


First HELIPORT Workshop

Monday, June 12, 2023 - Wednesday, June 14, 2023

Helmholtz-Zentrum Dresden-Rossendorf (HZDR)

Book of Abstracts

 **Editor:** Oliver Knodel, o.knodel@hzdr.de

 **Authors:** Alexander Kessler, Alexey Ponomaryov, Andrew K. Mistry, Anton Barty, Arie Irman, Astrid Schneidewind, Bernd Schuller, Boxing Gou, Brian Edward Marre, Bridget Murphy, Carina Becker, Carolin Hundt, Chien-Li Lee, Christian Gutt, Christiane Schneide, Claudia Engelhardt, David Pape, Florian Rau, Frank Maas, Frank Schreiber, Friedrich Bethke, Gerrit Guenther, Guido Juckeland, Gunnar Pruß, Hans-Peter Schlenvoigt, Jan-Christoph Deinert, Jan-Dierk Grunwaldt, Jeffrey Kelling, Joachim Hein, Johannes Sperling, Kilian Schwarz, Kristin Elizabeth Tippey, Leon Steinmeier, Lisa Amelung, Malte Christoph Kaluza, Mani Lokamani, Marc Hanisch, Martin Voigt, Michael Bussmann, Moritz Kurzweil, Nico Hoffmann, Nicole Wagner, Oliver Knodel, Oonagh Mannix, Patrick Ufer, Peter Baumgärtel, Ralph Müller-Pfefferkorn, Sebastian Baunack, Sebastian Busch, Sebastian Sachse, Sebastian Starke, Sergey Kovalev, Simone Vadilonga, Stefan Bock, Stefan Mueller, Susanne Schoebel, Thomas Gruber, Thomas Kluge, Tobias Unruh, Wiebke Lohstroh and Wolfgang Horn

DOI: 10.5281/zenodo.8064045

Organised by:




Contents

1 Workshop Introduction	1
Welcome	1
The Lifecycle and History of the HELIPOINT Project	2
2 Use-cases	3
HELIPOINT Ecosystem and Use Cases @ HI-Jena	3
Interactive workflows and data provenance with HELIPOINT	3
Using HELIPOINT in the Context of the HMC Project ALAMEDA	4
UNICORE: overview and recent developments	4
Implementing HELIPOINT for complex workflows at the TELBE user facility	5
Case study of data/metadata flow at Laser Particle Acceleration as use case of HELIPOINT	5
3 Poster Session	6
Data integration via Apache kafka, mongoDB and Grafana for online analysis	6
PUNCH4NFDI: an Update	6
DAPHNE4NFDI: DATA from PHoton and Neutron Experiments for NFDI	7
Workflow-centered analysis of data/metadata sources/flow at Laser Particle Acceleration	7
HELIPOINT — An Integrated Research Data Lifecycle	8
Enabling FAIRer Simulation and Experiment Data Cataloging at Laser Particle Acceleration	8
Enriching Scientific Research with Metadata: Electronic Lab Documentation at HZDR	9
IR of FAIR: Principles at the Instrument Level	9
Toward coherent and controlled data/metadata flow at Laser Particle Acceleration	10
4 Project Metadata	11
SaxFDM-DMP: Developing a DMP Service for Research Institutions in Saxony	11
Metadaten Standard for the HIL-Community	12
An Infrastructure for Harmonizing Semantics and Structures in Research Data	12
FAIR Data of Physical and Digital Beamlines	13
5 Research Data Management and Process Documentation	14
Research Data Management at GSI/FAIR: Progress and Future Prospects	14
Extending HELIPOINT for cleanroom process tracking at HZDR	14
6 Workflows	15
Integrated Data Workflow using HELIPOINT at TELBE	15
Complex workflows in modern computational science using AiiDA and HELIPOINT	16
Workflows and HELIPOINT: Why such a strong integration?	16
7 Develop and Deployment	17
gELBE Beamtime 21102205-ST	17
Keeping the HELIPOINT Code Base Maintainable: Our Tools and Approaches	17
Containerizing HELIPOINT - Development and Deployment with Docker	18
Getting Started With HELIPOINT Development	18
8 Wrap-up and summary	19

Chapter 1

Workshop Introduction

Welcome

 **Authors:** Oliver Knodel¹; Guido Juckeland¹

¹ HZDR

 **Corresponding Author:** o.knodel@hzdr.de

In our HELIPORT workshop, we provided insights into our project and share our results. In addition, we would like to provide a platform for the presentation of similar projects, as well as extensions or integrations from the surrounding research areas. The overall goal of the workshop is bringing together different institutions with similar challenges and establishing a community around our HELIPORT project.

