



# **CERN openlab Summer Student programme 2017**

## **FINAL REPORT**

**SEPTEMBER 2017**

**AUTHORS:**

Kristina Gunne  
CERN openlab Administration Officer

Marion Devouassoux  
CERN openlab Junior Administration Officer



## TABLE OF CONTENTS

---

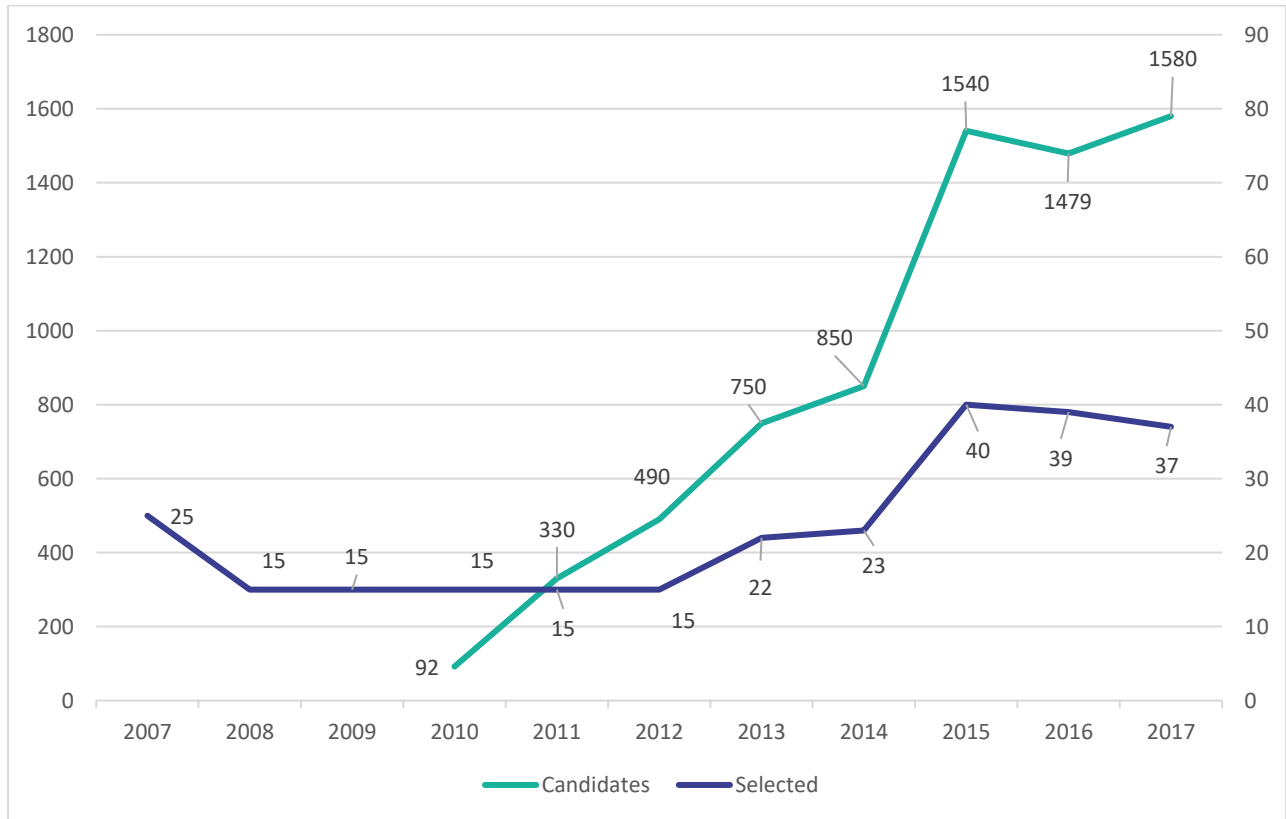
<b>INTRODUCTION</b>	<b>05</b>
<hr/>	
<b>SELECTED STUDENTS VS APPLICANTS</b>	<b>06</b>
<hr/>	
<b>BUDGET</b>	<b>07</b>
<hr/>	
<b>STUDENTS SURVEY RESULTS</b>	<b>10</b>
<b>BEFORE AND DURING THE STAY</b>	
<b>SUPERVISOR AND PROJECT</b>	
<b>LECTURES PROGRAMME</b>	
<b>FIELD TRIP</b>	
<b>VISITS</b>	
<b>HOSTING</b>	
<b>PAYMENT</b>	
<b>THE CERN OPENLAB SUMMER STUDENTS PROGRAMME IN GENERAL</b>	
<b>OVERALL ASSESSMENT</b>	
<hr/>	
<b>ACCEPTED PROJECTS AND STUDENTS</b>	<b>20</b>
<hr/>	



# 1. INTRODUCTION

Applicants in 2017: 1580 / Selected students 2017: 37

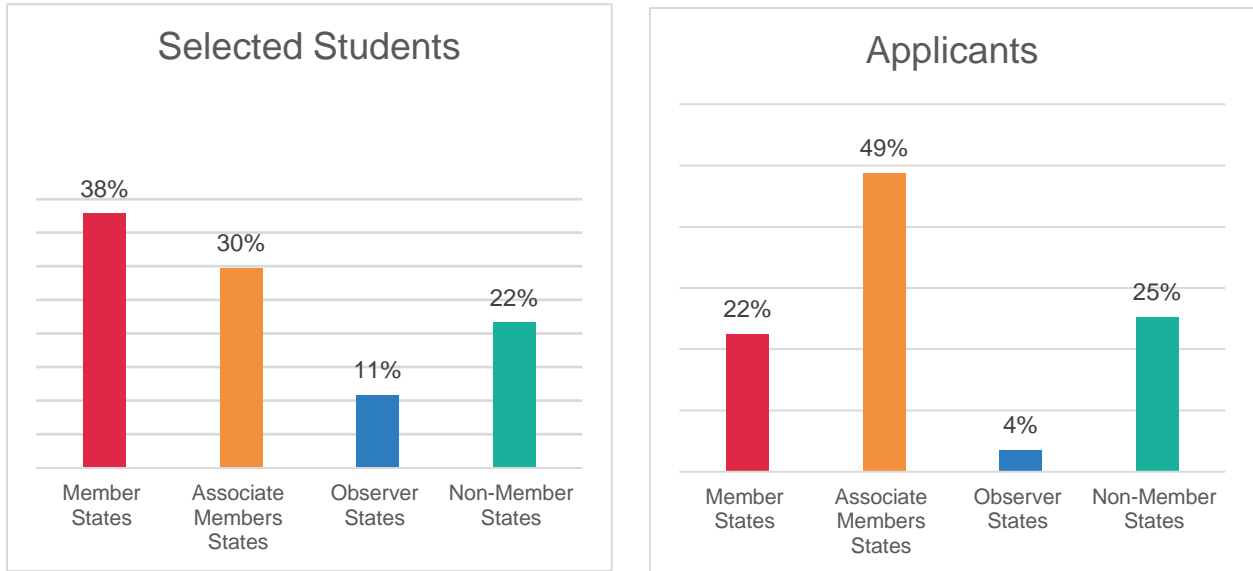
All CERN openlab summer student reports can be found here:  
<http://openlab.web.cern.ch/publications>





## 2. SELECTED STUDENTS VS APPLICANTS

### a. PERCENTAGE PER CERN ASSOCIATION STATUS



The number of applicants from the CERN Associate Members States increased by 30 percentage points between 2016 and 2017. This is mainly because India became an Associate Member States of CERN in November 2016.

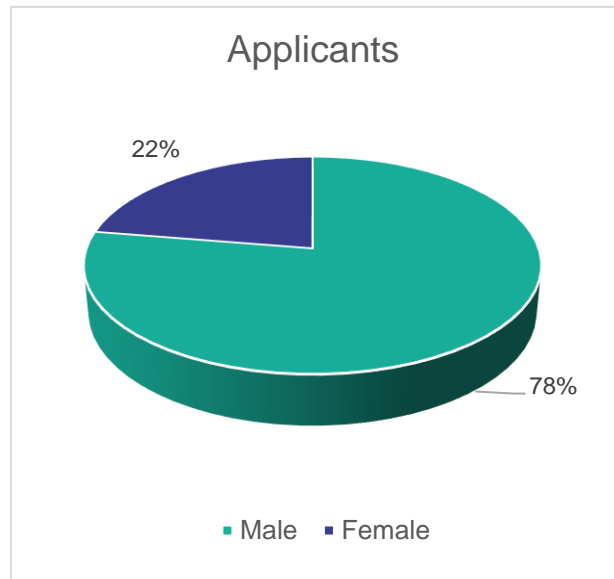
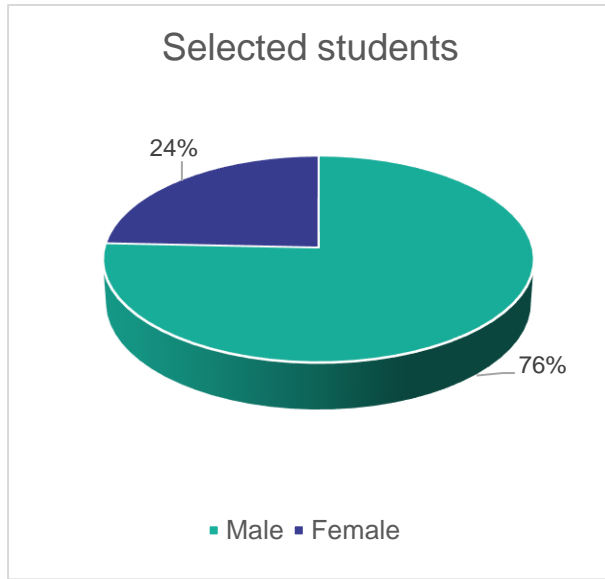
### b. NATIONALITIES OF SELECTED STUDENTS

Member States	Associate Member States	Observer States	Non-Member States
Finland	Pakistan (2)	Russian Federation	Algeria
Germany	India (7)	United States (3)	Bosnia and Herzegovina
Greece (3)	Serbia		Brazil (2)
Italy	Turkey		Guatemala
Norway			Indonesia
Poland (2)			Islamic Republic of Iran
Slovakia			Palestine
Spain (2)			
United Kingdom (2)			
<b>SUBTOTAL</b>			
14	11	4	8
<b>38%</b>	<b>30%</b>	<b>11%</b>	<b>22%</b>





### c. PERCENTAGE PER GENDER

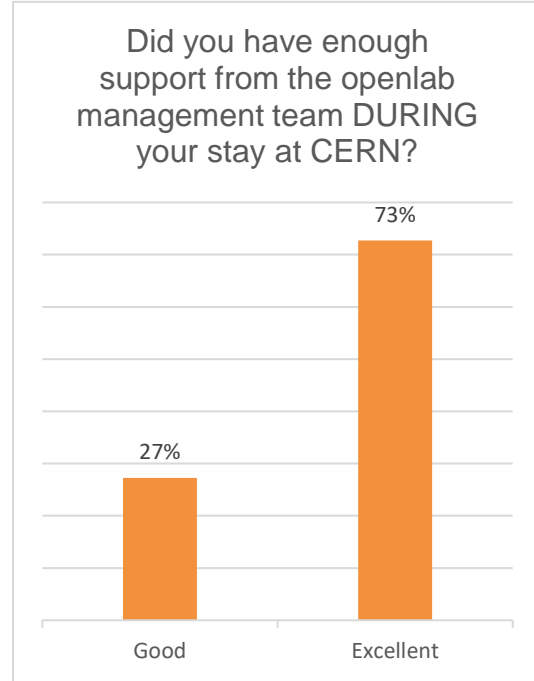
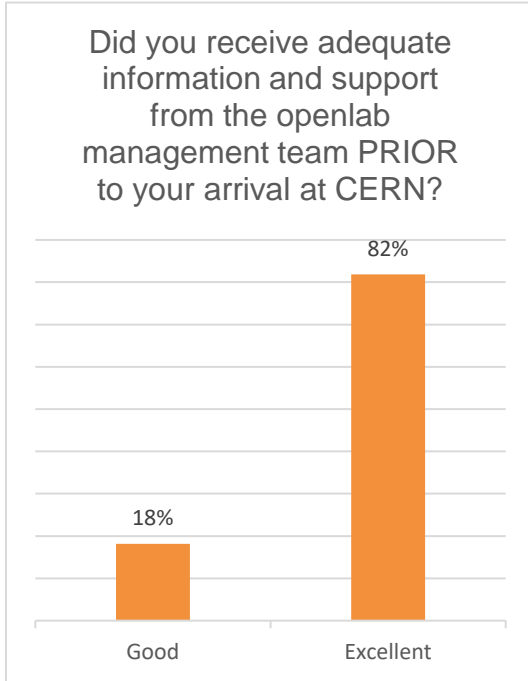




