PROCEEDINGS OF SCIENCE

Five years of Key4hep - Towards production readiness and beyond

---

Juan Miguel Carceller,[a] Wouter Deconinck,[h] Mateusz Fila,[a] Brieuc Francois,[a] Frank-Dieter Gaede,[b] Gerardo Ganis,[a] Benedikt Hegner,[a] Xingtao Huang,[g] Sylvester Joosten,[f] Sang Hyun Ko,[e] Weidong Li,[d] Teng Li,[g] Tao Lin,[d] Thomas Madlener,[b] Leonhard Reichenbach,[a,c] André Sailer,[a] Swathi Sasikumar,[a,*] Juraj Smiesko,[a] Alvaro Tolosa-Delgado,[a] Xiaomei Zhang[d] and Jiaheng Zou[d]

[a] *CERN, Geneva, Switzerland*
[b] *DESY, Hamburg, Germany*
[c] *University of Bonn, Germany*
[d] *IHEP, Beijing, China*
[e] *Seoul National University, Seoul, South Korea*
[f] *Argonne National Laboratory, Illinois, United States*
[g] *Shandong University, China*
[h] *University of Manitoba, Winnipeg, Manitoba, Canada*

E-mail: swathi.sasikumar@cern.ch

The Key4hep project aims at providing a complete software stack to enable complete and detailed detector studies for future experiments. It was first envisaged five years ago by members of the CEPC, CLIC, ILC and FCC communities and has since managed to attract contributions also from others, such as the EIC or the MuonCollider. Leveraging established community tools, as well as developing new solutions where necessary, the Key4hep software stack is reaching production readiness rapidly, and is already used for physics studies.

This presentation will give an overview of the status of the Key4hep project and the components that are developed within its context. We will also report on some key insights and experiences that we gained along the way, e.g. integrating communities and their existing tools into a coherent approach, or on our experiences with building and releasing the stack using spack. Finally, we briefly highlight currently ongoing developments and plans.

International Conference on High Energy Physics - ICHEP24,
17-24 July 2024
Prague Congress Centre, Prague, Czech Republic

---

[*] Speaker

For over two decades, tools for the generation of particle events, simulation, reconstruction, and analysis were developed for the International Linear Collider in `iLCSoft`. These tools have been rigorously tested and continuously optimised over the years, incorporating iterative feedback. Recently, there have been several propositions for future lepton colliders. However, the particle physics community struggles with the uncertainty of which lepton collider will be built next. These circumstances led to the development of a common turnkey software, Key4hep, for future colliders. Key4hep has a complete data processing framework from generation to data analysis. The components from different colliders are shared to reduce maintenance and development costs, allowing everyone to benefit from the improvements thus reducing the waste of human power. The Key4hep community is built with people from different future experiments e.g. FCC, ILC, CLIC, CEPC, EIC, Muon Collider, and others as illustrated in figure 1.

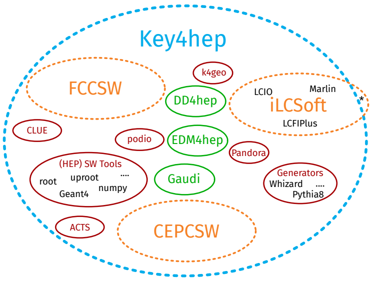


Figure 1: Schematic diagram showing Key4hep as a collaboration of different future collider communities.

1. An ensemble of state-of-the-art tools

Key4hep is based on state-of-the art tools, such as Gaudi [1] as the event processing framework, `DD4hep` [2] for geometry information, and `Podio` [3] to build its event data model `EDM4hep` [4].

1.1 The Key4hep Event Data Model

A well-defined Event Data Model (EDM) is one of the most important pieces of software in any high-energy physics processing framework. It defines the language used across the entire processing chain of HEP experiments, encompassing event generation, simulation, reconstruction, and analysis. `EDM4hep` is a common event data model for the Key4hep project shared within different future collider experiments. It is a shared EDM that is developing a common software stack for all future colliders and evolves through consensus among all stakeholders.

1.2 The `Podio` EDM toolkit

`Podio` (plain old data IO) is a toolkit for the creation of EDMs like `EDM4hep` [3]. Its main purpose is to have an efficiently implemented and thread-safe event data model. An EDM should be able to not only store data but also capture the relationships between different data types. Moreover, addressing all possible physics use cases across the diverse involved communities poses a significant challenge for `EDM4hep`. This is where `Podio` comes in. The key idea behind `Podio` is to start with a high-level description of the desired data types and their relationships, then automatically generate an efficient `C++` implementation of the EDM. This partial automatisation enables users to focus on designing better descriptions of the data models, free from implementation concerns [9]. Also to ensure that the new developments in the event data model do not break the existing workflows, the *schema evolution* [8] features of `Podio` are employed to ascertain backward compatibility with existing files.

