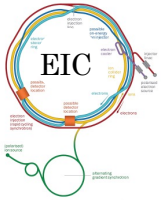
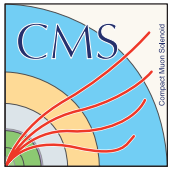
Training and onboarding initiatives in HEP





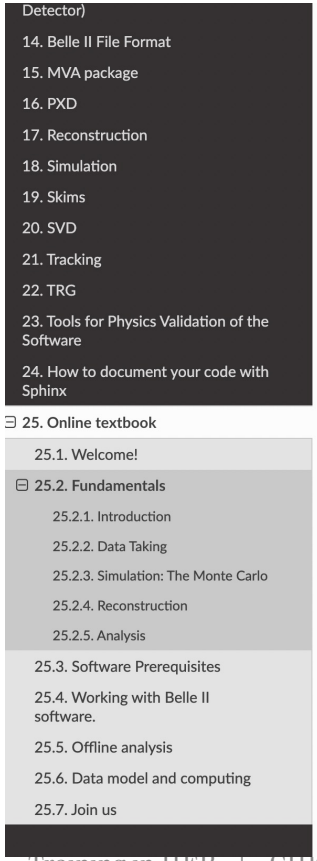
Allison Reinsvold Hall	(US Naval Academy)
Gabriele Benelli	(Brown University)
Ben Carlson	(Westmont College)
Claire David	(York University/FNAL)
Jonathan Davies	(U. of Manchester)
Wouter Deconinck	(U. of Manitoba)
Stephan Hageboeck	(CERN)
Kilian Lieret	(Princeton University)
Lera Lukashenko	(Nikhef)
Sudhir Malik	(U. of Puerto Rico)
Andy Morris	(Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France)
Heidi Schellman	(Oregon State University)
Nicole Skidmore	(U. of Manchester)
Jason Veatch	(Cal State, East Bay)
Michel Hernandez Villanueva	(DESY)

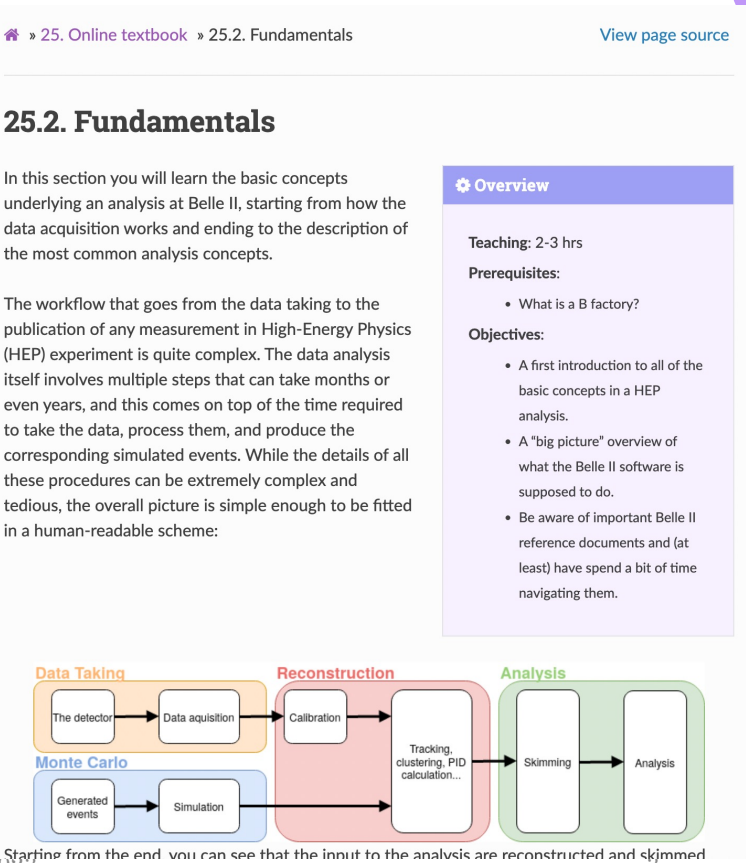
How do experiments teach analysis software?

