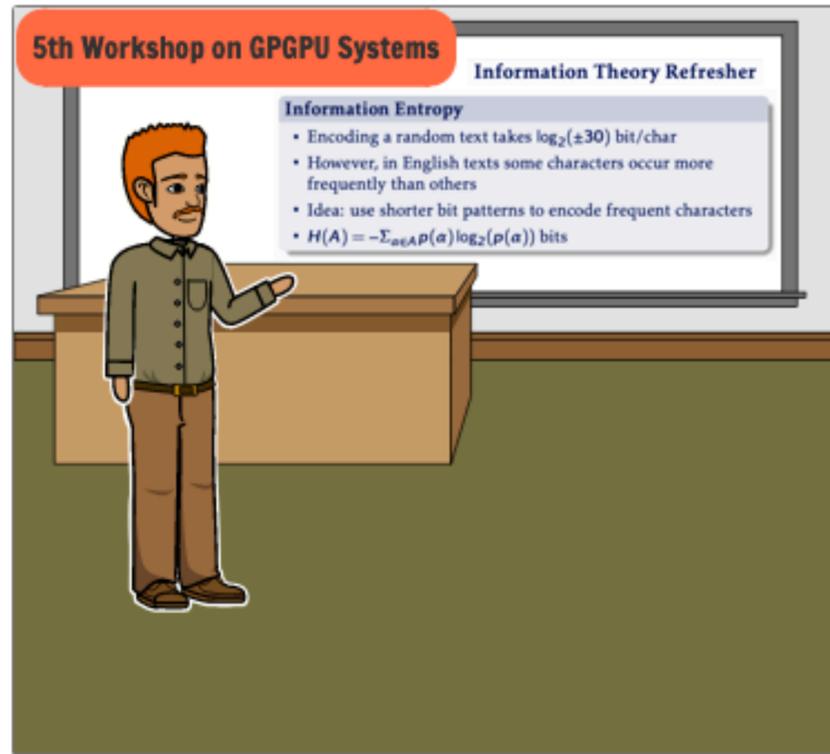
Names of people and institutions have been anonymised

# AN ALMOST TRUE STORY ON INTERACTION BETWEEN COMPUTER SCIENCE AND HEP



Names of people and institutions have been anonymised

# AN ALMOST TRUE STORY ON INTERACTION BETWEEN COMPUTER SCIENCE AND HEP



Names of people and institutions have been anonymised

# WHAT HAPPENED AND WHY ...

- ▶ Very good and welcome interaction between computer science and natural science
  - ▶ ... and these initiatives should continue by all means !

# WHAT HAPPENED AND WHY ...

- ▶ Very good and welcome interaction between computer science and natural science
- ▶ ... and these initiatives should continue by all means !

BUT

# WHAT HAPPENED AND WHY ...

- ▶ Very good and welcome interaction between computer science and natural science
  - ▶ ... and these initiatives should continue by all means !

BUT

- ▶ This also happened by chance, because ...
  - ▶ Raul's expertise in trigger software allowed him to extract information from a GPGPU programming talk for the field of DAQ network data compression
  - ▶ Raul was at the workshop and took the initiative to follow up with relevant people

# WHAT HAPPENED AND WHY ...

- ▶ Very good and welcome interaction between computer science and natural science
- ▶ ... and these initiatives should continue by all means !

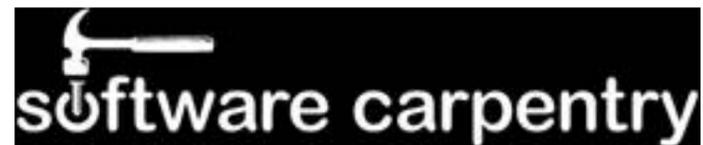
BUT

- ▶ This also happened by chance, because ...
  - ▶ Raul's expertise in trigger software allowed him to extract information from a GPGPU programming talk for the field of DAQ network data compression
  - ▶ Raul was at the workshop and took the initiative to follow up with relevant people

*Can we establish such interactions in a more structured way?*

# A NON EXHAUSTIVE LIST OF ALREADY EXISTING INITIATIVES TO SUPPORT SOFTWARE ENGINEERING

- ▶ Initiatives targeting mostly the major increases in computing needs for HL-LHC and future HEP experiments



Experiment hackathons, trainings, starterkits, ...

# A NEW IDEA: AN INSTITUTE CONNECTING ACADEMIA AND NATURAL SCIENCES

- ▶ Enable a strong interaction between **data intensive sciences** and **computer science departments / software engineering schools**
- ▶ Aim to **gather, curate and disseminate software engineering and computer science knowledge** in the long term

# WHAT ELSE WE WANT TO ACHIEVE

- ▶ Enable R&D resulting from collaborations of natural science researchers and computer scientists
- ▶ Establish a career path for scientists and engineers working in computing for natural sciences
- ▶ Recognition of software work and attracting / retaining software developers is a major concern e.g. in HEP
- ▶ Cross-fertilise between different science domains. Make knowledge available across domain boundaries
- ▶ Act as a lobbying organisation towards funding agencies
  - ▶ Raise awareness for the importance of software and computing in natural science

# MAIN CHARACTERISTICS OF THE INSTITUTE

## ▶ **Virtual**

- ▶ Loose connection of interested partners in natural sciences and academic institutions

