

# The Open Standard for Particle-Mesh Data

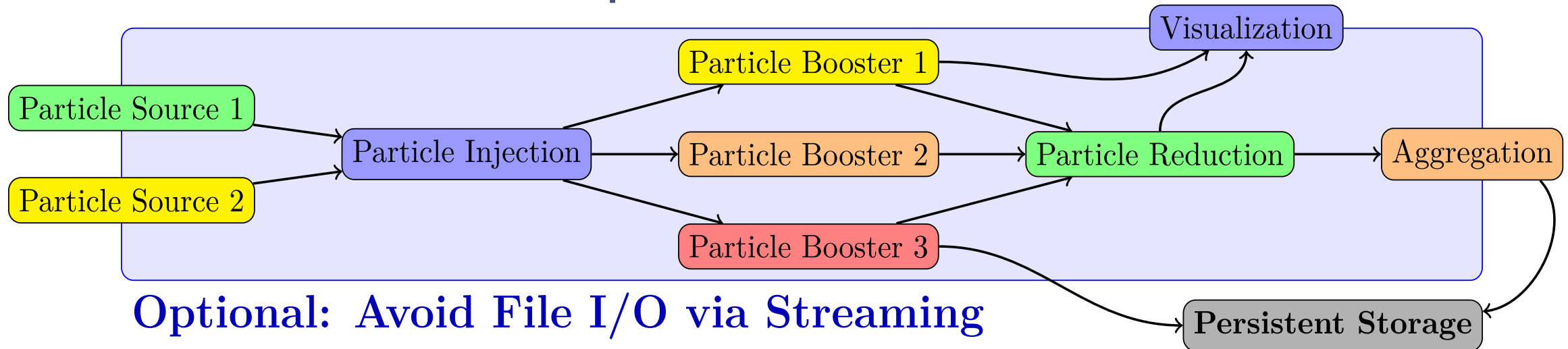
**Franz Poeschel (CASUS/HZDR)**  
2023 European HDF User Group (HUG)  
plugins and data compression summit

On behalf of the openPMD Community incl. content from  
Axel Huebl (LBNL), Lipeng Wan (GSU), Remi Lehe (LBNL)  
Norbert Podhorszki (ORNL), Junmin Gu (LBNL),  
Maxence Thévenet (DESY), Erik Schnetter (PITP),

Image:  
PIC simulation computed by PICongPU  
2<sup>nd</sup> prize Helmholtz Imaging Best Scientific Image Contest 2022

# Heterogeneity through Standardized Data

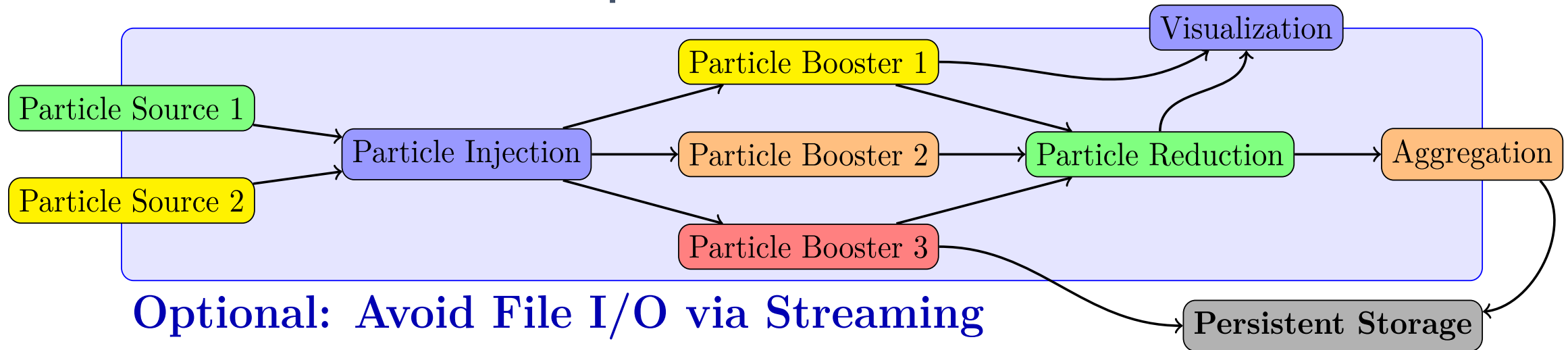
Scientific workflows are complex:



Optional: Avoid File I/O via Streaming

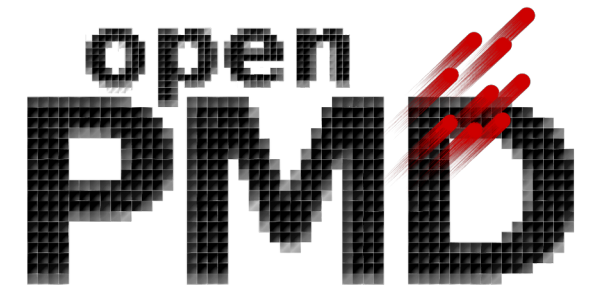
- need to span different **time** and **length scales**
- scientific modeling requires **multiple codes**, collaborating in a **data processing pipeline**
- **bridge heterogeneous models** by standardization of data

Scientific workflows are complex:

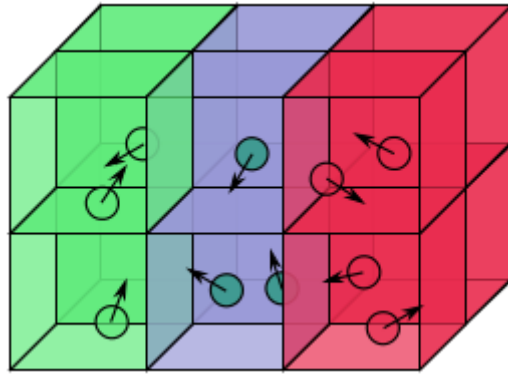


Optional: Avoid File I/O via Streaming

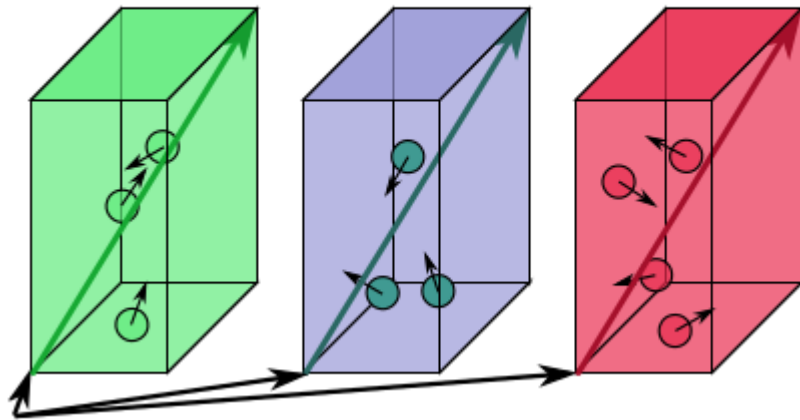
→ openPMD standard  
for **particle-mesh data**  
as communication layer



# What is particle-mesh data?



[0:3] particles [3:6] particles [6:10] particles



## Mesh

n-dimensional space,  
divided into discrete cells

- e.g. temperature:  
store a scalar number per cell
- e.g. electrical fields:  
store a 3D vector per cell

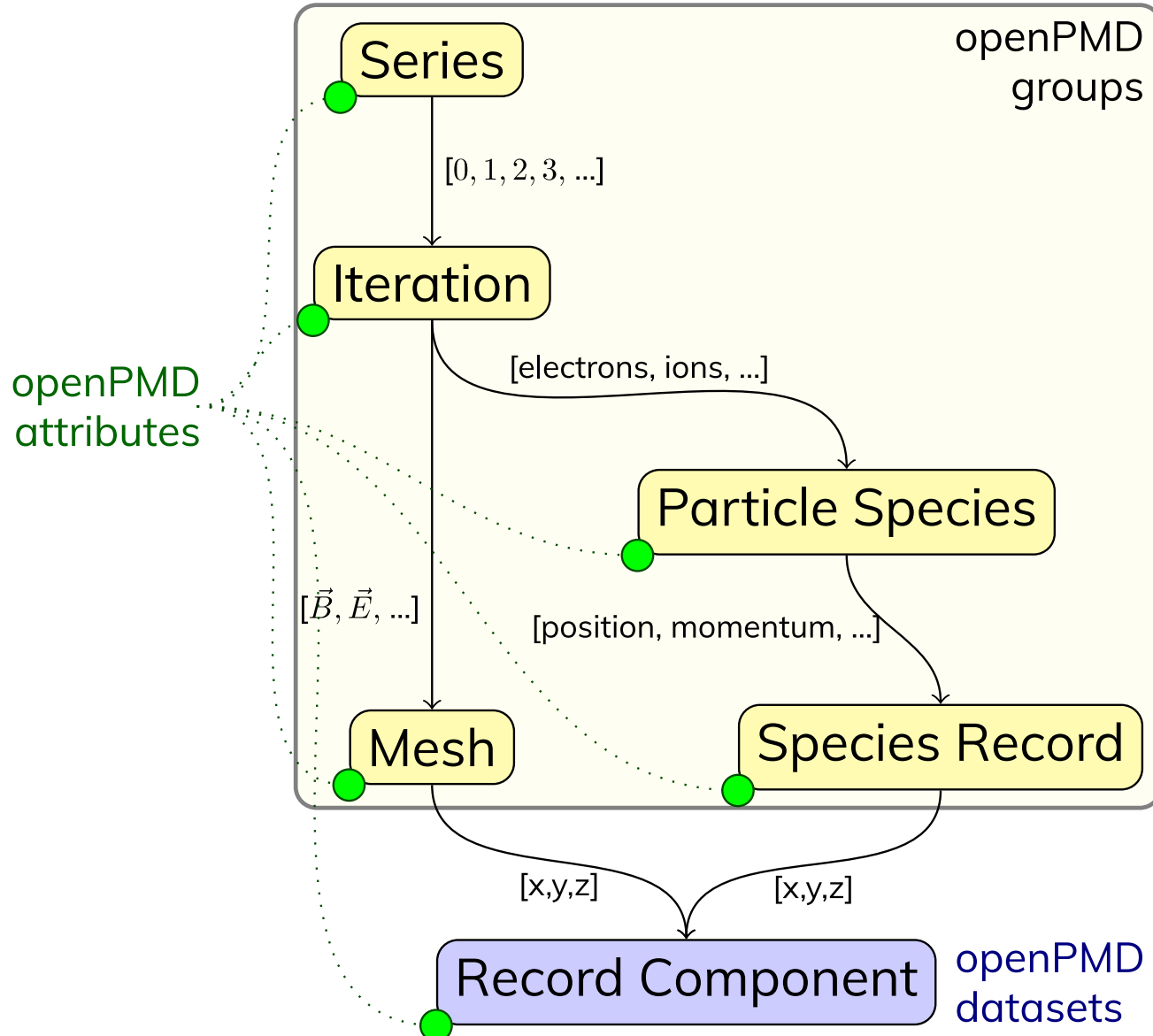
## Particles

A list of discrete objects,  
located on the mesh

- for each particle: list its position
- optionally: list charge, weight, ...

