

A Worldwide Software Collaboration?

David Lange Princeton University

Is a (dictated) worldwide software collaboration possible? Or even desirable?

2

- Discussion and planning of a common software stack for LHC was prevalent 5-10 years before LHC started taking data
- In the end, the Run 1 LHC experiments shared software tools, sometimes software builds, but operated independently.
- A lot has changed during the last decade of LHC running. Many of the "common projects" haven't stood the test of time, while other solutions have been adopted by multiple experiments after succeeding in one

When there is no common tool – instead an emphasis on sharing findings during design, development and evaluation

- Helps to identify commonalities avoid falling into the same traps as someone else even if the details of your implementations differ
- Learn about existing tools that solve the problem you have
- Get to know experts that can help your community
- At the same time contributions from scientific developers outside of HEP are more relevant than ever.

Worldwide information flow – a.k.a., collaboration







- Not about "software development for the sake of software development"
- Rather development needed to answer important questions [detector, technical, physics, etc]
- This takes your time (and therefore "in kind" money) to make happen

The "HEP Software Foundation" (HSF)



The HEP Software Foundation facilitates cooperation and common efforts in High Energy Physics software and computing internationally.

- The HSF (<u>http://hepsoftwarefoundation.org</u>) was created in early 2015 as a means for organizing our community to address the software challenges of future projects such as the HL-HLC. The HSF has the following objectives:
 - Catalyze new common projects
 - Promote commonality and collaboration in new developments to make the most of limited resources
 - Provide a framework for attracting effort and support to S&C common projects (new resources!)
 - Provide a structure to set priorities and goals for the work

Global R&D roadmap (For HL-LHC)

Community charge from the Worldwide LHC Computing Grid in July 2016:

- Anticipate a "software upgrade" in preparation for the HL-LHC
- Identify and prioritize the software research and developments investments
 - to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies
 - to enable new approaches to computing and software that could radically extend the physics reach of the detectors
 - to ensure the long term sustainability of the software through the lifetime of the HL-LHC
- In the US, NSF funded the S2I2-HEP Conceptualization Project (s2i2-hep.org/) in July 2016

Projections of HL-LHC compute needs were main motivating factor.

They suggest that the current approach doesn't scale



All CWP and S2I2 Workshops

- 26-27 Apr, 2018 Reconstruction, Trigger, and Machine Learning for the HL-LHC
 - Massachusetts Institute of Technology, Boston
- 26-29 Mar, 2018 Joint WLCG/HSF Workshop 2018
 - Centro Congressi Federico II, Naples, Italy
- 14 Dec, 2017 Mini-workshop on Building Collaborations for ML in HEP
 - Massachusetts Institute of Technology, Boston
- 28-29 Nov, 2017 S2I2/DOE-lab mini-workshop on HL-LHC Software and Computing R&D
 - Catholic University of America, Washington DC
- 16-17 Nov, 2017 Data Organisation, Management and Access (DOMA) in Astronomy, Genomics and High Energy Physics
 - Flatiron Institute (Simons Foundation), New York City
- 23-26 Aug, 2017 S2I2-HEP Workshop
 - University of Washington, Seattle
- 26-30 Jun, 2017 HEP Software Foundation Workshop
 - LAPP (Annecy)
- 5-6 Jun, 2017 CWP Event Processing Frameworks Workshop

 FNAL
- 22-24 May, 2017 HEP Analysis Ecosystem Retreat
 - Amsterdam

0

- 8-12 May, 2017 DS@HEP 2017 (Data Science in High Energy Physics)
 - FNAL
- 1-3 May, 2017 2nd S2I2 HEP/CS Workshop
 - Princeton University
- 28-30 Mar, 2017 CWP Visualization Workshop
 - CERN (and Vidyo)
- 23 Mar, 2017 Community White Paper Follow-up at FNAL
 FNAL
- 20-22 Mar, 2017 IML Topical Machine Learning Workshop
 - CERN, The workshop includes a CWP session on Machine Learning
- 9 Mar, 2017 Software Triggers and Event Reconstruction WG meeting
 - LAL/Orsay, session at Connecting The Dots workshop
- 8 Mar, 2017 S2I2-HEP/OSG/USCMS/USAtlas Panel at OSG All Hands Meeting
 - SDSC/UCSD
- 23-26 Jan, 2017 HEP Software Foundation Workshop
 - University of California at San Diego / San Diego Supercomputer Center
- 7-9 Dec, 2016 S2I2 HEP/CS Workshop
 - University of Illinois at Urbana-Champaign