We therefore encouraged our community to submit an abstract for a talk or poster. The submissions are dedicated to four thematic points:


- HELIPORT use-cases,
- Scientific project and metadata management,
- Experiment-specific and overall metadata and
- Scientific workflows.


HELIPORT (Helmholtz Scientific Project WORKflow PlaTform) itself is a project funded by the Helmholtz Metadata Collaboration, and runs from July 2021 until June 2023. HELIPORT aims to make the entire life cycle of a scientific project findable, accessible, interoperable and reusable according to the FAIR principles. In particular, our data management solution deals with the areas from the generation of the data to the publication of primary research data, the workflows carried out and the actual research results. For this purpose, a concept was developed which shows the various essential components and their connections.

Type:

Talk

The Lifecycle and History of the HELIPORT Project

 **Author:** Oliver Knodel¹

 **Co-authors:** David Pape¹; Guido Juckeland¹; Jeffrey Kelling¹; Mani Lokamani¹; Martin Voigt¹; Stefan Mueller¹; Thomas Gruber¹

¹ HZDR

 **Corresponding Author:** o.knodel@hzdr.de

The HELIPORT project aims to make the components or steps of the entire life cycle of a research project at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) and the Helmholtz-Institute Jena (Helmholtz-Institut Jena) discoverable, accessible, interoperable and reusable according to the FAIR principles. In particular, this data management solution deals with the entire lifecycle of research experiments, starting with the generation of the first digital objects, the workflows carried out and the actual publication of research results. For this purpose, a concept was developed that identifies the different systems involved and their connections. By integrating computational workflows, HELIPORT can automate calculations that work with metadata from different internal systems (application management, Labbook, GitLab, and further). This presentation will cover the first year of the project, the current status and the path taken so far in the life cycle of the project.

Type:

Talk

Chapter 2

Use-cases

HELIPORT Ecosystem and Use Cases @ HI-Jena

 **Author** Chien-Li Lee¹

 **Co-Author** Alexander Kessler¹

¹ *HI Jena*

 **Corresponding Author:** c.lee@hi-jena.gsi.de

The newly installed HELIPORT system in POLARIS laboratory is interfaced to POLARIS database, or known as SciCat. By using the SciCat's generic python library called Pyscat, both writing and reading via REST APIs are implemented. In the writing part, LabView programs will first collect data and metadata e.g. from experimental diagnostics and then call Pyscat to transfer this data to SciCat's database. In the reading part, users can manually store the URLs of data within the Heliport project to have easy access later and the users can also view their data in a "table-like" displays by our developed plug-in app within the HELIPORT system. With the help of HELIPORT and SciCat, the data are now one step further towards FAIR.

Type:

Talk

Interactive workflows and data provenance with HELIPORT

 **Author** Jeffrey Kelling¹

 **Co-authors:** David Pape¹; Sebastian Starke¹; Mani Lokamani¹; Oliver Knodel¹

¹ *HZDR*

 **Corresponding Author:** j.kelling@hzdr.de

Data processing or analysis workflows are generally understood as processes running without any user intervention where usually only a small set of parameters being provided upon workflow submission, adjustment of which is also limited by low turnaround rates of workflow runs due to scheduling alone. Many types of experimental data analyses require manual experimentation with parameters to succeed, necessitating interactivity and fast iteration. In

this talk we present examples of interactive workflow applications at HZDR, from data analysis and simulation; discuss challenges arising from differences to completely automated workflows and lay-out the related data-provenance and project-resource management features we envision for the HELIPORT workflow platform.


Type:

Talk

Using HELIPORT in the Context of the HMC Project ALAMEDA

 **Author** Gunnar Pruß¹

¹ *GFZ Potsdam*

 **Corresponding Author:** gunnar.pruss@gfz-potsdam.de

Modern Earth sciences produce a continuously increasing amount of data and metadata from observations, samples and analyses. In the HMC project ALAMEDA, we are developing a platform to manage, visualize and share metadata collected in laboratory and field experiments. We use HELIPORT in conjunction with Business Process Model and Notation (BPMN) to combine data from different software applications already used by the community to manage laboratory, sensor, sample and workflow data.

