HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

FAIR data storage for user experiments

Challenges and Opportunities

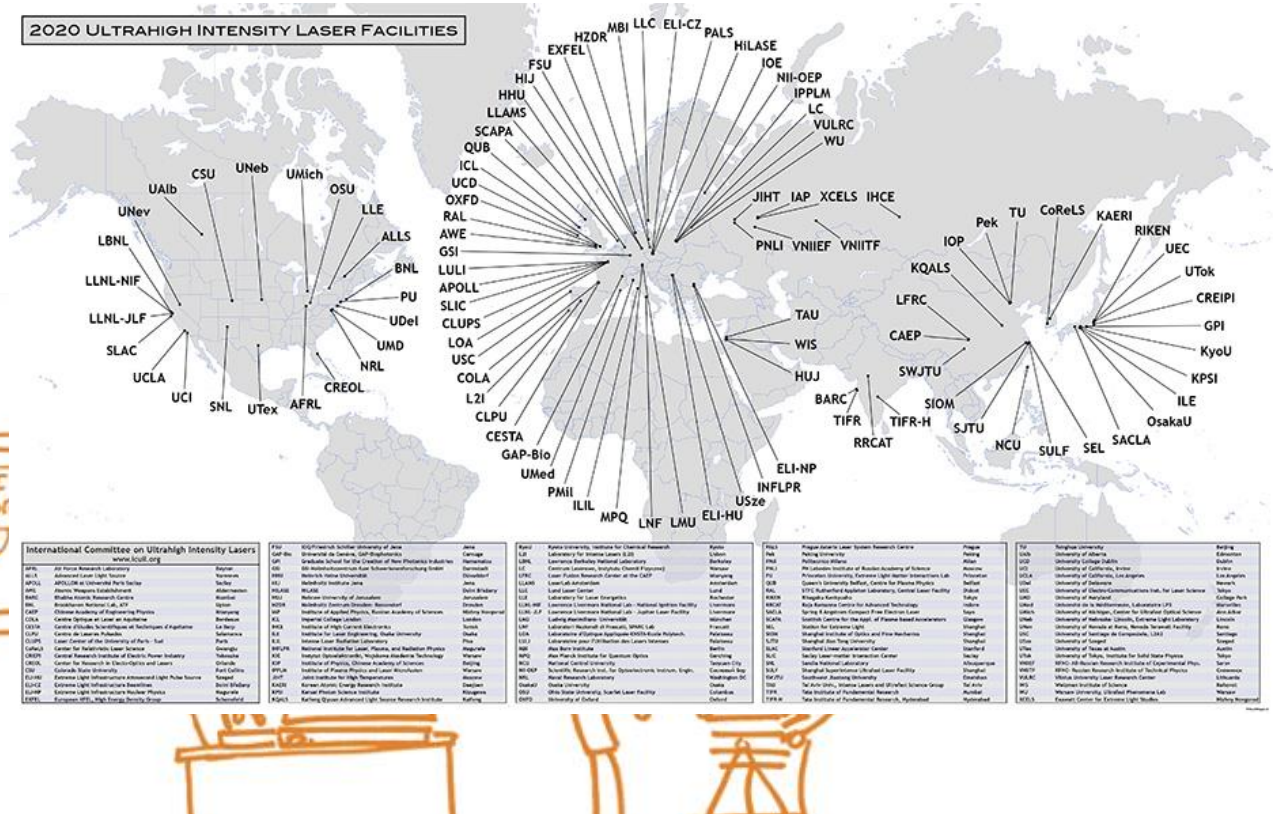
Institute of Radiation Physics · Laser-Particle Acceleration · Dr. Hans-Peter Schlenvoigt · h.schlenvoigt@hzdr.de · www.hzdr.de