Community White Paper

Now published in "Computing and Software for Big Science" <u>https://link.springer.com/article/10.1007/s41781-018-0018-8</u>

A Roadmap for HEP Software and Computing R&D for the 2020s

HEP Software Foundation¹

ABSTRACT: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

Contents 1 Introduction 2 2 Software and Computing Challenges 5 3 Programme of Work 11 3.1 Physics Generators 11 3.2 Detector Simulation 15 3.3 Software Trigger and Event Reconstruction 23 3.4 Data Analysis and Interpretation 27 3.5 Machine Learning 32Data Organisation, Management and Access 36 3.7 Facilities and Distributed Computing 41 Data-Flow Processing Framework 45 3.8 3.9 Conditions Data 48 51 3.10 Visualisation 3.11 Software Development, Deployment, Validation and Verification 53 3.12 Data and Software Preservation 58 3.13 Security 61 4 Training and Careers 65 4.1 Training Challenges 66 4.2 Possible Directions for Training 67 4.3 Career Support and Recognition 68 5 Conclusions 69 Appendix A List of Workshops 72Appendix B Glossary 74 References 80

S2I2 and US plans from CWP

- Development of **innovative algorithms** for data reconstruction and triggering;
- Development of highly performant analysis systems that reduce "time-toinsight" and maximize the HL-LHC physics potential; and
- Development of data organization, management and access systems for the community's upcoming Exabyte era.
- Integration of software and scalability for use by the LHC community on the Open Science Grid, the Distributed High Throughput Computing infrastructure in the U.S.

Intellectual Hub for the HEP Community





The plan for IRIS-HEP reflects a community vision developed by an international community process organized by the HEP Software Foundation (https://hepsoftwarefoundation.org). The S2I2-HEP conceptualization project (http://s2i2-hep.org) derived a Strategic Plan from the community roadmap which would leverage the strengths of the U.S. university community. IRIS-HEP aims to function as an intellectual hub for the national and international HEP community, through training, community workshops and the development of wider collaborations with the larger computer and data science communities.

IRIS-HEP Fellows Program

- Key Insight: we need to provide incentivized and explicit paths forward for enthusiastic students from the more advanced training schools (ESC/Bertinoro, CoDaS-HEP, MLHEP, etc.) or for people who become engaged with our software projects in other ways.
- **Project focused:** bring students into contact with "mentors" to work on a specific, predefined project, allowing them to grow their software skills and project experience. The fellow supports, when possible, travel and subsistence for a 3 month extended stays in the mentor's institution.

IRIS-HEP Fellows

Jun-Aug 2019



Jan-Mar 2020

Jun-Sep 2019

Linking IRIS-HEP to Future Collider Community

• While IRIS-HEP's mandate is HL-LHC, its goals include "outreach" beyond that community.

• We have found 2 potential pathways that match near-term interests

- Snowmass 2020 planning process
- HEP software training
- This idea is still at the proposal stage (unfortunately... but talk to me if you want to help with that:))

"Snowmass" is the US process for assessing future HEP community priorities

- Snowmass2013: Five intertwined scientific drivers were distilled out of the results of a yearlong community driven process
 - 1. Use the Higgs boson as a new tool for discovery
 - 2. Pursue the physics associated with neutrino mass
 - 3. Identify the new physics of dark matter
 - 4. Understand cosmic acceleration: dark energy and inflation
 - 5. Explore the unknown: new particles, interactions, and physical principles
- Snowmass2013 recognized the need for a global community The worldwide particle physics community can together address the full breadth of the field's most urgent questions by partnering to work on global facilities