Type:

Talk

UNICORE: overview and recent developments

 **Author** Bernd Schuller¹

¹ *Forschungszentrum Jülich GmbH*

 **Corresponding Author:** b.schuller@fz-juelich.de

UNICORE (UNiform Interface to COmputing REsources) provides tools and services for building federated systems, making high-performance computing and data resources accessible in a seamless and secure way for a wide variety of applications in intranets and the internet. UNICORE offers comprehensive RESTful APIs for batch job management, data access, data movement and computational workflows.

This talk gives an overview about the current feature set and overall state of the UNICORE ecosystem, presents recent developments and discusses different workflow options (UNICORE native and CWL).

Type:

Talk

Implementing HELIPORT for complex workflows at the TELBE user facility

Authors: Alexey Ponomaryov¹; Jan-Christoph Deinert¹; Sergey Kovalev¹

¹ HZDR

Corresponding Author: j.deinert@hzdr.de

The terahertz (THz) facility TELBE at HZDR provides intense THz sources uniquely suited for the study of nonlinear light-matter interactions. The TELBE THz sources are driven by the ELBE electron accelerator and provide frequency-tunable THz pulses with field strengths of several 100 kV/cm, pulse durations of a few picoseconds and a repetition rate on the order of 100 kHz. These light pulses can be used to excite multiple low-energy degrees of freedom in matter, such as spins, lattice dynamics or collective quasiparticle oscillations. The resulting dynamics are typically probed with femtosecond time resolution using optical lasers. Achieving this time resolution requires strategies to overcome the intrinsic time jitter between the accelerator-based THz pulses and the optical pulses generated by tabletop laser systems. We use a measure-and-sort approach that achieves the required time resolution, but requires the measurement of precise time stamps for each individual light pulse in the experiment. The corresponding high data rates of several GB/min require a fast network and computing infrastructure as well as sustainable concepts for data management and metadata generation. This is all the more important as TELBE is a user facility requiring rapid visualisation and sharing of data and associated metadata. A central hub for planning, monitoring, documenting and archiving the complex workflows at TELBE is therefore highly desirable. The implementation of HELIPORT at TELBE will therefore be a major improvement in terms of automation and meeting the requirements of the FAIR principle.

Type:

Talk

Case study of data/metadata flow at Laser Particle Acceleration as use case of HELIPORT

Authors: David Pape¹; Hans-Peter Schlenvoigt¹; Kristin Elizabeth Tippey¹; Nicole Wagner¹; Oliver Knodel¹; Stefan Bock¹; Thomas Gruber¹; Wolfgang Horn¹

¹ HZDR

Corresponding Author: h.schlenvoigt@hzdr.de

Within HELIPORT, providing guidance to scientific projects and workflows according to the FAIR principles within the entire research experiment lifecycle, the domain- and lab-specific workflows need to be embedded. To do so, we have analyzed the established scientific workflow during experiments in the context of laser-driven particle acceleration, with emphasis on data and metadata sources and their respective occurrences. Among the instances where either machines or humans generate or process data, we have identified where human input or interaction is mandatory and will prevail, as well as chances for automation – still under supervision and control by scientists. We present our strategy for a system of inter-connected databases and management software modules, interfacing to HELIPORT.

Type:

Talk

Chapter 3

Poster Session

Data integration via Apache kafka, mongoDB and Grafana for online analysis

Authors: Arie Irman¹; Friedrich Bethke¹; Kristin Elizabeth Tippey¹; Michael Bussmann¹; Nico Hoffmann¹; Patrick Ufer^{None}; Susanne Schoebel¹

¹ HZDR

Corresponding Author: k.tippey@hzdr.de

Managing high-throughput data incoming from different sources is a major challenge in achieving FAIR data. Experimental data currently traverses a complicated web of machines and python Jupyter notebooks that are used for on- and off-line analyses. This level of complexity in the data pipeline leads to difficulties making the data FAIR. We present a prototype for data integration of electron experimental data, using Apache kafka messaging protocol, MongoDB database, and Grafana + Plotly visualization to ease the data pipeline into being visualized in real-time in a flexible fashion, thereby facilitating more complete knowledge handling.

Type:

Poster

PUNCH4NFDI: an Update

Authors: Andrew. K. Mistry¹; Christiane Schneide²

¹ GSI Helmholtzzentrum für Schwerionenforschung GmbH(GSI)

² DESY

Corresponding Author: a.k.mistry@gsi.de

PUNCH4NFDI is the NFDI consortium for particle, astroparticle, astro-, nuclear and hadron physics. The work of the consortium is organised in seven task areas: management & governance, data management, data transformations, data portal, data irreversibility, synergies & services, education, training, and outreach & citizen science. Here, we give an update on current technical implementations and work-in-progress.