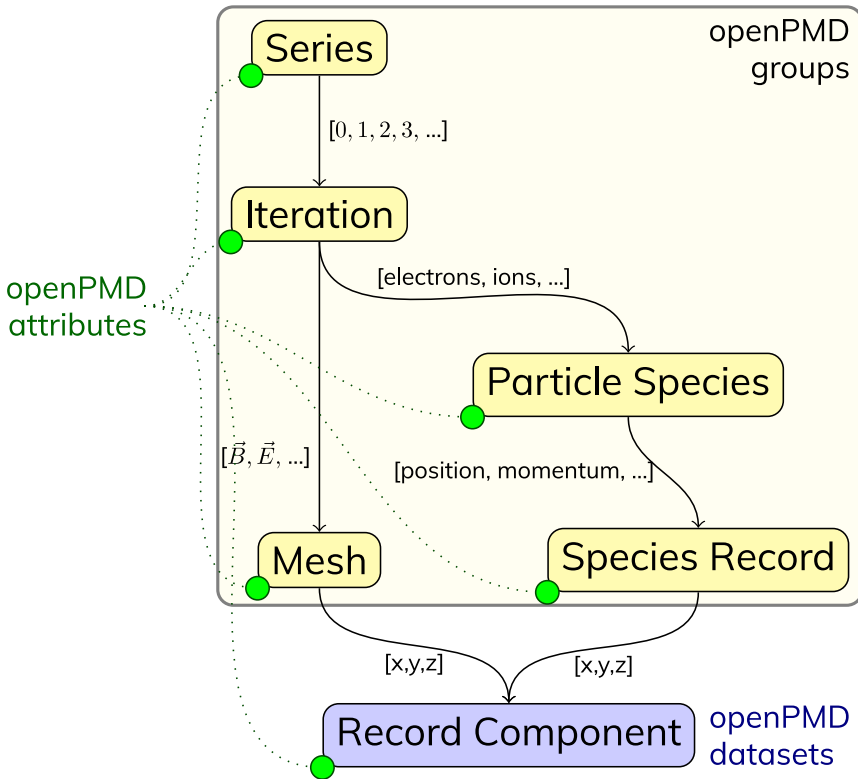


# openPMD hierarchy



- **Structure** for series & snapshots encoded as either:
  - **files** (one file per iteration)
  - **groups** (reuse files)
  - **variables** (reuse files & variables in ADIOS2)
- Records for **physical observables** constants, mixed precision, complex numbers
- **Attributes:** unit conversion, description, relations, mesh geometry, authors, env. info, ...

# Example dataset: HDF5 backend



Sample data  
created with PIconGPU

```

simData_000100.h5
├── data
│   └── 100
│       ├── fields
│       │   ├── B
│       │   └── E
│       │       ├── x
│       │       ├── y
│       │       └── z
│       ├── e_all_chargeDensity
│       ├── e_all_energyDensity
│       ├── i_all_chargeDensity
│       ├── i_all_energyDensity
│       ├── picongpu_idProvider
│       └── particles
│           ├── e
│           └── i
│               ├── charge
│               ├── mass
│               ├── momentum
│               ├── particlePatches
│               │   ├── extent
│               │   ├── numParticles
│               │   ├── numParticlesOffset
│               │   └── offset
│               ├── position
│               │   ├── x
│               │   ├── y
│               │   └── z
│               ├── positionOffset
│               └── weighting
    
```

Object Attribute Info    General Object Info

Attribute Creation Order:    Creation Order NOT Tracked

Number of attributes = 11

| Name             | Type   | Array Size | Value[50](...)                     |
|------------------|--------|------------|------------------------------------|
| axisLabels       | String | 3          | z, y, x                            |
| dataOrder        | String | Scalar     | C                                  |
| fieldSmoothing   | String | Scalar     | none                               |
| geometry         | String | Scalar     | cartesian                          |
| gridGlobalOffset | 64-bit | 3          | 0.0, 0.0, 0.0                      |
| gridSpacing      | 32-bit | 3          | 1.7416798, 1.7416798, 1.7416798    |
| gridUnitSI       | 64-bit | Scalar     | 5.3662849982E-8                    |
| position         | 32-bit | 3          | 0.0, 0.0, 0.0                      |
| timeOffset       | 32-bit | Scalar     | 0.0                                |
| unitDimension    | 64-bit | 7          | -3.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0 |
| unitSI           | 64-bit | Scalar     | 338590.78364382515                 |

# Unit System

## unitDimension

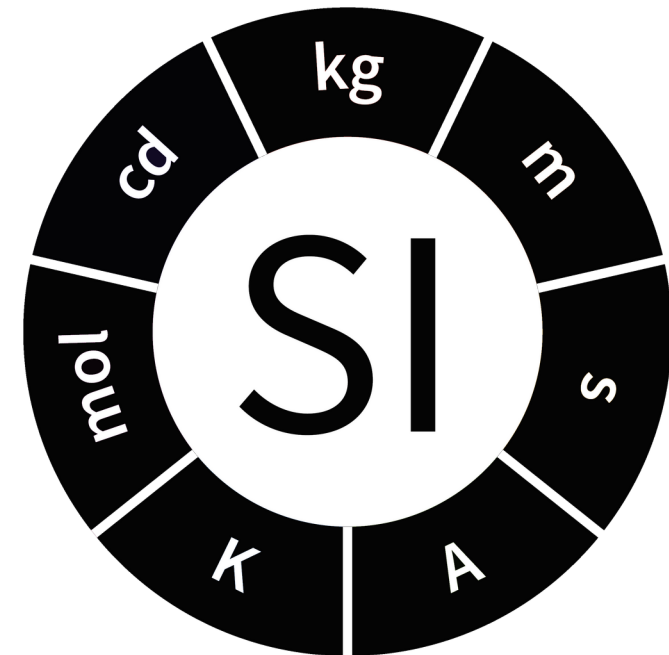
automated description of physical dimension  
only powers of base dimensions

length **L**, mass **M**, time **T**, electric current **I**,  
thermodynamic temperature **theta**,  
amount of substance **N**, luminous intensity **J**

Magnetic field:  $[B] = M / (I * T^2)$   
→ (0, 1, -2, -1, 0, 0, 0)

## unitSI (recommended)

relation to an absolute unit system

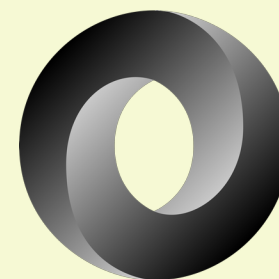


Wikimedia Commons

**Findable:** Standardized metadata to identify the data producer

```
string    /author          attr    = "franz"  
string    /software       attr    = "PIConGPU"  
string    /softwareVersion attr    = "0.5.0-dev"
```

**Accessible:** Open standard, implementable in various formats

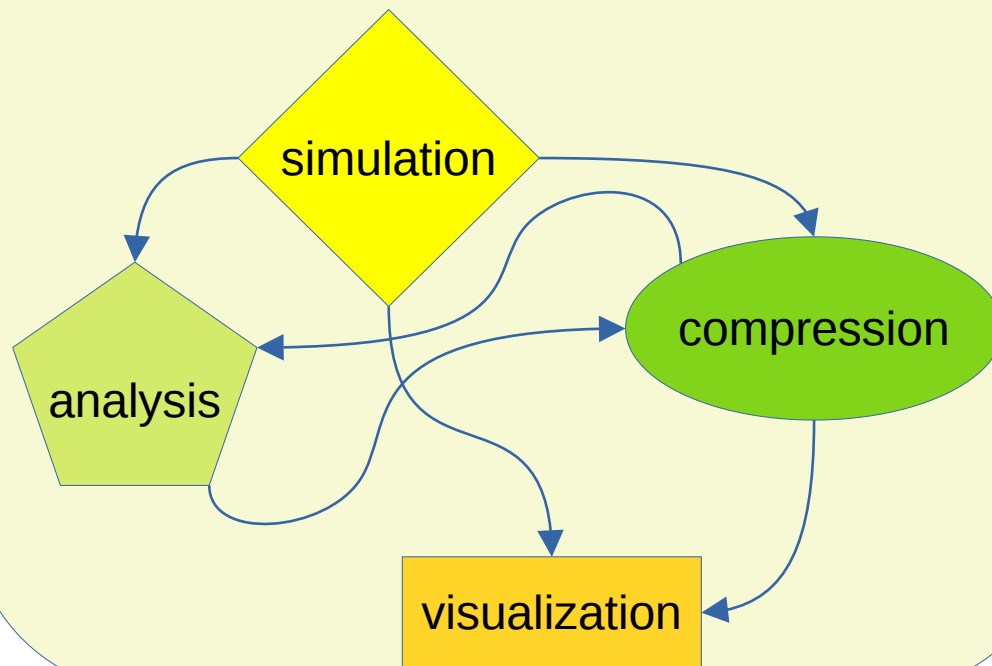


\*currently implemented,  
but not limited to



## Interoperable:

Data exchange spans applications, platforms and teams



## Reusable:

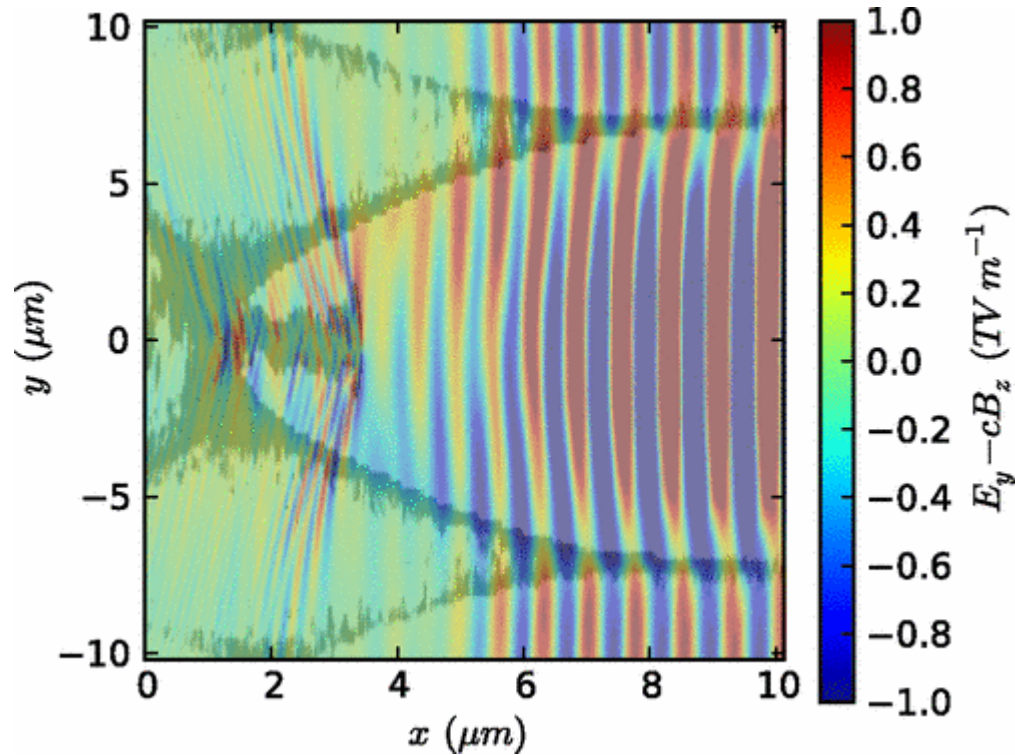
Rich and standardized description for physical quantities

| Name             | Value                         |
|------------------|-------------------------------|
| axisLabels       | [b'z' b'y' b'x']              |
| dataOrder        | b'C'                          |
| fieldSmoothing   | b'none'                       |
| geometry         | b'cartesian'                  |
| gridGlobalOffset | [0. 0. 0.]                    |
| gridSpacing      | [4.252342 1.0630856 4.252342] |
| gridUnitSI       | 4.1671151662e-08              |
| position         | [0. 0. 0.]                    |
| timeOffset       | 0.0                           |
| unitDimension    | [-3. 0. 1. 1. 0. 0. 0.]       |
| unitSI           | 15399437.98944343             |

“The FAIR Guiding Principles for scientific data management and stewardship” (Mark D. Wilkinson et al.)

