

# HEPMC Standards and the Path Forward

<http://hepmc.web.cern.ch/hepmc/>  
[hepmc-dev@cern.ch](mailto:hepmc-dev@cern.ch)

**Andrii Verbytskyi<sup>a</sup>, for the HepMC authors [1]**



GDR-QCD seminar: introduction to Rivet, HepMC and selected expertises,  
Saclay, July 01, 2021 (ZOOM)

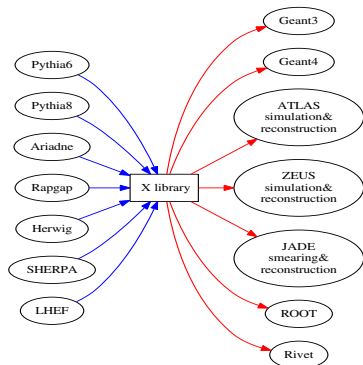
# What is HepMC3: Definition and application

- HepMC3 is a library designed to operate with Monte Carlo event records in High Energy physics (HEP), see Ref. [1].
- Event record contains all physical information on the initial, intermediate and final state particles in the simulated collision: particle flavours, momenta, production coordinates, etc.



- The library should be used to store and transfer Monte Carlo event records between different HEP software and/or disk.
- HEP software in focus: Monte Carlo event generators (e.g. Pythia8[2], Herwig7 [3], SHERPA [4]), simulation software (e.g. Geant4 [5], FLUKA [6]), analysis frameworks (ROOT [7], experiment-specific), plotting tools (e.g. Rivet [8], HZTOOL [9]) etc.

# What is HepMC3: Typical application example



- The typical task is to pass MC generated events between different generators, reconstruction programs and/or analysis frameworks.
- Before the HepMC3, the HepMC2 [10] partially performed these tasks.

# Motivation to update HepMC2

- **The communication with users was interrupted during the changes in the developers team and platform migrations in CERN.**
- Maintenance:
  - HepMC2 is not actively developed for a long time, last minor release is in **2012**, many users have to apply patches.
  - Many small changes would break compatibility but will not allow significant design improvements.
- Features:
  - Physics-related issues should be resolved to meet requirements of **modern physics**, e.g. Heavy Ion information [11], event weight treatment, etc.
  - I/O capabilities should be modernised, e.g. ROOT, LHEF [12], serialisation of custom information.

**HepMC3 is a natural successor of HepMC2. With an idea to keep what is good in HepMC2 and add more.**

## New in HepMC3 (vs. HepMC2)

- Consistent I/O for all popular formats of event records.
- The original LHEF routines by Leif Lönnblad.
- ROOT interface out of the box.
- Python2/Python3/PyPy interface out of the box.
- Extendable event record.
- Search engine to navigate in the event record.
- Multi-threading/thread safety.
- CMake instead of autotools.
- Automatic memory management with smart pointers.
- GitLab with CI and tickets.
- Test engine based on CTest.
- Extensive documentation and examples.
- Implemented interfaces in most modern MCEGs.

User	Type of SW	HepMC3 interface implementation
Athena@ATLAS	Main SW of experiment	3.2.3 in a separate branch
SHERPA-MC	MCEG	3.1+ in SHERPA-MC 2.2.5+
ThePEG	MCEG	3.1 in ThePEG 2.2.0
Herwig7	MCEG	3.1 via ThePEG
Pythia8	MCEG	3.0+ in Pythia8.3 and in HepMC3
Pythia6	MCEG	3.1+ in HepMC3 examples
Tauola	MCEG	3.1+ in Tauola 3.64 and in HepMC3
Photos	MCEG	3.1+ in Photos 1.1.8 and in HepMC3
WHIZARD	MCEG	3.1+ in Whizard 2.8.2+
EvtGen	MCEG	3.1+ in EvtGen 2.0.0+
GeantV	Simulation	3.0
ACTS	Tracking	3.2+ in multiple versions
Rappap	MCEG	3.1+, in a merge request to master
Cascade	MCEG	3.1+, in a merge request to master
Rivet	Analysis/Plotting	3.2.0 in Rivet 3.1.0+
MC-TESTER	Analysis/Plotting	3.1+ in MC-TESTER 1.25.1 and in HepMC3

HepMC3 3.1+ and HepMC2 can co-exist in one installation.