Type:

Poster

DAPHNE4NFDI: DAta from PHoton and Neutron Experiments for NFDI

Authors: Oliver Knodel¹; Bridget Murphy²; Anton Barty³; Astrid Schneidewind⁴; Sebastian Busch⁵; Jan-Dierk Grunwaldt⁶; Christian Gutt⁷; Wiebke Lohstroh⁸; Frank Schreiber⁹; Tobias Unruh¹⁰

¹ HZDR

² CAU Kiel

³ DESY

⁴ FZ Jülich

⁵ GEMS at MLZ, Helmholtz-Zentrum Hereon

⁶ KIT

⁷ Universität Siegen

⁸ TUM

⁹ Universität Tübingen

¹⁰ FAU

Corresponding Author: o.knodel@hzdr.de

The aim of DAPHNE4NFDI is to create a comprehensive infrastructure to process research data from large scale photon and neutron infrastructures according to the FAIR principles. Broadly, we will provide the following infrastructure for the community:

1. Improve metadata capture through consistent workflows supported by user-driven online log-books that are linked to the data collection;
2. Establish a community repository of processed data, new reference databases and analysis code for published results, linked, where possible, to raw data sources;
3. Develop, curate and deploy user-developed analysis software on facility computing infrastructure so that ordinary users can benefit from and repeat the analysis performed by leading power user groups through common data analysis portals.

Type:

Poster

Workflow-centered analysis of data/metadata sources and flow at Laser Particle Acceleration

Authors: Hans-Peter Schlenvoigt¹; Kristin Elizabeth Tippey¹; Nicole Wagner¹; Stefan Bock¹; Thomas Gruber¹

¹ HZDR

Corresponding Authors: h.schlenvoigt@hzdr.de, k.tippey@hzdr.de

We present the analysis of our established scientific workflow during experiments in the context of laser-driven particle acceleration. We focus on the occurrences and types of data generation and associated metadata. The workflow schematics are formatted abstract to allow for comparison to or application in other domains. We recognize that a substantial part of information cannot be entered automatically but relies on human input. Nevertheless, a lot of information, in particular metadata, is transferred manually by scientists, and cross-checks are not recorded. This leads to a high workload during experiments but allows for enormous flexibility.

Type:

Poster

HELIPORT — An Integrated Research Data Lifecycle

Authors: Alexander Kessler⁴; Bernd Schuller¹; David Pape²; Guido Juckeland²; Jeffrey Kelling²; Joachim Hein⁴; Malte Christoph Kaluza⁴; Mani Lokamani²; Martin Voigt²; Oliver Knodel²; Stefan Mueller²; Thomas Gruber²

¹ Forschungszentrum Jülich GmbH

² HZDR

⁴ Helmholtz-Institut Jena

Corresponding Author: d.pape@hzdr.de

HELIPORT is a data management solution that aims at making the components and steps of the entire research experiment's life cycle discoverable, accessible, interoperable and reusable according to the FAIR principles.

Among other information, HELIPORT integrates documentation, scientific workflows, and the final publication of the research results - all via already established solutions for proposal management, electronic lab notebooks, software development and devops tools, and other additional data sources. The integration is accomplished by presenting the researchers with a high-level overview to keep all aspects of the experiment in mind, and automatically exchanging relevant metadata between the experiment's life cycle steps.

Computational agents can interact with HELIPORT via a REST API that allows access to all components, and landing pages that allow for export of digital objects in various standardized formats and schemas. An overall digital object graph combining the metadata harvested from all sources provides scientists with a visual representation of interactions and relations between their digital objects, as well as their existence in the first place. Through the integrated computational workflow systems, HELIPORT can automate calculations using the collected metadata.

By visualizing all aspects of large-scale research experiments, HELIPORT enables deeper insights into a comprehensible data provenance with the chance of raising awareness for data management.

Type:

Poster

Enabling FAIRer Simulation and Experiment Data Cataloging at Laser Particle Acceleration

Authors: Brian Edward Marre¹; David Pape²; Hans-Peter Schlenvoigt²; Kristin Elizabeth Tippey²; Oliver Knodel²; Thomas Kluge²

¹ HZDR Laser Particle Acceleration/Computational Radiation Physics

² HZDR

Corresponding Author: k.tippey@hzdr.de


A major challenge in enabling FAIRer data and metadata is developing and deploying user interfaces that encapsulate clear and consistent metadata schemata. The FWKT Team is building tools for capturing metadata from simulation and experimental datasets through the combined use of stylized user-input forms and scraping of information from existing data structures. Captured information is then uploaded to queryable databases, with the current trials being focused on SciCat and mongoDB. Current sub-projects include metadata extraction toolkits for a variety of simulation codes, a SciCat extension to the laserdata-importer tool for uploading existing experimental data, and a Python-Flask-WTForms-mongoDB Shot-Sheet App for capturing and storing laser shot diagnostics.

