EPS-HEP 2017 (5–12 July 2017)

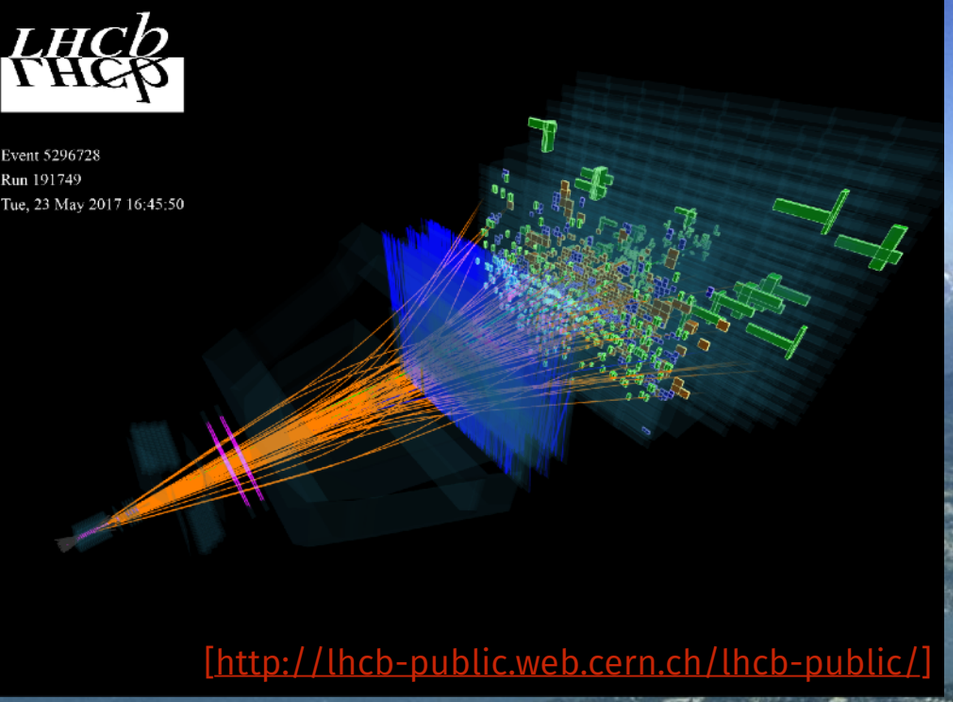
The LHCb Starterkit initiative

Albert Puig *on behalf of the LHCb collaboration*

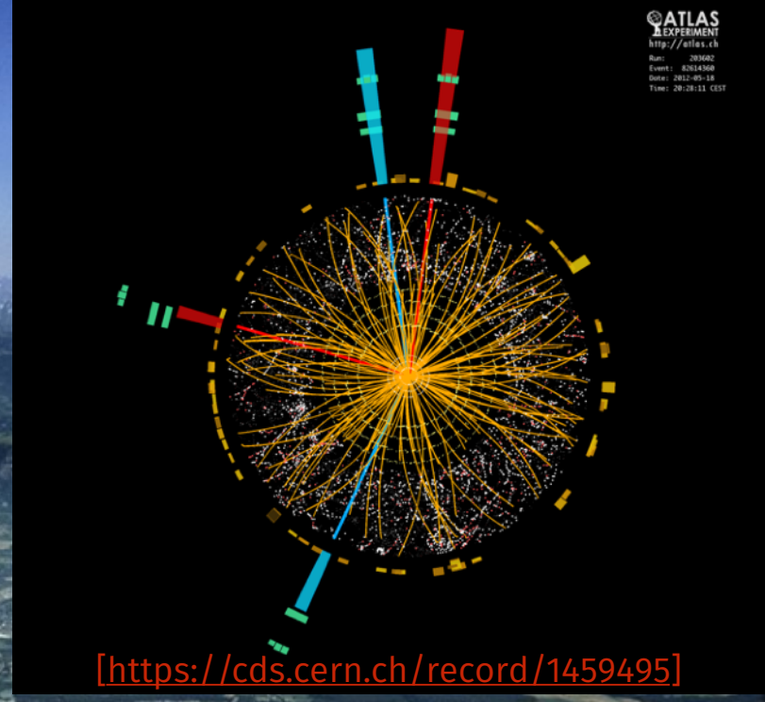
Universität Zürich[UZH]