## Selected new features: Attributes in HepMC3

- An attribute is class that holds information on event, particle or vertex and can be represented as a string.
- Every attribute class should inherit from `HepMC3::Attribute`. The attribute is handled all the time as a string (in memory and in I/O operations) till an explicit request to convert it to an object of some type.

Example of attributes are given below

```
1 GenCrossSection 2.64422551e+03 2.64422551e+03 -1 -1
```

```
1 GenPdfInfo 11 -11 9.97420767e-01 9.99999975e-01 9.18812775e+01 1.56824725e+01 2.82148362e+06 0 0
```

```
alphaQCD 0.129844
```

These attributes represent information of classes

```
1 HepMC3::GenCrossSection  
2 HepMC3::GenPdfInfo  
3 double
```

**One can add arbitrary complex information to event record.**

## Selected new features: separation of disk-resident formats and HepMC3::GenEvent

- The ‘HepMC3::GenEvent’ objects in memory are constructed from the information on disk, but **“Events on disk”**  $\neq$  **“Events in memory”**.
- The same input file processed with different ‘HepMC3::Reader’ can produce different events. If e.g. only the stable particles in the final state are needed – no need to invent new formats. A custom reader will do the job on the standard files.



# Selected new features: custom readers/writers for different formats

## HepMC3 I/O is easily extendable for different purposes

- Simple example to write into dot format and use with GraphViz.

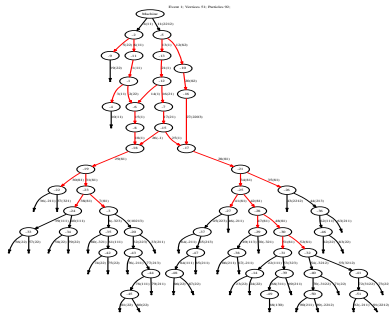


Figure:  $e^-p$  collision in Herwig 7.1.4

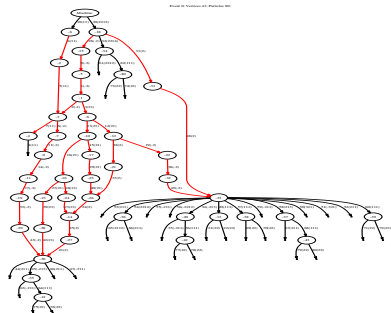


Figure:  $e^-p$  collision in Pythia 8.2.40

# Usage: the “classical” MCEGs and the simulations for old experiments

Because it is very easy to manipulate with different file formats in HepMC3, it is very easy to

- Interface a classical 20+ y.o. MCEG to modern tools.
- Interface a modern MCEG to 20+ y.o. detector simulation.

The typical workflow for personal QCD studies:

- Run MCEG with HepMC3/ROOT support and produce HepMC3/ROOTTree events.  $\approx 10^{7-8}$  events for  $e^+e^-/e^\pm p$  or  $10^{6-7}$  for  $pp$ . The ROOTTree output is compact and is stored locally or e.g. in dCache.
- Run ROOT analysis on the files (HepMC3 is NOT needed at this point) using PROOF. The input can be local or remote. With a 64 core machine, the typical QCD analysis runs at  $10^5$  events per second. I/O speed is not a problem. All the samples can be processed in  $\mathcal{O}(1)$  hours.
- Use the filled histograms.