### 3. STUDENTS SURVEY RESULTS

23 out of 37 students responded to the survey

#### a. BEFORE AND DURING THE STAY



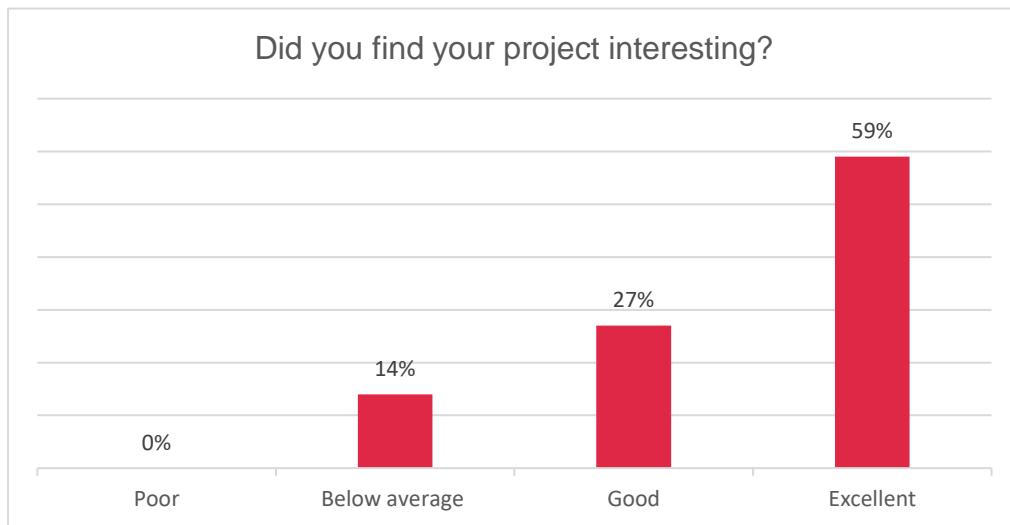
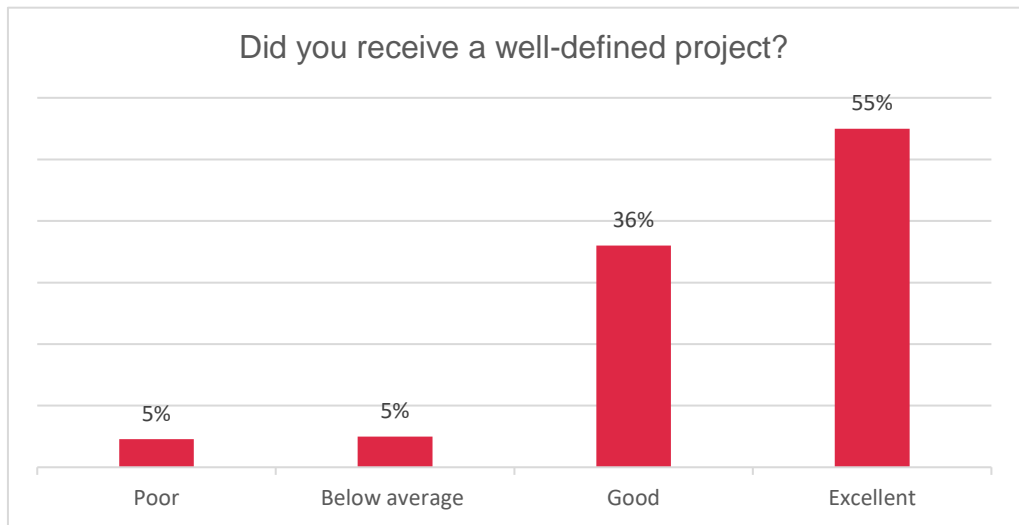
#### Do you have any suggestion or additional comment?

- More specific information about payments, currency conversion rates, cash withdrawal fees, etc.
- Helpful
- Please improve the quality of the bikes: most of them are pretty decent but a few are really bad in terms of gears slipping and have caused quite a few nasty falls.
- The CERN lectures could be scheduled either in the morning slot or maybe towards the end of the day. For students working in the Preessin site it becomes difficult to go for the lecture at 1:30 and return back at 3:30. It would be better if the lectures are in the morning, so they only have to go their once and then return to their offices.
- The two administrative students (Marion & Orestis) were extremely helpful.
- Bikes were bad. Report template does not work in any word processor I can use.
- The organization was perfect, really! We were prepared well before coming and also during the programme, thank you.





### b. SUPERVISOR AND PROJECT





### **Do you have any suggestion or additional comment?**

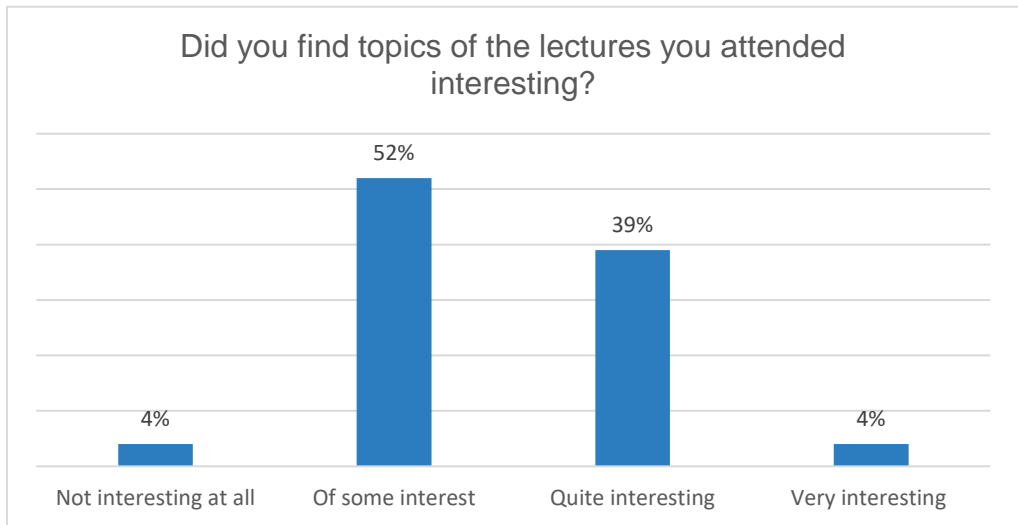
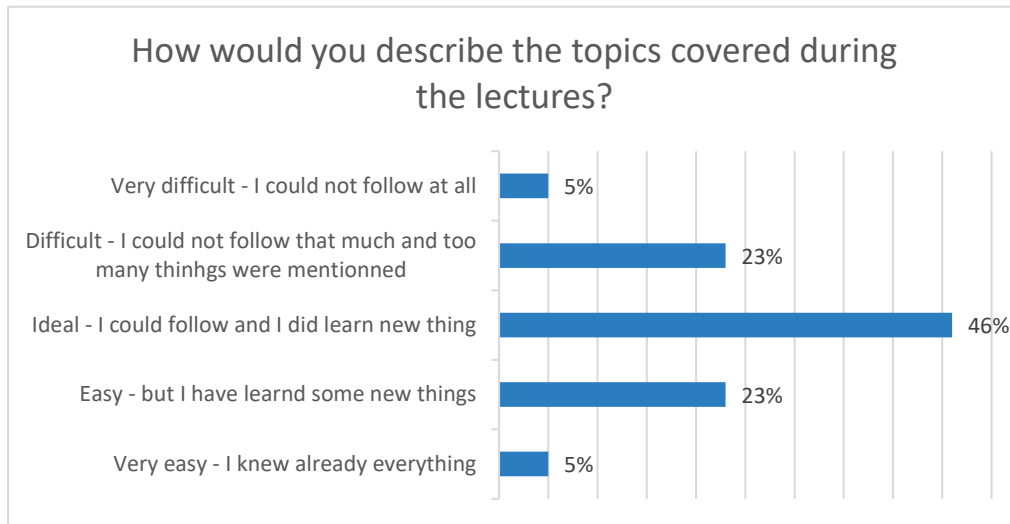
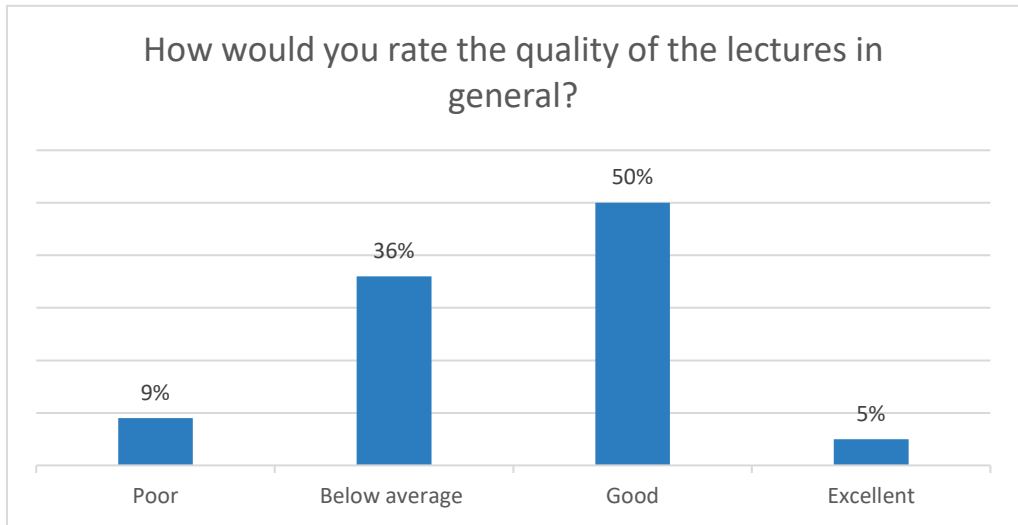
- Project was not well defined until 3rd or 4th week, which made it hard to finish on time.
- Really enjoyed it
- It was a brilliant experience overall!
- It was a wonderful experience working under Gianluca and Adrian. As I go back to my masters program I feel more enriched with knowledge and newfound skills. Thank you supervisors.
- I understand it is summer, and that people want to go on vacations, but we are sacrificing our entire summers for this, and the least the supervisors could do is make sure there is someone to replace them while they're away (if they're away more than a week), and that that person knows details of the project and can guide the students.
- I had the chance to work in IT-DB group, they were very warmfull with me. My supervisor help a lot during my project, it was a real pleasure to work with him and I learnt many things as well.