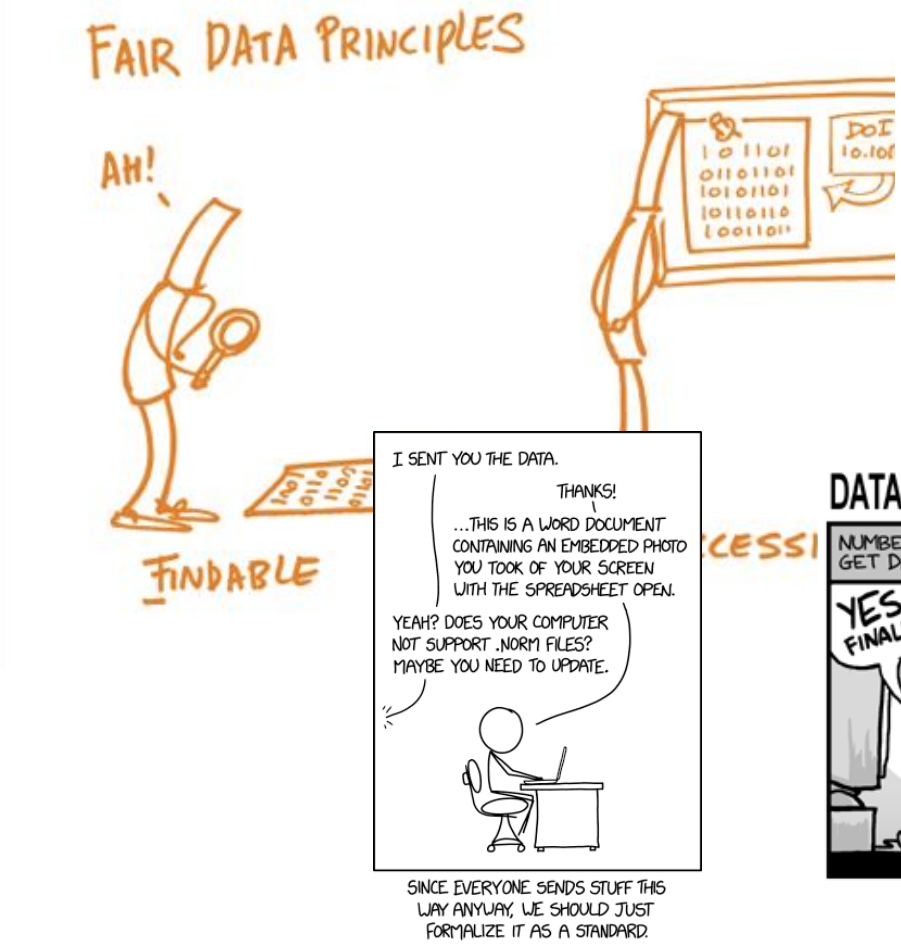
DRESDEN concept

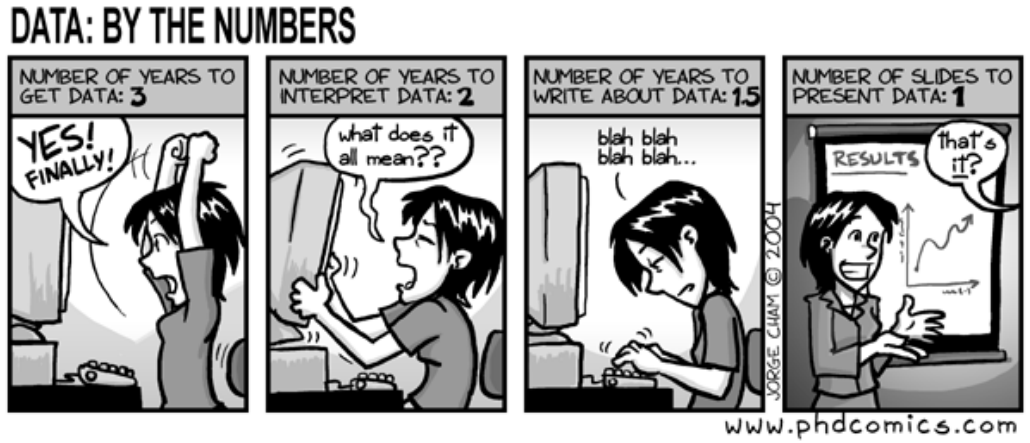
FAIR data – what is that?

https://helmholtz-metadaten.de/storage/369/conversions/Word-Art-(2)-preview-lg.jpg
https://www.openaire.eu/what-is-fair-data https://book.fosteropenscience.eu/

FAIR data and its benefits



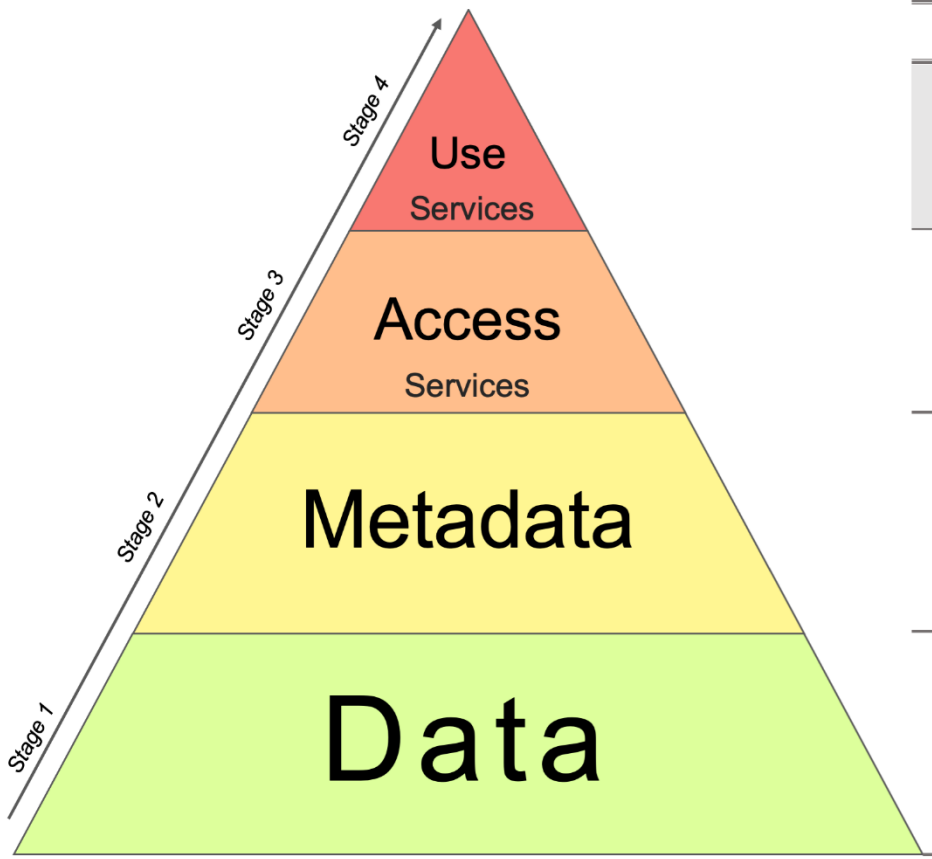




https://www.openaire.eu/what-is-fair-data https://book.fosteropenscience.eu/

xkcd #2116; Phd Comic #462 by Jorge Cham

Standard roadmap to FAIR data storage and services

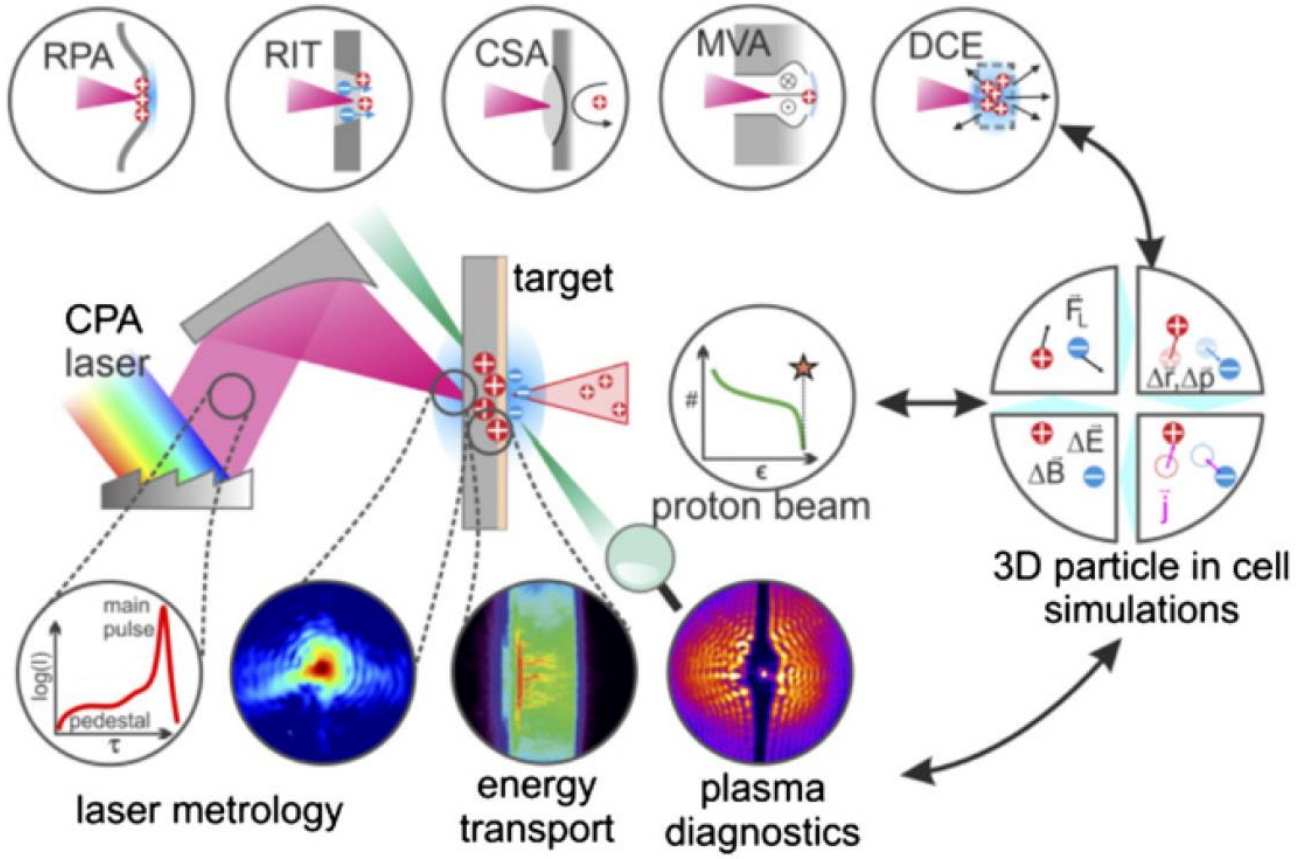


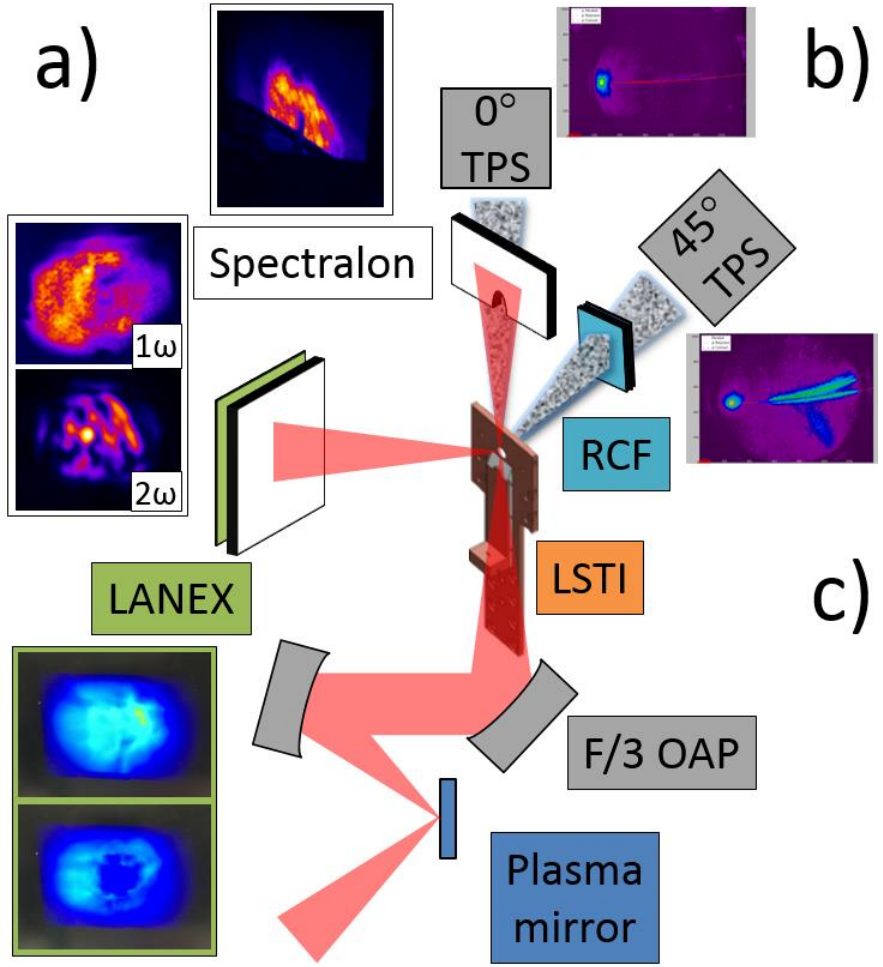
FAIR PRINCIPLES
A1. (meta)data are retrievable by their identifier using a standardized communications protocol. **A1.1.** the protocol is open, free, and universally implementable. **A1.2.** the protocol allows for an authentication and authorization procedure, where necessary. **F4.** (meta)data are registered or indexed in a searchable resource.
F1. Metadata are assigned a globally unique and eternally persistent identifier. **F2.** data are described with rich metadata. **F3.** metadata specify the data identifier. **F4.** metadata are registered or indexed in a searchable resource. **A2.** metadata are accessible, even when the data are no longer available. **I1.** metadata use a formal, accessible, shared, and broadly applicable. **I2.** metadata use vocabularies that follow FAIR principles. **I3.** metadata include qualified references to other metadata. **R1 (R1.1 - R1.2 - R1.3)** Metadata are richly described with a plurality of accurate attributes.
F1. Data are assigned a globally unique and eternally persistent identifier. **F4.** Data are registered or indexed in a searchable resource. **I1.** Data use a formal, accessible, shared, and broadly applicable language for knowledge representation. **I2.** Data use vocabularies that follow FAIR principles. **I3.** Data include qualified references to other (meta)data. **R1.1.** Data are released with a clear and accessible data usage license. **R1.3.** Data meet domain-relevant community standards.