# Usage: HepMC3 event analysis in ROOT (Headers omitted).

```
1 class SomeAnalysis{
2 public :
3     TChain          *fChain;    ///pointer to the analyzed TTree or TChain
4     Int_t           event_number;
5     Int_t           momentum_unit;
6     TBranch         *b_hepmc3_event_event_number;    ///!
7     TBranch         *b_hepmc3_event_momentum_unit;   ///!
8     TBranch         *b_hepmc3_event_length_unit;     ///!
9     TBranch         *b_hepmc3_event_particles_;      ///!
10    void Init(TChain *tree) {
11        if (!tree) return;
12        fChain = tree;
13        fChain->SetMakeClass(1);
14        fChain->SetBranchAddresses("event_number", &event_number, &b_hepmc3_event_event_number);
15        SomeAnalysis(const std::string& file) {
16            TChain* TempChain= new TChain("hepmc3_tree");
17            TempChain->Add(file.c_str());
18            Init(TempChain);
19        }
20    };
21    int main(){
22        TH1D H1("H1","Pt of pions or electrons;Events/100MeV;P_{T},GeV",1000,0,100);
23        SomeAnalysis* A= new SomeAnalysis("inputI04.root");
24        if (!A->fChain->GetEntries()) return 1;
25        for (int entry=0; entry<A->fChain->GetEntries(); entry++){
26            A->fChain->GetEntry(entry);
27            for (int i=0; i<A->particles_.size(); i++){
28                if (A->particles_status[i]==1&&(std::abs(A->particles_pid[i])==211||std::abs(A->particles_pid[i])==11))
29                    H1.Fill(std::sqrt(A->particles_momentum_m_v1[i]*A->particles_momentum_m_v1[i]+A->particles_momentum_m_v2[i]*
30                    A->particles_momentum_m_v2[i]) );
31            }
32            delete A;
33            H1.Print("All");
34            return 0;
35        }
36    }
37 }
```

- Reading file in ROOT format **without HepMC3.**

# Usage: Rivet with HepMC3

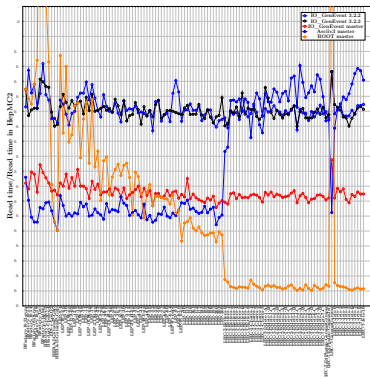
If compiled with HepMC3 > 3.2.0, Rivet uses standard HepMC3 functions

```
1  std::shared_ptr<HepMC3::Reader> deduce_reader(const std::string &filename);  
   std::shared_ptr<HepMC3::Reader> deduce_reader(std::istream &stream);
```

to deal with input, i.e. it is **disk format-agnostic** and the format of input will be deduced automatically. So, if HepMC3 can read some format, Rivet will be able to do it as well. No need for separate “converters” that use disk and Rivet can read ROOT.

# Developments

HepMC3 is stable and ready for use for a long time. But it does not mean the development is limited now to documentation improvements and bugfixes. **User issues should be addressed!** →



Improvements of I/O speed in master as a result of discussions

The first step to use Rivet with any MCEG is to assure the MCEG can produce an output readable by Rivet. **Means the MCEG should have HepMC interface.**

Right now, codes under review by the MCEG authors:

- HepMC3 interface for CASCADE
- HepMC3 interface for RAPGAP

**Once HepMC interface is accepted in the codebase, the implementation of Rivet interface is trivial.**

So, everything is great for DIS@Rivet?

Not really...

- DIS has specific issues related to the definitions of kinematic variables, which are not always easy to propagate to MC event record.
- The interface to actually used DIS MCEGs are missing.
- Even very good modern MCEGs have issues with certain aspects of *ep* physics, e.g. beam remnants.
- Expertise is scarce while the Rivet codes for HERA should be submitted voluntarily (v.s. obligatory for LHC).

**But, actually the invitation to this workshop stimulated a closer collaboration with Rivet authors. Thank you! This should work :)**



- If you have a good idea, feature request or bug report, please, create an issue in <https://gitlab.cern.ch/hepmc/HepMC3> or send a mail to [hepmc-dev@cern.ch](mailto:hepmc-dev@cern.ch) mailing list. **We do answer mails!** .
- If your idea was implemented, the issue was resolved, a bug was fixed – please give some **feedback**, close the resolved issues etc.

- HepMC3 3.2.4 will be released very soon. Another step towards full replacement of HepMC2, e.g. in ATLAS and MCEG community.
- Collecting feedback from LHC experiments, MCEG authors and interested individuals. **Discussions during this workshop will provide a crucial information.**

**Visit <http://hepmc.web.cern.ch/hepmc/>**

Write an e-mail to [hepmc-dev@cern.ch](mailto:hepmc-dev@cern.ch) or to any of authors.