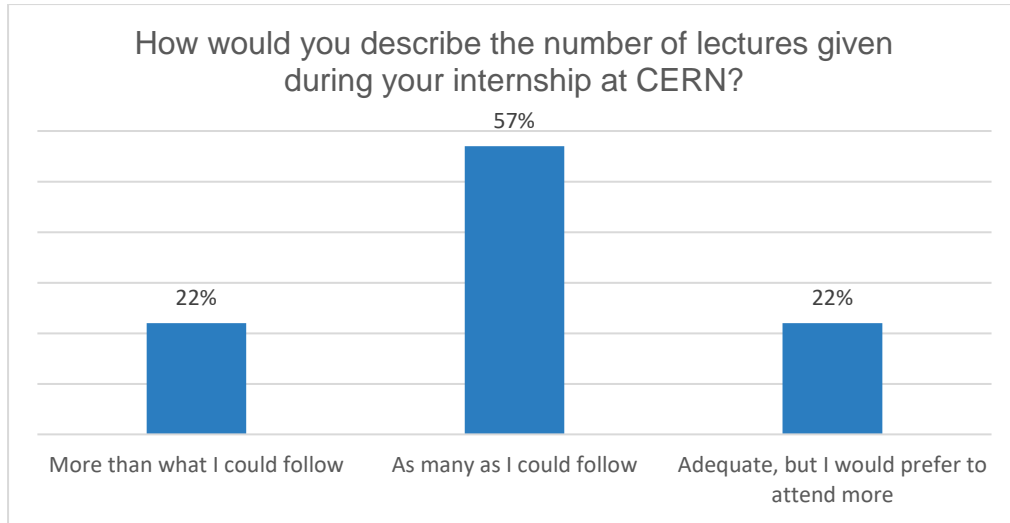






### c. LECTURES PROGRAMME





**Are there any other topic you would have been interested in?**

- Maybe something more about security. There were too many lectures about machine learning which is completely not interesting for me. I'm interested more in IT systems used in administration, but I understand that's not Cern's field.
- I think the topics were good - they often just started from a too complex level. They need to start simple and make it more hands on and interactive.
- Probably a deeper introduction to Machine Learning and Artificial Intelligence since Openlab has mostly Computing students.
- Computer security, System administration, Distributed computing
- More about hardware, the system itself.
- More security lectures Was a pity that the workshop was cancelled
- Maybe, Intelligence Artificial.

**Do you have any suggestion or additional comment?**

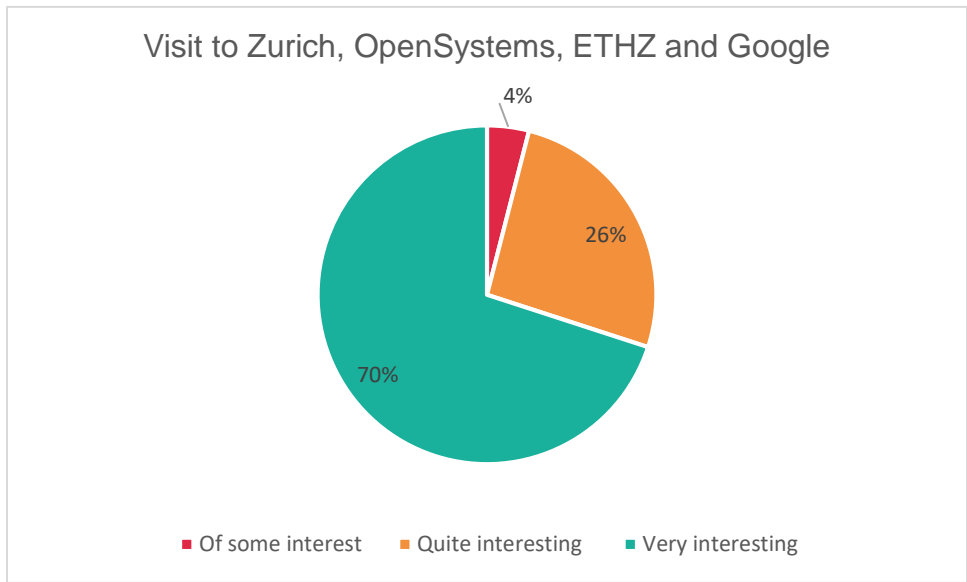
- Lecturers were knowledgeable on their respective fields but most were bad lecturers or presenters.
- No
- More hands-on workshops and less formal lectures would help increase participation.
- We had a lot of lectures which described the general way the LHC works, but no one went into more details. It would have been more useful if they did that instead of repeating same things over and over again.





- There were many lecturers, most of them gave excellent lectures, but some of them were not that good. Therefore, it is hard to give a general answer about a quality of lectures.
- Especially machine learning lectures started from beginner level but escalated quickly after third lecture. Also, for me it would have been better to get more code oriented lectures.
- I liked first presentation most. Rest of them covered mostly basics.
- In general, the lectures should be much more technical. Most of them are designed for physicists and it shows. The problem is that the majority of the openlab students are computer scientists. If the lectures cannot be made more technical they should at least explain, what CERN is doing and why it is important. For example, the machine learning lectures were quite easy but they would have been much more interesting if they presented some work done at CERN with it.
- I think that it is better to have the lecture in the morning because we are more concentrated than the afternoon.

**d. FIELD TRIP**



**Do you have any additional comment?**

- Highlight of the trip! OpenSystems was great, Google was amazing.
- Missed out on Google's original offices with the "cool rooms" but the new building was decent and we enjoyed it!



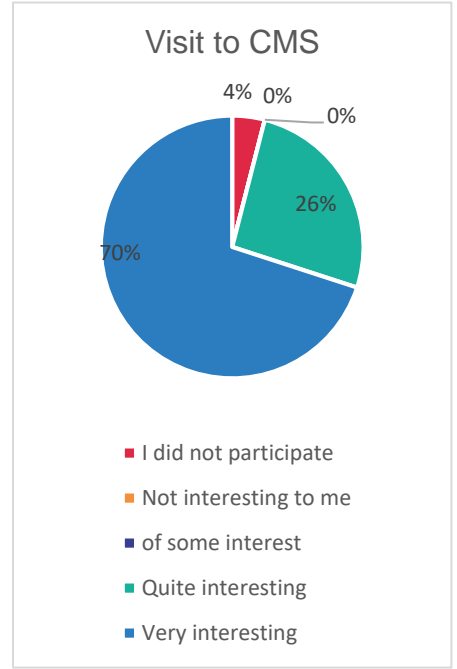
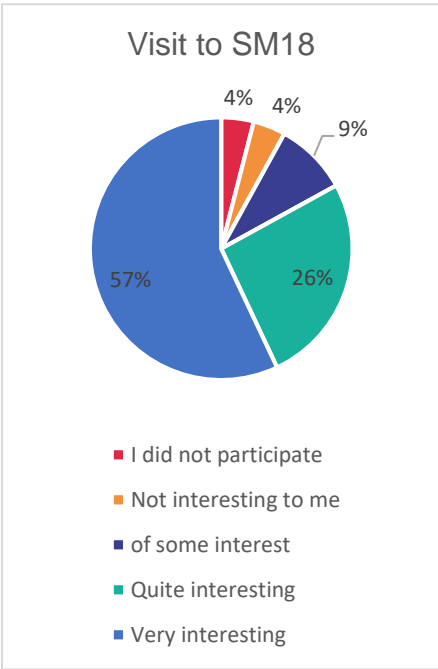
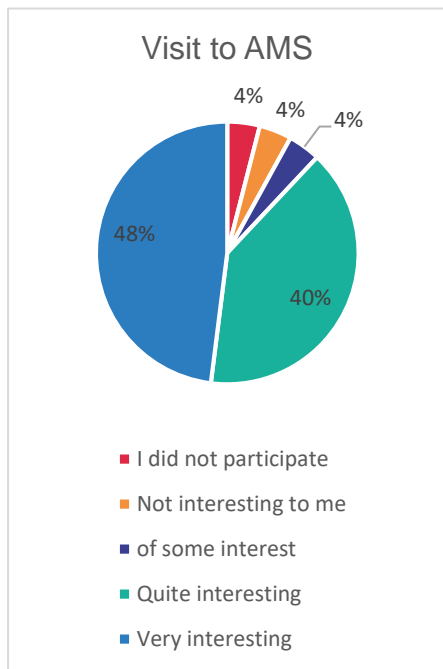
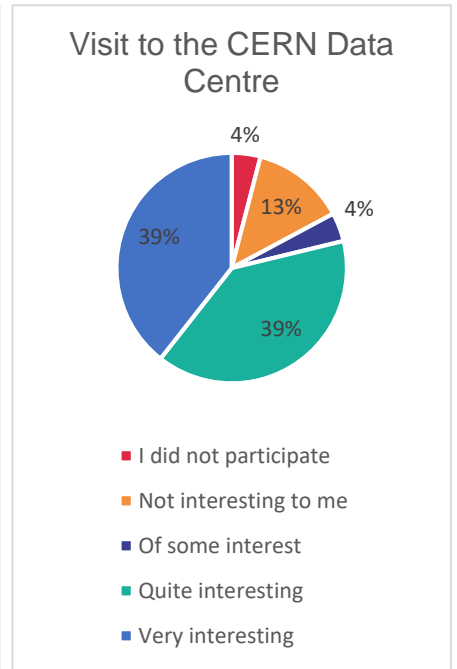
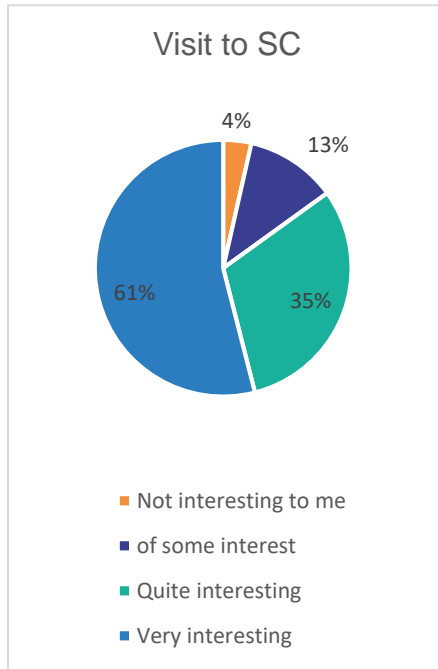
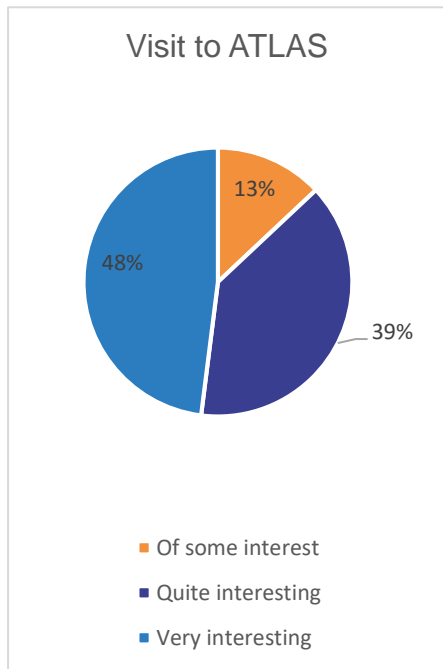


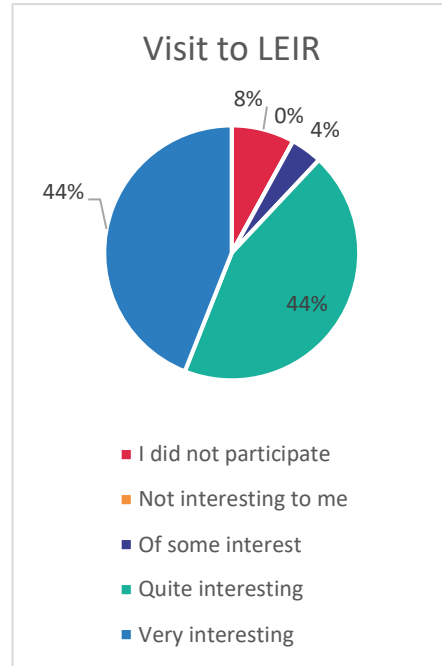
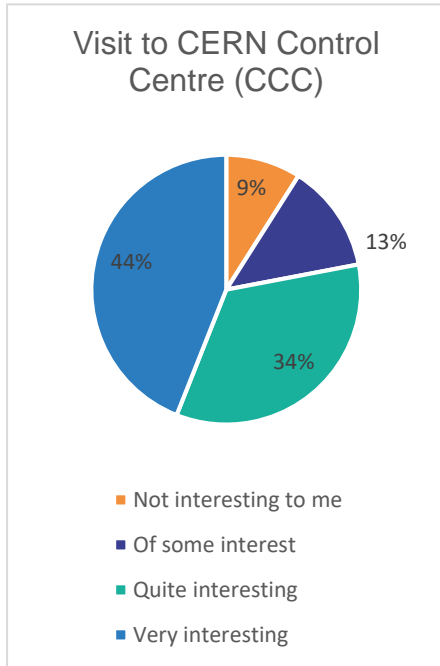
- I found time organisation especially on the second day too cautious. We had to be at the reception half an hour before bus departure, then we had an hour for the bus journey, which took less than 15 minutes, which resulted in another 40 minutes wait. The visit at Google was a disappointment for me with exception of the technical talk. HR talk showing us their careers website was poor and during the tour, they showed us a very small fraction of their facilities omitting any actual workspace. I would prefer to stay a bit longer in Zurich after the visits and have some free time couple of hours. I think it would be sufficient to come back to CERN around 9 to 10pm. We barely had time for lunch (partially because we have been at Google a lot longer than planned).
- Was the best event of the whole program Visit to SM18 was also awesome
- The visit to Zurich was really instructive. It was really nice to visit OpenSystems, ETHZ and Google. We learnt new things there and also how it works inside.



