PaNOSC

Photon and Neutron Open Science Cloud

H2020-INFRAEOSC-04-2018

Grant Agreement Number: 823852

Deliverable: 5.1 Prototype simulation data formats as openPMD domain specific extensions including example datasets

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

Project Deliverable Information Sheet

Project Reference No.	823852
Project acronym:	PaNOSC
Project full name:	Photon and Neutron Open Science Cloud
H2020 Call:	INFRAEOSC-04-2018
Project Coordinator:	Andy Götz (andy.gotz@esrf.fr)
Coordinating Organization:	ESRF
Project Website:	www.panosc.eu
Deliverable No:	D5.1 Prototype simulation data formats as openPMD domain specific extensions including example datasets
Deliverable Type:	Report
Dissemination Level:	Public
Contractual Delivery Date:	31/11/2019
Actual Delivery Date:	26/11/2019
EC project Officer:	Geert Vancraeynest

Document Control Sheet

Document	Title: D5.1 Prototype simulation data formats as openPMD domain specific extensions including example datasets
	Version: 1
	Available at:
	Files: 1
Authorship	Written by: Carsten Fortmann-Grote
	Contributors: Juncheng E, Aljosa Hafner, Mads Bertelsen
	Reviewed by: Jordi Bodera Sempere
	Approved: Andy Götz

List of participants

Participant No.	Participant organisation name	Country
1	European Synchrotron Radiation Facility (ESRF)	France
2	Institut Laue-Langevin (ILL)	France
3	European XFEL (XFEL.EU)	Germany
4	The European Spallation Source (ESS)	Sweden
5	Extreme Light Infrastructure Delivery Consortium (ELI-DC)	Belgium
6	Central European Research Infrastructure Consortium (CERIC-ERIC)	Italy
7	EGI Foundation (EGI.eu)	The Netherlands

Table of Content

1 Introduction

Standardized formatting and hierarchical organization of simulation data and associated metadata is paramount to support seamless exchange of data between simulation software in simulation pipelines and to benefit from 3rd party data visualization and analysis software that support these formats. By applying the openPMD metadata standard for particle and mesh data [1] to all simulated data in WP5, our workflows and results become accessible, inter–operable, and reusable, i.e. in line with the core concepts of FAIR Data Principles [2].

This report details the metadata standard extensions for simulation data developed in Work Package 5 (WP5) of the Photon and Neutron Open Science Cloud (PaNOSC). The openPMD standard in its current version 1.1 is the basis on which we have developed domain specific extensions for the simulation stages of start–to–end photon and neutron experiment simulations:

1. Coherent wavefront propagation (simulation of coherent lightsources and beam transport)
2. Photon raytracing (simulation of x–ray optical beamline components, incoherent sources)
3. Neutron raytracing (simulation of neutron beamlines)
4. Molecular dynamics simulations (target simulations, radiation–matter interaction)

All extensions are deposited under the WP5 github project at `https://github.com/PaNOSC-ViNYL/openPMD-standard` in the branch "upcoming-2.0" in fulfillment of the project deliverable D5.1 of PaNOSC. To expose our standard extensions to the wider community, we will propose to merge them into the base standard repository `https://github.com/openPMD/openPMD-standard` after the Deliverable has been accepted.

2 The openPMD base standard

OpenPMD stands for open particle and mesh data and was initially developed as a metadata and data hierarchy standard for particle–in–cell (PIC) simulations of high–power laser–matter interaction at the Helmholtz–Zentrum Dresden–Rossendorf. Due to its flexibility (allowing for adaptation and adjustments for other applications than PIC), independence from the actual file format (supporting e.g. hdf5 [3], netCDF [4], adios [5] and json [6] and more), and being maintained by an active open source community, the openPMD standard is widely adopted in numerous simulation codes, visualization codes, and simulation workflow platforms.

The base standard structures simulation data by the following characteristics:

1. The Series: The root object of the data hierarchy.
2. Iteration: Simulation time step (the time stamp of a given simulation snapshot)
3. Particle or mesh
4. Type of particle (e.g. electron, proton) or mesh (e.g. electric field, magnetic field).
5. Records (Physical variables):

- For particles: Position and momentum (velocity) vector components
- For meshes: Field values (components for vector fields)

In addition, the standard defines a number of mandatory and optional metadata information to define and specify the physical meaning of the data. E.g. for particle data, each record must provide information about the physical unit in which its data is written. A mesh must provide the grid spacing, dimensionality, extension, and geometry and, where appropriate, the corresponding units. An iteration must provide which physical time it corresponds to. Further metadata about the code that produced the data, the data author, the time and date of the simulation, the version of the standard that this data was formatted against, are provided as attributes of the Series.

3 OpenPMD extensions

3.1 Wavefront data

Coherent wavefront propagation plays an import role in the simulation of coherent light sources, such as free–electron lasers. Applications include studying the intensity and phase distributions at the experimental interaction point and effects of misalignments, pointing errors, fluctuations in the temporal, spatial, and spectral pulse structure on the latter.