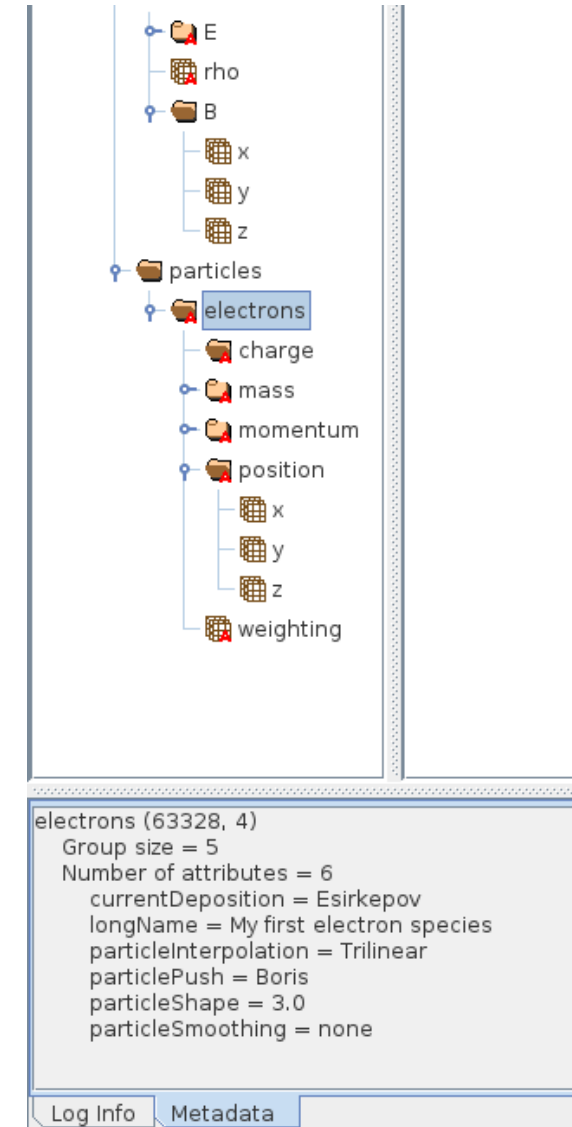
# Extensions: e.g. ED-PIC

**Field image** → field solver, smoothing



similar:

**Emittance** → particle push, field solver, shape



The screenshot shows a software interface with a hierarchical tree view on the left and a metadata panel on the right. The tree view includes nodes for 'E', 'rho', 'B', 'particles', and 'electrons'. The 'electrons' node is expanded, showing sub-nodes for 'charge', 'mass', 'momentum', 'position', and 'weighting'. The metadata panel displays the following information:

```
electrons (63328, 4)
Group size = 5
Number of attributes = 6
currentDeposition = Esirkepov
longName = My first electron species
particleInterpolation = Trilinear
particlePush = Boris
particleShape = 3.0
particleSmoothing = none
```

At the bottom of the metadata panel, there are two tabs: 'Log Info' and 'Metadata'.

# Ecosystem & Community

# openPMD powered Projects and Users

## Documents:

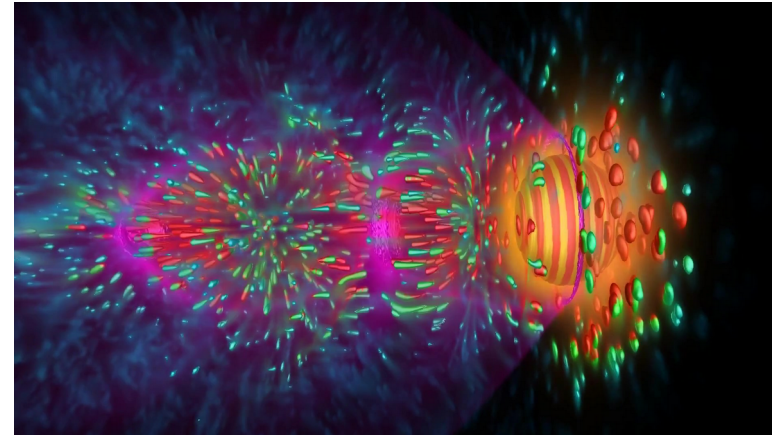
- **openPMD standard** (1.0.0, 1.0.1, 1.1.0)  
*the underlying file markup and definition*  
A Huebl et al., doi: 10.5281/zenodo.33624

## Language Binding:

- **openPMD-api** (HZDR, CASUS, LBNL)  
*reference API for openPMD data handling*  
maintainers: A Huebl, J Gu, F Poeschel et al.

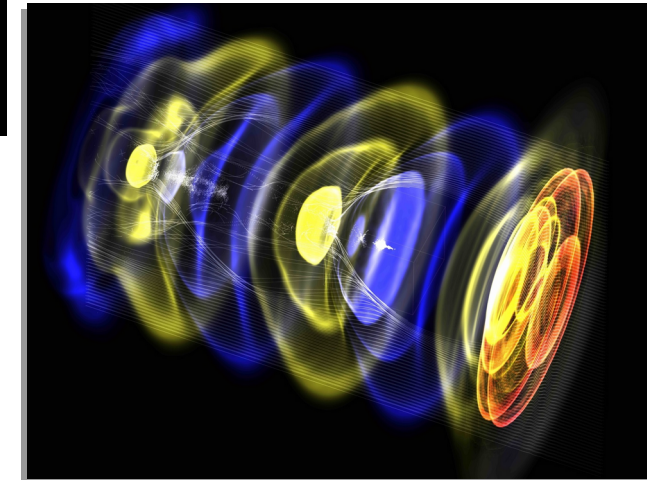
## Scientific Simulations:

- **PICongPU** (HZDR)  
*electro-dynamic particle-in-cell code*  
maintainers: R Widera, S Bastrakov, A Debus et al.
- **WarpX** (LBNL, LLNL)  
*electro-dynamic/static particle-in-cell code*  
maintainers: JL Vay, D Grote, R Lehe, A Huebl et al.
- **FBPIC** (LBNL, DESY)  
*spectral, fourier-bessel particle-in-cell code*  
maintainers: R Lehe, M Kirchen et al.
- **SimEx Platform** (EUCALL, European XFEL)  
*simulation of advanced photon experiments*  
maintainer: C Fortmann-Grote



**PICongPU+ISAAC** on Summit  
2<sup>nd</sup> prize Helmholtz Imaging  
Best Scientific Image Contest 2022  
Image credit: Felix Meyer/HZDR

**WarpX**  
PI: Jean-Luc Vay/LBNL



see also: <https://github.com/openPMD/openPMD-projects>



# openPMD powered Projects and Users

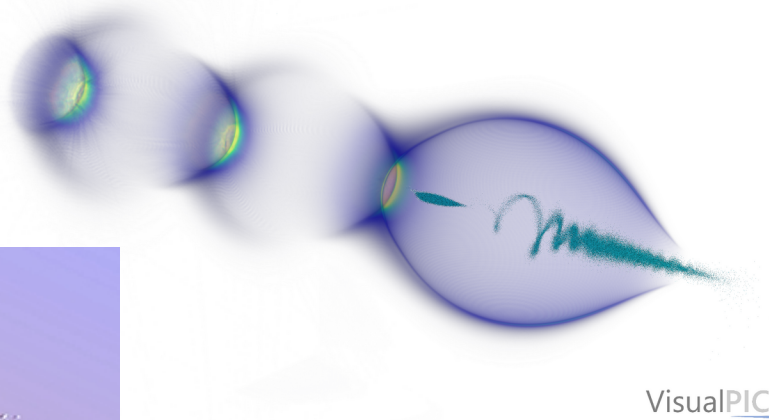
## Documents:

- **openPMD standard** (1.0.0, 1.0.1, 1.1.0)  
*the underlying file markup and definition*  
A Huebl et al., doi: 10.5281/zenodo.33624

## Language Binding:

- **openPMD-api** (HZDR, CASUS, LBNL)  
*reference API for openPMD data handling*  
maintainers: A Huebl, J Gu, F Poeschel et al.

**HiPACE++** → VisualPIC  
Credit: M.Thévenet  
& A. Ferran Pousa (DESY)



- **Wake-T** (DESY)  
*fast particle-tracking code for plasma-based accelerators*  
maintainer: A Ferran Pousa
- **HiPACE++** (DESY, LBNL)  
*3D GPU-capable quasi-static PIC code for plasma accel.*  
maintainers: M Thevenet, S Diederichs, A Huebl
- **Bmad** (Cornell)  
*library for charged-particle dynamics simulations*  
maintainers: D Sagan et al.
- **MALA** (CASUS, SNL)  
*ML models that replace DFT calculations in materials science*  
maintainers: Attila Cangi & Sivasankaran Rajamanickam
- and more...

**MALA** → ParaView  
Credit: A. Cangi (CASUS)

see also: <https://github.com/openPMD/openPMD-projects>

# Analysis and Visualization



openPMD/openPMD-viewer

```
In [1]: import numpy as np
import matplotlib notebook
# or `matplotlib inline` for non-interactive plots
# or `matplotlib widget` when using JupyterLab (github.com/matplotlib/jupyter-matplotlib)
import matplotlib.pyplot as plt
from openpmd_viewer import OpenPMDTimeSeries
```

```
In [2]: # Replace the string below, to point to your data
ts = OpenPMDTimeSeries('/home/franzpoeschel/singularity_build/pic_run/openPMD')
```

```
In [3]: # Interactive GUI
ts.slider()
```

- + iteration  1900

**Field type**

Field:

B E e\_all\_chargeDensity  
e\_all\_energyDensity picongpu\_idProvider

Coord:

x y z

▶ Particle quantities

▶ Particle selection

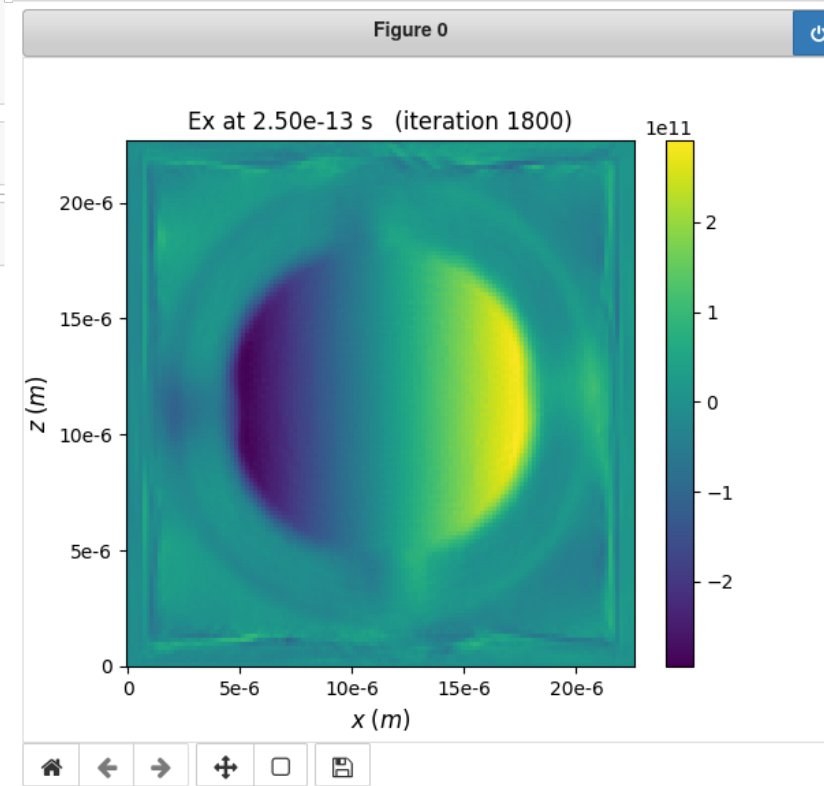
▶ Plotting options

Always refresh Refresh now!

▶ Slice selection

▶ Plotting options

Always refresh Refresh now!