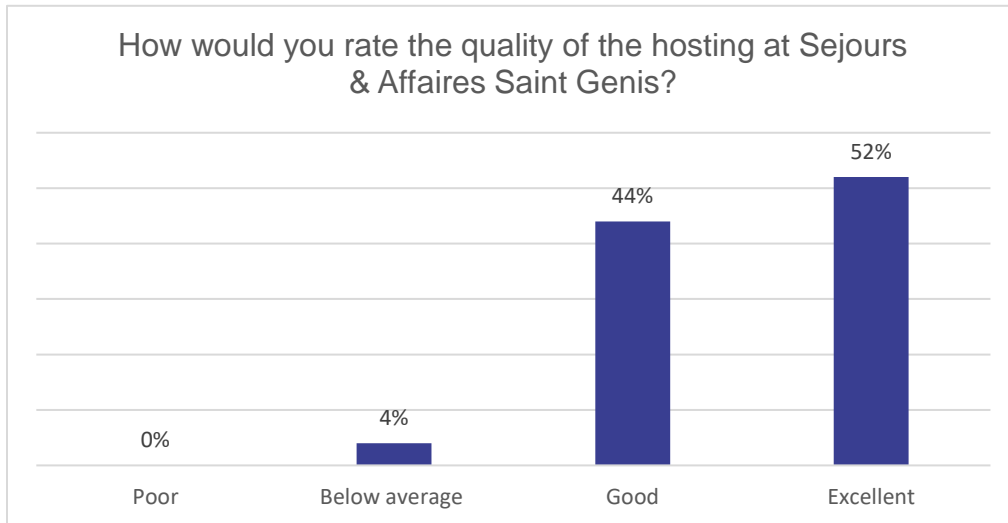


**e. VISITS**





**f. HOSTING**



**Do you have any suggestion or additional comment?**

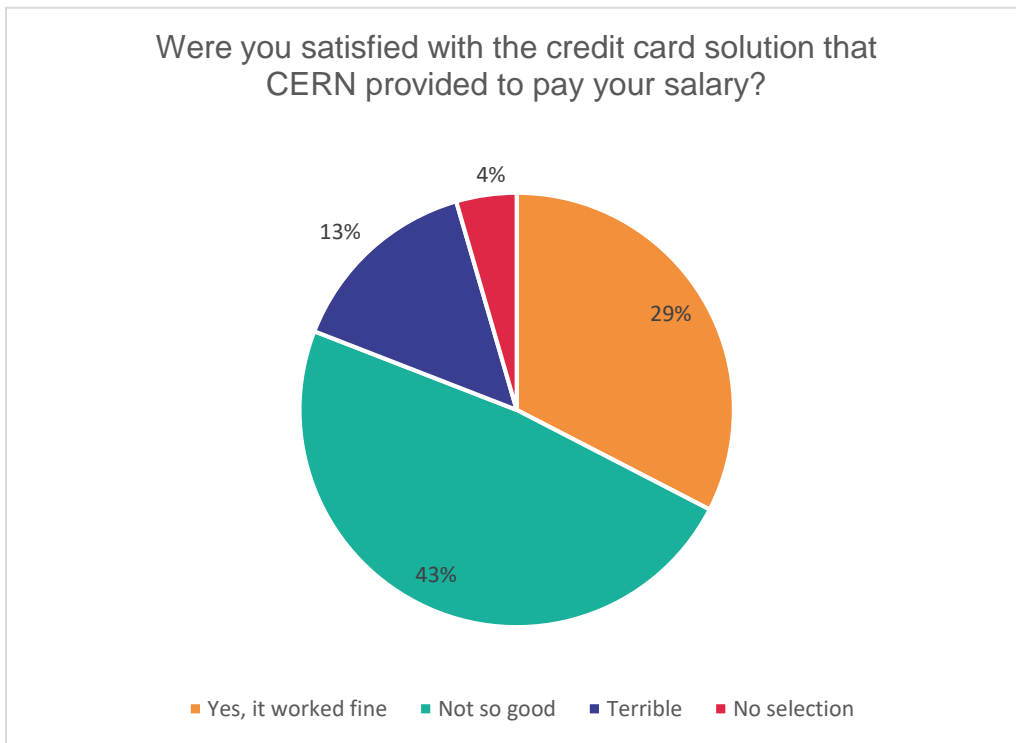
- Fair for the price. It would be better if we could know the things available at reception (vacuum cleaner, etc.) in advance.
- I think Sejours et Affaires was great, I liked all the services, room was big, clean, roommates were great, staff was nice





- Try to have more fans available in the hotel just in case it is a hot night (we had some of those but not too hot).
- Wi-Fi was definitely an issue, many a times bad enough to make a WhatsApp call back home.
- Maybe there should be less than four people per room.
- I had a private accommodation.
- It is a double-sided blade sharing a room with another person can be awesome or can be really bothersome. Was better if Sejours & Affaires was offering a cleaning schedule every 20 days at least or at least changing the bed sheets. Also was more comfortable if there was a place where we could put our valuables in.
- Good for friendship relations but it would have been better for me to stay in a room with less person, because we are staying with people we don't know from all over the world and there are some differences in cultures.
- Sincerely, our stay at Saint-Genis was really nice. We were four in the flat and we spend a great stay together. We were really comfortable.

**g. PAYMENT**





## Comments concerning the credit card or payment in general?

No
Currency conversion was a bit unfair and the monthly limit (and not knowing when the period ended) made me use my personal card more than a few times.
It was good, except the high fees for paying in Euro and withdrawing money. I used my American card because it had less fees. Overall, it was nice to have.
The conversion fees and processing fees were a slight disappointment. However the worst was that we could not change our phone number once we added it to the account. I cannot login because my Indian SIM is deactivated and the Indian number is the one that I used to login to the service for the first time, two months ago. Now I cannot update the number or receive the OTP via SMS so I cannot access my account.
In case there is some issue with the ban I think the student has to directly connect with the bank, which is difficult since they don't respond to mails and many students don't have a Swiss sim card here. Apart from that I feel it still is better than giving cash to students.
The credit card had a fees for almost everything, from withdrawal anywhere (Swiss or French side) to swiping it at French side. Since most of us were staying in the French side, it was always an additional charge. Similarly for paying at the hotel too. It would be great to have a credit card with less conversion fee or at least no charge on swiping whether on French or Swiss side.
It was kind of hard to keep track of expenses. Additional fees were added every time we withdrew money. Credit card's balance at the end of the month wasn't clear.
I find this solution to have very little benefits. The one benefit is that we were able to rent a car easily thank to the credit card which is often required there. Other than that the card was little use because of the high fees. If I would be using it freely, I would lose on average 4% of my salary, about 200 CHF (good deal for the bank...). We could not use it for buying lunches at CERN (only with withdrawn money with 5% commission), and most of the groceries and accommodation were paid in euros in France where we would incur 3% commission on top of the credit card exchange rate. The card is not usable for most of the purchases online because they require "3-D Secure" which can be registered only if the card is associated with an account number. I may be incorrect here, but simply I have not managed to pay even for the Geneva pass online with this card.
The online app for checking balance is weird, many students had difficulties paying sometimes. It's not ideal, but it worked good enough
ATM charges. Online balance check sucks, I and some other guys did not get the confirmation code on our phones and they blocked our online accounts after asking for codes more than few times. There are large number of currency conversion charges too. But obviously cash is







not a good choice either. So I think we should stuck with card but we need to find a more better option.
First of all the company is bad. Customer services, internet banking is terrible. I had to call customer services 5 times to enter to my online banking, and online banking is not helpful at all. I opened a new bank account recently and I sent a mail to CERN claims office that I want to transfer the rest of my belongings from my card to my new account and they told me that credit card company is dealing with that, So I did not even to bother call credit card company because I know it will be useless. The other weird thing is when I was withdrawing money from my card, it was charging me with a fee.
The credit card account login did not allow me to change my phone number for the login PIN SMS since I got a Swiss SIM later and wanted to use that instead.
The online app for checking balance is weird, many students had difficulties paying sometimes. It's not ideal, but it worked good enough
The website is not clear at all, rate for withdrawal are high but we are forced to do it because R1 and R2 only accept cash. It stopped working on the last day, but I still needed the Swiss card for the weekend.
From the statements or online bank you can understand nothing.
As I said before, we had many confusions with the credit card. For example, the amount the website didn't correspond the real amount available in the credit card.

**h. THE CERN OPENLAB SUMMER STUDENTS PROGRAMME IN GENERAL**

**What did you find most interesting in the CERN openlab Summer Student Programme?**

Definitely possibility to meet people from all over the world.
The other students and CERN community.
Being at CERN, my project, but the highlight of the trip for me was the Google Smart Reply talk.
The people, not only the other students but everyone at CERN was so helpful and interested in our work it felt really great to work here!
Trip to Zurich and fellow students whom I got a chance to know about.





Webfest and a lot more events like treasure hunt, picnics, project, and presentation.
My project indeed was very interesting, got to learn a lot both at implementation and documentation end.
The chance to work in my field of interests alongside really experienced people.
Visit to Open Systems
Probably the work itself, but I enjoyed other aspects of the program as well, such as visits, lectures and CERN itself.
Project that I worked with and the visits to the experiments
Meeting the new people, getting to know a new environment
I found my project very interesting, and through the project, I learnt a lot about technical challenges and physics at CERN. Also, I met people all around the world and I have never believed that we are so similar, but also so different.
It was really amazing learning all about the physics and interesting to know how scientific research is done at CERN (heavily relying on analysing data)
I enjoyed my project the most. I was not experienced in machine learning neither in satellite imagery so I learnt a lot of things. The only problem that I faced was for the first week I was understanding the project and stuff and then for next three weeks my supervisor was away for vacations so whenever I got stuck on small things I have to wait for the reply of my supervisor for quite some time or it took me a lot of time to search for solution of very small problems myself. Thus, I got the little time to complete my project. Otherwise, I really liked my project and I am planning to continue working on it after going back home. Further, because of this project and my stay here at CERN I have made my mind to pursue masters in machine learning after this. Further, the visits to CERN sites and Zurich was excellent.
Trip to Zurich and having a guide was fantastic. Also the trip to Google. Having an idea of how CERN works (mostly from trips and lectures). Getting to know people who work at CERN.
Opportunity to meet other people
The trips and the workshops were the most interesting part for me.
The enthusiasm people had at CERN to teach us more about anything and everything!
The exposure to an international research environment. I learned a lot and now I understand better what I want to do next
Opportunity to work on an interesting project in an international team, meet students from all over the world.





Everything: the project, the lectures, the visits, etc.
The visit to Zurich and the CERN's visit.

