A research data management workflow for applied plasma science

Markus M. Becker[1,*], Ihda Chaerony Siffa[1], Hans Höft[1], Fabian Hoppe[2,3], Detlef Loffhagen[1], Nick Plathe[1], Harald Sack[2,3], Volker Skwarek[4], Tabea Tietz[2,3], Simon Tschirner[4], and Laura Vilardell Scholten[1]

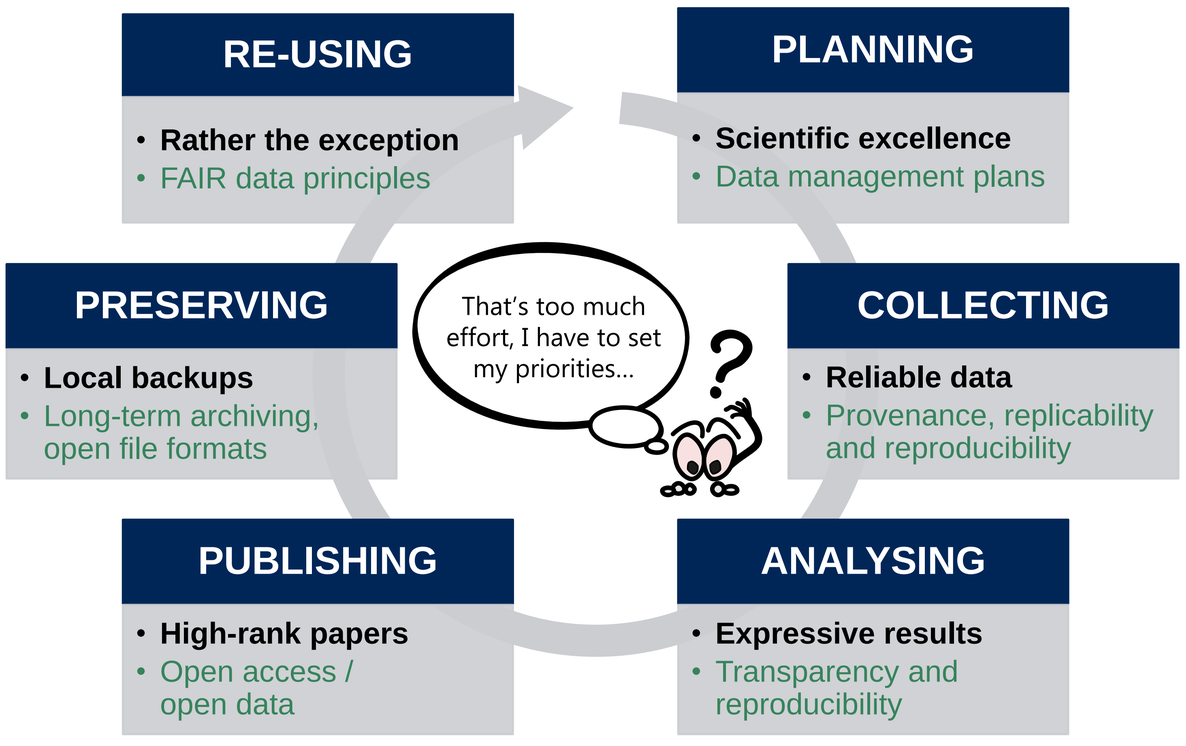
[1]Leibniz Institute for Plasma Science and Technology (INP), Felix-Hausdorff-Straße 2, 17489 Greifswald, Germany
[2]FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH, 76344 Eggenstein-Leopoldshafen, Germany
[3]Karlsruhe Institute of Technology (KIT), Institute AIFB, 76133 Karlsruhe, Germany
[4]Hochschule für Angewandte Wissenschaften Hamburg (HAW Hamburg), Berliner Tor 5, 20099 Hamburg, Germany

*E-mail: markus.becker@inp-greifswald.de

DOI: 10.5281/zenodo.5288495

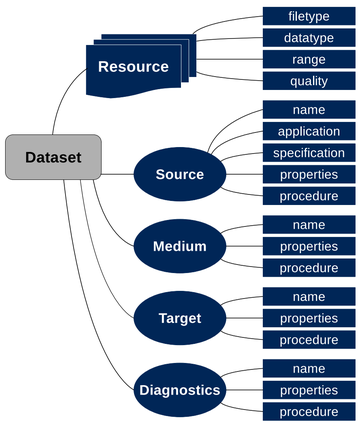
MOTIVATION

- Research data management (RDM) according to the FAIR data principles (Wilkinson 2016) poses new challenges to research processes and complements the primary goals of researchers in plasma science:



FIRST STEPS AND FUTURE REQUIREMENTS

- The plasma metadata schema, Plasma-MDS (Franke 2020) provides core elements for the uniform documentation of data in plasma science.
- Efforts are ongoing to build a common knowledge graph for plasma technology, which facilitates the implementation of the FAIR data principles.
- New workflows and implementations of data management tools are now required in the different labs to balance effort and benefit of RDM, see https://nfdi4phys.de.