Type:

Poster

Enriching Scientific Research with Metadata: Electronic Lab Documentation Landscape at HZDR

 **Author** Thomas Gruber¹

 **Co-authors:** Guido Juckeland¹; Oliver Knodel¹

¹ HZDR


 **Corresponding Author:** t.gruber@hzdr.de

The documentation at our centre is very diverse, which makes it difficult to choose a lab documentation tool, that fits the needs for most of the scientists. One key aspect to increase the usage and acceptance of a shared tool that highly depends on the infrastructure landscape of the research center is the degree of automation to reduce errors and save time. This poster provides some examples how to connect different meta data sources to import and export specific information for a better interaction. This provides a valuable basis of structured information for advanced searches, data processing and file exports, metadata catalogs or other publications. A functional rich API is essential to collect and provide the specific metadata to increase the FAIRness of the experiments.

Type:

Poster

IR of FAIR: Principles at the Instrument Level

 **Authors:** Boxing Gou¹; Frank Maas²; Gerrit Guenther³; Kilian Schwarz⁴; Oonagh Mannix³; Sebastian Baunack²

¹ *Institute of Modern Physics, Chinese Academy of Sciences*

² *Helmholtz-Institut Mainz, Johannes Gutenberg-Universität Mainz*

³ *Helmholtz-Zentrum Berlin*

⁴ *Deutsches Elektronen-Synchrotron DESY*

 **Corresponding Author:** gerrit.guenther@helmholtz-berlin.de

Awareness of the need for FAIR data management has increased in recent years but examples of how to achieve this are often missing. Focusing on the large-scale instrument A4 at the MAMI accelerator, we transfer findings of other projects to improve raw data, i.e. the primary output stored on long-term basis, according to the FAIR principles. Here, the instrument control software plays a key role as the central authority to start measurements and orchestrate connected (meta)data-taking processes. In regular discussions we incorporate the experiences of a wider community and engage to optimize instrument output through various measures from conversion to machine-readable formats over metadata enrichment to additional files creating scientific context.

Type:

Poster

Toward coherent and controlled data/metadata flow at Laser Particle Acceleration

Authors: David Pape¹; Hans-Peter Schlenvoigt¹; Kristin Elizabeth Tippey¹; Oliver Knodel¹; Stefan Bock¹; Wolfgang Horn¹

¹ *HZDR*

Corresponding Authors: h.schlenvoigt@hzdr.de, k.tippey@hzdr.de

Building upon the analysis of our established scientific workflow, we present ongoing improvements and additions for data and metadata handling, in particular a concept of a system of inter-connected databases, which will help to record data and metadata and thereby provide input for other databases. One example is the parsing tool of raw data files, regaining metadata being encoded in data file paths, facilitating data retrieval for analysis. We also present a concept of additional software modules which connect databases, data/metadata sources and data processing. They provide control interfaces for the scientists. As example, we show the prototype of the Draco Laser Shot Counter tool, which forwards data from one subsystem to others and can further software-trigger other processes.

Type:


Poster

Chapter 4

Project Metadata


SaxFDM-DMP: Developing a Data Management Plan Service for Research Institutions in Saxony

 **Author** Carina Becker ¹

 **Co-authors:** Carolin Hundt ¹; Claudia Engelhardt ²; Johannes Sperling ¹; Moritz Kurzweil ¹; Ralph Müller-Pfefferkorn ²

¹ Directorate – Library and Digital Services, Leibniz Institute for the History and Culture of Eastern Europe (GWZO), Leipzig

² Center for Information Services and High Performance Computing (ZIH), TU Dresden, Dresden