**What are the areas of improvement?**

For me we had too many lectures.
Lectures were on interesting topics but the presenter did a poor job delivering them
The lectures get boring. I think they should be shorter. I also think they should be more at the undergraduate level. Also, the CERN bikes are dangerous, they break, the pedals jam, the chain falls off, I wish you could improve that.
Probably the bikes, maybe more discounts for students at the restaurants (kidding, but we wouldn't mind, obviously!)
Everything was well arranged except the hackathons as R1 where it happens closes at 11 in the night.
Lectures CERN Bikes
Could the lectures be scheduled in the morning? since many departmental meetings take place during afternoon
The lectures could have been better.
Lecture speakers
<p>I felt like generation 3 failed to integrate into the group. It would be good to make some common events once everyone is at CERN. The credit card, see the box below. Webfest - this hackathon was very poorly organised. The amount of free food (2 lunches) was a joke. The minimum should be 3 meals a day and it is usual to offer snacks, coffee and other beverages through out the hackathon as well. The teams are working in their free time on projects which are usually connected to CERN, they should not be paying for it at least. In the description were mentioned 8 categories project, which should have been judged. There were completely ignored and 3 projects were awarded instead in which at least 2 fall into the same category. Judges should come from the informatics background. It seems like they were not able to appreciate the complexity of several projects and the argument "the only team who built something physical" sounds weird at the event which is called Webfest. Transportation - this is not related to openlab but it would be great if you could pass it to competent people. The large roundabout between St. Genis and CERN is very unpleasant and dangerous. Everyone</p>





<p>commuting to CERN from St Genis needs to pass 3 pairs of roads (soon there will be one more road) twice per day. I was once very close to a serious accident of being hit by a car in which if I would have been hit, I would be in the full right, but not sure if alive. Underpass or bridge should be built at this roundabout.</p>
<p>Payment.</p>
<p>Project definitions, lectures, bike availability</p>
<p>It was much easier to me when I got summer student teammate. Maybe putting summer students in the same team and letting them work together would be an interesting experiment.</p>
<p>The office I was working in had close to horrible condition. Especially on sunny days, the heat and the noise was too much.</p>
<p>You guys should improve the content of lectures. Because some lectures were not interesting at all, some were really easy and some were really difficult.</p>
<p>Presentations Working in an international environment Networking</p>
<p>None.</p>
<p>Change report template to other open format like. tex or odt. Current docx does not work in: Google Docs Word Online LibreOffice Calligra</p>
<p>Maybe do something about everyone falling off their bikes (partially our fault but also sometimes the bike's fault)!</p>
<p>The lectures can definitely be improved. Also the constraint to use word is not ideal for some kind of projects where you have to add equations or references in the report.</p>
<p>English skills, programming skills.</p>
<p>The overall organization of the programme was great. Maybe the lectures shouldn't be mandatory.</p>
<p>Maybe about the payment, we faced many confusions with the credit card.</p>

**Do you have any specific suggestion that would benefit the programme?**

<p>No</p>
<p>More visits like the one to Zurich would be great. That was one of the highlights of the program.</p>





Overall, the program is pretty great.
It's pretty much perfect. Kudos!
More interesting lectures but not in too detail. Just to define and explain the topic.
It would be good to improve the lectures. Since the lecture take up so much of our time here, it would be great if their quality is improved.
Everything was very well organized, a few suggestions that I had have already been written in the other sections. (but they are just minor glitches). Thank you for having me as a part of the program was very enriching.
The duration of the program could be a little longer (3 months for example), since getting acquainted in CERN's infrastructure takes some time and some of the projects are not so small in order to be completed in 2 months.
Pay with credit card @ restaurants Help find accommodation outside hostels Define summer student projects more concretely
Fees free salary payment, more food at Webfest, under pass at the St. Genis roundabout.
No.
Better defined projects, multiple supervisors (in case of vacations), prepared equipment which is adequate, so there's no waiting for PC's and similar
I was a participant of SDG summer school as well and it was a great experience because I got a chance to travel to China and how devices are manufactured. Maybe cooperation with similar summer schools would be interesting to other students as well.
With the projects, it could have been better defined or supported. Thank you very much for all your efforts. It was an amazing ride and without your immense efforts it wouldn't have been so.
Hmm No!
I think it would be more efficient to attend the lectures we like to. More lectures about code oriented would be better. First week I wasted a lot of time to search for the bike. So it would have been better to have a bike the very first day. Having the option to stay at a separate room in the hotel.
None.





More open source projects.
Partial cash and partial credit card payout makes it easier to use since it avoids extra charges for withdrawing cash.
The Intel modern code developer challenge was an interesting project but it had a really bad organization. In the end all of us put a lot of effort but didn't get much in return. For example the interview at the last day
Everything is cool, except credit cards :)
The Openlab Summer Student report template is ugly. Apart from that, it forces you to use Windows and Microsoft Word. At least in my case, it was a problem because I only use Linux so had to download a Windows ISO file from the hotel that took several days.
For me, it was a terrific experience and it's really sincere :) I liked everything, from my project, my team, the CERN's visit, the visit to Zurich and our trips to Switzerland's cities. I never forget the summer I spend there. Thank you openlab team, you did a perfect job.

**i. OVERALL ASSESSMENT OF THE CERN OPENLAB SUMMER STUDENTS PROGRAMME**

**Overall, what grade would you give to the CERN openlab summer student programme?**

9.1  
/ 10





#### 4. ACCEPTED PROJECTS AND STUDENTS

<b>Number</b>	1
<b>Partner/group</b>	Intel
<b>Project Title</b>	<b>Deep Learning for Fast Simulation: Development for Distributed Computing Systems</b>
<b>Description</b>	<p>Physicists widely use a software toolkit called GEANT4 to simulate what will happen when a particular kind of particle hits a particular kind of material in a particle detector. In fact, this toolkit is so popular that it is also used by researchers in other fields who want to predict how particles will interact with other matter: it's used to assess radiation hazards in space, for commercial air travel, in medical imaging, and even to optimise scanning systems for cargo security.</p> <p>An international team, led by researchers at CERN, is now working to develop a new version of this simulation toolkit, called GeantV. This work is supported by a CERN openlab project with Intel on code modernisation. GeantV will improve physics accuracy and boost performance on modern computing architectures.</p> <p>The team behind GeantV is currently implementing a 'deep-learning' tool that will be used to make simulation faster. The goal of this project is to write a flexible mini-application that can be used to support the efforts to train the deep neural network on distributed computing systems.</p>
<b>Supervisor(s)</b>	Sofia Vallecorsa, Andrei Gheata
<b>Student</b>	<b>Elena Orlova (Ms), Russia</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100 % openlab Intel
<b>Budget code(s)</b>	47503

<b>Number</b>	2
<b>Partner/group</b>	Intel -HTCC
<b>Project Title</b>	<b>Implementing a libfabric provider for DPDK</b>





<b>Description</b>	<p>Contents: libfabric is a new library in the OFED unifying access to high-speed, remote-DMA protocols such as InfiniBand or Intel OmniPath. Currently there is no high-performance implementation for Ethernet in libfabric. The topic of this project is to work on a provider for DPDK which is a high-speed, very low overhead Ethernet in user-space framework. The student will work with modern 100 Gigabit Ethernet Networks and have to think on how to efficiently mimic reliable messaging using Ethernet.</p> <p>Student profile: computer scientist / engineer</p> <p>Training value: high-speed local area networking</p> <p>Requisites: Knowledge of C and Linux systems programming, experience with Ethernet networking is an advantage</p>
<b>Supervisor(s)</b>	Niko Neufeld
<b>Student</b>	<b>Mohammad Hassan Zahraee (Mr), Iran</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100 % openlab Intel
<b>Budget code(s)</b>	47503

<b>Number</b>	<b>3</b>
<b>Partner/group</b>	Intel-HTCC/EP-LBC
<b>Project Title</b>	<b>Evolutionary optimization of LHCb software compilation</b>
<b>Description</b>	<p>Final code performance largely depends on four aspects: the algorithm's asymptotic complexity, the use of implementation techniques such as vectorization or multithreading, compiler tuning and optimizations, and finally the execution environment (ie. the hardware platform and the operating system). When implementing a new code the initial focus should be first on the used algorithms and data structures and second on the implementation.</p> <p>In this project, we would like to explore the third aspect of code performance: How can maximum performance be achieved on an existing code by tuning the compiler's optimization flags. Modern compilers offer a multitude of parameters that allow the</p>







	<p>developer to influence code optimization and generation. A few of these parameters (e.g. -O3) are today commonplace in development workflows, while many others remain often untouched. This is also because it is unclear in many cases what the best choice of parameter value is. The procedure of compiler parameter tuning can be regarded as the optimization of a multi-dimensional objective function with the code execution time as its score. One method to optimize such black-box functions are genetic algorithms. Genetic algorithms (or evolutionary optimization) apply the idea of evolution (ie. mutation and selection) to function optimization. In this case, each "generation" of compiler parameters are evaluated and the best are selected as the basis for the next generation, where some "mutation" allows for new parameter values to be explored. The final result is a set of compiler parameters that yield the best-performing code.</p> <p>The LHCb online framework offers a great opportunity for such an optimization where an improvement in execution time is of immediate value to the experiment. Here we would like to test a basic evolutionary optimization (EO) scheme to tune the compiler parameters with the objective of improving the framework's runtime. This can be easily implemented using a scripting language such as python, taking advantage of various existing libraries. The goal is to test a real production code on a state of the art hardware platform such as the Intel Xeon Skylake processor, using the latest Intel C++ compiler taking advantage of its many detailed tuning parameters.</p> <p>Training value: genetic algorithms, software development, compiler optimization, hardware architectures</p>
<b>Supervisor(s)</b>	Daniel Campora, Omar Awile
<b>Student</b>	<b>Leticia Freire de Figueiredo (Ms), Brazil</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100 % openlab Intel
<b>Budget code(s)</b>	47503

<b>Number</b>	4
<b>Partner/group</b>	Intel -HTCC/ IT-CF-FPP
<b>Project Title</b>	<b>Connecting the dots: Using machine learning to better identify the particles produced by collision events</b>





<b>Description</b>	<p>The particle detectors at CERN are like cathedral-sized 3D digital cameras, capable of recording hundreds of millions of collision events per second. The detectors consist of multiple ‘layers’ of detecting equipment, designed to recognise different types of charged particles produced by the collisions at the heart of the detector. As the charged particles fly outwards through the various layers of the detector, they leave traces, or ‘hits’.</p> <p>Tracking is the art of connecting the hits to recreate trajectories, thus helping researchers to understand more about and identify the particles. The algorithms used to reconstruct the collision events by identifying which dots belong to which charged particles can be very computationally expensive. And, with the rate of particle collisions in the LHC set to be further increased over the coming decade, it’s important to be able to identify particle tracks as efficiently as possible.</p> <p>Many track-finding algorithms start by building ‘track seeds’: groups of two or three hits that are potentially compatible with one another. Compatibility between hits can also be inferred from what are known as ‘hit shapes’. These are akin to footprints; the shape of a hit depends on the energy released in the layer, the crossing angle of the hit at the detector, and on the type of particle.</p> <p>This project investigates the use of machine-learning techniques to help recognise these hit shape more efficiently. The project will explore the use of state-of-the-art many-core architectures, such as the Intel Xeon Phi processor, for this work.</p>
<b>Supervisor(s)</b>	Omar Awile, Felice Pantaleo (EP-CMG), Luca Atzori
<b>Student</b>	<b>Antonio Carta (Mr), Italy</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100% openlab + partner
<b>Budget code(s)</b>	47503

<b>Number</b>	5
<b>Partner/group</b>	Intel/BE-ICS-CIC
<b>Project Title</b>	<b>Edge Computing: Integrating IoT Devices into the LHC Control Systems</b>





