WLCG/HSF Workshop 2025

Report of Contributions

https://indico.cern.ch/e/wlcg-hsf-2025

Contribution ID: **1** Type: **not specified**

Test

Contribution ID: **2** Type: **Talk**

Marionette: Data Structure Description and Management for Heterogeneous Computing

Wednesday 7 May 2025 11:00 (20 minutes)

Marionette is a header-only C++ library that was designed to allow the description of arbitrary data structures that can work across heterogeneous compute devices and on the host, providing complete interoperability and convenient interfaces with no impact on runtime performance. This is achieved by decoupling the description of the data to be held from the way in which data will be stored, which enables the generation of all memory allocations, transfers and deallocations at compile-time without requiring the end user to write any sort of boilerplate code, achieving the same performance as the equivalent hand-written alternatives. Furthermore, an expressive and intuitive object-oriented interface can be offered on both host and device(s), especially since arbitrary functions can be added to the interface of both the individual objects and the entire collection of them, without any runtime performance penalty as everything is resolved at compile-time. This user-extensible interface also means that the behaviour of pre-existing data structures can be replicated, allowing for immediate porting of pre-existing code with only minor adjustments to the data types. Furthermore, since the user can override and further specialize data transfers, gradual porting to the new data structures with minimal performance loss becomes possible as efficient ways to convert to and from pre-existing structures can be provided. With a focus on flexibility, customisability and extensibility without compromising expressivity, convenience or ease of use, Marionette is designed to offer a general solution for expressing data structures, catering to a wide variety of use cases and levels of expertise, with little to no runtime impact.

Requested talk length

Author: DOS SANTOS FERNANDES, Nuno (Laboratory of Instrumentation and Experimental Particle Physics (PT))

Presenter: DOS SANTOS FERNANDES, Nuno (Laboratory of Instrumentation and Experimental Particle Physics (PT))

Session Classification: HSF

Track Classification: HSF: Recognition of Sustainable Software

Contribution ID: **6** Type: **Talk**

GWOSC and gravitational-wave data analysis training

Wednesday 7 May 2025 14:35 (20 minutes)

Since their discovery in 2015, gravitational waves have become a hot topic in physics research. Gravitational-wave data produced by the LVK Collaboration, formed by the LIGO, Virgo and KAGRA collaborations, become fully public after a grace period; combined with the relative simplicity of the data themselves (one time series of the main signal channel per each interferometer, plus some simple data quality information), this created a large community of professional scientists, students and even citizen scientists analyzing them.
The Gravitational-Wave Open Science Centre (gwosc.org), besides maintaining a site hosting the public data and relevant software, developed a large amount of training and teaching materials such as tutorials and documentation, and organises periodical GW Open Data Workshops, "crash courses"in GW data analysis. This contribution describes such materials and activities.

Requested talk length

15

Author: Dr BAGNASCO, Stefano (Istituto Nazionale di Fisica Nucleare, Torino)

Presenter: Dr BAGNASCO, Stefano (Istituto Nazionale di Fisica Nucleare, Torino)

Session Classification: HSF

Track Classification: HSF: Training

Contribution ID: **7** Type: **Talk**

Julia: Sustainability and Efficiency

Wednesday 7 May 2025 11:20 (30 minutes)

There a number of studies of the general energy efficiency of different programming languages, however relatively few look at HEP specific examples. Here we present examples comparing energy efficiency of different jet reconstruction codes in different languages: specifically C++, Julia and Python. We also study the evolution of efficiency over recent releases of Julia and Python.

We also discuss general aspects of sustainability of code and show how the Julia language and ecosystem helps developers to write and maintain modular, interoperable codes that reduce the code maintenance burden.

We show that Julia is an excellent language choice, combining outstanding energy efficiency and human productivity, helping sustainability in all the most meaningful senses.