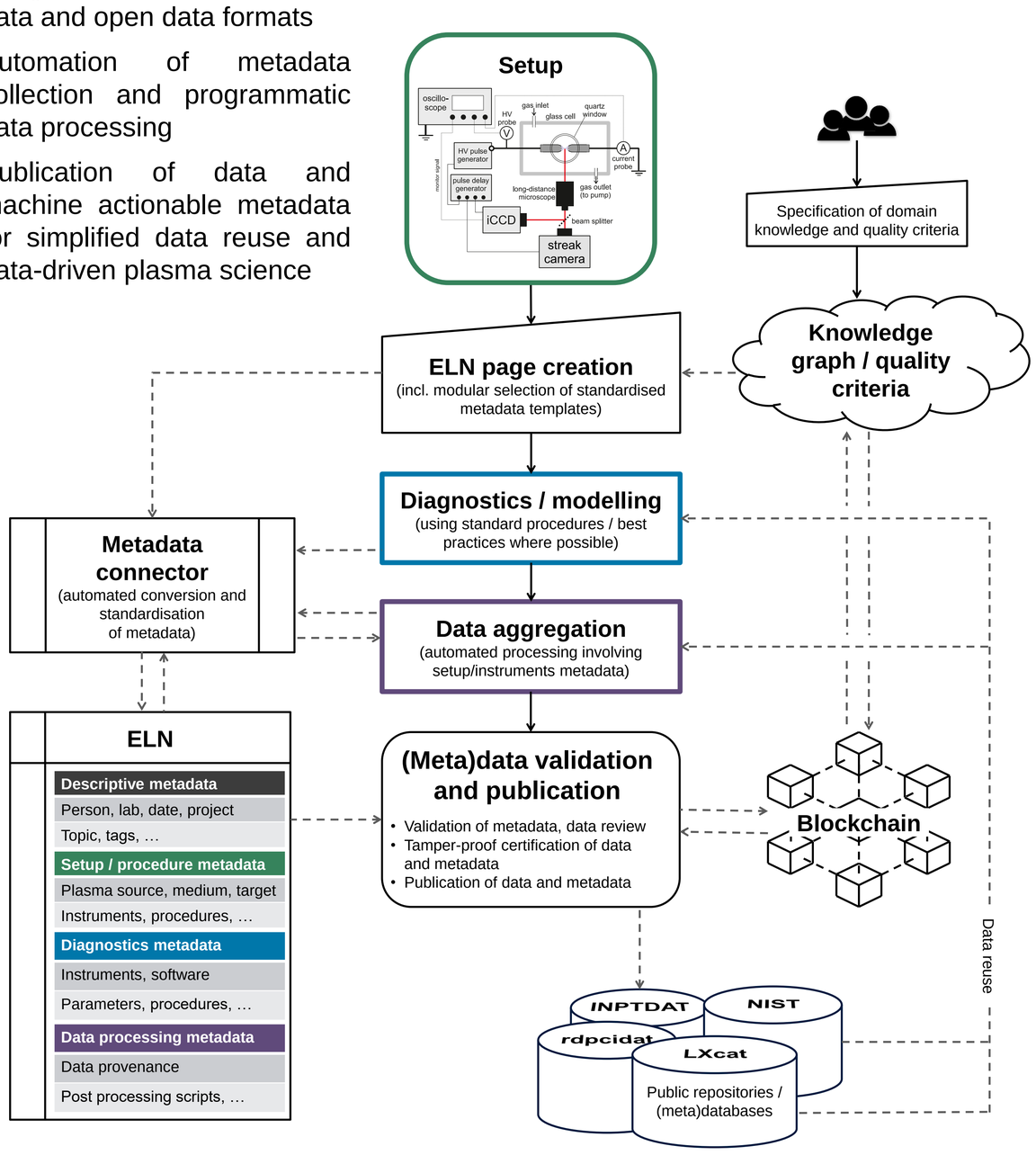


GLOSSARY OF RDM TOOLS

- The **FAIR data principles** are guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets.
- **Electronic lab notebooks (ELN)** are designed to replace paper lab notebooks and bring several advantages such as automated import/export of information.
- **Blockchain** technology provides a basis for tamper-proof time-stamping and certification of research data, hypotheses and methods.
- **Ontologies** serve to create a formal representation of entities in a knowledge graph.
- A **knowledge graph** is a semantic network of real-world entities, like research domains, authors, instruments, and datasets and defines the relationship between them.