https://zenodo.org/record/3299353#.Ye2Dhvwo-Uk

User experiments: Theory and reality

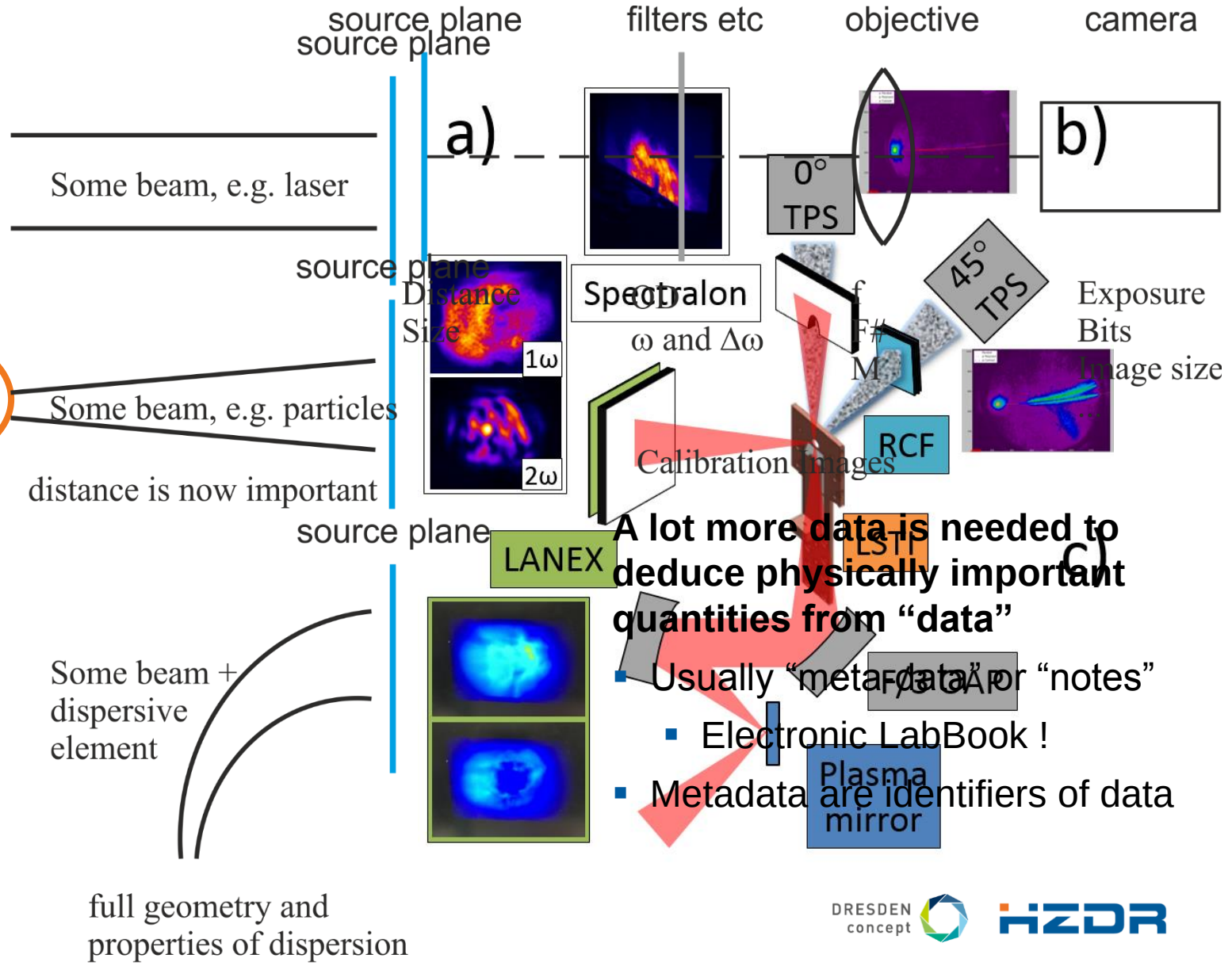
Example on solid-density plasma experiments





2020 roadmap on plasma accelerators (New J. Phys. 23 031101) , Courtesy of L. Obst-Huebl

User experiment breakdown



A lot more data is needed to deduce physically important quantities from "data"

- Usually "meta-data" or "notes"
  - Electronic LabBook !
- Metadata are identifiers of data

Brain dump

openPMD (openpmd.org)

- Meta-Standard for PIC data
- Developed for PIConGPU deployments at large-scale computing facilities
- Adopted by other PIC codes
- Software ecosystem

NeXus (nexusformat.org)

- International common data format for neutron, X-ray and muon science
- Meta-Standard
- Raw data, processed data, geometry data

DAPHNE (DAPHNE4NFDI)

- German project in Photon and Neutron science
- since 2022
- OPEN ! PostDoc position 3ys

PANOSC

- EU project in Photon and Neutron science community
- Part of HORIZON2020 (’14-’20)
- Connection to XFEL.eu

Expands.eu

- European Open Science Cloud (EOSC) Photon and Neutron Data Service

HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

I have a dream…

(not only Martin Luther King)

DRESDEN concept

The vision

FAIR data format

- Standard, for post-processing and archiving
- Both for simulation data and experimental data
- Re-use by others, cost-efficient meta-studies, automated processing (ML)
- Data remains localized, Processing as a service

Facility-dependent features

- Database for facility control, development, long-term studies
- Live data analysis and feedback

HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

To which extent are FAIR principles applicable to Laser-Plasma Experiments?
May it be too ambitious for our flexible use of diagnostics?

If one manages to analyze the own experiments, it is merely matter of data organization for that others could do as well. Standards, workflows and tools will ease that task. User facilities will profit from FAIR and can re-structure staff plans accordingly.

I am an experimentalist.

There is a lot of potential.

DRESDEN concept