 **Corresponding Author:** carina.becker@tu-dresden.de

Data Management Plans (DMPs) are crucial for a structured research data management and often a mandatory part of research proposals. The manual creation of DMPs can be very time-consuming, since many researchers have to start from scratch, are unsure about the required content and may run the risk of not meeting the funder requirements. By using tools, DMPs can be effectively developed and managed. There are a variety of tools to support the development of DMPs: from discipline-agnostic DMP tools, which can be used to generate a generic draft DMP, to discipline-specific DMP tools, which support the creation of a DMP in a specific research field, such as psychology, biodiversity, engineering, or the life sciences. Our aim is to develop a quick and easy to use DMP service for members of Saxon research institutions, building on existing work. In order to evaluate the existing tools and identify which one to choose as the basis for the Saxon tool, we defined 31 requirement parameters covering aspects with regard to basic functions, technical aspects and user-friendliness. Additionally, a weight factor between zero (not relevant) and three (high priority) was assigned to every parameter. Based on these parameters, we evaluated 18 mainly open access DMP tools to find the ones that meet our needs best. Experience from RDM consultancy showed that researchers find pre-fabricated text passages very helpful, which are automatically generated by the DMP tool based on their input. Accordingly, the corresponding requirement parameters are of high importance to us. The only tools generating pre-fabricated text passages while also providing the tool's source code were Data Stewardship Wizard, DataPLAN from NFDI4plants and the DMP tool provided by the TU Dresden Service Center Research Data. The highest total evaluation scores out of a maximum of 790 were reached by Data Stewardship Wizard (703.5), DMPTool (615.5) and RDMO NFDI4Ing (549.5). In a next step, we will check the feasibility of adapting each of the three tools according to our needs and estimate the respective workload. The most suitable DMP tool will then be customized for our requirements.

Type:

Talk

Metadaten Standard for the HIL-Community

 **Author** Alexander Kessler²

 **Co-authors:** Hans-Peter Schlenvoigt ¹; Malte Christoph Kaluza ²

¹ *HZDR*

² *Helmholtz-Institut Jena*

 **Corresponding Author:** a.kessler@hi-jena.gsi.de

HELIPORT tracks the life cycle of scientific data by linking PIDs, DOIs, SW repositories and many other things. If the sources are adequately described by metadata, a good provenance of the results and their findability is achievable. However, there is a problem with the “I” and “R” in the F.A.I.R. principles when a domain does not have an metadata standard. As part of our role as a test use case for HELIPORT, we have started to define an metadata standard for High Intensity Lasers (HIL) and associated experiments. This effort will be continued in the HMC project HELPMI. Based on this, the OpenPMD and NeXuS file formats will be extended and merged as far as possible.

Type:

Talk

An Infrastructure for Harmonizing Semantics and Structures in Research Data

 **Authors:** Florian Rau¹; Leon Steinmeier¹

¹ *Helmholtz Institute Freiberg*

 **Corresponding Author:** l.steinmeier@hzdr.de

One of the main enablers for Interoperability and Reproducibility of scientific research data could be the documentation and harmonization of semantics and data structures. We are developing a concept and prototype for an infrastructure and an end-user graph data editor that support these tasks.

Users of the editor will be able to enter their data and metadata in a graph while getting suggestions on existing semantics and structures. If necessary, they can also document their own semantics. The infrastructure provides a platform for publishing and subsequently harmonizing user-made semantics on both a local scale (e.g. within a project) and at a global scale (e.g. within the community of a scientific domain).

We are looking forward to define interfaces to HeliPort.

Type:

Talk

FAIR Data of Physical and Digital Beamlines

Authors: Gerrit Guenther¹; Oonagh Mannix¹; Peter Baumgärtel¹; Sebastian Sachse¹; Simone Vadilonga¹

¹ *Helmholtz-Zentrum Berlin*

Corresponding Author: gerrit.guenther@helmholtz-berlin.de

Simulations play a crucial role in instrument design as a digital precursor of a real-world object. To preserve the symbiosis of simulated and real-world instrument beyond commissioning we connect the two worlds at the NeXus file level. The instrument section of the produced NeXus file is enriched with detailed simulation parameters where the current state of the instrument is reflected. As a result, the enriched instrument description increases the reusability of experimental data in sense of the FAIR principles. The data is ready to be exploited by machine-learning techniques, such as for predictive maintenance applications, as it is possible to perform simulations of a measurement directly from the NeXus file.

Type:

Talk

Chapter 5

Research Data Management and Process Documentation

Research Data Management at GSI/FAIR: Progress and Future Prospects

 **Author** Andrew. K. Mistry¹

¹ *GSI Helmholtzzentrum für Schwerionenforschung GmbH(GSI)*

 **Corresponding Author:** a.k.mistry@gsi.de

The GSI accelerator facility faces challenges for the institutional management of research data due to the diverse nature of the generated datasets, and these challenges will intensify when the FAIR-facility comes online. The conceptualisation of RDM at GSI is advancing to achieve best practices, and to manage the specific challenges. In addition to the internal RDM goals, GSI is involved with several external open science projects. This presentation shows the development path of RDM at GSI, focusing on some tools that will be on offer. For example: the RDMO DMP tool; an institutional data repository to facilitate dataset-record generation, smaller dataset uploads and connectivity to larger datasets on a data lake; and a new logbook option to complement the existing ELog.