Requested talk length

30

Authors: DOGLIONI, Caterina (The University of Manchester (GB)); STEWART, Graeme A (CERN); MATO VILA, Pere (CERN); SKIPSEY, Samuel Cadellin

Presenter: STEWART, Graeme A (CERN)

Session Classification: HSF

Track Classification: HSF: Recognition of Sustainable Software

Contribution ID: **8** Type: **Talk**

Understanding the scale of HL-LHC physics analyses

Tuesday 6 May 2025 14:15 (15 minutes)

The existing roadmaps and computing model plans from ATLAS and CMS for the HL-LHC area are primarily focused on the centralized aspect of computing: those steps that lead up to sets of files made available to physicists for analysis. The general approaches, resources used, and software frameworks for the area of "end-user physics analysis", which starts from those files, are much less clearly defined, understood, or prescribed.

In order to better understand these aspects, IRIS-HEP, jointly with ATLAS AMG, CMS CAT and HSF DAAA, is designing a survey to capture the computational requirements from physics analysis use cases. This contribution will show first results from the survey and discuss the broader context of the effort. Following this survey, we aim to identify a set of physics analyses and define benchmark scenarios to extrapolate the concrete computing requirements to the HL-LHC era. The described workflows will provide more clarity about the role of analysis facilities and the kinds of services they should make available. It will allow for quantitative evaluation of analysis models and be a first step towards identifying what to do about analysis use cases that do not fit into the space set out by the benchmark examples.

Requested talk length

20

Authors: HELD, Alexander (University of Wisconsin Madison (US)); SHADURA, Oksana (University of Nebraska Lincoln (US))

Presenters: HELD, Alexander (University of Wisconsin Madison (US)); SHADURA, Oksana (University of Nebraska Lincoln (US))

Session Classification: Plenary

Contribution ID: **10** Type: **not specified**

ROOT: Taking stock of the Run 3 experience towards ROOT7

Tuesday 6 May 2025 09:00 (30 minutes)

ROOT is a unified software package for the storage, processing, and analysis of scientific data: from its acquisition to the final visualization in the form of highly customizable, publication-ready plots. Successfully used by experiments and thousands of physicists, the ROOT Project is preparing its seventh release cycle, sustained by intense R&D activities.
In this contribution, after briefly reviewing the status of the project, we'll focus on the results harvested along the R&D activities conducted so far and how those will shape the future of ROOT. The areas on which we'll concentrate will mainly be three. Firstly, we discuss how hardware accelerators can be exploited by users with ROOT and how these new features are relevant at present and future Analysis Facilities. Then, we review the current development status of RNTuple and its adoption by experiments, illustrating both advancements in terms of performance but also usability, in C++ and Python. Finally, we'll concentrate on the Python-C++ integration, at a low level, and the recent advancements in the interoperability of ROOT and Scientific Python, at the user-facing level.

Requested talk length

20

Author: PIPARO, Danilo (CERN)

Presenter: PIPARO, Danilo (CERN)

Session Classification: HSF

Track Classification: HSF: Common Software and Software Projects

Contribution ID: **11** Type: **Talk**

Introduction to the Virtual Research Environment: an end-user perspective

Tuesday 6 May 2025 10:00 (20 minutes)

One of the objectives of the EOSC (European Open Science Cloud) Future Project was to integrate diverse analysis workflows from Cosmology, Astrophysics and High Energy Physics in a common framework. This led to the inception of the Virtual Research Environment (VRE) at CERN, a prototype platform supporting the goals of Dark Matter and Extreme Universe Science Projects in compliance with FAIR (Findable, Accessible, Interoperable, Reusable) data policies. The goal of the project was to highlight the synergies between different dark matter communities and experiments, by producing new scientific results as well as by making the necessary data and software tools fully available. The VRE makes use of a common authentication and authorisation infrastructure (AAI), and shares the different experimental data (ATLAS, Fermi-LAT, CTA, Darkside, Km3Net, Virgo, LOFAR) in a reliable distributed storage infrastructure via the ESCAPE Data Lake. The entry point of such a platform (for an end-user) is a jupyterhub instance deployed on top of a scalable Kubernetes infrastructure, providing an interactive graphical interface for researchers to access, analyse and share data. The data access and browsability is enabled through API calls to the high level data management and storage orchestration software (Rucio). The VRE aims to streamline the development of end-to-end physics workflows, granting researchers access to an infrastructure that contains easy-to-use physics analysis workflow from different experiments. In this contribution, I will provide an overview of the VRE, highlight its use cases as an analyser for implementing and reproducing experimental analyses on a REANA cluster, and showcase the successful integration of an ATLAS experimental analysis workflow into the VRE platform.