- The software is different, but the **goals** and **challenges** are common
  - Train new collaboration members on analysis software tools to enable the physics goals of each experiment
- HSF Data Analysis Working Group (DAWG) and HSF Training Group hosted a discussion in summer 2022
  - Presentations from CMS, ATLAS, LHCb, Belle II, EIC, DUNE
  - Opportunity to learn from each other and share experiences
- Here focusing on experiment-specific training. See talk tomorrow for experiment-independent training initiatives in HSF

Belle II online textbook

- Focus on **self-study** as the primary training mode: https://training.belle2.org/
- Comprehensive onboarding experience, covering basic physics knowledge and collaborative tools to software prerequisites and grid job submissions





Belle II online textbook

- Hosted directly in the Belle II software framework repository, rendered via Sphinx
  - Plain text files in rst format
- Training materials are linked to relevant parts of the technical documentation (API documentation)
  - Teaching newcomers to use these resources is an objective in itself
  - Points intermediate learners to in-depth resources without sacrificing clarity
- **Unit testing** keeps materials up-to-date
  - No pull request can be merged without updating the relevant training material

LHCb StarterKit

- An introductory workshop aimed at young PhDs and advanced masters
  - No prior knowledge needed
  - 2 days: bash, gitlab and python (incl PyHEP)
  - 3 days : LHCb software
- Emphasizes **peer-to-peer instruction**: facilitators are mainly PhDs and young postdocs
  - Easier to ask questions to peers rather than senior collaboration members
  - Needs of newcomers are well-known to PhD students
  - Networking to improve the cohesion of early-career scientists within the collaboration
  - Fresh eyes each year helps keep materials up-to-date and relevant
- Pre-covid: 100 participants, in-person, including social events
- Post-covid: hybrid format is a hit, but networking is difficult for online-only participants

LHCb additional training resources

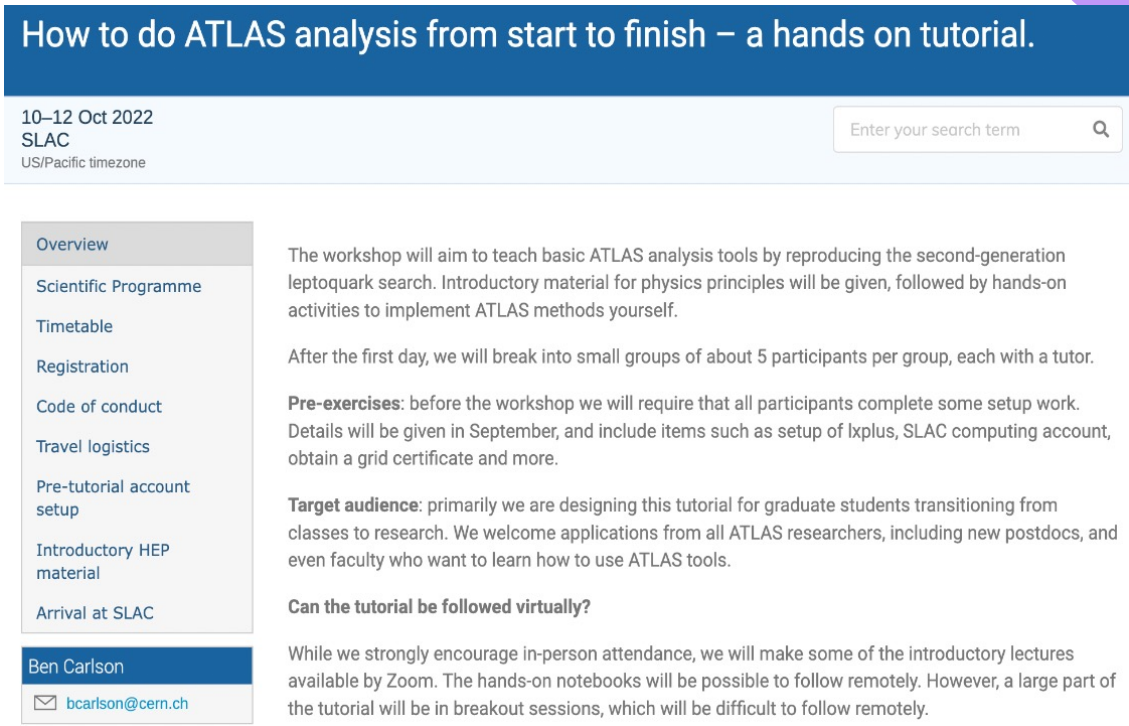
- ImpactKit: More advanced training, building on the Starterkit
  - 4 days: trigger lines development, local and custom MC, etc
  - Last held in 2021 + series of online talks in 2020
- Standalone lectures:
  - LHCb-UK student talks: explain a single aspect of the LHCb machine or LHCb analyses
  - StarterTalk: hour-long lecture on LHCb topics during LHCb collaboration week
- LHC Glossary: community-driven effort to define jargon and acronyms

ATLAS software tutorials

- Offered 3-4 times per year, coupled to the ATLAS Induction Day
- Targeting new PhD students, postdocs, and faculty
- Covers primary software necessary for data analysis
  - Setting up CERN computing accounts, grid credentials, etc.
  - Core analysis software framework, data formats, analysis workflows, tools for finding and accessing data, statistical tools
- Post-covid: moving to a **hybrid format**
  - Live lectures for those attending in-person or synchronously via Zoom
  - Videos/slides available for anyone attending asynchronously
  - Combined live Q&A session at 15:00 CEST
  - Hands-on exercise help conducted in-person and on Discord

ATLAS software tutorials: new format!

- **New tutorial design,** first tested in October 2022
- Interactive, hands-on and project-based structure with the aim of conducting end-to-end physics analysis
  - Working in teams (4-6 people) to perform a search for second-generation leptoquarks from start to finish
- See dedicated poster presentation



CMS Data Analysis Schools (DAS)

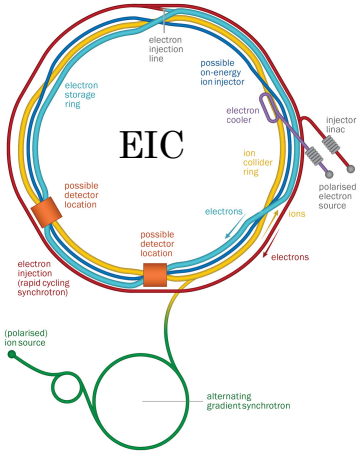
Weeklong schools held 2-3x per year since 2011, with ~50-70 students, 50 facilitators

Emphasizes hands-on exercises and networking

- **Pre-exercises:** mandatory setup before start of the school
- **Lectures:** introduction to LHC, CMS physics, detector, software/analysis etc.
- **Short Exercises:** about physics objects and basic analysis ingredients
- **Long Exercises:** students are assigned to teams and go through a full physics analysis
- **Mini-Symposium:** presentation of results from all teams

CMS Hands-on Tutorial Sessions (HATS)

- 15-20 per year since 2013, generally hybrid with video recordings and transcripts
- Focused tutorials:
  - Specific physics objects (jets, MET, electrons, etc.)
  - Tools for data analysis (Git, uproot, ML, etc.)
- Offered at Fermilab LPC (LHC Physics Center) in Spring/ Summer, but is a useful reference year-round
- Shifting to Carpentries format to simplify maintenance



EIC software development

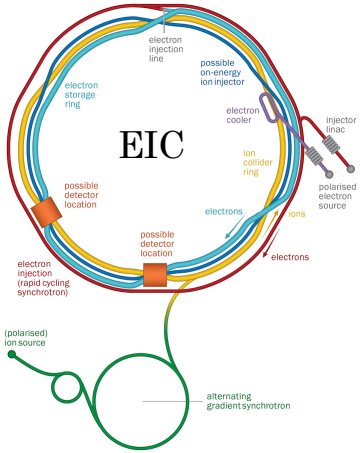
- Ongoing initiatives to organize EIC collaboration around a single software stack, focusing on **user-centered design**
- See EIC overview talk this afternoon!

User-Centered Design:
- State of Software Survey
- Follow-up Focus Groups
- Develop Testing Community

Discoverable Software:
- Single Point of Entry (~ key4hep)
- Feasible Option for >80% of EIC Simulations and Analyses

Workflows:
- Template Repositories for Key Analyses
- Template Repositories and Validation Workflows

Data and Analysis Preservation:
- User Analysis Code/Software Registry
- Tutorials on Reproducible Analyses



EIC software training

- Online interactive tutorials (recorded) at start of detector proposal development, ~30 users per tutorial per collaboration
- Online **interactive sessions** at the annual Center for Frontiers in Nuclear Science (CFNS) Summer School on the Physics of the Electron-Ion Collider, ~30 users per year
- Regular **office hours** (up to 3 times per week): inexpensive way to support and train users, get early feedback, and do an hour of work if no one shows
- Also developing **topical guides** for intermediate users to explore additional functionality

DUNE

- Currently running prototype detectors, with the full detector operation starting in 2029
- New users need to know how to run and modify the reconstruction and simulation algorithms, not just run final sample analysis
- Synchronous tutorials are offered 2-3 times per year
  - Using HSF Software Carpentries format
  - Serves a secondary role of allowing new members to form connections
  - Dedicated presentation for more details
- Need to expand infrastructure for advanced training in physics specific topics

Common challenges and varied solutions

- Balancing synchronous vs asynchronous training materials
  - Hybrid or self-study or in-person
- Maintaining up-to-date training materials
  - Unit tests or regular updates before tutorials
- Advertising documentation so it's clear where to look for answers to more advanced questions
  - Sphinx vs ReadTheDocs vs Doxygen vs Twikis
- Designing training materials *while* developing the software stack
- Keeping the training fun and engaging
  - Social events, teamwork, lots of hints and partial solutions

Lots to be learned from each other!

Thank you

Come join our discussions!

HSF Data Analysis working group: https://hepsoftwarefoundation.org/workinggroups/dataanalysis.html

HSF Training working group: https://hepsoftwarefoundation.org/workinggroups/training.html

See the dedicated talk tomorrow!