# Backup slides

# Obsolete and removed features (vs. HepMC2)

- General features and I/O:
  - Some duplicated functions.
  - Operators '«' and '»' for event input/output that violated class hierarchy.
  - Custom iterator-based access. Finally usage of STL is possible!
  - Flow and Polarisation classes are removed and the information should be represented by generic attributes. Event record is simplified!

Particle in standard HepMC2 representation (IO\_GenEvent)

```
P 10001 11 0.0e+00 0.0e+00 4.59999e+01 4.60000e+01 5.10999e-04 4 0 0 -1 0
```

Particle in standard HepMC3 representation (Asciiv3)

```
1 P 1 0 11 0.0e+00 0.0e+00 4.59999e+01 4.60000e+01 5.10999e-04 4
```

The difference stands for polarisation and event flow information  
→ rarely filled and meaningful.

```
0 0 -1 0
```

# Requirements for HepMC3 3.2.3+

## Minimal:

- Modern Linux, OSX, BSD, Solaris or Windows system
- C++11 compatible compiler (gcc/clang/Intel/Oracle SunPro/PGI/IBM XL)
- CMake 3.7+

## Recommended=Minimal+

- ROOT6

## Full=Recommended+

- doxygen, latex, graphviz for documentation
- valgrind, Pythia8, HepMC2, Photos [13], Tauola [14], MC-TESTER [15] for tests
- Fortran77 compiler for examples
- Python2/Python3/PyPy modules

**Contact developers in case something is not working.**

# HepMC3 installation

There is a peception that the MCEG-related software stack is too complex and hard to install. Not the case for HepMC3:

OS	Repository	HepMC3 versions	Recommended for installation?	Credits
Fedora	Standard	last	Yes	Mattias Ellert
CentOS	EPEL	last	Yes	Mattias Ellert
MacOS	homebrew-hep	last	Yes	Enrico Bothmann/AV
ArchLinux	AUR	last	Yes	Frank Siegert
openSUSE	science	last	Yes	Atri Bhattacharya
Windows10	PyPi	last	Yes	AV
Debian, Ubuntu	Standard	3.1.1	Outdated, install from source	Mo Zhou
Others		last	Install from source	

## One should be able to do

```
1 [username@fedora ~]$ dnf install HepMC3
or
3 [username@centos ~]$ yum install HepMC3
or
5 username-macbook% brew install HepMC3
```

**without knowing the requirements**

# Event analysis listing of ROOT HepMC3 tree, with HepMC3

```
1 #include "HepMC3/GenEvent.h"
2 #include "HepMC3/GenParticle.h"
3 #include "HepMC3/ReaderRootTree.h"
4 #include <TH1D.h>
5 int main()
6 {
7     TH1D H2("H2", "Pt of pions or electrons;Events/100MeV;P_{T},GeV", 1000, 0, 100);
8     HepMC3::ReaderRootTree inputA("inputID4.root");
9     if(inputA.failed()) return 10002;
10    while( !inputA.failed() )
11    {
12        HepMC3::GenEvent evt(HepMC3::Units::GEV, HepMC3::Units::MM);
13        inputA.read_event(evt);
14        if( inputA.failed() ) {printf("End of file reached. Exit.\n"); break;}
15        for(auto p: evt.particles() )
16        if ( std::abs(p->status()) == 1 && (std::abs(p->pdg_id()) == 211 || std::abs(p->pdg_id()) == 11) )
17            H2.Fill( p->momentum().perp());
18        evt.clear();
19    }
20    inputA.close();
21    H2.Print("All");
22    return 0;
23 }
```

Listing 10: Reading file in ROOT format with HepMC3





- The original Les Houches accord event format (LHEF) is designed for communicating between Matrix element generators and MCEG.
- It agreed upon in 2001 [16](see updates in Refs. [12][17][18]).
- **The routines to handle LHEF is a precious part of HepMC3.**

```
<LesHouchesEvents version="3.0">
<header>
<MG5ProcCard>
.....
MadGraph 5
.....
VERSI 2.0.0.beta4      2013-06-22
.....
The MadGraph Development Team - Please visit us at
https://server06.fymv.ucl.ac.be/projects/madgraph
.....