Requested talk length

Author: SINHA, Sukanya (The University of Manchester (GB))

Presenter: SINHA, Sukanya (The University of Manchester (GB))

Session Classification: HSF

Track Classification: HSF: Common Software and Software Projects

Contribution ID: **13** Type: **Talk**

The EVERSE/ESCAPE use cases: evaluating and improving the quality of HEP software

Wednesday 7 May 2025 09:00 (20 minutes)

This contribution will present how the EVERSE project interfaces with the European Open Science Clusters (ENVRI-FAIR for environmental sciences, Life Sciences RI, ESCAPE for Particle physics and astrophysics, PaNOSC for Photon and neutron science and SSHOC for social sciences and humanities) through use cases of software packages or infrastructures that are in current used by researchers.

It is through these use cases that EVERSE draws best practices from the different communities, and where the various elements of the software excellence framework are tested and implemented prior to their release to the wider community. In this talk, we will describe the three ESCAPE use cases, highlight the elements that these use cases have contributed to EVERSE so far, and outline the expected improvements and pathway to obtain them by the end of the project in 2027.

Requested talk length

10

Authors: DOGLIONI, Caterina (The University of Manchester (GB)); SMITH, James (The University of Manchester (GB)); SPARKS, Michael (BBC); FITSCHEN, Tobias (The University of Manchester (GB))

Presenters: SMITH, James (The University of Manchester (GB)); SPARKS, Michael (BBC); FITSCHEN, Tobias (The University of Manchester (GB))

Session Classification: HSF

Track Classification: HSF: Recognition of Sustainable Software

Contribution ID: **14** Type: **Talk**

The EVERSE Research Software Quality Toolkit

Wednesday 7 May 2025 09:20 (20 minutes)

The Research Software Quality Toolkit (RSQKit - https://everse.software/RSQKit/), developed by the EVERSE project, lists curated best practices in improving the quality of your research software. It is intended for use by researchers, research software engineers, as well as those running research infrastructures involving software or involved in research software-related policy and funding.

These practices are informed by software excellence and quality in the context of research; with a focus on FAIR software, Open Research, community development and software engineering practices at different tiers of research software (analysis scripts, prototype tools and research software infrastructure). RSQKit links to tools and resources which support best practices. It includes software quality dimensions and links to indicators and tasks to guide the usage of the best practice, as well as links to training resources and existing guides and materials.

This contribution will introduce RSQKit and its aims and architecture. It is presented in a way to attract and consider feedback from researchers who code from the particle physics community and from WLCG infrastructure experts.

Requested talk length

15

Authors: DOGLIONI, Caterina (The University of Manchester (GB)); STEWART, Graeme A (CERN); SPARKS, Michael (BBC); ROISER, Stefan (CERN)

Presenter: SPARKS, Michael (BBC)

Session Classification: HSF

Track Classification: HSF: Recognition of Sustainable Software

Contribution ID: **15** Type: **Talk**

The EVERSE training and recognition plan

Wednesday 7 May 2025 14:15 (20 minutes)

The EVERSE project aims to collect, enhance and curate training resources aligned with domain-specific practices, create a long-term training activity supported by community services and platforms and establish a framework for recognizing Trainers and RSEs.

This contribution will describe how EVERSE plans to collect and provide training, guidance and education to researchers, software developers, and other stakeholders in the research community to help them understand the importance of software and code quality, as well as how to apply established best practices and standards for assessing, verifying, and improving the quality of their software and code.
We will describe the registry of training initiatives compiled so far, and how this registry will be maintained and presented through the TESS infrastructure. We will also outline recent activities in terms of recognition of software activities and roles.

Our aim is to get feedback from the community at the workshop and to highlight missing aspects in training and recognition that we can bring back to the project.

Requested talk length

15

Authors: DOGLIONI, Caterina (The University of Manchester (GB)); STEWART, Graeme A (CERN); RIOJA, Kenneth Brian (IT-FTI); ROISER, Stefan (CERN)

Presenter: RIOJA, Kenneth Brian (IT-FTI)

Session Classification: HSF

Track Classification: HSF: Training

Contribution ID: **16** Type: **Talk**

Green software in HEP: benchmarks and studies on MC generators

Wednesday 7 May 2025 09:40 (20 minutes)

In this talk, we will describe the studies undertaken at the University of Manchester to estimate and improve the energy efficiency of computing hardware and software used by students and researchers.

The goal of these studies is to build an understanding of the environmental impact of paticle physics research focusing on two fronts:
1) the carbon cost of the hardware uses for high power computing hardware and the local computing cluster
2) the energy efficiency of data analysis software and machine learning models in "big data"-related scientific fields including as high-energy particle physics.

The focus of this contribution will be the energy efficiency of scientific software algorithms and MC generation packages, taking Herwig, ML data compression and top tagging algorithms as examples. We will discuss different tools and benchmarks and review their methodologies.

We will then describe our plans towards a lifecycle analysis for computing hardware, and work undergoing to estimate the power consumption of our local cluster more precisely.

Requested talk length

15

Authors: DOGLIONI, Caterina (The University of Manchester (GB)); SMITH, James (The University of Manchester (GB)); VILLAR, Luis (University of Manchester); FITSCHEN, Tobias (The University of Manchester (GB))

Presenters: VILLAR, Luis (University of Manchester); FITSCHEN, Tobias (The University of Manchester (GB))

Session Classification: HSF

Track Classification: HSF: Recognition of Sustainable Software

Contribution ID: **17** Type: **Talk**

Latest Developments in RooFit and Plans

Tuesday 6 May 2025 09:30 (30 minutes)

RooFit is a software package written in C++ for statistical data analysis that is part of ROOT. It is widely used in the High Energy Physics (HEP) community, with the most prominent users being the LHC collaborations. Recent RooFit development has focused on performance improvements and supporting new statistical analysis approaches to enable cutting-edge analyses, such as combined Higgs measurements with ATLAS or CMS. In this contribution, the development pillars that helped to achieve this goal are elaborated on. The first pillar is code optimization and refactoring to optimally use both CPU and GPU resources. Then, there is supporting Automatic Differentiation (AD) with Clad, a compiler plugin for Clang. Furthermore, RooFit now provides new Python interfaces to include ML models as likelihood surrogates, enabling Simulation-Based Inference (SBI). Finally, this contribution will also report on the development status of the Minuit2 library for numerical minimization since it is a key dependency of RooFit, and the two packages are developed hand-in-hand to implement performance-optimal statistical analysis workflows for HEP.

Requested talk length

30

Author: REMBSER, Jonas (CERN)

Presenter: REMBSER, Jonas (CERN)

Session Classification: HSF

Track Classification: HSF: Common Software and Software Projects

Contribution ID: **19** Type: **Talk**

Analysis at the HL-LHC: Data Delivery, ServiceX, and Addressing Our Analysis Challenges

As the HL-LHC era approaches, the scale and complexity of data present challenges for analysis workflows within ATLAS and other HL-LHC experiments. This contribution reports on recent developments in ServiceX, a cross experiment utility, and its role as a data delivery and transformation service within the analysis ecosystem. Designed to bridge the gap between centrally produced datasets and user-level analysis code, ServiceX now supports a broader range of input formats—including custom non-ROOT ATLAS data—and has seen significant improvements in reliability, scalability, and ease of use.

We highlight progress in several areas: improved bullet-proofing of the data transformation infrastructure to avoid errors at scale; new convenience utilities for introspecting datasets (e.g., listing available branches); and tighter integration with common ATLAS frameworks such as TopCPTools, enabling more standardized analysis environments. We also share results from recent scaling tests demonstrating performance at analysis-scale workloads, and discuss how ServiceX fits into the broader analysis software stack, helping to address long-standing challenges of reproducibility, data access, and user efficiency.