# Analysis and Visualization

```
In [1]: import numpy as np
import matplotlib notebook
# or '%matplotlib inline' for non-interactive plots
# or '%matplotlib widget' when using JupyterLab (github.com/matplotlib/jupyter-matplotlib)
import matplotlib.pyplot as plt
from openpmd_viewer import OpenPMDTimeSeries
```

```
In [2]: # Replace the string below, to point to your data
ts = OpenPMDTimeSeries('/home/franzpoeschel/singularity_build/pic_run/openPMD')
```

```
In [3]: # Interactive GUI
ts.slider()
```

iteration

**Field type**

Field:

B  E  e\_all\_chargeDensity

e\_all\_energyDensity  picongpu\_idProvider

Coord:

x  y  z

▶ Particle quantities

▶ Particle selection

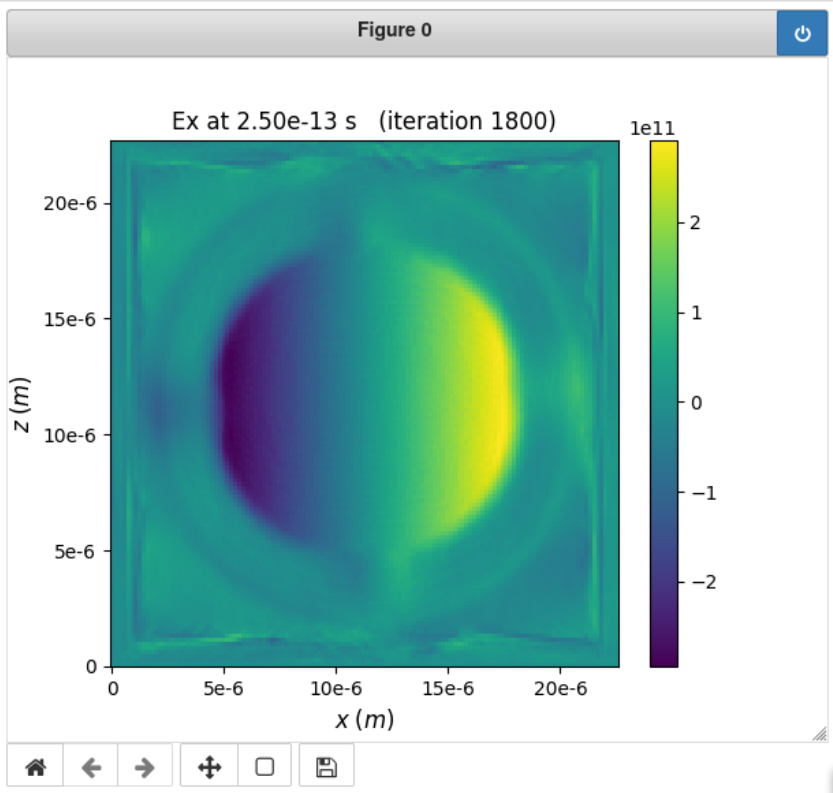
▶ Plotting options

Always refresh Refresh now!

▶ Slice selection

▶ Plotting options

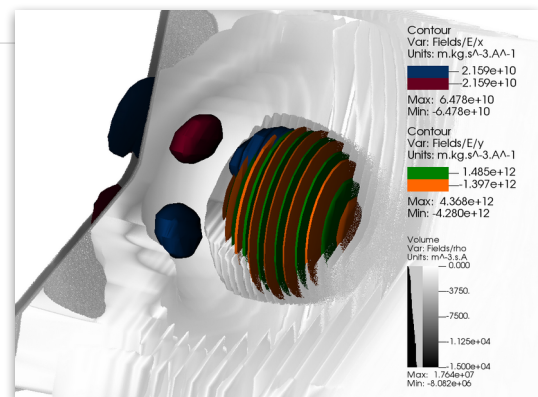
Always refresh Refresh now!



openPMD/openPMD-viewer

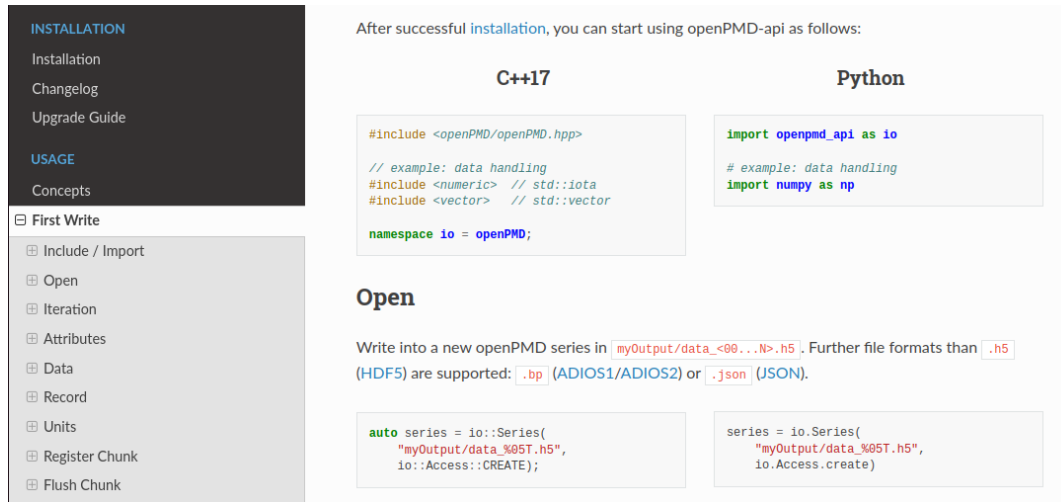


**Standardization of data**  
 → integration into modern scientific compute workflows



**Online Documentation:**  
[openpmd-api.readthedocs.io](https://openpmd-api.readthedocs.io)

**Open-Source Development & Tests:**  
[github.com/openPMD/openPMD-api](https://github.com/openPMD/openPMD-api)



After successful installation, you can start using openPMD-api as follows:

**C++17**

```
#include <openPMD/openPMD.hpp>

// example: data handling
#include <numeric> // std::iota
#include <vector> // std::vector

namespace io = openPMD;
```

**Python**

```
import openpmd_api as io

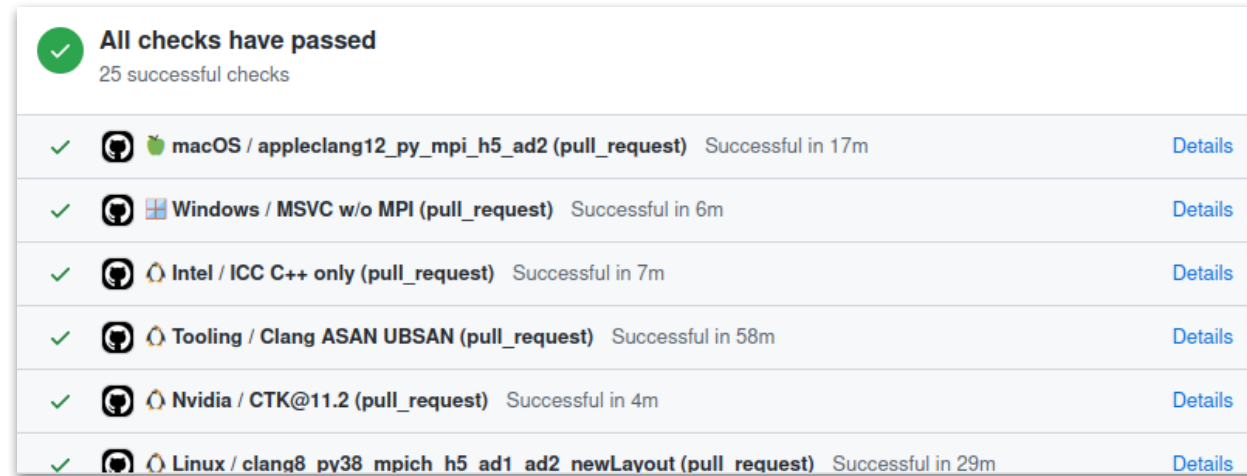
# example: data handling
import numpy as np
```

**Open**

Write into a new openPMD series in `myOutput/data_<00...N>.h5`. Further file formats than `.h5` (HDF5) are supported: `.bp` (ADIOS1/ADIOS2) or `.json` (JSON).

```
auto series = io::Series(
    "myOutput/data_05T.h5",
    io::Access::CREATE);
```

```
series = io.Series(
    "myOutput/data_05T.h5",
    io.Access.create)
```



**All checks have passed**  
25 successful checks

- macOS / appleclang12\_py\_mpi\_h5\_ad2 (pull\_request) Successful in 17m [Details](#)
- Windows / MSVC w/o MPI (pull\_request) Successful in 6m [Details](#)
- Intel / ICC C++ only (pull\_request) Successful in 7m [Details](#)
- Tooling / Clang ASAN UBSAN (pull\_request) Successful in 58m [Details](#)
- Nvidia / CTK@11.2 (pull\_request) Successful in 4m [Details](#)
- Linux / clang8\_py38\_mpich\_h5\_ad1\_ad2\_newLayout (pull\_request) Successful in 29m [Details](#)

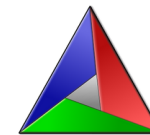
## Rapid and easy installation on any platform:



```
python3 -m pip install
openpmd-api
```



```
brew tap openpmd/openpmd
brew install openpmd-api
```



```
cmake -S . -B build
cmake --build build
--target install
```



```
conda install
-c conda-forge
openpmd-api
```



```
spack install
openpmd-api
```

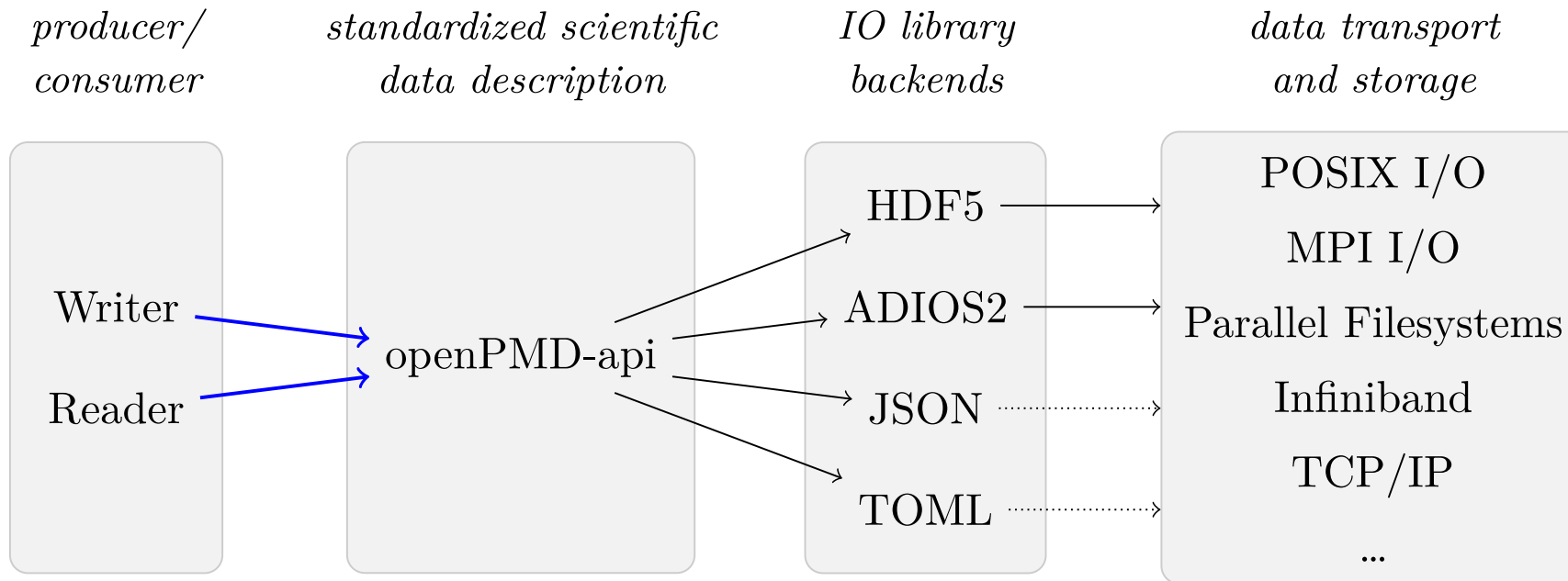


```
module load openpmd-api
```

A Huebl, F Poeschel, F Koller, J Gu, et al.  
 "openPMD-api: C++ & Python API for Scientific I/O with openPMD" (2018) DOI:10.14278/rodare.27



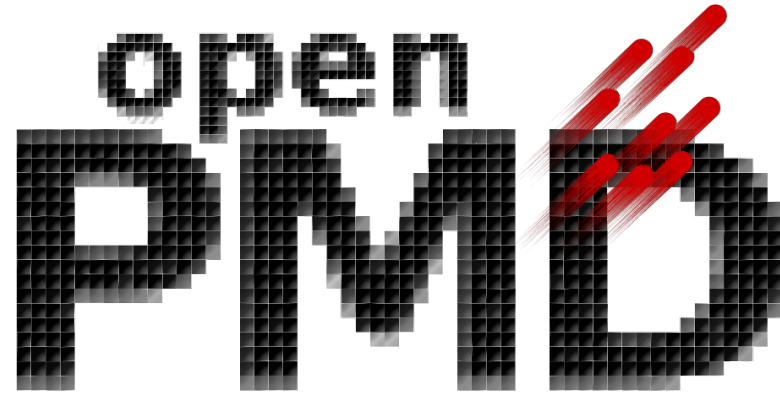
# openPMD-api – open stack for scientific I/O



