Survey

HSF IRIS-HEP Training Challenge 2 -- Nov 1 2021

- Event on November 1
- Live notes of previous experiment
- Editing privileges might be restricted to people registered for the event afterwards

👉 How does this work ❓

- **How can I add information about my experiment ❓**
  - Please check if a table for your experiment already exists.
    - If it does: Feel free to edit and add more information.
    - If not: Please copy the template and start filling it out :)
- **Can I add an additional question that I am interested in ❓**
  - Please add it to the bottom of the template and to all other “clones” so that people get a chance to fill them out. Keep in mind that we don’t want to overcrowd the tables for now.
- **I read something that I do not understand. What can I do ❓**
  - Please leave a comment and ask for clarification. Everything in this google document should be understandable for outsiders.
- **Any more questions ❓**
  - Write to kilian.lieret@lmu.de

DUNE (Heidi Schellman)

See below

Belle II (Kilian Lieret,)

Name of the experiment:	Belle II

Size of the experiment (approx number of people total):	~1000
Approximate newcomers to be trained per year:	50-100
What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc. If you have publicly accessible links to training events add these as well	We have StarterKit workshops that happen 1-3 times a year. Since recently all material is hosted on Sphinx in the form of lessons similar to the style of software carpentry or the HSF training modules. Rather than duplicating information about the software, the lessons directly let students search in the normal documentation (which is “more realistic” and much better maintenance). The recent StarterKits emphasize guided self study with the addition of Q & A sessions and small group mentoring sessions.
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	Everything is on sphinx bundled together with the main software and the API documentation.
How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	Most code snippets are included from external files that are run as unit tests. The sphinx documentation is bundled together with the main software, so breaking API changes cannot be merged without updating the examples. Currently one convener/librarian of the package (kilian.lieret@posteo.de) .
What are the current training challenges? Where are you struggling?	• **Getting more people to contribute**. One of the problems with sphinx and bundling with the main software is that it is a bit more technically involved to contribute (cloning repositories, compiling core software, modifying rst files, previewing changes, creating a PR etc.) • Encouraging **students to be more active** in interactive sessions in events
Can we augment your events with the curriculum we have from HSF? Are you already using some of the modules listed there? What is missing? How would you picture a collaboration?	We are using the git, bash and python modules from the curriculum/sw carpentry. Probably soon also uproot module. Most wanted would be training on medium level python, an expanded git tutorial that covers branches and pandas. Getting a common plotting framework together with a tutorial on that would be a nice to have.
What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email	For any suggestions regarding training, write to current training group convener: Kilian Lieret kilian.lieret@posteo.de and/or to the software training mailing list: software-doc@belle2.org.

addresses. Individual email addresses are encouraged over mailing lists.	If you want to advertise your event, you can also write Kilian to forward them or (if you are sure they fit) directly send them to software@belle2.org
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	• Perhaps by making it part of the git tutorial to contribute to one of the lessons on github. The SW Carpentry git lesson explains how to make forks already • Dedicated hackathon afterwards?
Is there any training material from your side that might be suitable for other collaborations as well? Who to contact and what steps would we need to make this happen?	We are working on making our SSH tutorial suitable for other collaborations as well. github repo
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	Will add soon.

LHCb (name?)

Name of the experiment:	LHCb
Size of the experiment (approx number of people total):	1200
Approximate newcomers to be trained per year:	Typically ~80 new students attending Starterkit

What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc.	• LHCb hackathons have between 30 and 60 attendees from all backgrounds (students to senior), happen 4 times per year, with varying tutorials for performance oriented programming (git, C++, CUDA, experiment specific), • We also host a 2nd stage (Impact) intermediate software workshop similar to starterkit. Smaller scale 50% of starterkit. Our last activity is a short hackathon to engage new students with our software stack. We also ask for lessons to be written at this time to keep our tutorials (both starterkit and impactkit) up to date.
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	StarterKit lessons
How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	Annual review before the workshops are held by the organizers. Maintenance by Chris Burr.
What are the current training challenges? Where are you struggling?	How to train people for Upgrade 1 without fixed software version.
Can we augment your events with the curriculum we have from HSF? How would you picture a collaboration?	Starterkit is quite full atm and we would prefer to leave it as it is. Even though the introduction is pretty much bash+python+git and is there to be sure newcomers are on the same page, the most of the event is quite LHCb specific. However, the Impactkit might be a great place to collaborate with HSF.
What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email addresses. Individual email addresses are encouraged over mailing lists.	• dorothea.vom.bruch@cern.ch. (Software developments and methods working group of LHCb) • Contact david.anthony.friday@cern.ch (ImpactKit) • martha.hilton@cern.ch • valeriia.lukashenko@cern.ch (Starterkit 2021) • lukas.calefice@cern.ch (Starterkit 2021) • lhcb-training-documentation@cern.ch - an official email list to discuss training and documentation
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	Analysis essentials

Is there any training material from your side that might be suitable for other collaborations as well? Who to contact and what steps would we need to make this happen?	We have out own snakemake tutorial, which is currently based on snakemake 6.0, but again gets updated annualy: analysis-automation-snakemake
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	Course pre survey Post hackathon survey