<b>Description</b>	<p>The Large Hadron Collider (LHC) accelerates particles to over 99.9999% of the speed of light. It is the most complex machine ever built, relying on a wide range industrial control systems for proper functioning.</p> <p>This project will focus on integrating modern ‘systems-on-a-chip’ devices into the LHC control systems. The new, embedded ‘systems-on-a-chip’ available on the market are sufficiently powerful to run fully-fledged operating systems and complex algorithms. Such devices can also be easily enriched with a wide range of different sensors and communication controllers.</p> <p>The ‘systems-on-a-chip’ devices will be integrated into the LHC control systems in line with the ‘internet of things’ (IoT) paradigm, meaning they will be able to communicate via an overlaying cloud-computing service. It should also be possible to perform simple analyses on the devices themselves, such as filtering, pre-processing, conditioning, monitoring, etc. By exploiting the IoT devices’ processing power in this manner, the goal is to reduce the network load within the entire control infrastructure and ensure that applications are not disrupted in case of limited or intermittent network connectivity.</p>
<b>Supervisor(s)</b>	Filippo Tilaro, Enrique Blanco Vinuela
<b>Student</b>	<b>Lamija Tupo (Ms), Bosnia</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100 % openlab Intel
<b>Budget code(s)</b>	47503

<b>Number</b>	6
<b>Partner/group</b>	<b>IT-DI-OPL</b>
<b>Project Title</b>	<b>Cells in the cloud: Running biological simulations more efficiently with cloud computing</b>
<b>Description</b>	<p>The BioDynaMo project is part of CERN openlab’s collaboration with Intel on code modernisation.</p> <p>It is a joint effort between CERN, Newcastle University, Innopolis University, and Kazan Federal University to design and build a scalable and flexible platform for rapid simulation of biological tissue development.</p>





	<p>The project focuses initially on the area of brain tissue simulation, drawing inspiration from existing, but low-performance software frameworks. Late 2015 and early 2016 saw algorithms already written in Java code ported to C++. Once porting was completed, work was carried out to optimise the code for modern computer processors and co-processors.</p> <p>In order to be able to address ambitious research questions, however, more computational power will be needed. Work will, therefore, be undertaken to adapt the code for running using high-performance computing resources over the cloud. This project focuses on adding network support for the single-node simulator and prototyping the computation management across many nodes.</p>
<b>Supervisor(s)</b>	Fons Rademakers, Lukas Breitwieser
<b>Student</b>	<b>Konstantinos Kanellis (Mr), Greece</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100% openlab Common Funds
<b>Budget code(s)</b>	47557

<b>Number</b>	7
<b>Partner/group</b>	IT-DI-OPL/openlab/UNOSAT
<b>Project Title</b>	<b>Disaster Relief Using Satellite Imagery</b>
<b>Description</b>	<p>UNOSAT is part of the United Nations Institute for Training and Research (UNITAR). It provides a rapid front-line service to turn satellite imagery into information that can aid disaster-response teams. By delivering imagery analysis and satellite solutions to relief and development organizations — both within and outside the UN system — UNOSAT helps to make a difference in critical areas such as humanitarian relief, human security, and development planning.</p> <p>Since 2001, UNOSAT has been based at CERN and is supported by CERN's IT Department in the work it does. This partnership means UNOSAT can benefit from CERN's IT infrastructure whenever the situation requires, enabling the UN to be at the forefront of satellite-analysis technology. Specialists in geographic information systems and in the analysis of satellite data, supported by IT engineers and policy experts, ensure a dedicated service to the international humanitarian and development communities 24 hours a day, seven days a week.</p>





	CERN openlab and UNOSAT are currently exploring new approaches to image analysis and automated feature recognition to ease the task of identifying different classes of objects from satellite maps. This project evaluates available machine-learning-based feature-extraction algorithms. It also investigates the potential for optimising these algorithms for running on state-of-the-art many-core architectures, such as the NVIDIA GPU.
<b>Supervisor(s)</b>	Taghi Alyiev
<b>Student</b>	<b>Muhammad Abu Bakr, (Mr) Pakistan</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	47503
<b>Budget code(s)</b>	

<b>Number</b>	8
<b>Partner/group</b>	<b>Oracle/IT-DB-DBR</b>
<b>Project Title</b>	<b>Building effective database backup &amp; recovery monitoring using Elastic Stack</b>
<b>Description</b>	<p>The aim of this project is to find an efficient way to handle all activity of our complex backup and recovery system and use data mining to enable pro-active monitoring. The selected candidate will publish logs from our Oracle Database Backup and Recovery system to the IT-DB-ELK (ElasticSearch, Logstash, Kibana) or equivalent infrastructure and extract potentially useful patterns or information from them.</p> <p>This project requires cross-section collaboration as the subject covers a broad range of technologies including Oracle RDBMS and Big Data Analytic. The candidate will gain knowledge on the backup and recovery system developed at CERN and the ELK stack, which is one of the most popular log analysis open-source solution.</p>
<b>Supervisor(s)</b>	Sebastien Masson
<b>Student</b>	<b>Yasmine Nasri (Ms), Algeria</b>
<b>Date</b>	03 July to 01 September





<b>Funding</b>	100% openlab Oracle
<b>Budget code(s)</b>	47504

<b>Number</b>	9
<b>Partner/group</b>	Oracle/IT-DB-SAS
<b>Project Title</b>	<b>Zenodo Keyword Auto-Suggest Using Parallel Graph Analytics</b>
<b>Description</b>	<p>Are you interested in acquiring solid knowledge in Big Data, data analytics, and stream data management and analysis? CERN's openlab data analytics team is seeking a motivated candidate to evaluate Oracle Stream Explorer analytics capabilities and integration with the CERN's Kafka Pilot Service, which automate data streaming management, analysis and ingestion from a wide range of data-services at CERN.</p> <p>During the summer you will work closely with CERN experts to access to specific data sources, integrate them within the Oracle Stream Explorer framework and perform data analytics task.</p>
<b>Supervisor(s)</b>	Manuel Martin Marquez
<b>Student</b>	<b>Alastair Paragas (Mr), USA</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	100% openlab Oracle
<b>Budget code(s)</b>	47504

<b>Number</b>	10
<b>Partner/group</b>	Oracle/IT-DB-IMS
<b>Project Title</b>	<b>Refactor of Rota application using modern developer techniques: Oracle Restful Data Services &amp; Oracle Jet</b>





<p><b>Description</b></p>	<p>The Rota application schedules the support activities of all the members of the CERN IT-DB group. This system is used by the database, middleware, storage and systems administrators. Nowadays the application is highly tied to the database due to its implementation (Oracle APEX).</p> <p>Oracle Restful Data Services (ORDS) makes it easy to develop modern REST interfaces for relational data in Oracle Databases. Oracle Javascript Extension Toolkit (JET) It's a collection of open source JavaScript libraries along with a set of Oracle contributed JavaScript libraries that make it as simple and efficient as possible to build applications that consume and interact with Oracle products and services.</p> <p>The goal of the project is to design a new web user interface for APP-X with these objectives in mind:</p> <ul style="list-style-type: none"> <li>- Improve the User experience.</li> <li>- Separate the back-end services from the user interface.</li> <li>- Make the application portable, modular and open to extensions.</li> <li>- Test and evaluate some of the latest Oracle technologies and frameworks.</li> </ul> <p>The student will get familiar with the full cycle of project development: from analyzing and getting the user requirements to the implementation and deployment of a final solution.</p>
<p><b>Supervisor(s)</b></p>	<p>Luis Rodríguez Fernández</p>
<p><b>Student</b></p>	<p><b>Rebekka Alsvaag (Ms), Norway</b></p>
<p><b>Date</b></p>	<p>19 June to 18 August</p>
<p><b>Funding</b></p>	<p>(100% openlab oracle)</p>
<p><b>Budget code(s)</b></p>	<p>47504</p>

<p><b>Number</b></p>	<p>11</p>
<p><b>Partner/group</b></p>	<p>Oracle/IT-DB-IMS</p>
<p><b>Project Title</b></p>	<p><b>Development of weblogic 12c management tools.</b></p>
<p><b>Description</b></p>	<p>Weblogic infrastructure (operated by the IT-DB group) is built of 250 clusters. It is used to deploy a wide range of critical application in administration and engineering sectors.</p>





	<p>Right now the group is preparing the upgrade to Weblogic 12c. This version offers few new interesting features like dynamic clusters, REST based management etc. In this context, we'd like the student to re-implement the weblogic domain creation process, plus some other supporting processes around clusters and partitioning.</p> <p>Experience required:</p> <ul style="list-style-type: none"> <li>- Linux - user level experienced</li> <li>- Python - simple scripting</li> </ul>
<b>Supervisor(s)</b>	Nicolas Bernard Marescaux
<b>Student</b>	<b>Saul Alonso Monsalve (Mr), Spain</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	(100% openlab)
<b>Budget code(s)</b>	47504

<b>Number</b>	<b>12</b>
<b>Partner/group</b>	<b>Oracle/IT-DB-SAS</b>
<b>Project Title</b>	<b>Anomaly Detection in Database Connections</b>
<b>Description</b>	<p>CERN IT database group has built a highly scalable, secure and central repository that stores consolidated audit data and optionally listener, alert and OS log events generated by the (oracle) databases. This central platform is used for reporting, alerting and security policy management. The database group want to further exploit the information available in central repository to build intrusion detection system to enhance the security of the database infrastructure. In addition build pattern detection models to flush out anomalies using the monitoring and performance metrics available in the central repository. The potential candidate will work on designing and developing these models with guidance and support from the supervisor.</p>
<b>Supervisor(s)</b>	Prasanth Kothuri , Daniel Lanza García
<b>Student</b>	<b>Swapneel Mehta (Mr), India</b>







<b>Date</b>	03 July to 11 August
<b>Funding</b>	100% openlab Oracle
<b>Budget code(s)</b>	47504

<b>Number</b>	<b>13</b>
<b>Partner/group</b>	<b>Siemens AG / CERN BE-ICS</b>
<b>Project Title</b>	<b>Web Reporting Framework for Control Data Analysis</b>
<b>Description</b>	The massive amount of data generated by the CERN control systems is continuously analyzed through specific custom made algorithms (machine learning, deep learning, NN, KNN ...). The project will focus on developing a web front-end able to provide these analysis results to the machine operators and experts. In order to achieve that this web service has to be integrated with the current analytical infrastructure based on Hadoop ecosystem (Spark, HDFS, Impala) and CEP engines. It will be an opportunity for the student to learn Big Data architecture concepts and contribute to the use of data analytics frameworks at CERN.
<b>Supervisor(s)</b>	Filippo Tilaro
<b>Student</b>	<b>Lauri Matthias Sainio (Mr), Finland</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	(i.e 100% openlab +partner, or 50/50 or 100% group budget)
<b>Budget code(s)</b>	47505

<b>Number</b>	<b>14</b>
<b>Partner/group</b>	<b>Siemens ETM / CERN BE-ICS</b>
<b>Project Title</b>	<b>Simplified Frontend for Data Generation and Testing Purposes in WinCC OA NextGen Archiver Project</b>





<b>Description</b>	<p>A huge number of critical systems for CERN accelerator and detectors are steered and supervised by their operators through "Control Systems". These are built using a commercial SCADA (Supervisory Control and Data Acquisition) system called WinCC OA, from the ETM company (subsidiary of Siemens). Measurements, such as cryogenic temperatures, pressures in vacuum systems, high voltages and currents are read out from sensors, displayed to operators but it is also essential to store their history in a database, such that it could be retrieved and plotted on the screen. This is the role of so-called "archiver" subsystem for any SCADA System.</p> <p>The Openlab Summer Student would join the already ongoing effort to provide a "Next Generation" archiver for WinCC OA, capable of storing huge amounts of data into a Hadoop-based storage system. Skills in C++ programming, Linux, large deployments as well as Big Data concepts will be on daily use.</p> <p>Collaboration with developers at both the Industrial Control and Safety group at CERN, at ETM as well as with other CERN groups would surely create a fascinating environment for development of various technical and personal skills. It will also be an opportunity to contribute to the essential part for the future of CERN operation, and enable the use of Data Analytics.</p>
<b>Supervisor(s)</b>	Jakub Guzik, Rafal Kulaga
<b>Student</b>	<b>Urishita Puri (Ms), India</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	(i.e 100% openlab + partner, or 50/50 or 100% group budget)
<b>Budget code(s)</b>	47505

<b>Number</b>	15
<b>Partner/group</b>	<b>Rackspace / IT-CM</b>
<b>Project Title</b>	<b>Cloud Log Data Analysis</b>
<b>Description</b>	In recent months the CERN OpenStack cloud team has built on the monitoring infrastructure to improve the collection and availability of its metrics and log data. This data is easily accessible via Elasticsearch and Influxdb, with multiple visualizations being built using tools like Grafana.





