HELMHOLTZ
Open Science

Helmholtz Open Science Briefing

5th Helmholtz Open Science Forum: Research Software

February 2024

Report

Impressum

Die Onlineversion dieser Publikation finden Sie unter:
https://doi.org/10.48440/os.helmholtz.074

Verfasser:innen

Christoph Bruch, Lena Messerschmidt, Martin Abbrent, Renē Caspart, Bernadette Fritzsch, Tobias Huste, Guido Juckeland, Uwe Konrad, Alexander Krimm, Christian Mee ßen, David Schäfer, Tobias Schlauch, Vladimir Voroshnin

Herausgeber

Helmholtz Open Science Office

Redaktion

Roland Bertelmann, Christoph Bruch, Lea Maria Ferguson, Marc Lange, Lena Messerschmidt, Heinz Pampel, Antonia C. Schrader, Paul Schultze-Motel

Kontakt

Helmholtz Open Science Office
c/o Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum GFZ
Telegrafenberg, 14473 Potsdam
E-Mail: open-science@helmholtz.de

Stand

18.03.2024. Version. 1.0

Lizenz

HELMHOLTZ

Open Science

1 Contents

2 Introduction

The relevance of research software in scientific work is increasing constantly. The employment of software for research broadens the spectrum of research methods and thus is the precondition of countless research projects. Concurrently, challenges are evolving, e.g. referring to reproducibility, detailed documentation for re-use and managing the various versions of software as a living research product. Obviously, the quality of software needs to be assured. The use of research software necessitates infrastructure for findability, testing, storing, managing. Also, research software should be as open as possible for reuse and further development which implies a suitable licensing.

Research software is therefore taken into account as a major pillar of the Helmholtz Open Science Policy[1] referring to the decision of the Helmholtz Assembly of Members that all Helmholtz Centers are to put in place a research software policy. In order to support the Centers in this endeavor Helmholtz Open Science Office and the Forum Research Software formed by the respective task group[2] of the working group Open Science and HIFIS has developed guidelines like the "Model Policy on Sustainable Software at the Helmholtz Centers"[3] and are regularly staging meetings to exchange about best practices.

The fifth meeting of this series took place 5-6 February 2024 at the Helmholtz-Centre for Environmental Research - UFZ. A key topic of this meeting was the formation of Open Source Program Offices. The term is used to describe a concerted effort for an overarching structure in research organizations where research software engineers, computer departments, research data management units, technology/knowledge transfer units, libraries and legal departments cooperatively aim to build an environment conductive to excellent research, excellent research software, open science and technology/knowledge transfer.

Previous Helmholtz Open Science Fora on Research Software event were held in May 2021[4], April 2022[5], November 2022[6], and May 2023.[7]

[1] https://os.helmholtz.de/en/open-science-in-helmholtz/open-science-policy/
[2] https://os.helmholtz.de/en/open-science-in-helmholtz/working-group-open-science/task-group-research-software/
[3] https://os.helmholtz.de/en/open-research-software/model-policy/
[4] https://os.helmholtz.de/veranstaltungen/foren/1-forum-forschungssoftware/
[5] https://os.helmholtz.de/veranstaltungen/foren/2-forum-forschungssoftware/
[6] https://os.helmholtz.de/en/events/fora/3rd-forum-research-software/
[7] https://os.helmholtz.de/en/events/fora/4th-forum-research-software/

3 Program

Table 1: Program on Monday, February 5, 2024

Time	Program	Speaker
12:45	Welcome, Organizational Matters	
13:00	Report from Helmholtz Open Science Forum "Open Science and Transfer"	Christoph Bruch, Helmholtz Open Science Office
13:10	OSPOs at Helmholtz Centers: Overview of state of play	Vladimir Voroshnin, HZDR
14:30	Break	
15:00	Helmholtz Incubator Software Award	Lena Messerschmidt, Helmholtz Open Science Office
15:15	Software Policy Development: Process and Results 2021-2023	Uwe Konrad, HZDR
15:30	-CANCELLED- From here to there - a long journey to an institute's software policy	Bernadette Fritzsch, AWI
15:45	EOSC-EVERSE: Paving the way towards a European Virtual Institute for Research Software Excellence	Guido Juckeland, HZDR
16:00	News from HIFIS	Tobias Huste, HZDR
16:20	Establishing Software Citation Practices at DLR	Tobias Schlauch, DLR
16:40	Helmholtz Platform for Research Software Engineering - Preparatory Study (HiRSE_PS)	René Caspart, KIT

Table 2: Program on Tuesday, February 6, 2024

Time	Program	Speaker
09:00	Helmholtz Research Software Directory: Status report and roadmap for 2024	Christian Meeßen, GFZ
09:30	Open Research Project Guidance System: HELIPORT	Tobias Huste, HZDR
10:00	Building Scalable Time Series Data Infrastructures: State-of-the-Art Solutions for Accessible and Interoperable Environmental Data	David Schäfer, UFZ
10:30	Break	
11:00	Overview about the EU Cyber Resilience Act	Tobias Schlauch, DLR
11:30	Open-Source Software Development with Industry	Alexander Krimm, GSI
12:00	Closing Discussion	

4 Presentations: Overview

4.1 Report from Helmholtz Open Science Forum "Open Science and Transfer"

Christoph Bruch (Helmholtz Open Science Office) reported from the Helmholtz Open Science Forum "Open Science and Transfer" which was staged online on 22 January 2024. The forum was an opportunity to showcase a wide variety of examples of technology/knowledge transfer. GSI invited local SMBs to use rack space at its data center. DESY, DZNE, and GFZ used the project SoftWert to create a suite of support offers in the context of technology transfer. At HZDR, the transfer policy was updated to better reflect needs of open science. The many citizen science projects within Helmholtz have lifted their cooperation to a new level by forming a Cooperation Across Research Fields (CARF). As an introduction to the following talk on open source program offices, Christoph Bruch referred to presentation "Open Source Program Office at CERN -Concept and Implementation" which vividly illustrated the link between open science. Open source, software and both technology transfer and knowledge transfer.