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

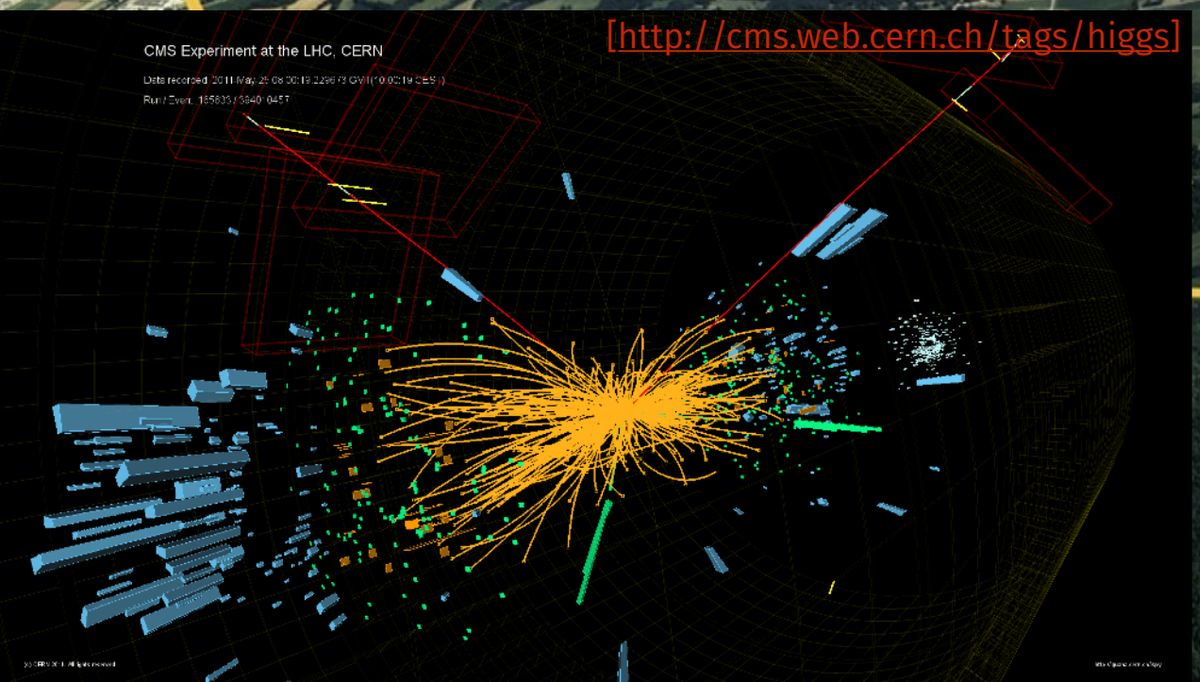
The LHC



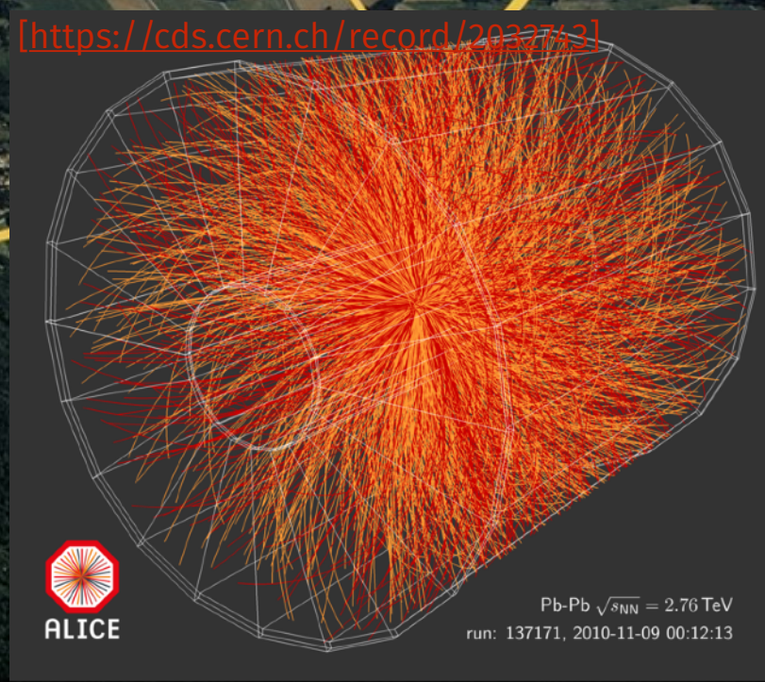
[http://lhcb-public.web.cern.ch/lhcb-public/]



[https://cds.cern.ch/record/1459495]



[http://cms.web.cern.ch/tags/higgs]



[https://cds.cern.ch/record/1233743]

From...

[LBNL Image Database 96602983]

… to

[CERN Computing]

Or rather

ROOT
Data Analysis Framework

python™

WOLFRAM
MATHEMATICA 11

git

NumPy

SciPy

matplotlib

IP[y]: IPython
Interactive Computing

pandas

scikit learn

And many others…
(in particular, LHC experiments' own software!)