2. Marlin Processor Wrapper

While Key4hep integrates tools and technologies from different future experiments, tools in `iLCSoft` developed for the ILC and CLIC exist in `MARLIN` framework. To integrate *processors* from `MARLIN` [7] framework and its corresponding data model `LCIO` [6], `k4MarlinWraper` [16] was created. The main idea of `k4MarlinWraper` is to convert the event data from `EDM4hep` to `LCIO` before executing a *processor* and then convert it back to `EDM4hep` after reconstruction. The idea of `k4MarlinWrapper` is illustrated in figure 2.

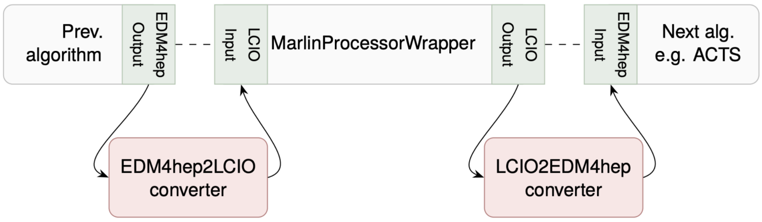


Figure 2: A schematic figure detailing the procedure of integration of `Marlin` processors into Gaudi workflows in Key4hep.

3. Recent developments in Key4hep

In addition to facilitating the sharing of existing experimental workflows, the Key4hep project prioritises the creation and integration of new software libraries that are independent of specific experiments. The ongoing studies such as `ACTS` [18], `CLUE` [10], PandoraPFA [11], and Open-DataDetector [12] demonstrate Key4hep's role as a flexible testing and development platform. Two of the recent developments in Key4hep are described in detail ahead.

3.1 Phoenix: a web-based visualisation tool

Phoenix is a web-based tool designed for the visualisation of detector data and events, enabling users to analyse events both independently of specific detectors and within the context of particular detector systems. Phoenix employs `JSROOT`, which facilitates the manipulation of `ROOT` files in a web environment. Here, the event data and detector descriptions are given using the methods in Key4hep. Events formatted in `EDM4hep` are initially stored in `ROOT` files and are then converted to `JSON` files without altering the `EDM4hep` data structure to be used by Phoenix. Detector descriptions are represented in `DD4hep` compact files formatted in XML. These XML files are converted into `ROOT` files to be utilised by `JSROOT`. Figure 3 shows the event visualisation for full simulation in the CLD detector.

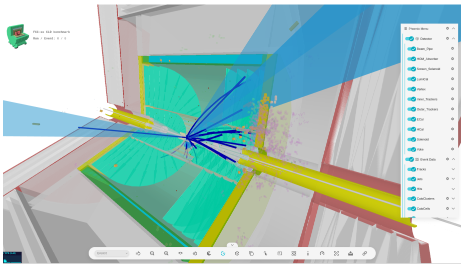
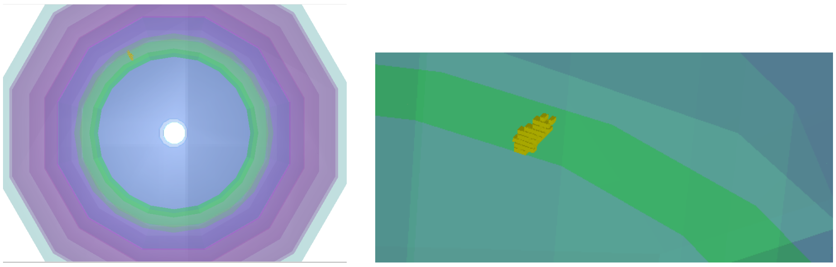


Figure 3: The event visualisation in Phoenix for full simulation of events in CLD detector.

3.2 Pandora Particle Flow in Key4hep

Several studies have established that particle flow reconstruction is one of the most effective methods for obtaining optimal jet energy resolutions in future Higgs factory experiments. Pandora [11] is a particle flow algorithm developed to study particle flow calorimetry for high-granularity sandwich calorimeters [17](HG-SiW-CAL) at linear colliders. However, PandoraPFA is an `iLCSoft` independent algorithm that was employed using a `MARLIN` interface named `DDMarlinPandora`. To integrate such a robust tool into Key4hep, it is important to make it available across detector models. For this, the PandoraPFA is tested for the first time on a non-HG-SiW-CAL calorimeter. A noble Liquid Argon (LAr) calorimeter has fundamentally a very different geometry as compared to an HG-SiW-CAL [14]. Pandora requires specific material properties, such as radiation length and interaction length, to accurately calculate the depth of a particle shower. These properties are provided through the `DD4hep` class, `Layered Calorimeter Data`. However, in a calorimeter with Pb/Steel absorbers and readouts immersed in liquid argon at a 50° angle, it becomes difficult to predict which section of the calorimeter a particle will first encounter. To address this, the `DD4hep` class `Material Manager` is employed. `Material Manager` computes the averaged material properties over a specific range within the detector, effectively evaluating the characteristics of the mixed materials. This approach allows for the dynamic determination of material properties, ensuring they remain applicable across different detector models, independent of their specific designs. With all relevant material properties provided to Pandora, 500 events of 10 GeV photons were simulated in the LAr calorimeter and subsequently reconstructed using the Pandora algorithm. As illustrated in figure 4, the successful reconstruction of a photon cluster



(a) **(b)**

Figure 4: (a) The event display for CLD-LAr with a photon reconstructed (yellow) in LAr (in green) towards the top mid-left of the detector. (b) Close-up version of the left event display.

within the LAr calorimeter highlights the effectiveness of this approach. More details of this study can be found in [13, 15].