## ▶ **Fundamental**

- ▶ Aiming at work which is not specific to one natural science domain

## ▶ **Multi domain**

- ▶ Domain agnostic work allows multiple sciences to profit

## ▶ **Multi national**

- ▶ Connect natural science activities and academic institutions in many countries

# WHERE WE ARE

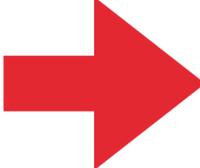
- ▶ At the very start of the journey

# WHERE WE ARE

- ▶ At the very very start of the journey

# WHERE WE ARE

- ▶ At the very very start of the journey
- ▶ Engaging with interested parties
  - ▶ University institutes, schools, labs, natural sciences
    - ▶ So far very positive responses received from universities and labs

 **First event, workshop with Alpaka team on accelerator agnostic code translation, <https://indico.cern.ch/event/858758/>**

## A Plan to Set Up an Institute for Software in Data Intensive Sciences

Data processing using software is a fundamental part of today's scientific research. Planning for future experiments, e.g. in high energy physics (HEP), shows that significant improvements in software is needed to fully exploit the physics potential within a realistic computing budget [1]. This is compounded by a rapidly evolving hardware landscape, which needs to be adapted to. According to a study from 2014 in the UK [2], 92% of academics use research software and 56% develop software. Despite these numbers, software engineering for science has not yet gained a high reputation in the academic world, putting the careers of scientists who engage in software engineering at risk. Another article [3] suggests that only 8% of scientists scrutinise the software they use and that "Most scientists [...] continue to emerge from natural science training without formal training in computational methods and software development and/or engineering".

To address these problems we propose to create an European Institute with a mission to promote excellence in software engineering and best practice in natural sciences based on a strong collaboration with computer science departments and software engineering schools. To ensure a sustainable impact the institute will gather, curate and disseminate software engineering and computer science knowledge in the long term.

In addition the institute also aims to:

- Encourage R&D resulting from collaborations of natural science researchers and computer scientists.
- Promote a career path for data scientists within scientific collaborations, communities and academia by raising awareness on technical, sociological and political levels.
- Cross-fertilize between different science fields, make knowledge retrievable and accessible across domain boundaries.
- Provide complex, large applications and data from natural sciences to computer scientists to serve as input for innovative engineering techniques or studies.
- Act as a lobbying forum for software engineering for natural sciences on national and international levels.

Following a Trans-European strategy, the institute will also help to link and bridge gaps amongst different European national and international initiatives which already have been funded or are being proposed in various countries such as the [Software Sustainability Institute](#) (UK), [CDCS](#) (DE), [IRIS-HEP](#) (US), [HSE](#), etc.

On the path forward, the institute will first enter a conceptualisation phase where the above ideas will be implemented in the scope of HEP and astrophysics, reaching out to a limited set of interested countries, experiments and computing and software scientists. A series of topical workshops will be held to initiate the process and gather feedback on the most useful areas for the institute to be active in. At the same time concepts for funding and governance shall be developed. The conceptualisation phase shall not last longer than 2 years from the start of the initiative. In a second phase, after having proven to be successful, the institute shall reach out to additional data intensive sciences and operate in a sustained mode. More details can also be found in [4].

[1] The HEP Soft. Foundation, A Roadmap for HEP Software and Computing R&D for the 2020s, [DOI:10.1007/s41781-018-0018-8](https://doi.org/10.1007/s41781-018-0018-8)

[2] S.J. Hettricket et al, UK Research Software Survey 2014, [DOI:10.5281/zenodo.1183562](https://doi.org/10.5281/zenodo.1183562)

[3] L.N. Joppa et al, Troubling Trends in Scientific Software Use, [DOI:10.1126/science.1231535](https://doi.org/10.1126/science.1231535)

[4] I. Bird et al, Memorandum on a Software Institute for Data Intensive Sciences, <http://cern.ch/go/f7lq>

# A ONE PAGE STATEMENT EXPLAINING MOTIVATION AND GOALS

DOI 10.5281/zenodo.3466587

# SUMMARY

- ▶ We propose a new initiative for a **strong interaction of data intensive sciences and academic research institutions in software engineering and computer science**
- ▶ ... with a focus on **multi-national basic research across natural science domains**



- ▶ We are at the very start of the initiative reaching out to interested partners
- ▶ Please come and talk to any of us if you are interested to join

A statement on the initiative is available at <https://doi.org/10.5281/zenodo.3466587>

# HOW TO GET IN TOUCH

- ▶ [stefan.roiser@cern.ch](mailto:stefan.roiser@cern.ch), +41 75 411 5334,  sroiser
- ▶ Ian Bird, Simone Campana, Pere Mato, Markus Schulz, Graeme Stewart, Andrea Valassi (<first>.<second>@cern.ch)

# LINKS

- ▶ HEP Software Foundation

- ▶ <https://hepsoftwarefoundation.org/>

- ▶ IRIS-HEP

- ▶ <https://iris-hep.org/>

- ▶ CERN Openlab

- ▶ <https://openlab.cern/>

- ▶ AIDA 2020

- ▶ <http://aida2020.web.cern.ch/>

- ▶ LHCb Starterkit

- ▶ <https://lhcb.github.io/starterkit/>

- ▶ Computing Schools

- ▶ CERN <https://csc.web.cern.ch/>

- ▶ Bertinoro <http://cs.unibo.it/projects/biss2019/index.html>

- ▶ Erum

- ▶ <https://bit.ly/2mlFD5j>

- ▶ IRIS

- ▶ <https://www.iris.ac.uk/>