During a wavefront simulation, the horizontal and vertical polarization components of the complex electric field is calculated at a given position in the path of the light pulse in three dimensions: horizontal and vertical position in a plane perpendicular to the beam axis and time. From this data, all observables such as intensity, phase, polarization, divergence, pointing, energy spectrum and other, can be calculated. The information is complete in the sense, that the propagation can be continued further down the beamline from a simulation that stopped at an upstream position. The temporal sampling resolves the slowly varying pulse envelope but not every single oscillation of the the laser fundamental frequency.

The electric field components as defined above are the required data records for the wavefront domain extension. Besides, the standard requires information about the radius of curvature in horizontal and in vertical direction, its standard deviation, as well as the standard photon energy to be given as attributes of the Series. Furthermore, it must be specified whether the data represents the pulse in the time or in the frequency domain, and in the cartesian, or in the reciprocal space.

This domain extension can be found at https://github.com/PaNOSC-ViNYL/openPMD-standard/blob/upcoming-2.0.0/EXT_WAVEFRONT.md. An example dataset is published on zenodo at https://dx.doi.org/10.5281/zenodo.3524710 [7].

Example python code how to write and read can be found in the python simulation library SimEx https://github.com/eucall-software/simex_platform/blob/openpmd/Sources/python/SimEx/Utilities/wpg_to_opmd.py.

3.2 Raytracing data

3.2.1 Neutrons

The field of neutron scattering relies heavily on Monte Carlo raytracing simulations for design of instrumentation. The low brilliance of neutron sources means that larger beams are required to achieve the necessary intensity, and thus the point approximations used in most analytical calculations are inadequate. Raytracing simulations provides a phase-space description of the neutron beam at the sample position, but it is also possible to do virtual experiments that include the sample physics resulting in expected intensity on the detector.

In Monte Carlo raytracing simulations a large number of random initial neutron states are generated from realistic random distributions or loaded from a file. The neutron state includes position, velocity, polarization, weight and time. The physical unit of the weight is neutrons/s, as the expected intensity of the source is split into the total weight of simulated rays. These rays are propagated through an instrument, and the neutron weights are lowered to account for absorption or reflectivity losses. The weights can also be manipulated to provide variance reduction, for example by focusing all scattering towards the detector and reducing the weight with the appropriate factor based on the solid angle of the detector. The intensity expected in a pixel of a detector is then the sum of the weights of neutrons that arrived in that pixel.

The openPMD extension for neutron raytracing simulations contains the necessary fields for describing neutron state, facilitating transfer of neutron rays between different simulation tools. The different simulation packages include different capabilities, and thus it can be useful to switch from one package to another to get the highest accuracy possible.

The proposed domain extension for neutron scattering can be found at https://github.com/PaNOSC-ViNYL/openPMD-standard/blob/upcoming-2.0.0/EXT_NRAYTRACING.md. An example mock data set is published on Zenodo along with the code used to generate the dataset [8].

3.2.2 Photons

Raytracing is extensively used in the design and performance assessment of X-ray beamlines. It allows for simulation of different properties of the beam, such as the spot size, divergence and generally propagate rays through the beamline. To do this, a number of rays N are first initialized according to spatial and energy distributions, corresponding to the source (bending magnet, undulator, etc.) properties. Each ray is then propagated from one to the next optical element (plane mirror, elliptical mirror, etc.) and the position and divergence of the beam are stored for each position. This way, a full description of the beam can be had at any point downstream and allows for easy modifications to the beamline, making use of the previously calculated unchanged parts.

One of such codes is called ShadowOui and is a part of the OASYS framework [9]. In the PaNOSC project, other potentially interoperable codes (e.g. Simex) where identified and a common data format, allowing for linking the output of one code and the input to another, has been proposed and devised. The openPMD extension for X-ray/photon raytracing contains all the necessary

information for that. Additionally to the position and the direction, it includes the following information: the energy of the rays, phases for P- and S- polarized photons and the P- and S-polarized components.

The extension can be found at https://github.com/PaNOSC-ViNYL/openPMD-standard/blob/upcoming-2.0.0/EXT_PRAYTRACE.md and its example data file and the code used to generate it has been published on Zenodo [10].

3.3 Molecular Dynamics data

Molecular dynamics (MD) is a long–established computer simulation method to analyze atom/molecular movements by solving Newton's equations of motion assuming a given potential energy function. It is broadly used in statistical physics, physical chemistry, structural biology, material science, and crystallography. MD simulation is an important tool to predict the response of a molecular system to external fields as well as to probes such as photon or neutron beams.

