PROCEEDINGS OF SCIENCE

# The LHCb Starterkit

**Lennaert Bel**
*Nikhef*
*Science Park 105, The Netherlands*
*E-mail:* l.bel@cern.ch

The LHCb Starterkit is an initiative by members of the LHCb collaboration, providing up-to-date tutorials and interactive workshops for new collaboration members to get acquainted with the software used. Centrally organizing this has improved the efficiency of new collaboration members, and made it easier for people to quickly start developing code.

*38th International Conference on High Energy Physics*
*3-10 August 2016*
*Chicago, USA*





PoS(ICHEP2016)334

## 1.Introduction

When the field of high-energy physics (HEP) started out in the previous century, the process of analyzing data looked vastly different than it does today. Events were counted by hand, and discoveries were made using oscilloscopes and photographs. Physical tools, such as rulers, were used to determine properties from photographed tracks. Up to 50 years ago, universities employed teams of people to study such photographs.

There is nothing inherently wrong with these methods – indeed, lots of active research still proceeds very similarly – but over the course of the 20$^{th}$ century, the analysis techniques evolved immensely. Manual methods were replaced, on a large scale, by digital readout using computers. Consequently, the physicists working on the experiments had to become computer experts. The size of the datasets collected are huge, and large amounts of data requires large amounts of software to be processed efficiently. Computer programs, in turn, are ever-increasing in complexity. From rulers to such complex software is a big step, and for people to take that step confidently, training is indispensable.

It is much too common, however, that training in high-energy physics experiments is lacking. Many will recognize the pain of working through broken tutorials and the effort of using outdated, or even completely missing, documentation. Hours are wasted by people running into the same types of problems, and experts answering the same questions, over and over again. This can be mitigated by adopting a centralized setup to provide working tutorials and answers to frequently asked questions. The LHCb Starterkit aims to accommodate such a setup.

## 2.The LHCb Starterkit

The LHCb Starterkit is an initiative with the aim to provide software training, in the broadest sense, by and for collaboration members. Software includes general tools, such as Python, as well as collaboration-specific packages. Training entails ensuring the availability of up-to-date tutorials on those subjects, as well as organizing workshops where participants obtain hands-on experience with these tutorials, under the supervision of more experienced collaborators.

### 2.1Goals

The specific goals of the Starterkit are the following:

- **Provide working tutorials.** The tutorials [1] are freely accessible to everyone. They are open-source [2], and anyone can submit requests to update them. Many experts and users have contributed. This decentralized way of keeping the tutorials up-to-date ensures that they are always in optimal shape.
- **Organize interactive workshops.** Each year, an interactive workshop is organized at CERN, also dubbed "Starterkit". Participants follow the tutorials, while teachers and helpers, often comprising past workshop participants, are present to help them out and answer additional questions.
- **Improve software literacy.** New collaboration members come from many different backgrounds and have varying levels of experience, and both the tutorials and workshops help to get everyone on the same level. This, in turn, makes communication and problem-solving easier for everyone.
- **Teach good practices.** There are many ways to write software, and for each there are numerous reasons to pick and eschew them. By getting people on the same track in this regard, collaborating becomes more straightforward.
- **Socialization.** The workshops allow collaboration members to get to know each other, which is especially valuable for new members. This goal is further effectuated through social events during those workshops.

### 2.2Organization

While one of the main goals of the Starterkit initiative is to provide a central repository for the creation and maintenance of teaching and self-help material, the organization itself is entirely decentralized. Most of the organizers, as well as the target audience, are PhD students. Being on the same level as the audience helps writing suitable material, but it also implies that after some time, organizers leave. To ensure the survival of the setup, each workshop is organized by a different group of people. There is no strict hierarchy within the group of people working on the Starterkit – decisions are made democratically. Because there is no single person in charge, which could be damaging to the project were they to leave the collaboration, the work can always continue with people present at any particular time. Knowledge and experience is continuously transferred, and the project endures.

## 3.Content

### 3.1Tutorials

The tutorials are freely accessible, both within and outside of the LHCb collaboration. They're hosted as GitHub webpages, and the source code from which they are built is open source. The following topics are covered:

- **Bash.** Using a shell allows one to quickly navigate folders and perform (scriptable) operations on files. Some basic bash scripting can also help to invoke a program multiple times, with different options.
- **Git.** Modern version control is paramount to the success of large software packages. LHCb recently adopted Git, and teaching people the ins and outs of this version control system helps them develop and share code more efficiently.
- **Python.** This language is widely used in LHCb, be it mostly to invoke C++ to do the underlying work. Python 3 is used as a version of choice, with discussion on backwards compatibility with Python 2.
- **LHCb Software stack.** The basics of the software stack: from raw data to offline data, ready for processing, and everything in between. This is the largest body of material, and comprises many sub-packages of LHCb software, including running code and managing storage on globally distributed computing systems.

The tutorials are inspired by the well-established Software Carpentry suite of tutorials [3], both in structure and content.

### 3.2The Starterkit workshop

Once per year, a Starterkit workshop is held at CERN. The target audience of the workshop are PhD and Master's students who are new who are new to the collaboration, with a capacity of about 40 students. The LHCb collaboration has about 80 new students each year, so the workshop includes about half the students – enough to make a noticeable impact on the software literacy of the collaboration as a whole. The workshop takes four days, divided in two sections of two days each: the first section covers the general tools (bash, git and python, as written above), while the second section is about LHCb-specific software. The workshops follow a hands-on approach: the participants follow a main instructor in completing each of the tutorials, and can ask for help at any time from the 3-4 experts in the room. The atmosphere is generally very informal: the aim is to create an atmosphere where it is encouraged to ask questions. Splitting the participants into two smaller groups of twenty people each also helps accommodate such a setting.

During each of the workshops, a social event is hosted, where participants get to socialize with each other and the instructors and experts, while enjoying food and beverages. These social events help establish contact amongst peers, and also serve to make new members feel at home in the collaboration.

### 3.3The Impactkit workshop

In 2016, by the way of experiment, the first Impactkit workshop was hosted. This was a three-day follow-up workshop, focused on more advanced LHCb-specific software use cases. Twenty participants were given two days of advanced lectures on LHCb topics, much in the same manner as during the Starterkit. On the last day, they sat down in groups of 2-4 people in a mini-hackathon, where they were tasked with solving several challenges involving the LHCb software over the course of one day. The issues ranged from solving longstanding bugs in the software to writing missing documentation for important packages and creating new scripts and interfaces – all with the aim to be actually used in production. The hackathon proved successful, with most of the issues being solved and many solutions now being used in practice. This has lead to the Impactkit joining the Starterkit in being organized annually.

## 4.Reception

The Starterkit initiative, as a whole, has received overwhelmingly positive feedback from across the collaboration. The tutorials are being actively followed (and updated) by many, and the workshops are always oversubscribed. Workshop participants have hailed the clarity and usefulness of the tutorials and workshops, and many have since joined in organizing successive workshops. All in all, the need for a centrally organized, functioning educational system within HEP collaborations could not have been more evident.

## 5.Conclusion

In conclusion, the Starterkit initiative has been a success, as judged by over 100 students who have participated in the workshops. It has enabled new collaboration members to kickstart their career within high-energy physics, and prevented people from feeling lost in the myriad of software that are being used every day. In short, the efficiency with regard to software has increased collaboration-wide – an enhancement that is beneficial to everyone.

## References

[1] lhcb.github.io/starterkit

[2] lhcb.github.com/starterkit

[3] software-carpentry.org