```

LHEF is basically XML with extra rules.

Listing 14: First lines of LHEF header

```
<event>
6 66 0.50109093E+02 0.88344669E+02 0.75663862E-02 0.12114027E+00
2 -1 0 0 502 0 0.00000000E+00 0.00000000E+00 0.62728157E+03 0.62728166E+03 0.33000000E+00 0.0000E+00 0.0000E+00
4 5 -1 0 0 501 0 0.00000000E+00 0.00000000E+00 -0.44481443E+02 0.44739678E+02 0.48000000E+01 0.0000E+00 0.0000E+00
6 6 2 1 2 501 0 -0.59350663E+02 0.7244533E+02 0.21037840E+03 0.28804239E+03 0.17311146E+03 0.0000E+00 0.0000E+00
8 24 1 3 3 0 0 0.28747403E+02 0.66692808E+02 0.11644985E+03 0.15832494E+03 0.80388000E+02 0.0000E+00 0.0000E+00
10 5 1 3 3 501 0 -0.80928080E+02 0.55516749E+01 0.94928843E+02 0.12971745E+03 0.48000000E+01 0.0000E+00 0.0000E+00
12 8 1 1 2 502 0 0.59350663E+02 -0.7244533E+02 0.37242173E+03 0.38397894E+03 0.33000000E+00 0.0000E+00 0.0000E+00
14 #MCatML1 1 6 2 0 0 0.00000000E+00 0.00000000E+00 0 0 0 0.10000000E+01 0.86321181E+03 0.11022720E+01 0.00000000E+00 0.00000000E+00
16 #MCatML1 1 6 2 0 0 0.00000000E+00 0.00000000E+00 0 0 0 0.10000000E+01 0.91292678E+03 0.10493312E+01 0.00000000E+00 0.00000000E+00
18 #MCatML1 1 6 2 0 0 0.00000000E+00 0.00000000E+00 0 0 0 0.10000000E+01 0.91292678E+03 0.10493312E+01 0.00000000E+00 0.00000000E+00
20 #MCatML1 1 6 2 0 0 0.00000000E+00 0.00000000E+00 0 0 0 0.10000000E+01 0.91292678E+03 0.10493312E+01 0.00000000E+00 0.00000000E+00
</event>
<wgt id="1001"> 0.50109E+02 </wgt>
<wgt id="1002"> 0.43258E+02 </wgt>
<wgt id="1003"> 0.55234E+02 </wgt>

```

More:

Listing 15: First lines of some LHEF files for GDR-QCD -

# I/O: ROOT Tree format

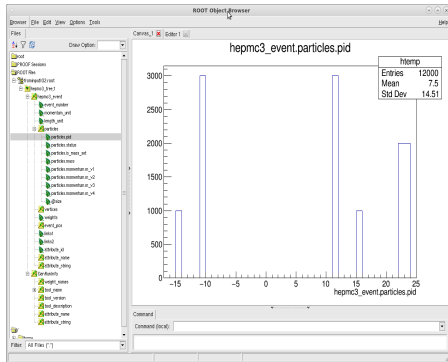
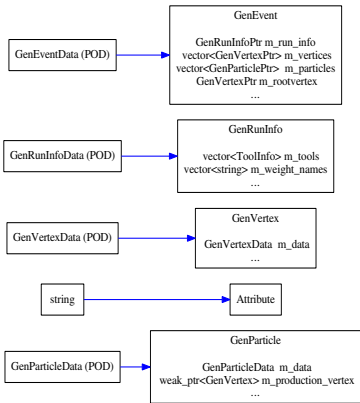


Figure: HepMC3 ROOT file in TBrowser.