Type:

Talk

Extending HELIPORT for cleanroom process tracking at HZDR

 **Author** Martin Voigt¹

¹ *HZDR*

 **Corresponding Author:** m.voigt@hzdr.de

Currently, nanofabrication processes at the HZDR cleanroom are recorded on a manually compiled paper printout and later transferred to a wiki. This hinders search, automated information processing, and reuse of process parts. As an improvement, a paperless approach is being developed. We integrate sample ancestry and processing history stored in MediaWiki into HELIPORT by developing custom modules. This will allow HELIPORT to collect metadata while guiding through the manual workflows that represent cleanroom processes.

At the current point of progress, we estimate that it may well be feasible and useful to generalize the approach and thereby contribute to possible future efforts of representing samples or manual workflows in HELIPORT.

Type:


Talk

Chapter 6

Workflows

Intergrated Data Workflow using HELIPORT at TELBE

 **Author** Thomas Gruber¹

 **Co-authors:** David Pape¹; Guido Juckeland¹; Jeffrey Kelling¹; Mani Lokamani¹; Martin Voigt¹; Oliver Knodel¹; Stefan Mueller¹

¹ *HZDR*

 **Corresponding Author:** t.gruber@hzdr.de

At the High-Field High-Repetition-Rate Terahertz facility @ ELBE (TELBE), ultrafast terahertz-induced dynamics can be probed in various states of matter with highest precision. The TELBE sources offer both, stable and tunable narrowband THz radiation with pulse energies of several microjoules at high repetition rates and a synchronized coherent diffraction radiator, that provides broadband single-cycle pulses. The measurements at TELBE are data intensive, which can be as high as 20GB per experiment, that can lasts up to several minutes. As a result, the current data aquisition and data analysis stages are decoupled, where in the first step the primary data is processed and stored at HZDR and in a later step, restricted data access is made available to the user for post-processing.

In this presentation, we present an integrated workflow for post-processing of the experimental data at TELBE with in-built exchange of metadata between the experiment control software LabView and the workflow execution engine UNICORE. LabView manages the data storing exchanges metadata with the electronic lab notebook for automated documentation. We also present the guidance system HELIPORT which manages the metadata of the associated project proposal and job information from UNICORE, and integrates with the electronic lab notebook (MediaWiki), providing a user-friendly interface for monitoring the actively running experiments at TELBE.

Type:

Talk

Complex workflows in modern computational science using AiiDA and HELIPORT

 **Author** Mani Lokamani¹

+ **Co-authors:** David Pape¹; Guido Juckeland¹; Jeffrey Kelling¹; Oliver Knodel¹

¹ HZDR

 **Corresponding Author:** m.lokamani@hzdr.de

AiiDA is an open-source Python infrastructure for devising complex workflows associated with modern computational science and streamlining the four core pillars of the ADES model: Automation, Data, Environment, and Sharing. In this contribution, we showcase features of AiiDA like workflow-forging, high-throughput capability and data provenance as implemented in the AiiDA-FLEUR plugin. Finally, we address the possibility of managing AiiDA-projects through HELIPORT.

Type:

Talk

Workflows and HELIPORT: Why such a strong integration?

 **Author** Oliver Knodel¹

+ **Co-authors:** David Pape¹; Jeffrey Kelling¹; Mani Lokamani¹; Martin Voigt¹

¹ HZDR

 **Corresponding Author:** o.knodel@hzdr.de

In the initial HMC application, CWL workflows had a special role in heliport. This contribution explains why this seemed necessary at the time and illustrates what has changed since then. The role of UNICORE as a central component that provides access to our computing resources and also supports workflows has emerged over the course of the project.

It is challenging to support everything that modern workflow languages can support. Instead, we focus in the future on the work with provenance information (visualisation, ...). The execution of CWL will slowly be phased out and moved to stand-alone third-party tools (e.g. Rabix, Toil, UNICORE, ...) which are able to generate provenance metadata according to established standards (wf4ever wfdesc, W3C PROV, ...). This would also allow us to import workflows from other custom software stacks (e.g. spekNG, Eupraxia Notebooks, ...) and move the implementation into the tools themselves instead of creating more HELIPORT plugins.