2-year process to define science goals and then priorities



International involvement in Snowmass is important

- As this workshop illustrates, future collider R&D is global. Needed expertise is widespread in important areas such as
 - Generators
 - Geometry
 - Tracking
 - Simulation
 -

 We hope to facilitate this – just as US researchers were active participants in the EU Strategy discussions



- Software skills are integral to being a successful experimental HEP physicist, and are key transferable skills for career evolution for people trained in HEP.
- The software/computing systems are key subsystems of our experiments, and arguably the lifetime costs are larger than many others. Maximizing science from the hardware investments relies critically on software. Training came up during the CWP process over and over.
- As a field have medium/long term challenges both from planned facilities and from technology evolution. But this also means we can take a long term view to how we develop software capabilities.



Not all funding agencies, institutions and funded projects have the same priority for training and education (e.g. DOE vs NSF in the US) relative to other goals like building/operating experiments, physics analysis, etc. Training activities are not always valued relative to other activities in making career steps.



Despite this many individuals do get enthusiastic about training others, but often only in specific career phases and often as a side "hobby" project. How do the activities then scale up? How are they sustainable?

Training challenges



Technology evolution means that training materials need to evolve, too. Separating "local" specifics (e.g. computing environments or experiment-specific bits) from generally usable material is important, but doesn't always happen.



Are training materials a common good or an individual product? Even if individuals *do* want to contribute to a common good, how do they do so?

Can we build a sustainable and scalable training framework that grows skills through multiple stages of people's careers?



CERN Software Carpentries Event - 27-29 Nov 2019

Software Carpentry at CERN

27-29 November 2019 Europe/Zurich timezone



Overview

Timetable Registration

Participant List

Past training events

Photos

Contact us

Mainter Mainte

We are very excited to announce the Software Carpentry Workshop at CERN, organised through the HEP Software Foundation

Over three days we will cover the fundamentals of: Unix (e.g. shell, bash and scripting); Git and Github – how to version control your code

Python - fundamentals of using the Python language

Jupyter Notebooks and SWAN

Python for analysis – how to combine Python with ROOT to start analysing data (i.e. pyroot and uproot)

This training is almed at those who are new to HEP and want a fast-track to competency with software fundamentals, as well the non-expert self-taught who wish to ensure they do not have gaps in their knowledge.

The first two days are covered by The Carpentries. The third day will be taught by tutors expert in HEP software. Interactive hands-on sessions lead by the tutor will be supported by a number of helpers to ensure all participants are able to follow and understand the material.



Advanced schools on HEP software development



http://codas-hep.org

The CoDaS-HEP school aims to provide a broad introduction to these critical skills as well as an overview of applications High Energy Physics. Specific topics to be covered at the school include:

- Parallel Programming
- Big Data Tools and Techniques
- Machine Learning Technology and Methods
 Practical skills: performance evaluation, use of ait for version control

The program includes both lectures and practical hands-on exercises.



CoDaS-HEP 2019

INFN International School on

Architectures, Tools and Methodologies for Developing Efficient large Scale Scientific Computing Applications

GridKa School 2019 - The Art of Data

GridKa School 2019

CSC 2020



Main school 23 Aug - 5 Sep 2020

AGH University of Science and Technology and Institute of Nuclear Physics (IFJ PAN) Kraków | Poland

What are we proposing in context of future detectors?

- Enabling collaboration facilitating common tools when it's the right approach
- Support travel for useful workshops.
 - Topics determined "bottoms up" not "top down"
 - Formats like this one and/or ones with a more focused theme, or more hands on (hackathons are typically successful in the LHC community)
- Targeted software training programs rather than schools
 - Beginner and expert. We can build on beginner programs that are already happening at CERN and elsewhere.
 - Some clear candidates for topical training courses would be interested in your opinions