DUNE (Claire, Heidi)

Name of the experiment:	DUNE
Size of the experiment (approx number of people total):	> 1000 collaborator and growing
Approximate newcomers to be trained per year:	around 60 registering at each tutorial x 2 tutorials / year ⇒ > 120 people
What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc.	Tutorials usually occur before each collaboration week (end of September, end of January, end of May). They have been ‘re-branded’ since January 2021. January version: one day, 4 hours. “DUNE Getting Started Tutorial”. May version: see below.
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	As of May 2021, we put the training public (good influence from our HSF friends). GitHub “Lesson” page using the Software Carpentry format (awesome). Interactive sessions spanning 2-3 days, see schedule on the first page here: https://dune.github.io/computing-training-basics/index.html is the current form of the tutorial. It covers basics of getting an account, submitting jobs and modifying code.

How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	after the tutorial we put the links of the videos (YouTube embedded). See e.g. https://dune.github.io/computing-training-basics/index.html Experts prepare and vet the materials and then iterate for the next time.
What are the current training challenges? Where are you struggling?	Keeping up with changes to the compute environment. Balancing detail with an overview. Extremely wide range of experience in participants, annoying security environment making large scale deployment of things like Jupyter notebooks problematic. . Lack of funding mechanism for commercial solutions. Lack of time to do it all.
Can we augment your events with the curriculum we have from HSF? How would you picture a collaboration?	We do refer to some of HSF modules (beginner lessons) to our newcomers. Unix shell and git are the main ones. Others that we suggest but don't use in formal training are cmake and docker and the language courses.
What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email addresses. Individual email addresses are encouraged over mailing lists.	The training are advertised as internal communication. Computing Consortium Lead: Heidi Schellman (Heidi.Schellman@oregonstate.edu) Team Training & Documentation: Claire David (cdavid@fnal.gov) , Michael Kirby (kirby@fnal.gov), David DeMuth (david.demuth@vcsu.edu) , Tom Junk (trk@fnal.gov), Ken Herner (kherner@fnal.gov) Steve Timm (timm@fnal.gov)
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	at DUNE we have a special mailing list for training provision that DUNE members can subscribe to. We also make extensive use of slack and have a training and a beginners channel.
Is there any training material from your side that might be suitable for other collaborations as well? Who to contact and what steps would we need to make this happen?	For now, our training is DUNE specific as it relates to the basics of the DUNE software suite. However we may organize additional tutorials on Machine Learning tools that could interest other collaboration. We will notify HSF in advance for timely advertisement.
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	There are some confidentiality issues (especially regarding the comments) so I just sent the survey questions and results per email to the address above. Kilian: We'll update a version with just the questions later

CMS (Sudhir Malik)

Name of the experiment:	CMS
Size of the experiment (approx number of people total):	5000
Approximate newcomers to be trained per year:	300
What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc.	(1)CMS Data Analysis School - one week long, 2-3 times per year globally - Physics Object and 1 Physics analysis per group, 70 students and 50 educators (2) HATS (Hands-on Advance Tutorials) - dedicated topics like GitHub, Machine Learning< Trigger etc throughout the year about 15-20 events Events in person and also remote participation https://lpc.fnal.gov/programs/schools-workshops/index.shtml
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	Most work is carried out at CMS linux cluster where everyone has accounts. We also use Notebooks, Gitlab and Github. Most links are internal. The link to CMS Workbook is here: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBook
How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	The material instructions are available via WorkBook twikis that can be used anytime irrespective of training. Collaborators are given EPR (Experimental Physics Responsibilities) credits (commonly called as service work for the experiment). Each collaborator has to contribute 4 months of service work. Best way to maintain this is to have repeated training on a topic. In case of issues users provide feedback via hypernews emails.
What are the current training challenges? Where are you struggling?	Not that I know of.
Can we augment your events with the curriculum we have from HSF? How would you picture a collaboration?	CMS software and analysis training has already been a role model for other experiments to follow. We already are collaborating with HSF.

What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email addresses. Individual email addresses are encouraged over mailing lists.	Cms.Secretariat@cern.ch, malik@fnal.gov, Gabriele.Benelli@cern.ch tonjes@fnal.gov (Marguerite Tonjes)
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	At the moment we are not using the HSF material but have used its style (for example - https://hsf-training.github.io/hsf-training-cicd/ or carpentry style)
Is there any training material from your side that might be suitable for other collaborations as well? Who to contact and what steps would we need to make this happen?	Please have a look here: https://lpc.fnal.gov/programs/schools-workshops/hats.shtml Mostly it would how to adapt it experiment agnostic training
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	Feedback for each event is via google form that are internal to the collaboration.

Electron-Ion Collider (Markus Diefenthaler)

Name of the experiment:	Electron-Ion Collider (EIC)
Size of the experiment (approx number of people total):	The international community organized itself late in 2015 as the EIC Users Group (EICUG), which now consists of almost 1300 physicists at more than 250 institutions in 35 countries worldwide, working together to realize and utilize the EIC. The EICUG membership roughly doubled over the last four years, and growth is expected to continue with the EIC Project progressing. We estimate that the EIC user community at EIC science turn-on in the early 2030s will have doubled again and grown to above 2500 members.