This talk will provide a view of the design of ServiceX, the technical and usability gains made in recent months, and what's coming next as we prepare for the HL-LHC's analysis demands.

Requested talk length

30

Authors: CORDEIRO OUDOT CHOI, Artur (University of Washington (US)); GALEWSKY, Benjamin (Univ. Illinois at Urbana Champaign (US)); WATTS, Gordon (University of Washington (US)); VUKOTIC, Ilija (University of Chicago (US)); CHOI, Kyungeon (University of Texas at Austin (US)); ONYISI, Peter (University of Texas at Austin (US)); JANUSIAK, Roger (University of Washington)

Presenter: WATTS, Gordon (University of Washington (US))

Session Classification: Plenary

Track Classification: Analysis at scale

Contribution ID: **21** Type: **Talk**

AdePT - Offloading electromagnetic showers in Geant4 simulations to GPU

Tuesday 6 May 2025 11:00 (20 minutes)

The Geant4 simulation throughput of LHC experiments is limited by increasing detector complexity in the high-luminosity phase. As high-performance computing shifts toward heterogeneous architectures such as GPUs, GPU-accelerated particle transport simulations offer a potential way to improve performance. Currently, only electromagnetic showers can be offloaded to GPUs, making an efficient CPU–GPU workflow essential. In this contribution, we present state-of-the-art detector simulations for LHC experiments using GPUs, outline the outstanding challenges, and discuss future directions.

Requested talk length

20

Authors: GHEATA, Andrei (CERN); MORGAN, Benjamin (University of Warwick); APOSTOLAKIS, John (CERN); HAHNFELD, Jonas (CERN & Goethe University Frankfurt); GONZALEZ CAMINERO, Juan (CERN); NOVAK, Mihaly (CERN); DIEDERICHS, Severin (CERN); HAGEBOECK, Stephan (CERN); POKORSKI, Witold (CERN)

Presenter: DIEDERICHS, Severin (CERN)

Session Classification: HSF

Track Classification: HSF: Common Software and Software Projects

Contribution ID: **22** Type: **Talk**

Accelerating HEP detector simulations using G4HepEm

Tuesday 6 May 2025 11:20 (20 minutes)

Geant4 based detector simulations make a significant contribution to the overall computing budget of the LHC experiments. The individual experiments have been investing considerable effort in making their simulations more and more efficient. These performance optimisations are now even more important in order to cope with the special computing challenges of the HL-LHC era.

G4HepEm is one of the R&D projects that have been launched with the goal of contributing to this effort. It provides an efficient simulation of the electromagnetic shower, tailored for HEP detector simulations, in the form of a Geant4 extension. A significant performance improvement (˜20 %) of the ATLAS and CMS full detector simulations has been achieved recently after integrating G4HepEm into the ATLAS Athena and CMS-SW frameworks while preserving the accuracy of the results. The motivations, ideas and results obtained for ATLAS and CMS will be presented.

Requested talk length

Author: NOVAK, Mihaly (CERN)

Co-authors: MORGAN, Benjamin (University of Warwick (GB)); HAHNFELD, Jonas (CERN & Goethe University Frankfurt)

Presenter: NOVAK, Mihaly (CERN)

Session Classification: HSF

Track Classification: HSF: Common Software and Software Projects

Contribution ID: **23** Type: **Talk**

Scikit-HEP project news and future directions

Wednesday 7 May 2025 10:00 (20 minutes)

Scikit-HEP is a community-driven and community-oriented project with the goal of providing an ecosystem for particle physics data analysis in Python fully integrated with the wider scientific Python ecosystem. The project provides many packages and a few "affiliated"packages for data analysis. It expands the typical Python data analysis tools for particle physicists, with packages spanning the spectrum from general scientific libraries for data manipulation to domain-specific libraries. An overview of where the project is will be presented. Future developments and matters of sustainability will be discussed.

Requested talk length

Author: RODRIGUES, Eduardo (University of Liverpool (GB))

Presenter: RODRIGUES, Eduardo (University of Liverpool (GB))

Session Classification: HSF

Track Classification: HSF: Recognition of Sustainable Software