HepMC3 uses custom streamers + POD types for ROOT I/O.  
**HepMC3 ROOT files are readable with standard ROOT, NO**

```
1 #include "HepMC3/GenEvent.h"
2 #include "HepMC3/WriterAscii.h"
3 #include "HepMC3/ReaderAsciiHepMC2.h"
4 int main()
5 {
6     HepMC3::ReaderAsciiHepMC2 inputA("inputI01.hepmc");
7     if(inputA.failed()) return 1;
8     HepMC3::WriterAscii outputA("frominputI01.hepmc");
9     if(outputA.failed()) return 2;
10    while( !inputA.failed() )
11    {
12        HepMC3::GenEvent evt(HepMC3::Units::GEV,HepMC3::Units::MM);
13        inputA.read_event(evt);
14        if( inputA.failed() ) {printf("End of file reached. Exit.\n"); break;}
15        outputA.write_event(evt);
16        evt.clear();
17    }
18    inputA.close();
19    outputA.close();
20    return 0;
21 }
```

Listing 16: Reading file in HepMC2 format (IO\_GenEvent) and writing in HepMC3 (AsciiV3)

- **Note: Works in the same way for all formats in different combinations.**

- [1] A. Buckley, P. Ilten, D. Konstantinov, L. Lonnblad, J. Monk, W. Porkorski, T. Przedzinski and A. Verbytskyi,  
The HepMC3 event record library for Monte Carlo event generators.  
Comput. Phys. Commun. **260**, 107310 (2021).  
[arXiv:1912.08005](#).
- [2] T. Sjöstrand, S. Mrenna and P.Z. Skands,  
A brief introduction to PYTHIA 8.1.  
Comput. Phys. Commun. **178**, 852 (2008).  
[arXiv:0710.3820](#).
- [3] J. Bellm et al.,  
Herwig 7.0/Herwig++ 3.0 release note.  
Eur. Phys. J. **C76**, 196 (2016).  
[arXiv:1512.01178](#).
- [4] T. Gleisberg et al.,  
Event generation with SHERPA 1.1.  
JHEP **02**, 007 (2009).  
[arXiv:0811.4622](#).
- [5] GEANT4, S. Agostinelli et al.,  
GEANT4: A Simulation toolkit.  
Nucl. Instrum. Meth. **A506**, 250 (2003).
- [6] A. Ferrari et al.,  
FLUKA: A multi-particle transport code (Program version 2005).  
(2005).
- [7] I. Antcheva et al.,  
ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization.  
Comput. Phys. Commun. **182**, 1384 (2011).

- [8] A. Buckley et al.,  
Rivet user manual.  
Comput. Phys. Commun. **184**, 2803 (2013).  
arXiv:1003.0694.
- [9] J. Bromley et al.,  
HZTOOL: A package for Monte Carlo-data comparison at HERA (version 1.0),  
Future physics at HERA. Proceedings, Workshop, Hamburg, Germany, September 25, 1995-May 31, 1996.  
Vol. 1, 2.  
(1995).
- [10] M. Dobbs and J.B. Hansen,  
The HepMC C++ Monte Carlo event record for High Energy Physics.  
Comput. Phys. Commun. **134**, 41 (2001).
- [11] J.G. Milhano et al.,  
Lisbon Accord: a standard format for heavy-ion event generators.  
(2017).
- [12] J. Alwall et al.,  
A Standard format for Les Houches event files.  
Comput. Phys. Commun. **176**, 300 (2007).  
arXiv:hep-ph/0609017.
- [13] E. Barberio and Z. Was,  
PHOTOS: A Universal Monte Carlo for QED radiative corrections. Version 2.0.  
Comput. Phys. Commun. **79**, 291 (1994).
- [14] S. Jadach, J.H. Kühn and Z. Was,  
TAUOLA: a library of Monte Carlo programs to simulate decays of polarized tau leptons.  
Comput. Phys. Commun. **64**, 275 (1990).

# Bibliography III

- [15] P. Golonka, T. Pierzchala and Z. Was,  
MC-TESTER: A Universal tool for comparisons of Monte Carlo predictions for particle decays in high-energy physics.  
Comput. Phys. Commun. **157**, 39 (2004).  
[arXiv:hep-ph/0210252](#).
- [16] E. Boos et al.,  
Generic user process interface for event generators.  
(2001).  
[arXiv:hep-ph/0109068](#).
- [17] J.M. Butterworth et al.,  
THE TOOLS AND MONTE CARLO WORKING GROUP Summary Report from the Les Houches 2009 Workshop on TeV Colliders.  
(2010).  
[arXiv:1003.1643](#).
- [18] J.R. Andersen et al.,  
Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report.  
(2014).  
[arXiv:1405.1067](#).