Type:

Talk

Chapter 7

Develop and Deployment

gELBE Beamtime 21102205-ST

 **Authors** Oliver Knodel¹; Stefan Mueller¹

¹ HZDR

 **Co-authors:** Martin Voigt¹; Robert Ufer¹

 **Corresponding Author:** o.knodel@hzdr.de

One of our first experiments, which we more or less randomly supported in the summer of 2021 at the HZDR, was a detector test as part of our Mu2e collaboration with FERMILAB. We evaluated the performance and data acquisition system of two detectors that will be used to monitor the stop target for the upcoming Mu2e experiment at Fermilab: the high-purity germanium (HPGe) and lanthanum bromide (LaBr) detectors in the presence of the pulsed gamma ray at the gELBE beamline of the ELBE facility at HZDR. The documentation of the experiment in the different systems was a first field test for the deployment of HELIPORT and everything started with the proposal 21102205-ST. This contribution is a short demonstration of the documentation of the gELBE experiment as one of our first HELIPORT use cases.

Type:

Talk

Keeping the HELIPORT Code Base Maintainable: Our Tools and Approaches

 **Author** David Pape¹

¹ HZDR

 **Co-authors:** Martin Voigt¹; Oliver Knodel¹

 **Corresponding Author:** d.pape@hzdr.de

In this talk, different approaches to keeping the HELIPORT code base maintainable will be presented. It will discuss both the tools used to automate various aspects of development and operation of HELIPORT, as well as how certain aspects of development are approached and how the choice of libraries and tooling helps these aspects.


Type:

Talk

Containerizing HELIPORT - Development and Deployment with Docker

 **Author** Marc Hanisch¹

¹ *GFZ Potsdam*

 **Corresponding Author:** marc.hanisch@gfz-potsdam.de

The workshop contribution provides a comprehensive overview of how Docker can be used as a tool for containerizing the HELIPORT software. The aim is to offer developers a fast and easy way to launch a local instance of HELIPORT in development mode and start coding directly. Additionally, a container configuration is provided for productive deployment.

For this purpose, the individual components of HELIPORT are isolated in separate Docker containers. The required software dependencies for the software operation are already installed within the underlying container images, relieving developers from this task. The use of Docker also ensures software portability across various operating systems. Furthermore, the Docker configuration allows for individual customization for productive operation of HELIPORT, accommodating specific requirements such as certificates or secure access to specific areas of the application. The process isolation and targeted resource management provide advantages, particularly in productive environments, for secure software operation.

Type:

Talk

Getting Started With HELIPORT Development

 **Author** David Pape¹; Martin Voigt¹

¹ *HZDR*

 **Corresponding Authors:** d.pape@hzdr.de, m.voigt@hzdr.de

An interactive session to get people started with HELIPORT development.

We will show participants the way around the HELIPORT repository, its structure rooted in how Django applications work, show where to find what you're looking for and where to find the documentation. Then we will set up a local development environment allowing you to run the tests and build the documentation and finally start the development of your own HELIPORT apps.

Please bring you laptop with a working Python (≥ 3.8), Poetry¹ (≥ 1.2) and Yarn² installation if you would like to participate.

¹) <https://python-poetry.org/>

²) <https://yarnpkg.com/>

Type:

Talk

Chapter 8

Wrap-up and summary

Speaker: Oliver Knodel¹

Co-authors: David Pape¹; Guido Juckeland¹; Jeffrey Kelling¹; Mani Lokamani¹; Martin Voigt¹; Stefan Mueller¹; Thomas Gruber¹

¹ HZDR

Corresponding Author: o.knodel@hzdr.de

Our first "HELIPORT Community Workshop 2023" took place at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR).

Three days full of lectures, posters, tutorials and exciting discussions around data management, metadata and our guidance system HELIPORT!

Our workshop marks the end of the official HELIPORT project. Many thanks to the Helmholtz Metadata Collaboration for making this wonderful project possible. But this end is only the beginning of the next part of our journey with HELIPORT. We are on our way building a HELIPORT community together with our colleagues from HIJ, FZJ, GFZ, GSI and the HMC. We look forward to working together to improve and expand HELIPORT.

The final workshop agenda, including abstracts and materials is available in our Indico event.

We thank all participants 🧑🏻 for their attendance and contributions!

Type:

Talk

🧑🏻 Made possible thanks to...



HELMHOLTZ

SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN