

Contribution ID: 86

Type: TALK

# Progress and lessons learned from Data and Metadata management in laser-particle acceleration at HZDR

Monday 4 November 2024 11:45 (20 minutes)

Metadata is a key element in data management when taking account of the F.A.I.R.(findable, accessible, interoperable and reusable) principles, answering the need for better data integration and enrichment. In the field of high-intensity laser-plasma physics, numerical simulations and experiments go hand in hand, complementing each other. While simulation codes are well documented and output files can follow some standards (like openPMD or Smilei/happi), input files were often neglected. Experiment documentation is typically diverse, containing the description of manually executed setup steps (with photos or hand drawings), tabular data of experiment execution (parameters and observations) alongside with actual detector data. Often, data from the driving laser system is better organized but poorly connected to the experiment.

We will report on recent progress of data management in the field of high-intensity laser-plasma physics at HZDR by means of the center's data strategy and external projects like "HELPMI" by HMC, "DAPHNE4DNFI" by NDFI and others.

HELPMI is an HMC project aiming towards a data standard for laser-plasma (LPA) experiment data, making data interoperable (I) and reusable (R). While openPMD is an open standard for simulations in that domain, NeXus is an open standard for experimental data in Photon and Neutron sciences. HELPMI has identified benefits and challenges when adopting NeXus for the LPA domain and extended openPMD for arbitrary data hierachies like NeXus. Alongside, a domain-specific glossary is being developed, where the community must be involved.

DAPHNE4DNFI supports metadata capture and data enrichment activities at HZDR. One major achievement is a web app for manual experiment logging, i.e. taking the above-mentioned tabular data of parameters and observations. This app is highly configurable, following the changes and improvements of experimental setup and techniques. It can connect to other electronic lab documentation resources (like the Mediawiki system deployed at HZDR) in order to directly re-use metadata. Data is stored in a central database instead of multiple spreadsheet files and can be directly plotted, also against historical data.

Another important outcome of DAPHNE4DNFI is metadata capturing and cataloging of simulation input data. This way, tables of simulation input data can be generated, allowing to re-use input files. Of course, output data and analysis scripts are also linked, thus the in-house re-use of data is strongly enhanced.

On top of that, but also as necessary tool, DAPHNE4DNFI has strongly promoted the use of metadata catalogues, in particular SciCAT. Even though daily usage is automated via scripts, a web interface to browse and search for data and metadata is extremely helpful. Such metadata catalogues complement data repositories by enabling the F (findable) but require a lot of effort in data enrichment.

### Please specify "other"

#### In addition, please add 3 to 5 keywords.

metadata capture, enrichment, metadata catalogue

## For whom will your contribution be of most interest?

Data professionals who provide and maintain data infrastructure

## Please assign yourself (presenting author) to one of the following groups.

Researchers

Primary author: SCHLENVOIGT, Hans-Peter (HZDR)

**Co-authors:** DEBUS, Alexander (Helmholtz-Zentrum Dresden-Rossendorf); KESSLER, Alexander; POESCHEL, Franz (CASUS/HZDR); HORNUNG, Johannes (GSI Helmholtzzentrum für Schwerionenforschung); TIPPEY, Kristin Elizabeth (HZDR); KALUZA, Malte Christoph (Helmholtz-Institut Jena); SCHWAB, Matthew Bradley (Friedrich-Schiller-Universität Jena); BUSSMANN, Michael (HZDR); KNODEL, Oliver (Helmholtz-Zentrum Dresden-Rossendorf); KLUGE, Thomas (HZDR); BAGNOUD, Vincent

Presenter: SCHLENVOIGT, Hans-Peter (HZDR)

Session Classification: Session B1

**Track Classification:** Connecting research data: 6. Interoperable semantics at domain and application level