We're trained as physicists

We're asked to be programmers

We need training

The problem

Many new students feel overwhelmed by all the software they have to learn, the experiment conventions, etc

Tutorials/documentation are often broken, incomplete or inexistent

Many hours are wasted

- Students trying to use/understand/modify “inherited” scripts from their supervisors
- Students also need to figure out who to contact to ask questions
- Experts answering trivial or repeated questions in mailing lists

[https://lhcb.github.io/starterkit/]

The solution

In 2015, the LHCb Starterkit initiative was started to

- Give a solid starting point to newcomers in the most used software in LHCb (both general and specific)
- Improve software literacy in the experiment
- Teach good practices
- Help newcomers socialise and integrate in the collaboration

This is done through the preparation of detailed, up-to-date tutorials and the organisation of periodic workshops of different levels

What makes it special?

The project was started by early career scientists (mostly at PhD and MSc level), and it is still run by them

Organisation, teaching, and tutorial writing is done on a voluntary basis (no institute commitment)

Democratic approach to decisions

What makes it special?

Since the project is run by non-permanent staff, good knowledge transfer is ensured with industry-standard tools

- Slack/Mattermost chat
- Use sharing tools such as Google Docs, Github issues, etc, to keep track of as many organisational details as possible

Heavy rotation of teachers, helpers and organisers

- Attendees of previous workshops are typically invited to help in the subsequent ones
- More experienced helpers volunteer to organise a workshop

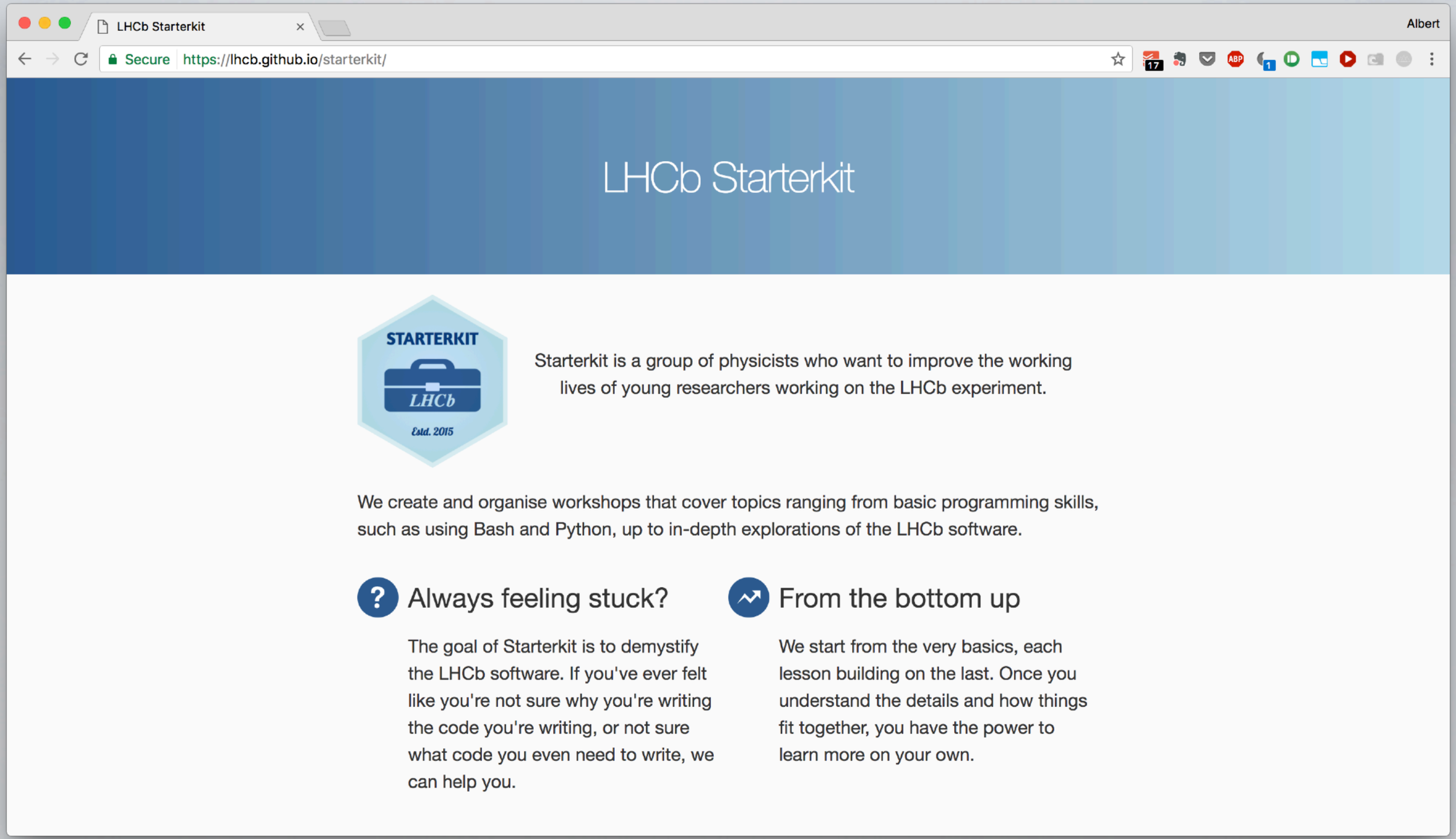
Conclusions (TLDR)

The Starterkit started as a mostly PhD- and MSc-led project to address common problems in software literacy within the LHCb collaboration

Two years after its start, the Starterkit “brand” is consolidated within LHCb

- Workshops happen regularly and are usually oversubscribed
- Tutorials widely used within the collaboration (even by senior members), have mostly replaced “old” tutorials
- Tutorials kept up to date thanks to a huge collaborative effort

[https://lhcb.github.io/starterkit/]



[https://lhcb.github.io/starterkit/]

So, what do we do?

Online tutorials

Interactive workshops

Tutorials

Inspired by the well-established Software Carpentry tutorials, use same template and techniques

- Currently moving to GitBook, since SC style was limiting us

Freely accessible, Github hosted tutorials/lessons

- Regularly, collaboratively updated (~50 contributors)
- People following the tutorials on their own encouraged to submit issues (>250 in the last two years) and, possibly, merge requests with a fix
- Use the Fork & Pull Request model to lower the entry barrier (less intimidating than editing collaboration-wide Twikis)

Tutorials

Tutorials only cover LHCb software, starting from the basics and moving on, step by step, to more complicated material

- Link to Software Carpentry tutorials as a guide for common tools such as git, bash, python, etc

Philosophy of the tutorials is not to provide recipes of how to do things within the LHCb software, but to provide the basic knowledge and building blocks that allow to build complex workflows

- Full examples of working code provided, updated when software changes

Workshops

Inspired by the Software Carpentry workshops

Held at CERN with few basic rules:

- Small fee charged to avoid no-shows, reinvested in social event
- Small groups of people (20) with 1 instructor and 4–5 helpers
- No copy-paste, all code and commands are written in real time, leaving enough room for students to try on their own, find problems, etc
- Very interactive: questions encouraged, problems are given for the students to solve, use colour coded stickers to track progress
- Use stickers to get feedback at the end of every lesson

Starterkit workshop

Entry-level introduction to LHCb software, targeted at new members of the collaboration (PhD, MSc)

- Introduction to how the software works, basic tools, etc

Held once a year, around October/November

40 participants (out of ~80 new students every year)

4 days, 8–12 instructors, ~15 helpers

- First 1.5 days devoted to general tools, using SC tutorials
- Last 2.5 days dedicated to LHCb software

Impactkit workshop

More advanced tutorials LHCb software, targeted at more experienced members

- Content changes according to the needs/interests of attendees, gauged beforehand with a survey

Held once a year, in summer

20 participants, 3 days, 4–5 instructors, ~10 helpers

- First 2 days devoted to advanced software topics
- Last day consists in a hackathon were the students attack (in groups) short computing projects that can be useful to the collaboration

Social aspect: networking

Social event is very important for newcomers, allow to start a network outside their home institute

Problems

Since all work is done on a voluntary basis by early career scientists, sometimes it's hard to keep the project going

- People have a lot of obligations/work

Sometimes hard to recruit new helpers for workshops, since economic resources are limited

- Try to mitigate the effect by scheduling the workshops around LHCb-wide events
- Cannot make larger workshops

Even with great help from the LHCb secretariat, its's hard to find suitable rooms in CERN, even months in advance

[https://lhcb.github.io/starterkit/]

Conclusions

A lot of work has been put into consolidating the project

Improvements are being fed back into the lessons and content is continuously updated, but structure and methodology are quite stable now

If it works, don't touch it (too much)

Haven't found a solution yet for outstanding issues (organisation, attendance...), but these are not showstoppers

The Starterkit is a very successful project

An idea

Even if we've heard of some interest, we're not aware of any similar initiatives in other LHC experiments

It would be interesting to collaborate in doing something similar in other LHC experiments to try to grow the organisation in a transversal way

Collaboration between experiments would be beneficial

- Share some teaching/tutorial writing burdens (common software, such as git, bash, etc), widen topic range.
- Expand the professional network of the members of the team outside their collaboration

An idea

Even if we've heard of some interest, we're not aware of any similar initiatives in other LHC experiments

It would be interesting to collaborate in doing something similar in o e organisatio

Anybody interested?

Collaboration between experiments would be beneficial

- Share some teaching/tutorial writing burdens (common software, such as git, bash, etc), widen topic range.
- Expand the professional network of the members of the team outside their collaboration

Thank you