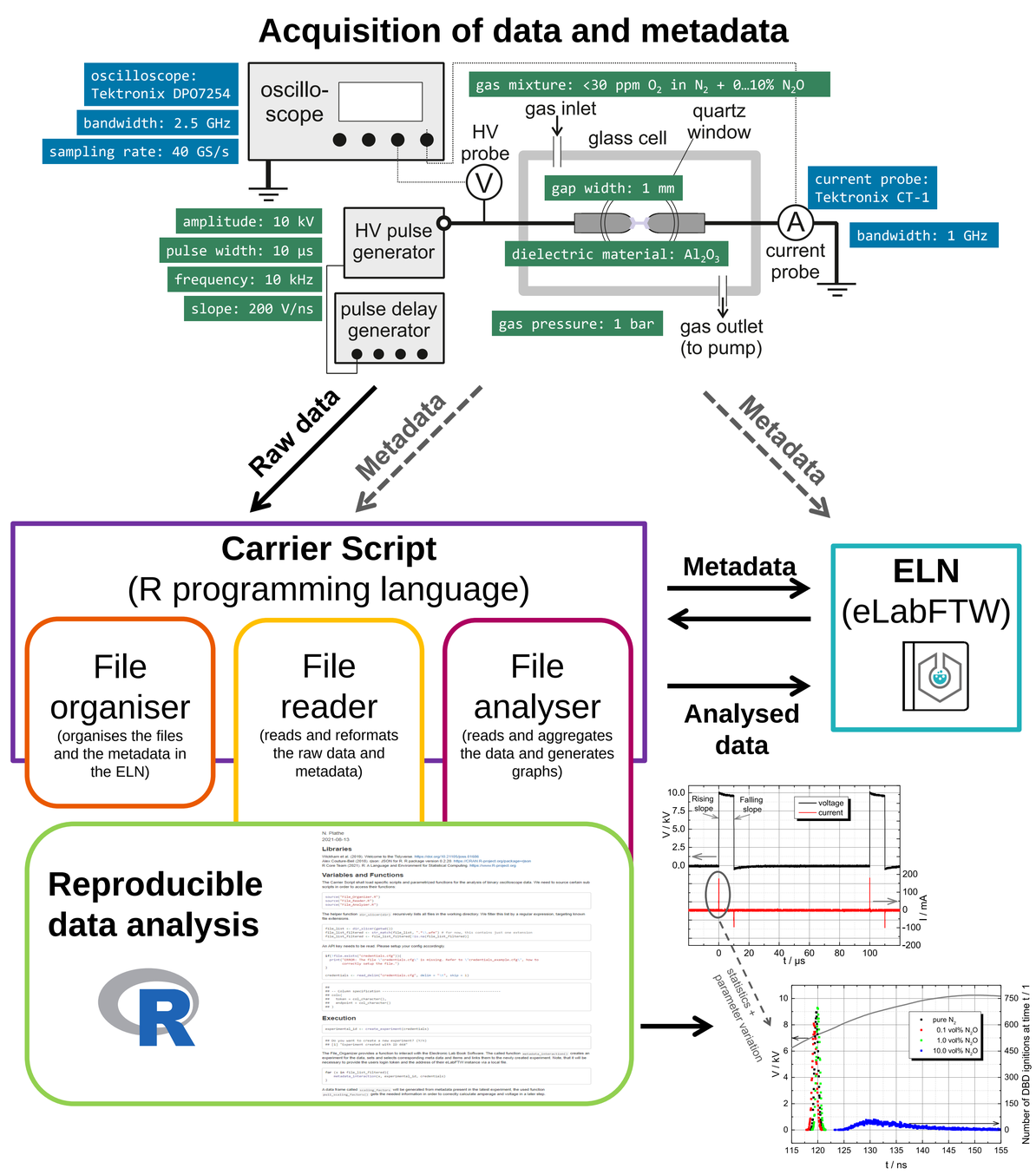
RESEARCH DATA MANAGEMENT WORKFLOW

- Usage of standardised meta-data and open data formats
- Automation of metadata collection and programmatic data processing
- Publication of data and machine actionable metadata for simplified data reuse and data-driven plasma science



WORKFLOW IMPLEMENTATION FOR DBD LAB

- Current measurements with large statistics are used to characterise pulsed dielectric barrier discharges (DBD) at atmospheric pressure.
- Here, the proposed workflow particularly supports the automated and reproducible aggregation of thousands of single-shot measurements.



H. Höft *et al.*, *J. Phys. D: Appl. Phys.* **53** (2020) 025203

References:

M. D. Wilkinson *et al., Scientific Data* **3** (2016) 160018.
St. Franke *et al., Scientific Data* **7**, (2020) 439.
INPTDAT data platform, https://www.inptdat.de.
RDPCIDAT data platform, https://rdpcidat.rub.de.
LXCat plasma data exchange project, https://lxcat.net.
eLabFTW, https://www.elabftw.net.
R: A language and environment for statistical computing, https://www.R-project.org/.

SPONSORED BY THE
Federal Ministry of Education and Research
Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation
This work was supported by the BMBF under grants 16QK03A, 16QK03B, 16QK03C and the DFG under grant 408777255.