	<p>The student will explore this data using recent ML techniques, namely clustering, with the aim of:</p> <ul style="list-style-type: none"> <li>• detecting patterns in the behavior of the cloud services</li> <li>• identifying correlations triggering common anomalies in the services</li> <li>• exposing these results to the cloud operators</li> </ul> <p>For this analysis we aim at using medium sized container clusters.</p> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>• Knowledge of basic ML techniques such as clustering</li> <li>• Spark (or similar tools, tensorflow, keras) data analysis</li> </ul>
<b>Supervisor(s)</b>	Ricardo Rocha
<b>Student</b>	<b>Shweta Oak, Ms, India</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	100% Rackspace
<b>Budget code(s)</b>	47559

<b>Number</b>	16
<b>Partner/group</b>	Rackspace / IT-CM
<b>Project Title</b>	<b>Stateful Services in Containers</b>
<b>Description</b>	<p>Until now most of the container workloads have involved stateless services, so that container instances can be easily migrated between nodes without disruption.</p> <p>Recently Kubernetes has added support for StatefulSets, where containers are more tied to the underlying node, making them usable for use cases with higher persistency and locality requirements such as databases.</p> <p>The student will investigate the current state and support for running stateful applications in containers, taking as a starting use case a database like MySQL or Cockroachdb.</p> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>• Good sysadmin skills</li> </ul>





<b>Supervisor(s)</b>	Spyros Trigazis
<b>Student</b>	<b>Markus Sommer (Mr), Germany</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100% Rackspace
<b>Budget code(s)</b>	47559

<b>Number</b>	<b>17</b>
<b>Partner/group</b>	<b>Yandex /EP-CMG</b>
<b>Project Title</b>	<b>Machine Learning for Data Certification</b>
<b>Description</b>	Machine Learning can be used to build a binary classifier aiming to distinguish good from bad data. The project starts from the strategy set up by the Yandex data school team on the CMS 2010 Open data. The work consists in porting this work into the data certification system of CMS and to develop it further, applying semi-supervised techniques (e.g., autoencoders) to define a classification strategy that doesn't assume previous knowledge of failure modes.
<b>Supervisor(s)</b>	Giovanni Franzoni / Federico De Guio
<b>Student</b>	<b>Lukas Danev (Mr), Slovakia</b>
<b>Date</b>	26 June to 25 August
<b>Funding</b>	openlab common funds
<b>Budget code(s)</b>	47557

<b>Number</b>	<b>18</b>
<b>Partner/group</b>	<b>IDT/IT-CF-FPP</b>





<b>Project Title</b>	<b>A Monitoring and Management Utility for RapidIO Clusters</b>
<b>Description</b>	<p>The objective for the Openlab collaboration with IDT is to test and evaluate the suitability of IDT’s low-latency RapidIO interconnect technology for a number of use-cases stretching from LHC Data Acquisition and Triggering to Data Analytics for the data center monitoring and operations.</p> <p>Within the project there are a number of activities for streamlining the benchmarking and management of the project clusters. These activities involve C programming, web development, scripting and understanding the technology</p> <p>The student may, depending on their skills and personal interests, get involved in any of these activities. Possible directions are tools and web interfaces for remote switch management, automation of post-installation sanity checks and benchmarks using for example Jenkins and other automation tools.</p>
<b>Supervisor(s)</b>	Sima Baymani
<b>Student</b>	<b>Scott Fitsimones (Mr), USA</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	100% IDT
<b>Budget code(s)</b>	47556

<b>Number</b>	<b>19</b>
<b>Partner/group</b>	<b>Brocade/IT-CS</b>
<b>Project Title</b>	<b>Network Automation with Brocade Workflow Composer</b>
<b>Description</b>	<p>CERN is investigating the suitability of commercial or open-source platforms for evolving and improving its network configuration management tools. An extensible configuration management framework is essential for enabling rapid network services provisioning without time-consuming, manual interventions.</p> <p>The selected candidate will investigate the network automation services support Brocade’s Workflow Composer (BWC) platform [1]. Once familiar with the BWC software, the candidate will aim to develop a proof of concept solution for automating a concrete configuration scenario required in CERN’s networks. After validating the workflow using Brocade switches, the candidate will investigate BWC’s extensibility</p>





	for supporting third party network devices. The project requires good Linux system administration and Python scripting skills. Knowledge of computer networks concepts is considered a plus.  [1] <a href="http://www.brocade.com/en/products-services/automation/workflow-composer.html">http://www.brocade.com/en/products-services/automation/workflow-composer.html</a>
<b>Supervisor(s)</b>	Stefan Stancu, Adam Krajewski
<b>Student</b>	<b>Muhammad Abdullah (Mr), Pakistan</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	100% openlab
<b>Budget code(s)</b>	47509

<b>Number</b>	20
<b>Partner/group</b>	EP-CMG
<b>Project Title</b>	<b>Particle-Level Event Classification for LHC Trigger with End-to-End Sorting RNNs</b>
<b>Description</b>	This projects aims at building a real-time monitoring system for a particle detector. Using the information from status sensors (pressure, temperature, etc) and modern Machine Learning techniques, the candidate will define algorithms capable of spotting in real time the occurrence of an anomalous behavior. In addition, the possibility of predicting failure patterns will be investigated. Such a project will be developed on CMS data. The technique has a wide class of use cases, from particle detectors to the accelerator complex.
<b>Supervisor(s)</b>	Maurizio Pierini , Jean-Roch Vlimant
<b>Student</b>	<b>Daniel Weitekamp (Mr), USA</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	





<b>Budget code(s)</b>	47557
-----------------------	-------

<b>Number</b>	21
<b>Partner/group</b>	EP-CMG
<b>Project Title</b>	<b>Anomaly Detection Using Machine Learning for Data Quality Monitoring in the CMS Experiment</b>
<b>Description</b>	Using anomaly detection techniques, based on Machine Learning, one can perform tests on incoming data to assess their quality, through a comparison with known good data. In this project, the student will apply this technology to real workflows, related to the online DQM system of the CMS experiment
<b>Supervisor(s)</b>	Adrian Pol, Gianluca Cerminara
<b>Student</b>	<b>Agrima Seth (Ms), India</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	openlab
<b>Budget code(s)</b>	47557

<b>Number</b>	22
<b>Partner/group</b>	Huawei/IT-CM-RPS
<b>Project Title</b>	<b>Openstack Cloud Application Catalog</b>
<b>Description</b>	The OpenStack project Murano allows users to spawn applications from a pre defined catalog. This can be useful for unexperienced users that can get a pre configured application with one click. Typical applications already available are listed at <a href="https://apps.openstack.org/#tab=murano-apps">https://apps.openstack.org/#tab=murano-apps</a> but these can be further enhanced with local applications.





	<p>The project would be to investigate the use of Murano (and the OpenStack artifacts store, Glare) for potential future use on the CERN cloud.</p> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>• Linux</li> <li>• Python</li> </ul>
<b>Supervisor(s)</b>	Belmiro Moreira
<b>Student</b>	<b>Victor Araujo Soto (Mr), Guatemala</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	
<b>Budget code(s)</b>	47952

<b>Number</b>	23
<b>Partner/group</b>	Huawei
<b>Project Title</b>	<b>Job scheduling with Kubernetes</b>
<b>Description</b>	<p>Kubernetes has the notion of a Job, a higher level abstraction which adds to the underlying Pods the notions of retries, scheduling and some error handling.</p> <p>This makes it easy to implement workflow engines, or to use Kubernetes as a batch scheduler. The student will take a basic batch use case and explore how much of the required functionality is already available in Kubernetes, and identify the missing functionality.</p> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>• Linux</li> </ul>
<b>Supervisor(s)</b>	Ricardo Rocha
<b>Student</b>	<b>Clenimar Filemon Souza (Mr), Brazil</b>
<b>Date</b>	03 July to 01 September







<b>Funding</b>	Huawei
<b>Budget code(s)</b>	47952

<b>Number</b>	24
<b>Partner/group</b>	IT-DI-OPL
<b>Project Title</b>	<b>First Results of Storing Genomics Data in ROOT</b>
<b>Description</b>	Genomics data is traditionally being stored in the BAM format ( <a href="http://genome.sph.umich.edu/wiki/BAM">http://genome.sph.umich.edu/wiki/BAM</a> ), however this format is not optimal for large data access by analysis tools. At CERN all scientific experimental data is stored in the ROOT format which is optimized for very large data sizes and for efficient data analysis. In this project the summer student will work on understanding the BAM and ROOT file formats and on how BAM files can be converted into ROOT files in a format that is most efficient for later data analysis. The student should be good in C++ programming. Having some understanding of genomics is a plus. The student will learn about the leading I/O formats in genomics and particle physics and the importance of efficient I/O for data analytics.
<b>Supervisor(s)</b>	Fons Rademakers
<b>Student</b>	<b>Jose Gonzalez Ortiz (Mr), Spain</b>
<b>Date</b>	19 June to 11 August 2017
<b>Funding</b>	100 % openlab
<b>Budget code(s)</b>	47557

<b>Number</b>	25
<b>Partner/group</b>	University of Geneva
<b>Project Title</b>	<b>Gigabit Ethernet for Volunteer Computing</b>





<b>Description</b>	The objective of this project is to collaborate with a team in Kansas City which has received funding from Mozilla Foundation to test new ways to exploit Gigabit-ethernet-to-the-home for distributed volunteer computing, using LHC@home as a test case. Such Gigabit Ethernet is being rolled out on an experimental basis by Google in certain neighbourhoods of a few US cities. Specific LHC@home jobs that require large I/O could provide a compelling science and education application for such high bandwidth connections. Building on a previous summer student project, this project will gather data on the performance of the infrastructure, and use this to model what sort of science applications could benefit, both for LHC@home and more broadly for volunteer computing projects in areas like climate change modelling or epidemiology.
<b>Supervisor(s)</b>	Laurence Field (IT-CM-IS), Ben Segal (IT), Francois Grey (CCL)
<b>Student</b>	<b>Aleksandra Starzec (Ms), Poland</b>
<b>Date</b>	19 June to 11 August
<b>Funding</b>	100% covered by University of Geneva
<b>Budget code(s)</b>	47557

<b>Number</b>	26
<b>Partner/group</b>	<b>University of Geneva</b>
<b>Project Title</b>	<b>Open Cosmic Ray Detector Data Standards</b>
<b>Description</b>	Over the last two years, a team at CERN under James Devine has built a first version of a low-cost, open source cosmic ray detector. The objective of this project is (1) to explore how data from this detector could be combined with data from other educational cosmic ray detector initiatives, to provide reliable cosmic ray detection on a planetary scale and (2) to understand how this data might inform efforts to track cloud coverage, in collaboration with experts from the CLOUD experiment.
<b>Supervisor(s)</b>	James Devine (EN-EL-ENP), Antti Onella (EP-DT-CO), Francois Grey (CCL)
<b>Student</b>	<b>Darko Lukic, Mr, Serbia</b>
<b>Date</b>	03 July to 11 August