With MD simulation, the structural evolution of the experimental sample can be modeled. Depending on the computing power, the time scale of the model can range from fs to ns, and the length scale can be up to several hundred nm. The fact that the scales are close the the experimental scales and the atomic-level simulation solution make it possible to calculate scattering signals from the MD atomic configuration which is comparable to the experimental sample. This kind of simulations has already been used in the researches of deformation under high pressure[11, 12, 13, 14], crystallization [15], and crystallography [16] to predict diffraction patterns for experimental design and interpretation.

Radiation damage induced by high–intensity X-ray is assumed to be one of the limiting factors for achieving subnanometer X-ray imaging of single particles[17]. With MD simulation codes considering ionization process like GROMACS[18] and XMDYN[19], the effects of radiation damage on the reconstruction resolution of single particle imaging can be investigated.

The necessary output of MD simulation is the coordinate of each atom for each atom species in the sample structure. Besides that, the velocity of each atom, the simulation box and the force fields implemented should be recorded if a restart of previous simulation is asked for. Simulation box defines the extent of the coordinates of simulated atoms. Dimension and boundary are the attributes of box required for MD simulation. Boundary defines the periodicity of each axis of the box. Edge defines the shape of simulation box and allows the box to be triclinic.

This domain extension can be found at https://github.com/PaNOSC-ViNYL/openPMD-standard/blob/upcoming-2.0.0/EXT_MD.md. An example dataset is published on zenodo at https://dx.doi.org/10.5281/zenodo.3525951 [20].

Example python code to convert the output of XMDYN to comply the MD domain extension can be found here https://github.com/ejcjason/MDDomainExtension/blob/master/xmdyn_to_opmd.py.

References

[1] Huebl, A. *et al.* Openpmd 1.0.0: A meta data standard for particle and mesh based data (2017). URL https://dx.doi.org/10.5281/zenodo.33624.

[2] Wilkinson, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016). URL https://dx.doi.org/10.1038/sdata.2016.18.

[3] The HDF Group. Hierarchical Data Format, version 5 (1997). URL http://www.hdfgroup.org/HDF5/.

[4] Rew, R. *et al.* Unidata netcdf (1989).

[5] Liu, Q. *et al.* Hello ADIOS: the challenges and lessons of developing leadership class I/O frameworks **26**, 1453–1473 (2014).

[6] json. The json website (2019). URL https://www.json.org/.

[7] Fortmann-Grote, C. Example dataset for openpmd conform wavefront propagation data (wavefront domain extension) (2019).

[8] Bertelsen, M. Example dataset with neutron raytracing data conforming to openpmd standard (neutron raytracing domain extension) along with the script used to generate the data. (2019).

[9] Rebuffi, L. & Sánchez del Río, M. ShadowOui: a new visual environment for x-ray optics and synchrotron beamline simulations **23**, 1357–1367 (2016).

[10] Hafner, A. Example openpmd dataset for photon raytracing (2019).

[11] Kimminau, G. *et al.* Simulating picosecond x-ray diffraction from shocked crystals using post-processing molecular dynamics calculations. *Journal of Physics: Condensed Matter* **20**, 505203 (2008).

[12] Tang, M. X., E, J. C., Wang, L. & Luo, S. N. Loading-path dependent deformation of nanocrystalline Ta under single-and double-shock, and quasi-isentropic compression. *Journal of Applied Physics* **121**, 115901 (2017).

[13] Tang, M. X., Zhang, Y. Y., E, J. C. & Luo, S. N. Simulations of X-ray diffraction of shock-compressed single-crystal tantalum with synchrotron undulator sources. *Journal of Synchrotron Radiation* **25**, 748–756 (2018).

[14] Zhang, Y. Y., Tang, M. X., Cai, Y., E, J. C. & Luo, S. N. Deducing density and strength of nanocrystalline Ta and diamond under extreme conditions from X-ray diffraction. *Journal of Synchrotron Radiation* **26**, 413–421 (2019).

[15] E, J. C., Wang, L., Cai, Y., Wu, H. A. & Luo, S. N. Crystallization in supercooled liquid Cu: Homogeneous nucleation and growth. *The Journal of Chemical Physics* **142**, 064704 (2015).

[16] Coleman, S. P., Spearot, D. E. & Capolungo, L. Virtual diffraction analysis of Ni [0 1 0] symmetric tilt grain boundaries. *Modelling and Simulation in Materials Science and Engineering* **21**, 055020 (2013).

[17] Caleman, C. *et al.* On the Feasibility of Nanocrystal Imaging Using Intense and Ultrashort X-ray Pulses. *ACS Nano* **5**, 139–146 (2011).

[18] Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Molecular modeling annual* **7**, 306–317 (2001).

[19] Jurek, Z., Son, S.-K., Ziaja, B. & Santra, R. *XMDYN* and *XATOM* : Versatile simulation tools for quantitative modeling of X-ray free-electron laser induced dynamics of matter. *Journal of Applied Crystallography* **49**, 1048–1056 (2016).

[20] E, J. C. Example dataset for openpmd-conform molecular dynamics data (md domain extension) (2019).