Approximate newcomers to be trained per year:	150–200
What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc.	The Software Working Group (SWG) has organized tutorials, originally in-person and online and now only online. The organization of the tutorials have been ad-hoc, typically to prepare the EIC for initiatives, in particular the Yellow Report. The material has been created by developers / volunteers from the Software Working Group. In addition to the tutorials, we reached out to the working groups and provided guidance and help with software. The EICUG currently works on detector collaboration proposals (ATHENA, CORE, ECCE), each of them maintaining their own software stack and providing specific tutorials for the tutorials. For an EIC summer school, we provided for the first time combined tutorials by SWG and the ATHENA and ECCE proposals and will continue them beginning of the year.
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	We have used Docker and singularity containers for the tutorials and provided recordings of some of the tutorials on the YouTube channel of the EICUG. For some of the tutorials we have take advantage of Jupyter notebooks on Binder or Google Colab, e.g.: https://github.com/eic/python-analysis-bootcamp
How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	The EIC project is at an early stage and the software is evolving as rapidly as the community and the project itself. There is at this moment neither a common set of tutorials nor maintainers. We are in the process of changing this.
What are the current training challenges? Where are you struggling?	The EIC project is at an early stage and the software is evolving as rapidly as the community and the project itself. There is a diverse software set with only a few standards in the community. We can at this moment not even predict if one of two collaborations will form next years.
Can we augment your events with the curriculum we have from HSF? How would you picture a collaboration?	We would like to collaborate with the HSF. The basic courses of your curriculum are very helpful and we would like to take advantage of it instead of developing our own basic courses. A lot of new students start in the summer and we would like to provide intensive software & computing tutorials when the students “arrive”. Ideally, this event would be organized annually in the week after Memorial Day (and the week after the July 4 long weekend in 2021). Recognizing the geographic distribution of the EICUG, the tutorials will always be virtual, but labs may provide a satellite location with assistant instructors. We hope to organize the tutorials jointly with the HSF.

What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email addresses. Individual email addresses are encouraged over mailing lists.	You can reach the SWG conveners via email: Andrea Bressan (andrea.bressan@ts.infn.it), Markus Diefenthaler (mdiefent@jlab.org), and Torre Wenaus (wenaus@gmail.com). We can distribute announcements via the mailing list of the EICUG and the EICUG SWG.
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	Surveys, directly after the tutorials, appear to be a good way for collecting feedback and suggestions.
Is there any training material from your side that might be suitable for other collaborations as well? Who to contact and what steps would we need to make this happen?	Not yet but this will hopefully change over the future, e.g., for our Geant4 tutorials.
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	We have conducted a state of software survey (https://eic.github.io/activities/ucd.html) and will repeat the survey at the end of this year / early next year.

ATLAS (Jason Veatch)

Name of the experiment:	ATLAS
Size of the experiment (approx number of people total):	~3000

Approximate newcomers to be trained per year:	300-400
What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc.	Induction Day - full day of talks (in-person or remote depending on Covid) covering all aspects of the collaboration and research Software Tutorial - 4-5 days of lectures from experts, discussion, and hands-on tutorials. Covers core ATLAS software for analysis, grid tools, and statistical tools.
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	Various documents on internal twikis
How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	We have quarterly tutorials and we make sure to test the tutorial material in advance of each one
What are the current training challenges? Where are you struggling?	Participant retention, relevance to specific analysis There are many variations of tools available (e.g., statistical analysis tools). We have trouble balancing giving time to each variation while being able to dive deeply enough into the tools to be useful.
Can we augment your events with the curriculum we have from HSF? How would you picture a collaboration?	
What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email addresses. Individual email addresses are encouraged over mailing lists.	
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	
Is there any training material from your side that might be suitable for other collaborations as well?	

Who to contact and what steps would we need to make this happen?	
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	

Template (Add your name here…)

Name of the experiment:	
Size of the experiment (approx number of people total):	
Approximate newcomers to be trained per year:	
What are the existing training/induction activities? Please describe the form of the event, how it is organized, how the training material is created, etc.	
In what form are training materials available for self guided study (e.g. “Jupyter notebooks”, “lessons on sphinx”, …). Please enter URLs if publicly available.	
How is the correctness and maintenance of these materials ensured? E.g. are there specific maintainers, are there unit tests, etc.	

What are the current training challenges? Where are you struggling?	
Can we augment your events with the curriculum we have from HSF? How would you picture a collaboration?	
What are the possible administrative connections? E.g. How do we distribute information about training opportunities? Please give names and email addresses. Individual email addresses are encouraged over mailing lists.	
If the existing HSF material is being used in experiment-organized events, how can we encourage (and by what mechanisms) feedback or improvements (PRs)?	
Is there any training material from your side that might be suitable for other collaborations as well? Who to contact and what steps would we need to make this happen?	
We are trying to standardize surveys to merge information of different events and collaborations. What surveys are conducted in order to measure the efficacy of the training? Ideally please provide a link to a full set of your questions (e.g. via URL to google form, or link to a PDF version/printout of your form etc.)	