<b>Funding</b>	100% covered by University of Geneva
<b>Budget code(s)</b>	47557

<b>Number</b>	27
<b>Partner/group</b>	IT-ST/openlab
<b>Project Title</b>	CERNBox
<b>Description</b>	<p>The IT-ST group at CERN runs and evaluates innovative cloud storage technologies for their application to big data problems in high-energy physics research. One of them is CERNBox, the CERN cloud synch and share platform which is build on top of ownCloud (open source software) and uses EOS, the CERN multi-Petabyte disk storage built from commodity hardware heavily used as well by LHC and non-LHC experiments.</p> <p>CERNBox provides an easy access to users' data, at anytime, from anywhere, on any device.</p> <p>In the scope of improving the current offers for mobile users the candidate will:</p> <ul style="list-style-type: none"> <li>• Successfully build the CERNBox application for Android and iOS</li> <li>• Develop bug-fixes and enhancement for the mobile applications</li> <li>• Evaluate alternative access methods for Windows phones</li> <li>• Investigate the integration of a test suite for mobile application</li> </ul> <p>The successful candidate will have knowledge of scripting languages (e.g. python/bash) as well as mobile programming languages for Android and iOS platform, experience with Linux and web services, and will enjoy working in a multinational team.</p> <p>The student will gain working experience with a large-scale data storage system and exposure to several other user+support-facing IT services; contributions will be running in production - project will include roll-out procedures/cycles and user feedback.</p>
<b>Supervisor(s)</b>	Luca Mascetti
<b>Student</b>	Khaled Abushammala (Mr), Palestine
<b>Date</b>	26 June to 15 September





<b>Funding</b>	
<b>Budget code(s)</b>	47507

<b>Number</b>	28
<b>Partner/group</b>	IT-CDA/openlab/UNOSAT
<b>Project Title</b>	<b>Developing Augmented Reality 3D Model of Ab Berek Landslide on Microsoft Hololens</b>
<b>Description</b>	CERN openlab and the IT-CDA Group in collaboration with UNOSAT are investigating novel technologies for facilitating satellite maps navigation and comparisons using augmented or virtual reality. The Microsoft HoloLens device provides a programmable environment to create such augmented reality applications. The objectives of this initial investigation are to build familiarity with the hardware and the programming environment, use, adapt, or develop a simple map manipulation application, understand the requirements for importing UNOSAT satellite maps into the application and implement a minimal set of manipulation functions able to demonstrate the potential of the approach. The ideal candidate will have knowledge of programming languages for the Windows Modern environment (preferably C++ or C#, and XAML), experience with Windows and/or mobile app development, experience with 3D modelling frameworks such as Unity, some knowledge of map technologies and formats.
<b>Supervisor(s)</b>	Bruno Silva De Sousa, Lars Bromley
<b>Student</b>	<b>Daniar Heri Kurniawan (Mr), Indonesia</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	openlab
<b>Budget code(s)</b>	47557





## IT Groups

<b>Number</b>	29
<b>Partner/group</b>	IT-CDA
<b>Project Title</b>	<b>URL Shortening Microservice</b>
<b>Description</b>	<p>Creation of a simple open source, easy-to-deploy, multi-domain URL shortening service. The service in question will take over from go.cern.ch and provide concise and printable URLs that redirect to full length URLs. The generated URLs can be linked from presentations and e-mails, as well as printed and shared on social media. It will provide a web UI as well as an http API which will allow easy integration by external applications like Indico.</p> <p>Training value: Web UI, python, SQL, large open-source software development</p>
<b>Supervisor(s)</b>	Pedro Ferreira, Bruno Silva De Sousa
<b>Student</b>	<b>Joshi Varun (Mr), India</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	50/50
<b>Budget code(s)</b>	47612/47507

<b>Number</b>	30
<b>Partner/group</b>	IT-CDA
<b>Project Title</b>	<b>Statistics visualisation for the CERN Document Server (CDS)</b>
<b>Description</b>	<p>The CERN Document Server (CDS) is going through a transition phase, in which the entire underlying software stack is being updated. One of the features highly requested by our users is statistics, either in the form of individual charts per records, or in the form of dashboards for collections. Taking advantage of the new framework used for building the new CDS (Invenio v3.0) and of the Elasticsearch cluster that is currently gathering CDS user statistics, the student will work on ways of visualising this data. The</p>





	<p>implementation will be deployed on the new version of CDS as an individual Invenio v3.0 module.</p> <p>Training value: Web UI, python, PostgreSQL, large open-source software development</p>
<b>Supervisor(s)</b>	Jose Benito Gonzalez Lopez
<b>Student</b>	<b>Ioannis Androulidakis (Mr), Greece</b>
<b>Date</b>	26 June to 25 August
<b>Funding</b>	50/50
<b>Budget code(s)</b>	47158/47507

<b>Number</b>	<b>31</b>
<b>Partner/group</b>	<b>IT-CDA</b>
<b>Project Title</b>	<b>Automatic editing tool for Virtual Visit recordings</b>
<b>Description</b>	<p>With the advent and success of virtual visits in the LHC experiments and soon CERN at large, automation is sought in all surrounding processes, including the production and publishing of visit recordings. The student will develop a system based on the ffmpeg open source tool which will automatically crop a given video file and add intro and outer slides to the video. This system should have a web interface exposed to the users to enter the parameters for editing and designating the source file.</p> <p>Training value: modern stack web development, ffmpeg, video editing, scripting automation</p>
<b>Supervisor(s)</b>	Marek Domaracky
<b>Student</b>	<b>Lucarelli Pierpaolo (Mr), Italy</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	50/50





<b>Budget code(s)</b>	47612/47507
-----------------------	-------------

<b>Number</b>	32
<b>Partner/group</b>	IT-CM
<b>Project Title</b>	<b>Quality Assurance Framework for Linux Releases Delivery</b>
<b>Description</b>	<p>We need to automate in a better way the Continuous Integration for Linux cloud images and base OS installation to cover a higher number of test cases.</p> <p>The project would be to investigate, test and select a product that would fit our needs building on CERN IT infrastructure and tools.</p> <p>Some open source solution exists and may need to be tailored :</p> <ul style="list-style-type: none"> <li>Automated testing for operating systems <a href="https://openga.opensuse.org/">https://openga.opensuse.org/</a> - <a href="https://github.com/os-autoinst/openQA">https://github.com/os-autoinst/openQA</a></li> <li>Beaker is open-source software for managing and automating labs of test computers. <a href="https://beaker-project.org/">https://beaker-project.org/</a></li> </ul> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>Good knowledge of Red Hat installation process (kickstart)</li> <li>Programming skills and scripting</li> <li>Good communication skill to get involved in the community</li> <li>Hardware skills a plus (bios upgrade, pxe).</li> </ul>
<b>Supervisor(s)</b>	Thomas Oulevey
<b>Student</b>	<b>Pawel Dybiec (Mr), Poland</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	IT-CM Group
<b>Budget code(s)</b>	47952-47507

<b>Number</b>	33
---------------	----





<b>Partner/group</b>	IT-CM
<b>Project Title</b>	<b>Analysis and Machine Learning on Logs of the Monitoring Infrastructure</b>
<b>Description</b>	<p>The CERN IT monitoring team handles every day millions of monitoring events from CERN data centres and WLCG sites. Today this data is provided to users through a number of different tools and dashboards.</p> <p>This project aims at exploring with practical examples new logging, data transport and deployment techniques in order to extend the current infrastructure and also apply such techniques to IT services and the monitoring service itself</p> <p>The candidate will work in a DevOps environment and acquire practical experience with modern logging and transport technologies, such as LogStash, Flume, Kafka. The candidate will be exposed the several infrastructure deployment and management technologies such as Docker containers, OpenStack virtualization and Puppet configuration.</p> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>• Linux (basic)</li> <li>• Python or Java</li> </ul>
<b>Supervisor(s)</b>	Borja Garrido Bear, Alberto Aimar
<b>Student</b>	<b>Mert Ozer (Mr), Turkey</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	IT-CM Group
<b>Budget code(s)</b>	47952/47507

<b>Number</b>	<b>34</b>
<b>Partner/group</b>	IT-CM
<b>Project Title</b>	<b>Database as a Service OpenStack investigation</b>
<b>Description</b>	The OpenStack project includes a self service database project (Trove) to support requesting a variety of different database types from a self service interface.







	<p>The project will be to investigate the deployment and configuration of the Trove component in CERN's OpenStack cloud and test the suitability for a production service.</p> <p>Desired Skills:</p> <ul style="list-style-type: none"> <li>• Linux</li> <li>• Programming skills and scripting</li> <li>• Good communication skill to get involved in the community</li> <li>• Open Source databases</li> </ul>
<b>Supervisor(s)</b>	Arne Wiebalck (IT-CM) and Ignacio Coterillo (IT-DB)
<b>Student</b>	<b>Thomas Hartland (Mr), UK</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	IT-CM Group
<b>Budget code(s)</b>	47804/47507

<b>Number</b>	35
<b>Partner/group</b>	IT-DI-LCG
<b>Project Title</b>	<b>APIs for Computing Resource Information Catalog (CRIC)</b>
<b>Description</b>	<p>The student will take part in the project aimed to architect, develop and deploy the WLCG Configuration system - Computing Resource Information Catalog (CRIC). CRIC will describe the WLCG distributed resources and their topology. From the implementation point of view, CRIC represents re-factored and extended reincarnation of ATLAS GRID Information System (AGIS) developed in the Django framework. Work will include</p> <ul style="list-style-type: none"> <li>- modeling of the topology concepts , like for example storage object with multiple protocols, quota nodes and permissions</li> <li>- implementation of the user access permission policies and federated identity integration</li> <li>- development of the UI provided for various user categories</li> </ul> <p>Required skills: django</p>
<b>Supervisor(s)</b>	Julia Andreeva, Alessandro Di Girolamo





<b>Student</b>	<b>Ravin Kohli (Mr), India</b>
<b>Date</b>	03 July to 01 September
<b>Funding</b>	(i.e 100% openlab +partner, or 50/50 or 100% group budget)
<b>Budget code(s)</b>	47541/47507

<b>Number</b>	36
<b>Partner/group</b>	<b>IT/DI/CSO</b>
<b>Project Title</b>	<b>Internet of Things (IoT) Security</b>
<b>Description</b>	CERN uses ADFS for its Authentication and Authorisation needs. This implementation of a SAML2 Identity Provider (IdP) does not offer the standard Enhanced Client or Proxy ECP Endpoint that is required to access some non-web services. For example, Fermilab is setting up a command line based access mechanism to their job submission that required a SAML token - any Identity Provider wishing to use this must be able to use the ECP protocol ( <a href="https://wiki.shibboleth.net/confluence/display/CONCEPT/ECP">https://wiki.shibboleth.net/confluence/display/CONCEPT/ECP</a> ).
<b>Supervisor(s)</b>	Hannah Short
<b>Student</b>	<b>Sharad Agarwal (Mr), India</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	(i.e 100% openlab +partner, or 50/50 or 100% group budget)
<b>Budget code(s)</b>	47611/47507

<b>Number</b>	37
<b>Partner/group</b>	<b>IT/DI/CSO</b>
<b>Project Title</b>	<b>John the Ripper - Parallel and Distributed</b>
<b>Description</b>	All CERN critical interactive systems (e.g. LXPlus, LXBatch, SWAN) are equipped with 'execlog/netlog' ( <a href="https://github.com/CERN-CERT/activity_klog">https://github.com/CERN-CERT/activity_klog</a> ). This technology is





	based on kernel modules (on the 'kprobe' technology) which means "tainting" the kernel and complex maintenance. Auditd is an upstream solution, which provides with audit logs for a large number of actions within the kernel. Auditd has not been considered so far mostly because its output is horrible and need to be processed (several lines have to be grouped and can be unordered, etc). Slack has released a tool that process auditd and produce usable data: <a href="https://github.com/slackhq/go-audit">https://github.com/slackhq/go-audit</a> . This project is to evaluate if go-audit can be a replacement of netlog/execlog (same features?) and if so, to write the puppet configuration and rpm spec files to deploy it at CERN
<b>Supervisor(s)</b>	Vincent Brillault
<b>Student</b>	<b>Konstantinos Dalianis (Mr), Greece</b>
<b>Date</b>	19 June to 18 August
<b>Funding</b>	(i.e 100% openlab +partner, or 50/50 or 100% group budget)
<b>Budget code(s)</b>	47611/47507