4. Conclusion

Key4hep has evolved into a collaborative platform that brings together multiple projects, enabling not only the sharing of tools and software but also the joint development of new solutions. Recent developments within Key4hep include the integration of both existing and emerging tools, further strengthening its capabilities and versatility. Built upon well-established frameworks such as `Geant4`, `DD4hep`, and Gaudi, along with newer technologies like EDM4hep and Podio, Key4hep is advancing toward a comprehensive and robust framework tailored for future Higgs factory experiments.

5. Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under GA no 101004761. This work benefited from support by the CERN Strategic R&D Programme on Technologies for Future Experiments (https://cds.cern.ch/record/2649646/, CERN-OPEN-2018-006)

References

[1] Barrand, G. et al. GAUDI — A software architecture and framework for building HEP data processing applications. *Computer Physics Communications*. **140**, 45-55 (2001), https://www.sciencedirect.com/science/article/pii/S0010465501002545, CHEP2000

[2] Frank, M. et al. DD4hep: A Detector Description Toolkit for High Energy Physics Experiments. *Journal Of Physics: Conference Series*. **513**, 022010 (2014,6), https://dx.doi.org/10.1088/1742-6596/513/2/022010

[3] Gaede, F., Hegner, B. & Mato, P. PODIO: An Event-Data-Model Toolkit for High Energy Physics Experiments. *Journal Of Physics: Conference Series*. **898**, 072039 (2017,10), https://dx.doi.org/10.1088/1742-6596/898/7/072039

[4] Gaede, F. et al. EDM4hep and podio - The event data model of the Key4hep project and its implementation. *EPJ Web Conf.*. **251** pp. 03026 (2021), https://doi.org/10.1051/epjconf/202125103026

[5] Sailer, A. et al. The Key4hep software stack: Beyond Future Higgs factories. (2023), https://arxiv.org/abs/2312.08151

[6] Gaede, F., Behnke, T., Graf, N. & Johnson, T. LCIO - A persistency framework for linear collider simulation studies. (2003), https://arxiv.org/abs/physics/0306114

[7] Gaede, F. Marlin and LCCD—Software tools for the ILC. *Nuclear Instruments And Methods In Physics Research Section A: Accelerators, Spectrometers, Detectors And Associated Equipment*. **559**, 177-180 (2006), https://www.sciencedirect.com/science/article/pii/S0168900205022643, Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research

[8] Fernandez Declara, P. et al. Of Frames and schema evolution – The newest features of podio. (2023), https://arxiv.org/abs/2312.08199

[9] Gaede, F. et al. EDM4hep - a common event data model for HEP experiments. *PoS*. **ICHEP2022** pp. 1237 (2022), https://cds.cern.ch/record/2869508

[10] Brondolin, E. & CMS collaboration CLUE: a clustering algorithm for current and future experiments. *Journal Of Physics: Conference Series*. **2438**, 012074 (2023,2), https://dx.doi.org/10.1088/1742-6596/2438/1/012074

[11] Thomson, M. Particle flow calorimetry and the PandoraPFA algorithm. *Nuclear Instruments And Methods In Physics Research Section A: Accelerators, Spectrometers, Detectors And Associated Equipment*. **611**, 25-40 (2009,11), http://dx.doi.org/10.1016/j.nima.2009.09.009

[12] Allaire, C. et al. (2022). OpenDataDetector (Version v2). Zenodo. https://doi.org/10.5281/zenodo.6445359

[13] Sasikumar, S., Sailer, A., Francois, B. First Look at Particle Flow in a LAr Calorimeter Using Pandora in the Key4hep Framework. To be submitted to *POS ICHEP 2024*

[14] Aleksa, M. Calorimeters at FCCee. *The European Physical Journal Plus*. (2021)

[15] Key4hep Nobel Liquid ECal source file, `https://github.com/key4hep/k4geo/blob/559f41185a8f166519a392ff7cea3ef0add41b0b/detector/calorimeter/ECalBarrel_NobleLiquid_InclinedTrapezoids_o1_v01_geo.cpp#L564C1-L632C1`

[16] k4MarlinWrapper Repository, https://github.com/key4hep/k4MarlinWrapper

[17] Pöschl, R. The CALICE SiW ECAL Technological Prototype—Status and Outlook. *Instruments*. **6**, 75 (2022,11), http://dx.doi.org/10.3390/instruments6040075

[18] Ai, X. et al. A Common Tracking Software Project. *Computing And Software For Big Science*. **6** (2022,4), http://dx.doi.org/10.1007/s41781-021-00078-8