- MPI support at all levels
- Implemented in C++17
- Bindings in C++17, Python and (dev version only) Julia

```
import openpmd_api as io

# pick backend by filename extension
series = io.Series("simOutput.h5", io.Access.create)
series = io.Series("simOutput.bp", io.Access.create)
series = io.Series("simOutput.sst", io.Access.create)
series = io.Series("simOutput.json", io.Access.create)
```

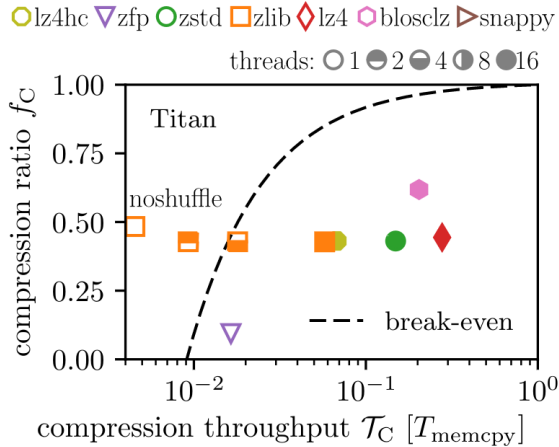


The **openPMD standard** is co-authored by [Axel Huebl](#), [Rémi Lehe](#), Jean-Luc Vay, David P. Grote, Ivo F. Sbalzarini, Stephan Kuschel, David Sagan, Frédéric Pérez, Fabian Koller, [Franz Poeschel](#), Carsten Fortmann-Grote, Ángel Ferran Pousa, Juncheng E, [Maxence Thévenet](#), and Michael Bussmann.

The authors are thankful for the **community contributions** to libraries, software ecosystem, user support, review and integrations. Particularly, thank you to Yaser Afshar, Lígia Diana Amorim, James Amundson, Weiming An, Igor Andriyash, Ksenia Bastrakova, [Jean Luca Bez](#), Richard Briggs, Heiko Bureau, Jong Choi, Ray Donnelly, Dmitry Ganyushin, Marco Garten, Lixin Ge, Berk Geveci, Daniel Grassinger, Alexander Grund, [Junmin Gu](#), Marc W. Guetg, Ulrik Günther, Sören Jalas, Manuel Kirchen, John Kirkham, Scott Klasky, Noah Klemm, Fabian Koller, Mathieu Lobet, Christopher Mayes, Ritiek Malhotra, Paweł Ordyna, Richard Pausch, [Norbert Podhorszki](#), David Pugmire, Felix Schmitt, [Erik Schnetter](#), Dominik Stańczak, Klaus Steiniger, Michael Sippel, Frank Tsung, Lipeng Wan, René Widera, and Erik Zenker!

# Ongoing projects

# Performance: Data Layouts and no-file I/O

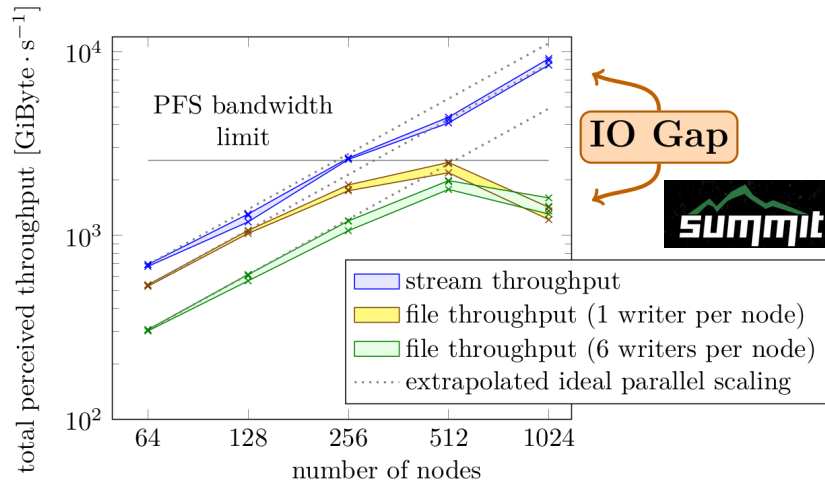


$$\frac{\mathcal{T}_R \times (1 - f_R)}{1 - \mathcal{T}_R} > \mathcal{T}_{out}$$

**Fast Compressors Needed:**  
[DOI:10.1007/978-3-319-67630-2\\_2](https://doi.org/10.1007/978-3-319-67630-2_2)  
 by A Huebl et al., ISC DRBSD-1 (2017)



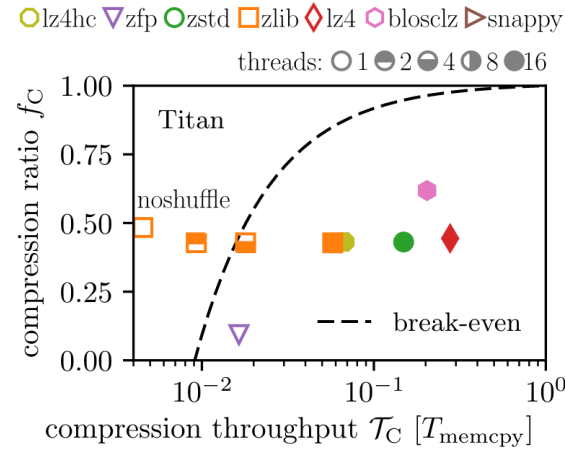
# Performance: Data Layouts and no-file I/O



## Streaming Data Pipelines:

[DOI:10.1007/978-3-030-96498-6\\_6](https://doi.org/10.1007/978-3-030-96498-6_6)

by F Poeschel, A Huebl et al., SMC21 (2022)



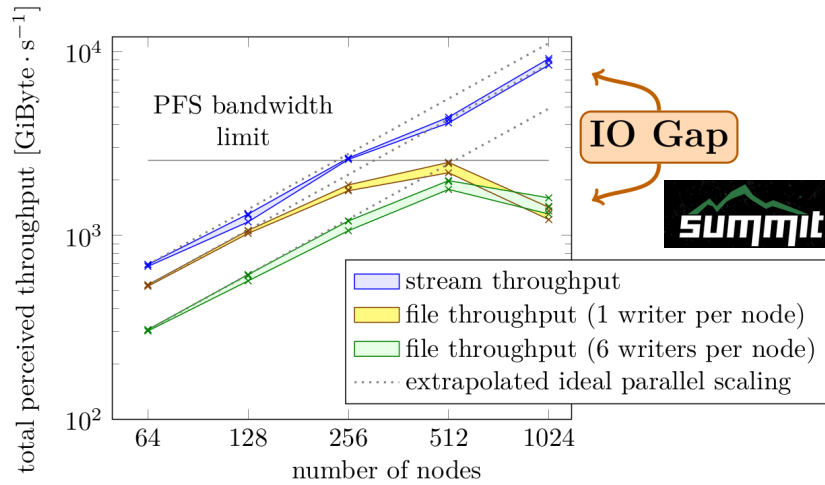
## Fast Compressors Needed:

[DOI:10.1007/978-3-319-67630-2\\_2](https://doi.org/10.1007/978-3-319-67630-2_2)

by A Huebl et al., ISC DRBSD-1 (2017)

$$\frac{\mathcal{T}_R \times (1 - f_R)}{1 - \mathcal{T}_R} > \mathcal{T}_{out}$$

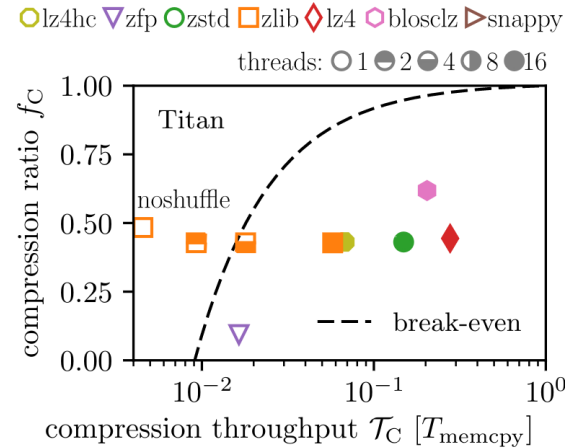
# Performance: Data Layouts and no-file I/O



## Streaming Data Pipelines:

[DOI:10.1007/978-3-030-96498-6\\_6](https://doi.org/10.1007/978-3-030-96498-6_6)

by F Poeschel, A Huebl et al., SMC21 (2022)

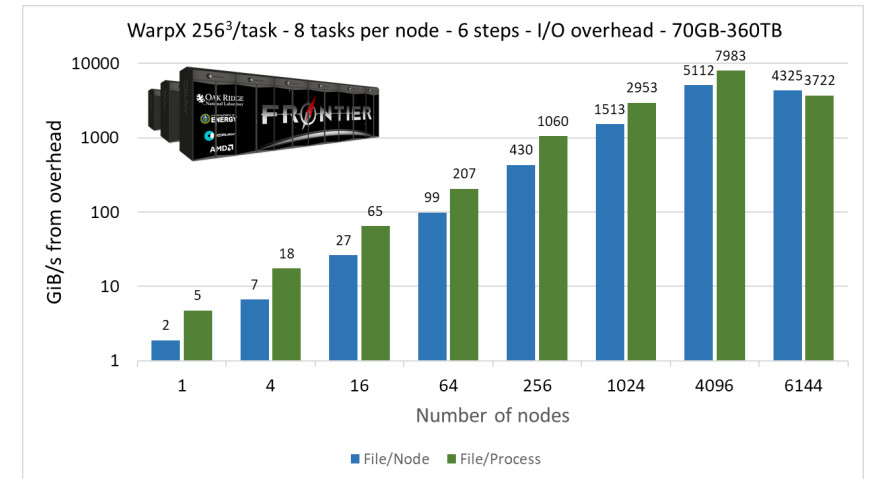


## Fast Compressors Needed:

[DOI:10.1007/978-3-319-67630-2\\_2](https://doi.org/10.1007/978-3-319-67630-2_2)

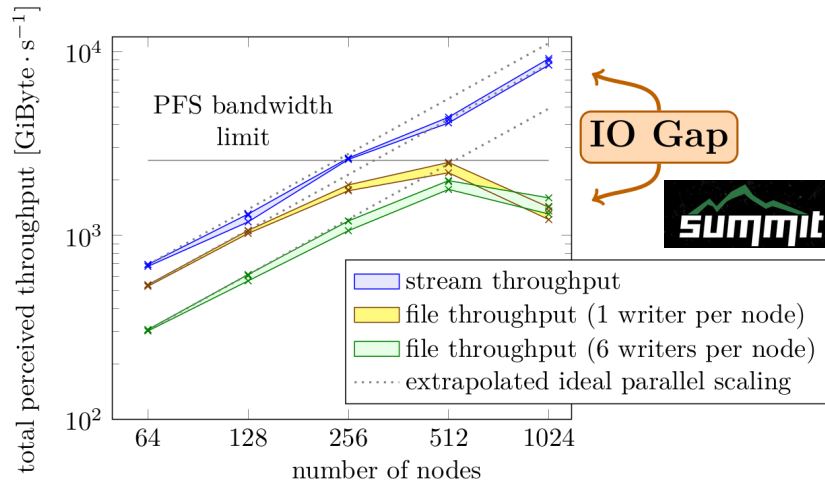
by A Huebl et al., ISC DRBSD-1 (2017)

## ADIOS openPMD-api w/ WarpX



>5.5TB/s FS BW: two-tier lustre w/ high-performance storage & progressive files

# Performance: Data Layouts and no-file I/O



## Streaming Data Pipelines:

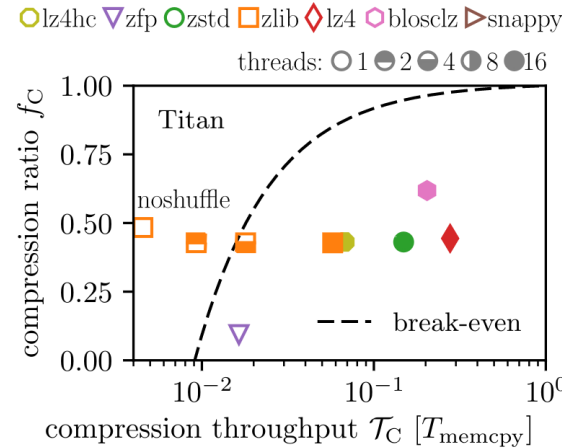
[DOI:10.1007/978-3-030-96498-6\\_6](https://doi.org/10.1007/978-3-030-96498-6_6)

by F Poeschel, A Huebl et al., SMC21 (2022)

## Online Data Layout Reorganization:

[DOI:10.1109/TPDS.2021.3100784](https://doi.org/10.1109/TPDS.2021.3100784)

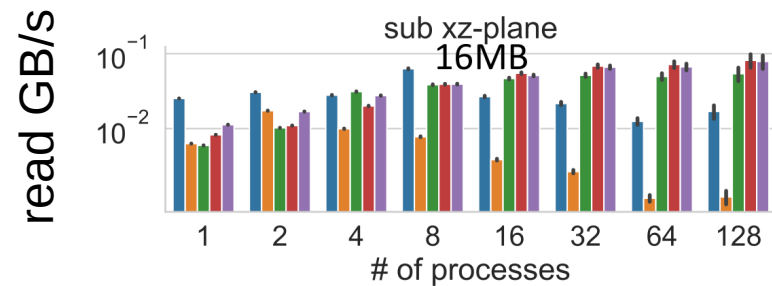
by L Wan, A Huebl et al., TPDS (2021)



## Fast Compressors Needed:

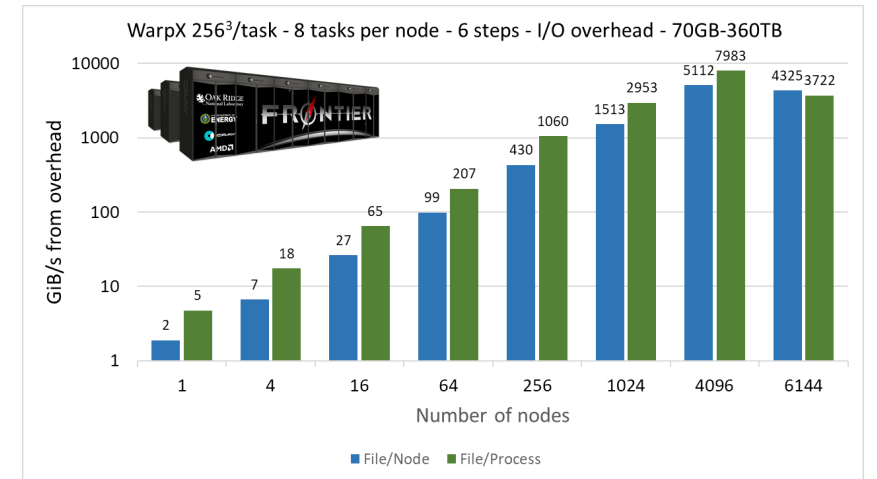
[DOI:10.1007/978-3-319-67630-2\\_2](https://doi.org/10.1007/978-3-319-67630-2_2)

by A Huebl et al., ISC DRBSD-1 (2017)

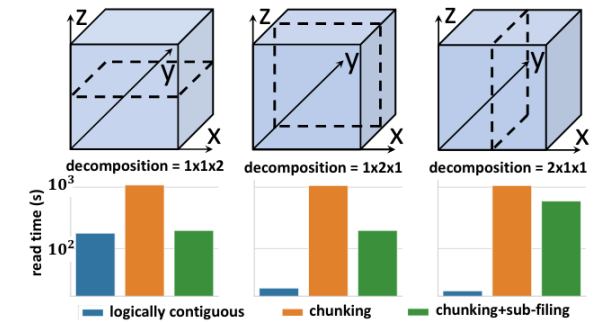


logically contiguous   chunking   chunking+sub-filing  
 chunking+sub-filing+intra-process-merging   chunking+sub-filing+intra-node-merging

## ADIOS openPMD-api w/ WarpX



>5.5TB/s FS BW: two-tier lustre w/ high-performance storage & progressive files



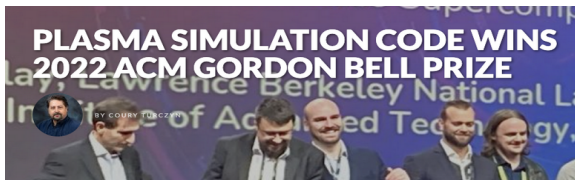
Impact of decomposition schemes when reading

# openPMD: Block-Structured Mesh-Refinement (WarpX)


Slide by Axel Huebl (LBNL)

## WarpX Particle-in-Cell Code

implemented on AMReX  
structured fields + many particles



WarpX was first PIC code to win prestigious ACM Gordon Bell award, the “Oscars” of Supercomputing.

open source   
[ece-warp.github.io](https://github.com/ece-warp)

## Multi-Node parallelization

- MPI: 3D domain decomposition
- dynamic load balancing



*>100TB per output  
Multi-PB per sim*

## On-Node Parallelization

- GPU: CUDA, HIP and SYCL
- CPU: OpenMP

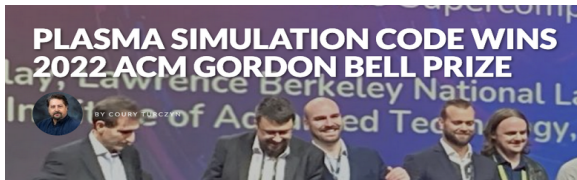


# openPMD: Block-Structured Mesh-Refinement (WarpX)

Slide by Axel Huebl (LBNL)

## WarpX Particle-in-Cell Code

implemented on AMReX  
structured fields + many particles



WarpX was first PIC code to win prestigious ACM Gordon Bell award, the “Oscars” of Supercomputing.

open source   
[ece-warp.github.io](https://github.com/ece-warp)

## Multi-Node parallelization

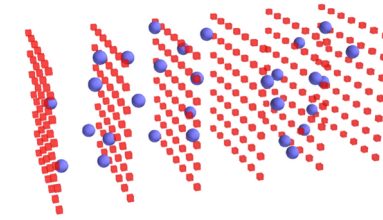
- MPI: 3D domain decomposition
- dynamic load balancing

## On-Node Parallelization

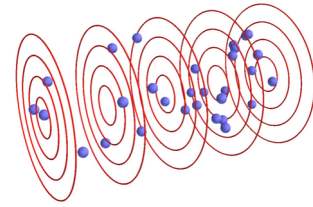
- GPU: CUDA, HIP and SYCL
- CPU: OpenMP

## Geometries

- 1D3V, 2D3V, 3D3V and RZ (quasi-cylindrical)



3D Cartesian grid



Cylindrical grid (schematic)



>100TB per output  
Multi-PB per sim



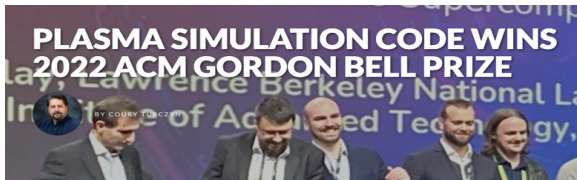


# openPMD: Block-Structured Mesh-Refinement (WarpX)


Slide by Axel Huebl (LBNL)

## WarpX Particle-in-Cell Code

implemented on AMReX  
structured fields + many particles



WarpX was first PIC code to win prestigious ACM Gordon Bell award, the “Oscars” of Supercomputing.

open source   
[ece-warp.github.io](https://github.com/ece-warp)

## Multi-Node parallelization

- MPI: 3D domain decomposition
- dynamic load balancing

## On-Node Parallelization

- GPU: CUDA, HIP and SYCL
- CPU: OpenMP

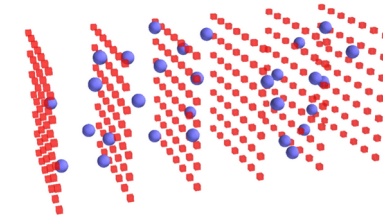


>100TB per output  
Multi-PB per sim

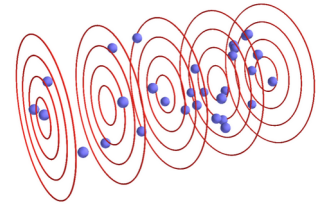


## Geometries

- 1D3V, 2D3V, 3D3V and RZ (quasi-cylindrical)



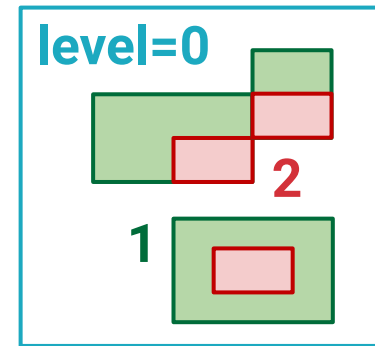
3D Cartesian grid



Cylindrical grid (schematic)

## Block-Structured Mesh-Refinement in openPMD

- ADIOS2:
  - variable: 1 per level
  - fill with partial blocks
- HDF5:
  - one variable per patch, N per level
  - overhead...? see: Elena's *sparsity* proposal
- alternative: AMReX HDF5  
1 variable per level, but loses HDF5 self-description (concatenated patches)



# Bridge the gap between openPMD and NeXus?

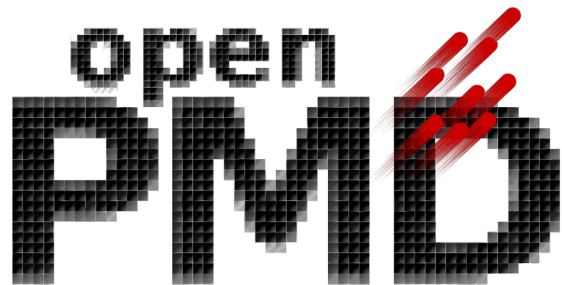
## The HELPMI project

### openPMD

---

focus on simulations

background:  
laser-plasma physics



### NeXus

focus on experiments

background:  
photon and neutron physics





# Bridge the gap between openPMD and NeXus?

The HELPMI project

## openPMD

focus on simulations

background:  
laser-plasma physics

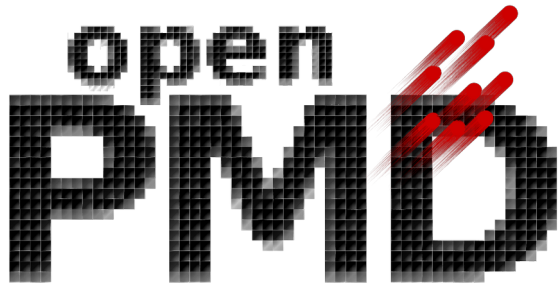


## NeXus

focus on experiments

background:  
photon and neutron physics

**and in between?**



# Bridge the gap between openPMD and NeXus?

## The HELPMI project

### openPMD

focus on simulations

background:  
laser-plasma physics

### HELPMI

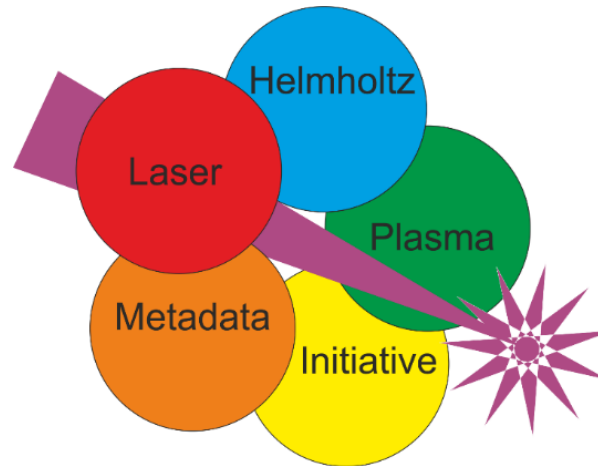
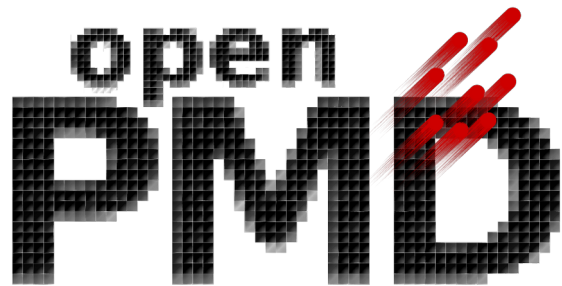
→ focus on experiments →

background:  
← laser-plasma physics ←

### NeXus

focus on experiments

background:  
photon and neutron physics



# Bridge the gap between openPMD and NeXus?

## The HELPMI project

### openPMD

### HELPMI

### NeXus

- data00000255.h5
  - data
    - 255
      - fields
        - rho
      - E
      - particles
        - Hydrogen1+
          - weighting
          - positionOffset
          - charge
          - momentum
          - position
          - mass
        - electrons
          - id
          - weighting
          - positionOffset
          - charge
          - momentum
          - position
          - mass

- waaterdroplett experiment
  - 220914
    - Farfield
    - Focus
    - GCCD
    - MCP
    - Nearfield
    - Probe1
    - Probe2
    - Transmission
    - Example SciCat Instrument Definition.json
    - example shot.json
    - Experimentator DB UI.xlsx
    - Screenshot 2023-06-20 at 15-31-32 POLARIS shot ...

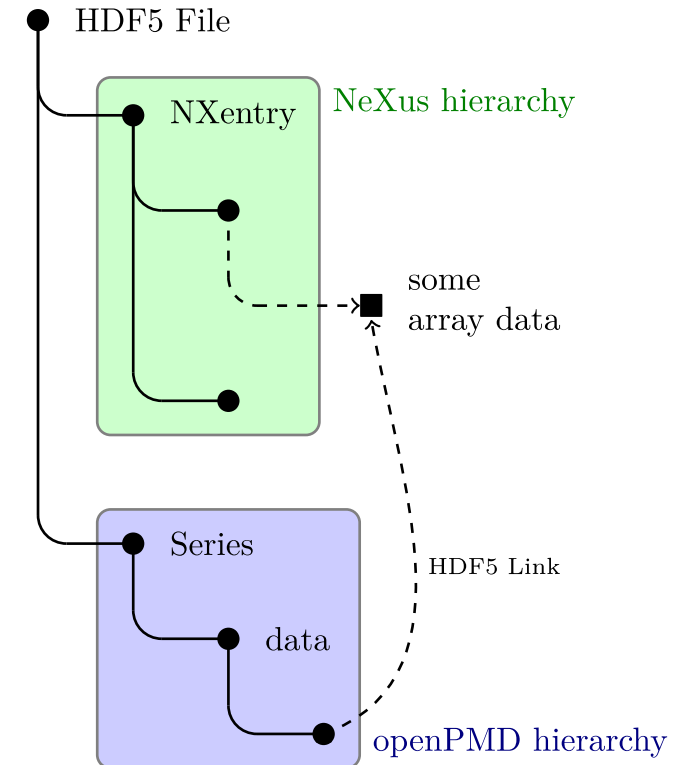


Metadata?  
Setup Models?  
Calibration?

- AgBehenate\_228.hdf5
  - entry
    - program\_name
    - run\_cycle
    - start\_time
    - definition
    - end\_time
    - title
    - AD\_template\_ID
    - data
      - data
      - model
      - make
      - local\_name
      - description
    - user1
      - name
      - proposal\_number
    - control
      - integral
      - preset
      - mode
    - instrument
      - name
      - aperture
      - detector
      - source
      - 15ID-D metadata
      - collimator
      - monochromator
    - link\_rules
      - link
    - sample
      - thickness
      - aequatorial\_angle
      - name

# HELPMI – a Helmholtz Metadata Collaboration project

- Primary goal: Develop a user-driven NeXus extension proposal for laser-plasma experiments
- Develop a glossary for LPA experiment data and infer the ontology for automated validation and processing
- openPMD is an existing standard for laser-plasma *simulations*:
  - aim for interoperability between both standards
  - example: openPMD “view” into NeXus data to compare experimental and simulation data



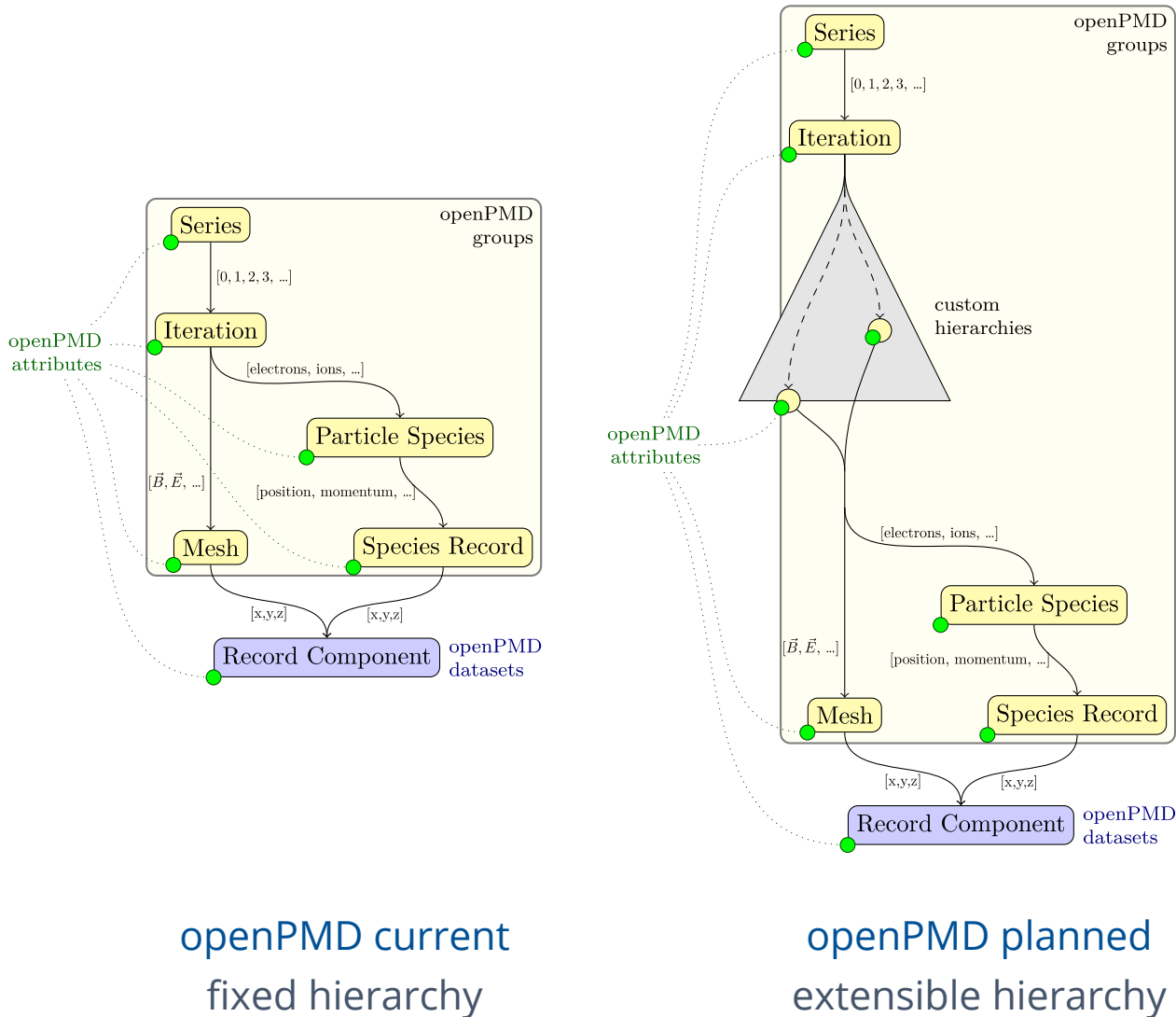
# HELPMI – a Helmholtz Metadata Collaboration project

- Primary goal: Develop a user-driven NeXus extension proposal for laser-plasma experiments
- Develop a glossary for LPA experiment data and infer the ontology for automated validation and processing
- openPMD is an existing standard for laser-plasma *simulations*:
  - aim for interoperability between both standards
  - example: openPMD “view” into NeXus data to compare experimental and simulation data

## Contact and time frame

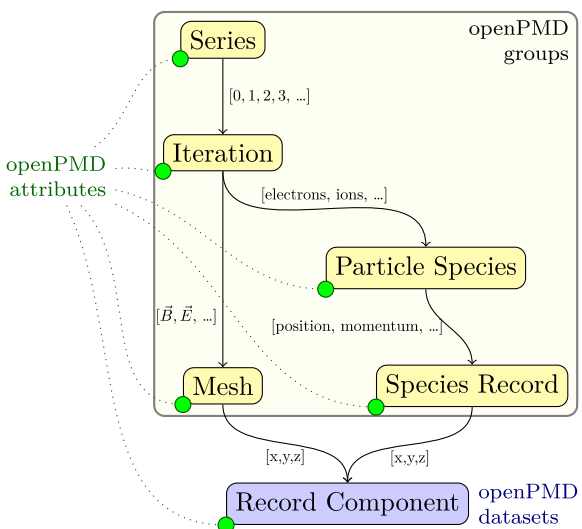
- Project from April 2023 to April 2025
- Upcoming: Helpmi Workshop at GSI (Darmstadt) Nov 13-14
- [helpmi@hzdr.de](mailto:helpmi@hzdr.de)

# openPMD: improve interoperability with other HDF5 standards

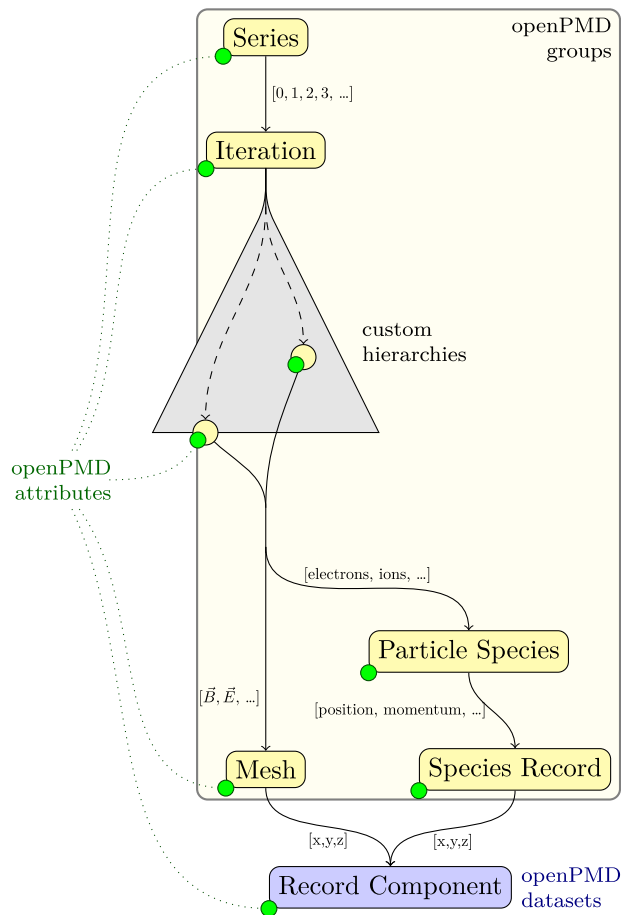


- openPMD current:  
does not forbid custom data,  
but does not make use of it either
- openPMD planned:  
use particle-mesh markup  
within a user-defined custom hierarchy
- for data reuse:  
support HDF5 hardlinks in future  
and for non-HDF5 backends:  
support softlinks

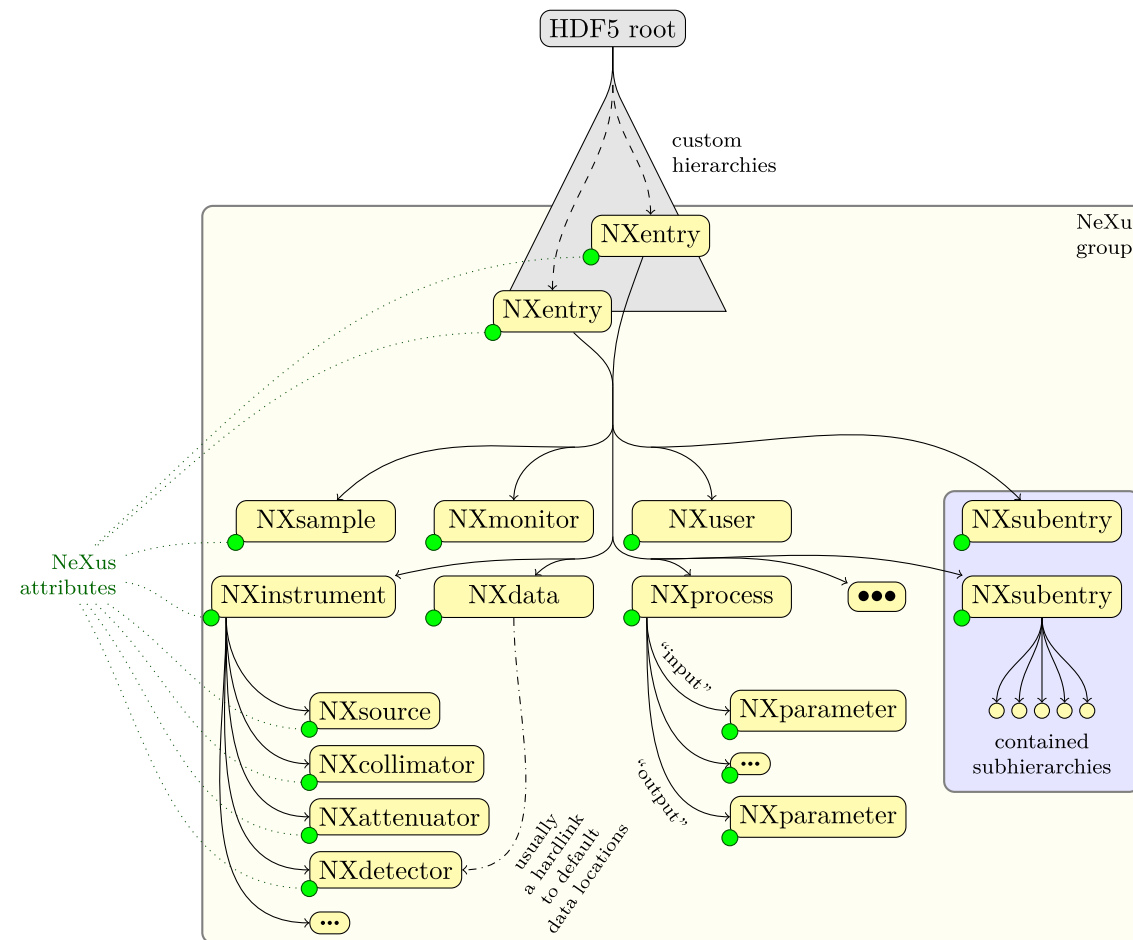
# openPMD: improve interoperability with other HDF5 standards



openPMD current  
fixed hierarchy



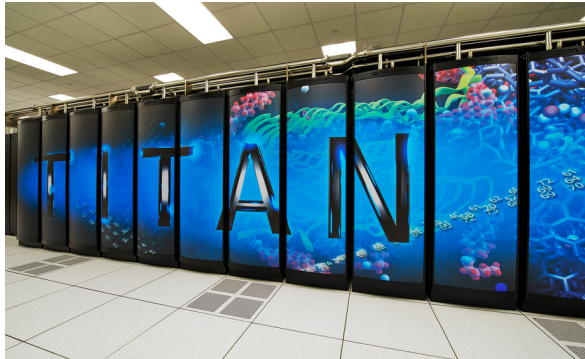
openPMD planned  
extensible hierarchy



comparison: NeXus  
hierarchy via NXentry and application definitions



# Similar challenges in simulations and experiments?



|                          | <b>Titan</b> | <b>Summit</b> | <b>Frontier</b> |
|--------------------------|--------------|---------------|-----------------|
| <b>Peak Performance:</b> | 27 Pflop/s   | 200 Pflop/s   | 1.6 Eflop/s     |
| <b>FS Throughput:</b>    | 1 TiByte/s   | 2.5 TiByte/s  | 5~10 TiByte/s   |
| <b>FS Capacity:</b>      | 27 PiByte    | 250 PiByte    | 500~1000 PiByte |

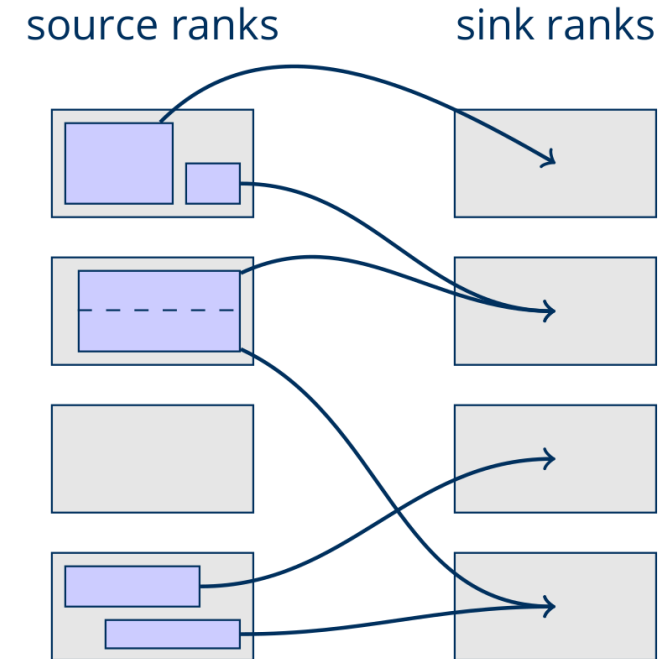
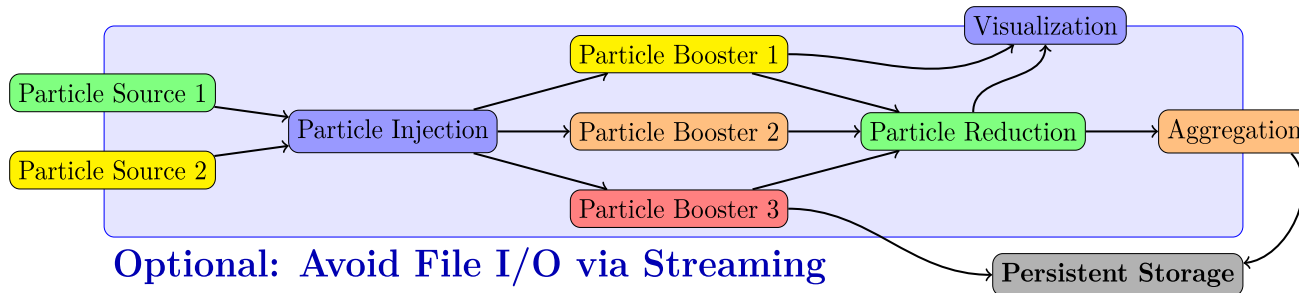
**Growth Factor**  
~60  
5~10  
18~37

- **parallel bandwidth** insufficient for HPC at full scale
- **filesystem capacity** insufficient for HPC at full scale

Same trend in **experiments?**

- Increasing **camera resolutions and data rates**

# Streaming: Don't touch the Filesystem at all



- Data processing pipelines and increasingly experiments setups have large I/O usage
- Scalable alternative: Streaming  
e.g. via Infiniband (on HPC systems)  
or wide area networks (in lab settings)


## Challenge:

Compute a **balanced, aligned, local** mapping between two applications that remains useful in the problem domain

# First Steps







→ head to <https://github.com/openPMD/>



 **openPMD**  
Open Standard for Particle-Mesh Data Files  
<https://www.openPMD.org> [axelhuebl@lbl.gov](mailto:axelhuebl@lbl.gov)

**Repositories** 17   Packages   People 50   Teams 5   Projects

## Pinned repositories

|   |  |  |
|---|--|--|
| <p> <b>openPMD-standard</b><br/>Open Standard for Particle-Mesh Data Files<br/>☆ 41   🍴 17</p> | <p> <b>openPMD-projects</b><br/>Overview on Projects around openPMD<br/>☆ 4   🍴 4</p>                   | <p> <b>openPMD-viewer</b><br/>Python visualization tools for openPMD files<br/>Jupyter Notebook   ☆ 35   🍴 26</p> |
| <p> <b>openPMD-api</b><br/>C++ &amp; Python API for Scientific I/O<br/>C++   ☆ 55   🍴 30</p> | <p> <b>openPMD-visit-plugin</b><br/>Plugin allowing VisIt to read openPMD files<br/>C   ☆ 8   🍴 3</p> | <p> <b>openPMD-example-datasets</b><br/>HDF5 Example Files<br/>Python   ☆ 5   🍴 1</p>                           |

...and of course <https://openpmd-api.readthedocs.io/>

## Acknowledgements

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration). Supported by EC through Laserlab-Europe, H2020 EC-GA 871124. Supported by the Consortium for Advanced Modeling of Particles Accelerators (CAMPA), funded by the U.S. DOE Office of Science under Contract No. DE-AC02-05CH11231. This work was partially funded by the Center of Advanced Systems Understanding (CASUS), which is financed by Germany's Federal Ministry of Education and Research (BMBF) and by the Saxon Ministry for Science, Culture and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament. The HELmholtz Laser Plasma Metadata Initiative (HELPMI) project (ZT-I-PF-3-066) was funded by the "Initiative and Networking Fund" of the Helmholtz Association in the framework of the "Helmholtz Metadata Collaboration" project call 2022.



# Summary and Outlook

- openPMD is a **F.A.I.R. standard for scientific metadata**
  - bridge scientific models and domains by **common markup language**
  - Large **open-source ecosystem**: documentation, example data, validation, scripts, integration via plugins and converters, reference libraries
- Reference **implementation**:
  - Easy to use I/O for scientific data
  - **Scalable I/O** at the Exascale and Pbyte-scale
  - Scalable from small workstation via parallel in-transport data processing to file-less RDMA workflows
- **Outlook**
  - Complex data layouts such as **mesh refinement**
  - Bridge towards **experimental data acquisition systems**
  - Transfer HPC solutions to experiments challenges



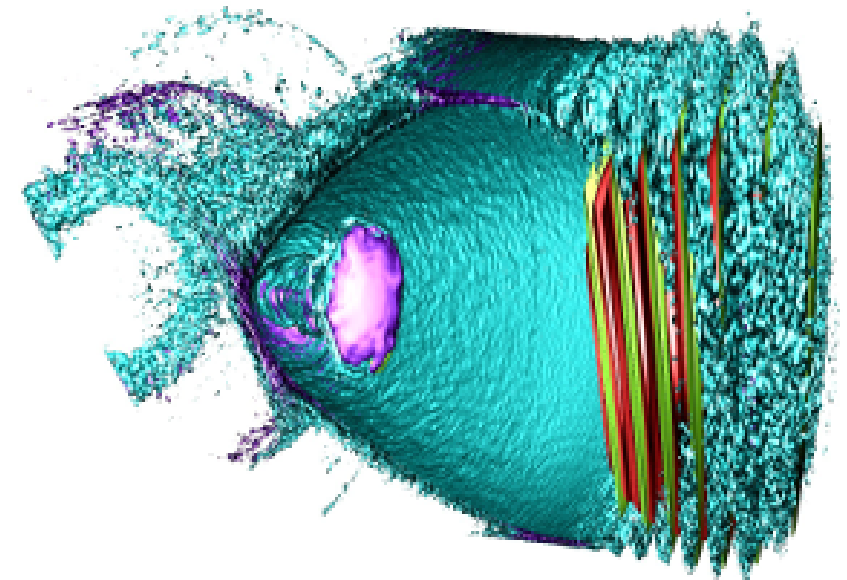
openPMD



openpmd.slack.com



openpmd.org



Picture: LWFA simulation in PConGPU