4.2 OSPOs at Helmholtz Centers: Overview of state of play

Vladimir Voroshnin (Helmholtz-Zentrum Dresden-Rossendorf, HZDR) reported findings from the project OpenTransfer[8], a cooperation of GSI, HZDR, and Leibniz Institute of Photonic Technology, sponsored by the Federal Ministry of Education and Research. He started his talk by explaining scenarios of the use of software across projects, research entities, and companies, thus analyzing the concept open source. He emphasized that open source creates opportunities to align social impact and economic goals. Clearly this is a key motivation for the cooperation of software engineers and researchers in developing and maintaining free open source software. The term open source policy office can refer to a physical unit within an organization, but it can just as well be understood as a label for a bundle of services. This approach fits the necessity to align the development of support services for the production of free an open source software with greatly varying environments such as the individual Helmholtz centers.

4.3 Helmholtz Incubator Software Award

As was announced in the previous Helmholtz Open Science Forum on Research Software, an initiative of Helmholtz Federated IT Services[9] (HIFIS), the Helmholtz Information & Data Science Academy[10] (HIDA), the project „Helmholtz Platform for Research Software Engineering - Preparatory Study"[11] (HiRSE_PS), and the Helmholtz Open Science Office[12] have been collaborating to realize the Helmholtz Incubator Software Award.[13]

[8] https://www.gsi.de/en/work/administration/stabsabteilungen_stellen/technology_transfer/open-transfer
[9] https://www.hifis.net/
[10] https://www.helmholtz-hida.de/en/
[11] https://www.helmholtz-hirse.de/
[12] https://os.helmholtz.de/en/
[13] https://os.helmholtz.de/en/open-research-software/helmholtz-incubator-software-award/

Lena Messerschmidt (Helmholtz Open Science Office) has recapped the developments up to this point. After the publication of the Call for Nominations initiated the two-phase application period in Juli 2023, the Helmholtz Centers were asked to provide up to three nominations each. The nominees were then asked to apply via the Helmholtz Initiative and Networking Fund.

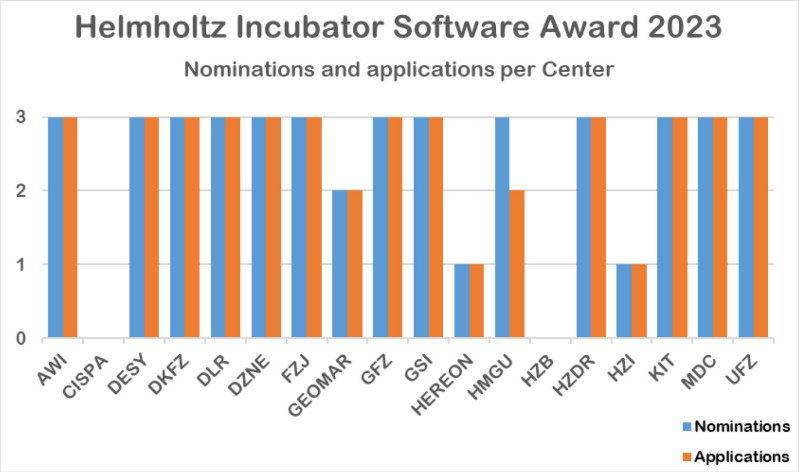


Figure 1. Almost every Helmholtz Center submitted up to three nominations.

Currently, the software entries are being reviewed by an external group of national and international experts in the field of Research Software Engineering. A concluding panel meeting will be held to determine the winners in the three categories "Scientific Originality", "Sustainability", and "Newcomer".

Noteworthy is also the strong connection to the Helmholtz Research Software Directory[14] (RSD). For a software to qualify for the award, it needs to be registered in the RSD with a well-maintained profile. The publication of the Call for Applications to the Helmholtz Incubator Software Award led to a strong increase in software entries in the RSD, as Christian Meeßen reported later on in his contribution on the RSD.

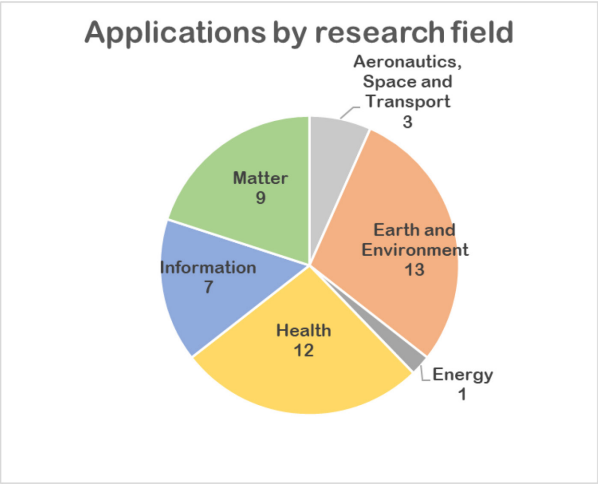


Figure 2. Applications to the Helmholtz Incubator Software Award by research field.

[14] https://helmholtz.software/

4.4 Software Policy Development: Process and Results 2021-2023

In his talk, Uwe Konrad (Helmholtz-Zentrum Dresden-Rossendorf, HZDR) illustrates the process of developing a software policy at HZDR, spanning about two years, which resulted in publication in December 2023. The first draft started out with the Model Policy on Sustainable Software at the Helmholtz Centers[15], and the final policy in its current form also builds upon the software policy[16] published by ForschungsZentrum Jülich (FZJ) and the model of software application classes[17] as proposed by the German Aerospace Center (DLR). The policy combines obligations and support offers with strategic guidance for the choice of an open-source software license in the form of a decision tree. The decision tree has been presented and discussed in this Forum and can be found in its full form in the annex.

4.5 From here to there – a long journey to an institute's software policy

Research software guidelines help ensure the quality of software as a basis for scientific research. On the one side, establishing policies and guidelines provides a reliable basis for developing shared processes and infrastructures, as well as standardized ways of dealing with research software. On the other side, policies are an opportunity to empower developer teams with autonomy based on clarity, direction and defined boundaries. However, the process of implementing a policy in an institute can be challenging. In this talk, Bernadette Fritzsch will discuss some experiences with this process at the Alfred Wegener Institute (AWI, Helmholtz Center for polar and marine research).

[The talk had to be canceled, but the slides are provided in the annex.]

4.6 EOSC-EVERSE: Paving the way towards a European Virtual Institute for Research Software Excellence

Guido Juckeland (Helmholtz-Zentrum Dresden-Rossendorf, HZDR) introduced the upcoming project (scheduled start March 1, 2024) "EVERSE"[18], which is an acronym for "European Virtual Institute for Research Software Excellence". The project is a European endeavor of HZDR along with partners across ten countries. Their aim is to create a European network of Research Software Quality by ensuring best practices for research software curation, quality, and preservation, as well as adopting a three-tier model for research software that captures the varying complexity of software and its development. Additional emphasis will be placed on credit and recognition for developers and software as essential components of the scientific process.

To leverage existing tools and resources, the work will build on existing practices and standards across research communities as represented by the EOSC Science Clusters.[19] Because the Helmholtz Association covers four of the five research fields defined by the EOSC Science Clusters, participation

[15] https://os.helmholtz.de/en/open-research-software/model-policy/
[16] http://hdl.handle.net/2128/33259
[17] https://zenodo.org/records/1344612
[18] https://everse.software/
[19] https://zenodo.org/records/3675081

and representation of Helmholtz in "EVERSE" is especially meaningful. Further details regarding the project structure and expected outcomes can be found in the slides in the annex.

4.7 News from HIFIS

The team of Helmholtz Federated IT Services - HIFIS[20] is distributed across eleven centers and organized into three clusters: backbone, cloud, and software. However, their work affects nearly everyone working at Helmholtz, for example by providing unified user authentication via the Helmholtz ID or the Helmholtz Cloud[21]. In his presentation, Tobias Huste (Helmholtz-Zentrum Dresden-Rossendorf, HZDR) provided insight into the core topics of HIFIS:

- Education and Training: courses, material and workshops related to the Research Software Engineering (RSE) practice
- Community: building communities supporting the cultural change related to RSE, building and maintaining a Helmholtz-wide Research Software Directory[22]
- Consultation: contact points and collections[23] to address RSE-related questions
- Technology: sustainable and well-integrated infrastructure services to support the software development lifecycle

Future topics are also planned to include aspects of cyber- and software security.

4.8 Establishing Software Citation Practices at DLR

Compared to most of the other Helmholtz Centers, the German Aerospace Center (DLR) already has a history of software guidelines due to the nature of space applications. The DLR Software Engineering Guidelines[24] and its Open Source Brochure[25] are available online, but Tobias Schlauch (DLR) and his team are also working on an extensive update of the existing support system. Their goal is to establish a more binding open-source policy that closely follows the Model Policy on Sustainable Software at the Helmholtz Centers[26] that has been published by the Helmholtz Open Science Office in 2019.

Furthermore, their focus is on enabling practices of software citation at DLR that are in line with the guiding principles of software citation.[27] The two essential steps for this are providing citation metadata for the software in a CITATION.cff file and publishing the software in publication repository. Zenodo[28] is a good example for a public publication repository as it provides storage of standardized metadata, archived artifacts, and DOIs as persistent identifiers, but closed-source software should be published in an organizational publication repository as well. To help simplify the process of software publication, the Helmholtz project HERMES[29] develops automated workflows to publish research software with rich metadata.

[20] https://hifis.net/
[21] https://helmholtz.cloud/
[22] https://helmholtz.software/
[23] https://github.com/hifis-net/awesome-rse
[24] https://rse.dlr.de/01_guidelines.html
[25] https://www.dlr.de/de/medien/publikationen/broschueren/opensource-software_dlr_2022.pdf
[26] https://os.helmholtz.de/en/open-research-software/model-policy/
[27] https://peerj.com/articles/cs-86/
[28] https://zenodo.org/
[29] https://project.software-metadata.pub/

4.9 Helmholtz Platform for Research Software Engineering – Preparatory Study (HiRSE_PS)

Software as a research infrastructure is a key component for scientific in all research fields of Helmholtz. The project "Helmholtz Platform for Research Software Engineering - Preparatory Study"[30] HiRSE_PS) is aimed at establishing central activities in Research Software Engineering (RSE) at Helmholtz by providing a testbed for structural RSE support. The tree pillars of their work include so-called Community Software Infrastructure (CSI) groups, a central support and consulting unit, and the on-going HiRSE Seminar series on varying topics of RSE.

René Caspart of Karlsruhe Institute of Technology (KIT) went on to explain the objectives of the different work packages and how they are being implemented in the Helmholtz infrastructure. The CSI groups for example are designed to answer the question of how RSE should best be embedded into research groups. Apart from structural implications, this also heavily relies on establishing a technological basis as well as enabling processes.

The HiRSE Seminar series[31] has been on-going since April 2022 and since then has accumulated over 25 talks, the recordings of which can be accessed on Youtube.[32] Caspart also announced that the KIT with be hosting the first German Summer School on Research Software Engineering in September 2024.

4.10 Helmholtz Research Software Directory: Status report and roadmap for 2024

After introducing the Helmholtz Research Software Directory[33] (RSD) at the previous Helmholtz Open Science Forum on Research Software in May 2023, Christian Meeßen (Helmholtz Centre Potsdam GFZ) gave a status update for the HIFIS Software Community project. Within the past nine months the usage of the RSD has grown significantly, with the number of software entries more than doubled and almost all Helmholtz Centers contributing their software. From looking at the growth timeline it can also be derived that the opening of the call for the Helmholtz Incubator Software Award has influenced the stark increase. An entry in the RSD is necessary for a software to qualify for the Award.

Meeßen also talked over some new features of the platform that have recently been added, like automatic mention scraping and links to personal ORCiD[34] pages. Future plans include a standardized API, a license consultation service, and a better representation of the Helmholtz research fields.

4.11 Open Research Project Guidance System: HELIPORT

In this talk, Tobias Huste (Helmholtz-Zentrum Dresden-Rossendorf, HZDR) introduced the project HELIPORT[35]. "HELIPORT" stands for Helmholtz Scientific Project Workflow Platform and aims to embed the complete scientific lifecycle into a platform to create FAIR and comprehensible project descriptions. It is funded by the Helmholtz Metadata Collaboration and the project members amongst HZDR are

[30] https://www.helmholtz-hirse.de/
[31] https://www.helmholtz-hirse.de/series.html
[32] https://www.youtube.com/@Helmholtz_Platform_for_RSE
[33] https://helmholtz.software/
[34] https://orcid.org/
[35] https://heliport.hzdr.de/

Forschungszentrum Jülich (FZJ) and Helmholtz Institute Jena. It grew out of the challenge to support the many steps of different research experiments within the research fields Matter, Energy, and Health with tools for electronic lab books, interactive analysis, publication of datasets, workflow management, and handle generation and management. HELIPORT's goal is to link these different resources and thereby provide a uniform access to and between all services and systems.

HELIPORT is designed to answer scientists' questions like "How to automate recurring processes and keeping track of status and data products?", "Which data and software can be published, and how?", or "How to onboard new team members into to project lifecycle and associated tools?". HELIPORT does this by describing and collecting all metadata from all services and systems involved in a scientific experiment. The code and documentation are publicly available in the repository.[36] An extensive overview of the architecture and features can be found in the slides in the annex.

4.12 Building Scalable Time Series Data Infrastructures: State-of-the-Art Solutions for Accessible and Interoperable Environmental Data

David Schäfer of UFZ Helmholtz Center for Environmental Research presented his team's approach to building scalable time series data infrastructure. The approach is built around the need for FAIR sensor data in environmental earth sciences. It consists of three components which are separate systems to handle metadata, data, and quality control markers respectively. These systems are collectively hosted inside a Docker environment.

The system for integrated time series management, time.IO[37], as well as the other two components, has been developed by Schäfer and his team at UFZ. It provides distribution, transfer, storage, and visualization of time series data in accordance with FAIR standards. First pilot projects using time.IO are already running at UFZ, with production use planned for this year. In general, it is meant to be part of upcoming large-scale infrastructures like the German National Research Data Infrastructure (NFDI)[38] and the European Open Science Cloud (EOSC)[39] by embedding into the DataHub of the Research Field Earth and Environment[40].

4.13 Overview about the EU Cyber Resilience Act

The Cyber Resilience Act[41] (CRA) is an upcoming EU legislation which establishes general cybersecurity rules for placing software and hardware products on the EU market. Among general cybersecurity rules, the main elements of the law regard obligations for manufacturers, distributors, and importers, requirements, harmonized standards, conformity assessments, reporting, and market surveillance. In its initial draft, the CRA was formulated in a way that could potentially highly disrupt the research software ecosystem. Incalculable risks could have led to unavailability of important open-source projects developed within Helmholtz and beyond.

[36] https://codebase.helmholtz.cloud/heliport/heliport
[37] https://codebase.helmholtz.cloud/ufz-tsm/
[38] https://www.nfdi.de/?lang=en
[39] https://eosc-portal.eu/
[40] https://datahub.erde-und-umwelt.de/en/
[41] https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

After rounds of feedback and re-evaluation, the proposal has been re-worked and now considers open-source specific development models more differentiated. In his presentation, Tobias Schlauch (German Aerospace Center, DLR) highlights the key elements of the new proposal and in which ways it touches open-source software. He also provides a decision graph to determine whether or not an open-source software project is covered by the CRA.

The final text of the CRA is currently under preparation and its final publication is expected for mid-2024. After that, three years will be granted to prepare for full application to the rules.

4.14 Open-Source Software Development with Industry

In his talk, Alexander Krimm of GSI Helmholtz Center for Heavy Ion Research talked about how researchers at GSI use open-source paradigms to collaborate with companies in the private industry. The reason for this need for collaboration with external partners is a very high demand for sustainable software development that cannot be covered by current in-house resources.

After they became dissatisfied with the previous contract model for industry collaboration, Krimm and his team at the FAIR particle accelerator facility[42] developed a new 'agile' and 'lean' framework contract to enable faster and more lightweight dispatch of projects to a pre-qualified pool of industry partners. The selection process for potential partners is guided by a public open-source track record. When beginning a new cooperation, the short-form legal contract is accompanied by an agile process description that focuses on small and re-usable components, live iterations and incentives for efficiency. The guiding principles of this approach are focusing on function over form, building upon existing projects if possible, co-development, and shared code ownership. Most importantly however, collaborations must be based on trust and open and transparent communication.

5 Outlook

The Meeting was concluded with a general discussion on future activities within Helmholtz supporting research software development. One key outcome of this conversation was consensus concerning the importance of the HIFIS services and based on this the necessity to engage in advocacy for continued financial support for the platform.

The report concerning the conflict of objectives resulting from the EU Cyber Resilience Act's aim to increase cyber security and the open source software community's aim to avoid overly burdening software maintenance obligations and liability risks illustrated that support for open source software production includes engagement with law makers. This was seen as a potential task for open source policy offices or comparable structures in Helmholtz.

There was also agreement that the importance of cyber security has greatly increased and that this trend is likely to continue. The consideration in HIFIS to address this issue more prominently was unanimously supported.

[42] https://www.gsi.de/en/researchaccelerators/fair

6 Annex: Presentation Slides

6.1 Report from Helmholtz Open Science Forum "Open Science and Transfer"

HELMHOLTZ
Open Science

Report from

Helmholtz Open Science Forum: Open Science and Transfer

Dr. Christoph Bruch

Helmholtz Association
Helmholtz Open Science Office

Helmholtz Open Science Forum: Research Software — 5-6 February 2024, Leipzig
staged by: Helmholtz Task Group Research Software & Helmholtz Open Science Office
hosted by: Helmholtz Centre for Environmental Research

Various colors of transfer – many stakeholder groups

- Knowledge/technology transfer can be achieved in many ways.
- Open science is one way.
- Since the development of “open policies” concerning publications, data, software (hardware may be next) and trying out new forms of societal engagement, e.g. citizen science, Helmholtz Open Science Office has been creating opportunities for exchange of different stakeholder groups within and outside Helmholtz.
- The latest example is the Helmholtz Open Science Forum: Open Science and Transfer (22 January 2024)

Helmholtz Open Science Forum: Open Science and Transfer

- Open Source Program Office at CERN - Concept and Implementation
  Sünje Dallmeier-Tiessen, CERN

- Das Digital Open Lab der GSI
  Tobias Engert, GSI

- SoftWert – Verwertung von Forschungssoftware erfolgreich gestalten
  Zahra Saleh, DESY; Janine Fischer, DESY; Lisa Wenzel, GFZ

- Citizen Science und Partizipation @Helmholtz
  Julia von Gönner, UFZ; Christin Liedtke, Helmholtz-Gemeinschaft, Geschäftsstelle Berlin

- OpenTransfer – Open Science and Transfer Icebreaker: Open Toolbox for Technology Transfer
  Vladimir Voroshnin, HZDR

- OpenFlexure Microscope - Open Hardware in practice
  Julian Stirling, Freelancer

CERN: Open Source Program Office (OSPO)

What will the OSPO do?

- The OSPO develops and **drives Open Source practices**, **promotes CERN as a contributor** to Open Source development; **asserts CERN as** competent and conscious **user** of Open Source; and **enables CERN's due diligence** related to Open Source.
- The OSPO is an **entry point to CERN's expertise** in Open Source.
- The OSPO is an open and inclusive service, **working internally** across departments with and **for the** entire CERN **community**, and **externally as** a visible **interface** to potential partners and the interested public.
- The OSPO may provide or organise **events, communication, support and training** related to Open Source.

GSI: Digital Open Lab

Digital Open Lab
Real-Labor im Green IT-Cube

Gewinner des Datacenter Strategy Awards 2022 im Bereich Innovation

Damit wurde die Strategie von GSI/FAIR zur Nutzung des Green IT Cube als Reallabor zur Entwicklung neuer Ideen und Innovationen in Zusammenarbeit mit Startups, KMU`s und Forschungsinstituten gewürdigt.
Dabei setzt GSI/FAIR auf die Open Innovation Strategie und die Co-Innovation Strategie

Gemeinsame Entwicklung

- Angebot zur Bereitstellung der Infrastruktur und IT-Kompetenzen von GSI und FAIR für die gemeinsame Entwicklung rund um die Themen HPC, Big Data und ultraschnelle Datenerfassung, einschließlich Softwareentwicklungen und Produkte

Kollaborationen

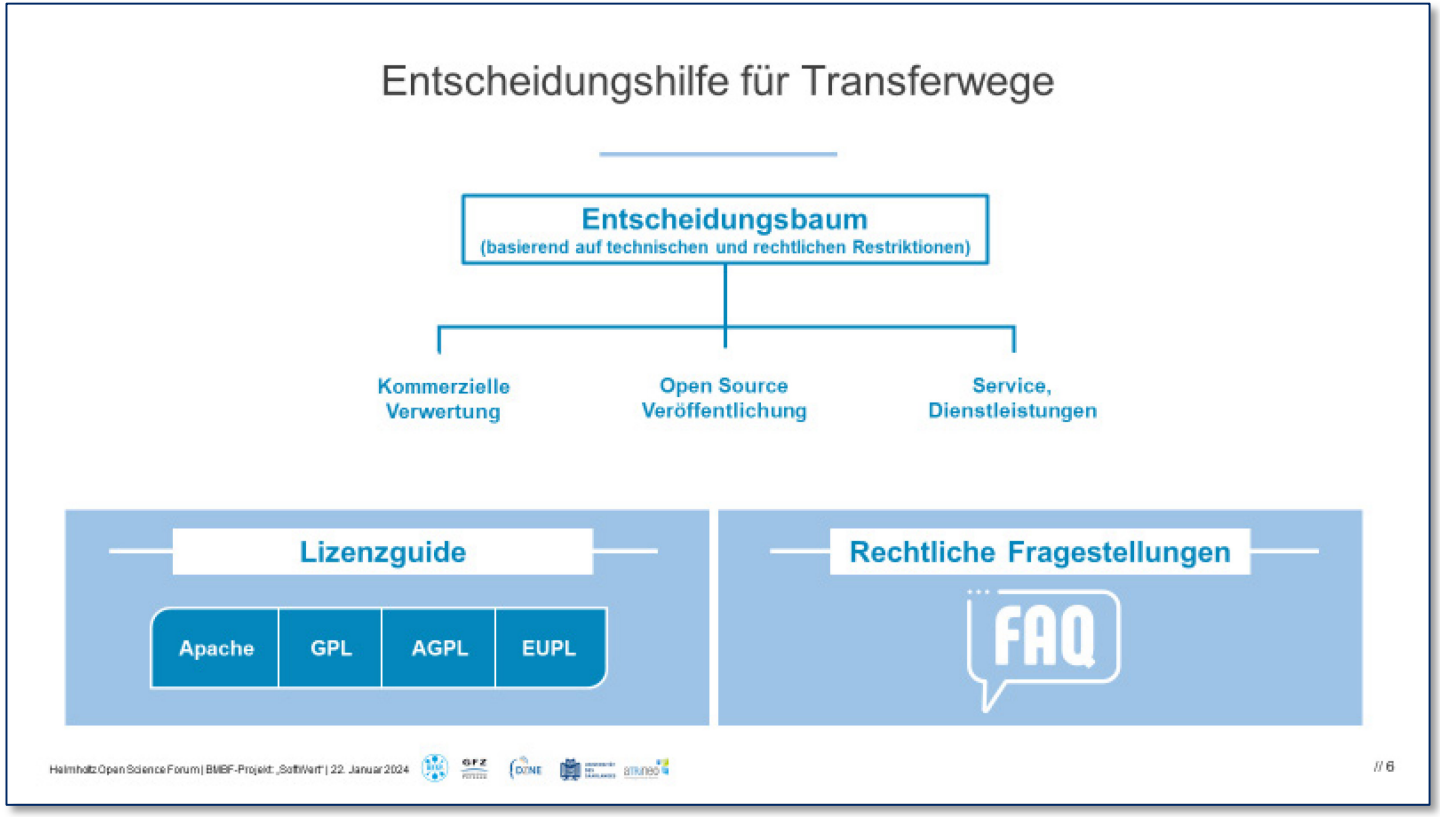
- Zugang zu HPC-Systemen und Projekten für externe Partner über Kooperationsprojekte

Bereitstellung von Rackspace

- Angebot von Dienstleistungen im Rechenzentrum

FAIR GmbH | GSI GmbH Dr. Tobias Engert

DESY/DZNE/GFZ: SoftWert



HZDR: Open Transfer

Open Transfer: Why?

- Miss out opportunities related to OPEN
- Often unnecessarily effort & recourse on protecting knowledge
- Collaboration and synergies bottlenecks

- Take perspective of the end-product
- Do what is necessarily

New OPEN tool box needed!

Citizen Science @ Helmholtz

Presentation at Research Field
Earth and Environment

Start CS@Helmholtz
Feb 17

Helmholtz CS
online survey

WT AK
Mai & Sep 18

CS@Helmholtz ThinkCamp
Mar 19
900K Helmholtz call CS projects

Annual CS@Helmholtz ThinkCamps
2020 2021 2022 2023

Development CARF CitizenScience@Helmholtz

Green Paper Citizen Science Stragegy 2020 for Germany
International conference
May 16

Helmholtz Thinktank
May 18

WT AK Conf
Dec 19

White Paper Citizen Science Stragegy 2030 for Germany
Launch & CS festival
April 22

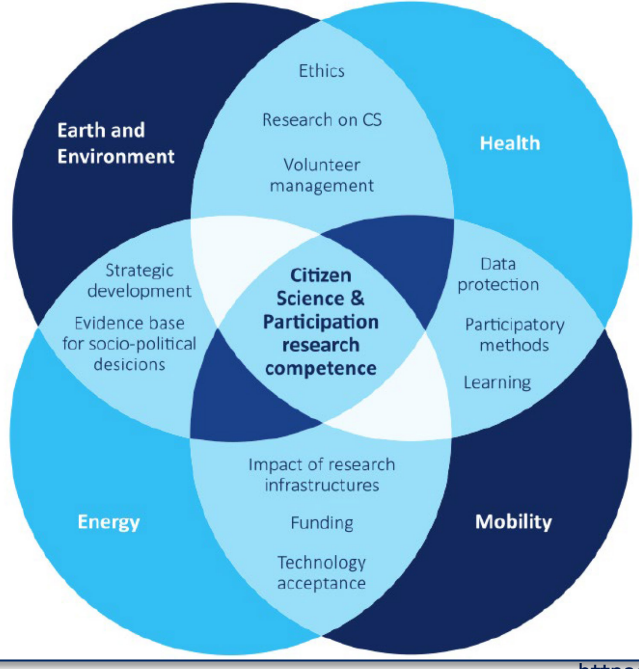
Citizen Science @ Helmholtz

HELMHOLTZ
Citizen Science

Citizen Science and Participation@Helmholtz
Cooperation Across Research Fields (CARF)

HELMHOLTZ
Open Science

- Themenübergreifende Forschungskooperationen zur Stärkung der CS-Kompetenzen & Synergien @Helmholtz
- Strategische Positionierung & Innovation im Bereich Citizen Science
- Citizen Science Qualitätsstandards, Methodik & Indikatorik
  - Wissenschaftliche Exzellenz
  - Gesellschaftliche Wirkung
  - Aktionsplan zur CS-Whitepaper-Strategie



OpenFlexure Microscope - Open Hardware in practice

There is a sliding scale of openness

- Everything Patented
- Design files private
- Trade secrets

- Openly licensed
- Full history of original design files is available
- Design intent explained
- **Move the conversation online**
- Fully documented
- Fully open toolchain

Why build another medical microscope?

How much does a commercial microscope cost?

Purchase cost ~£20,000

Maintenance: Parts cost + **engineer travel**

Thanks for your attention!

HELMHOLTZ
Open Science

- E-Mail: open-science@helmholtz.de
- Mastodon: @HelmholtzOpenScienceOffice@helmholtz.social
- Website: https://os.helmholtz.de
- Mailing list for members of Helmholtz:
  Helmholtz Open Science Professionals
- Helmholtz Open Science Newsletter

6.2 OSPOs at Helmholtz Centers: Overview of state of play

Open Transfer

GSI

Leibniz ipht
LEIBNIZ-INSTITUT für PHOTONISCHE TECHNOLOGIEN

HZDR
HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Presenter: Dr. Vladimir Voroshnin

OSPOs at Helmholtz Centers: Overview of state of play

Plan:

- OSPO Introduction,
- Our ideas
- Discussion on how to we converge?

Published at zenodo.org/communities/opentransfer

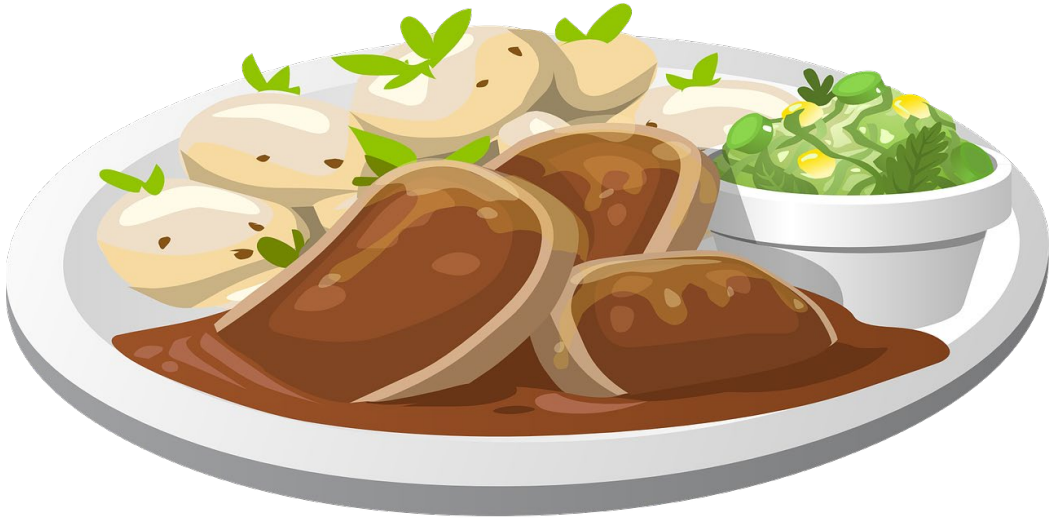
OSPO: Open Source Project Office

On open source is about

- **code, text, creative materials (copyright law)**
- **collaboration and co-development culture**
- **open licensing**
- **partnerships and collaborations (also commercial)**

Metaphor: Recipe vs Meal

vs



OSPO: Open Source Project Office

Managing an organization's open source efforts

-> widely known in industry

- **Open source policies, guidelines, instructions, …**
- **Ensuring compliance with licenses**
- **Aligning with commercial and social impact goals**
- **Fostering community engagement and synergies**
- **Promoting the use (utilization) of open source**

OSPO: Reason for existence / added value

How to tackle when it comes to the details?

- Open source is a tool to be used (e.g., better science, commercial, …)
- Processes and administrative support
- Decision making support
- Legal support
- Business and partnership support

- Connecting and building expertise
- Creating efficient and structured process

OSPO: parts to be covered

How to tackle when it comes to the details?

- **Software (already do it)**
- **Hardware (to be developed)**
- **Data (not quite copyrighted)**
- **Materials (e.g., educational)**

What is the goal and what should be done?

OSPO: actors

Combining different perspectives

- Software developers
- Software and data infrastructure developers
- Researchers (content producers)
- Administration (financing)
- Legal (legally clean)
- Technology Transfer (utilization)

OSPO: two main perspectives

- **How To: constancy, education, tools…**
- **Decisions, outcomes and responsibility**

OSPO: at HZDR

We anyway do it, now it's more transparent and structured

- **GitLab**
- **Life documentation**
- **Involving different perspectives**

- **Software**
- **Hardware**
- **Data**
- **Content**

OSPO: Structured discussion

How to reach synergy (1+1 > 2)

- What is the need from our centers regarding OSPO?
- What we already do regarding open?

  -> what efforts exist in centers

  -> How to make it more systematic
- What we can share from what we already have?
- What we don’t have?
- What we need to make it together

6.3 Helmholtz Incubator Software Award

No slides

6.4 Software Policy Development: Process and Results 2021-2023

HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

Software Policy Development

Process and Results 2021-2023

Uwe Konrad, HZDR Policy Team

Information Services and Computing · Dr. Uwe Konrad · u.konrad@hzdr.de · www.hzdr.de/fwc

HZDR- Software Policy - Timeline

Policy Development Process

Initial phase
» Decision by the HZDR board
» Set up WG
» Invitation of the expert team

Create first draft
» Define scope
» Start with **Helmholtz policy template**
» Discuss **guidelines and legal aspects**

Review first draft
» Broad review of all stakeholders

Create second draft
» Use **FZJ Policy** as starting point
» Discuss **DLR application classes**
» Create open source **license guideline**

Review second draft
» Final Review of all Stakeholder
» Involve HZDR **Digitalization Team**

Signature
» Signature by science and administration

Aug. 2021 | **Sept. 2021 - Feb. 2022** | **Feb.-May 2022** | **May 2022 – May 2023** | **Jun.-Okt. 2023** | **Dez. 2023**

Kick-off
» Define Goal, Roles and Process
» **Invite Team:**
SW Development,
SW Engineering,
Transfer Office,
Legal Office,
IT and Library

Result
» The policy should **be more specific**
» Classification according to **application classes** missing
» Improve guidelines **for license selection**
» SW **publication process** to be described in detail

Result
» The policy is **well accepted**
» To describe the **support services is** important
» The **digital publication process** is analogous to the publication of articles and data

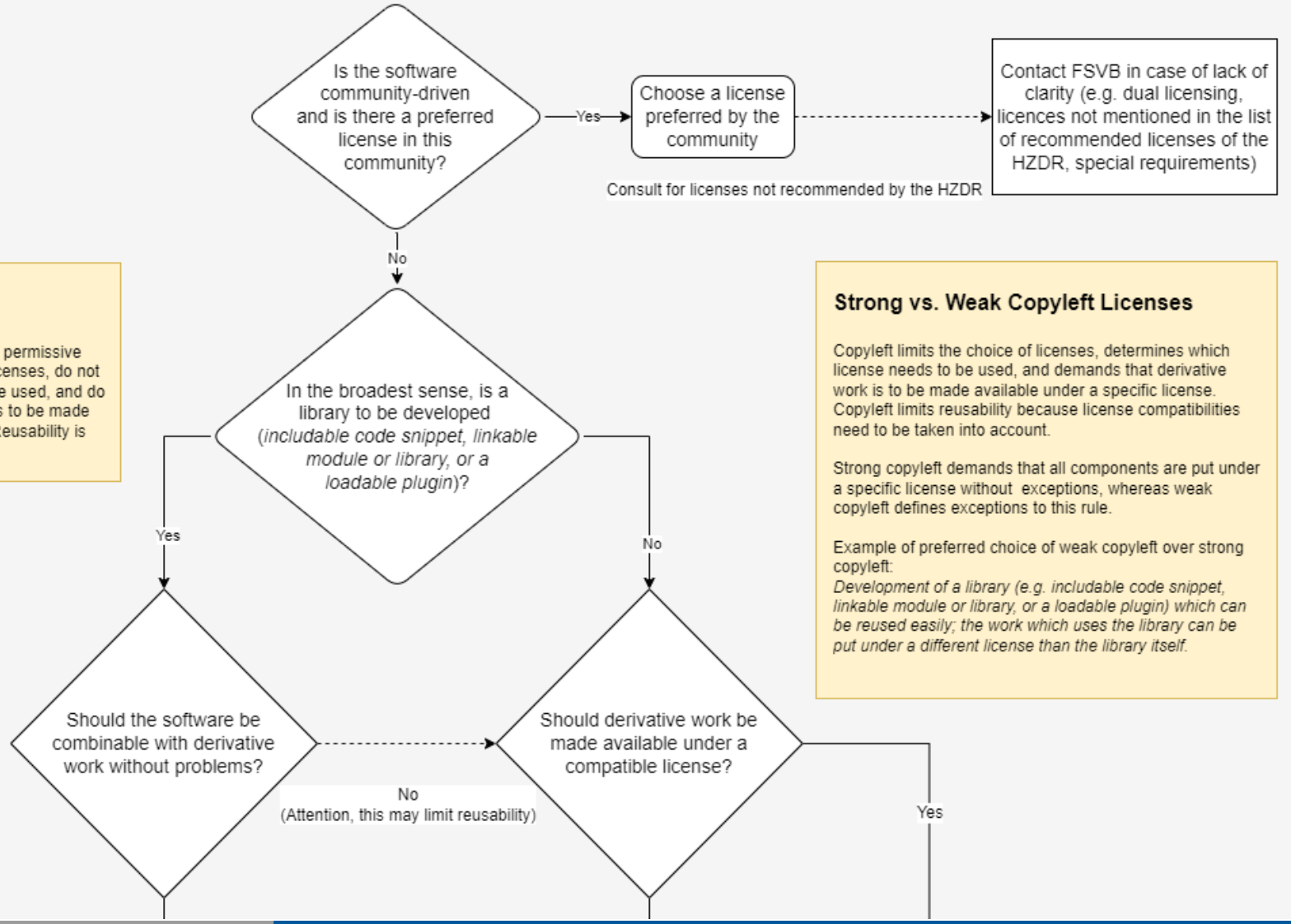
The HZDR Software Policy

Inhaltsverzeichnis

Anlagenverzeichnis

HZDR Software Policy – OS License Choice (1)

Strategic guidance for the choice of an open-source software license



Permissive Licenses

In contrast to licenses with copyleft, permissive licenses do not limit the choice of licenses, do not determine which license needs to be used, and do not demand that a derivative work is to be made available under a specific license. Reusability is greatest with permissive licenses.

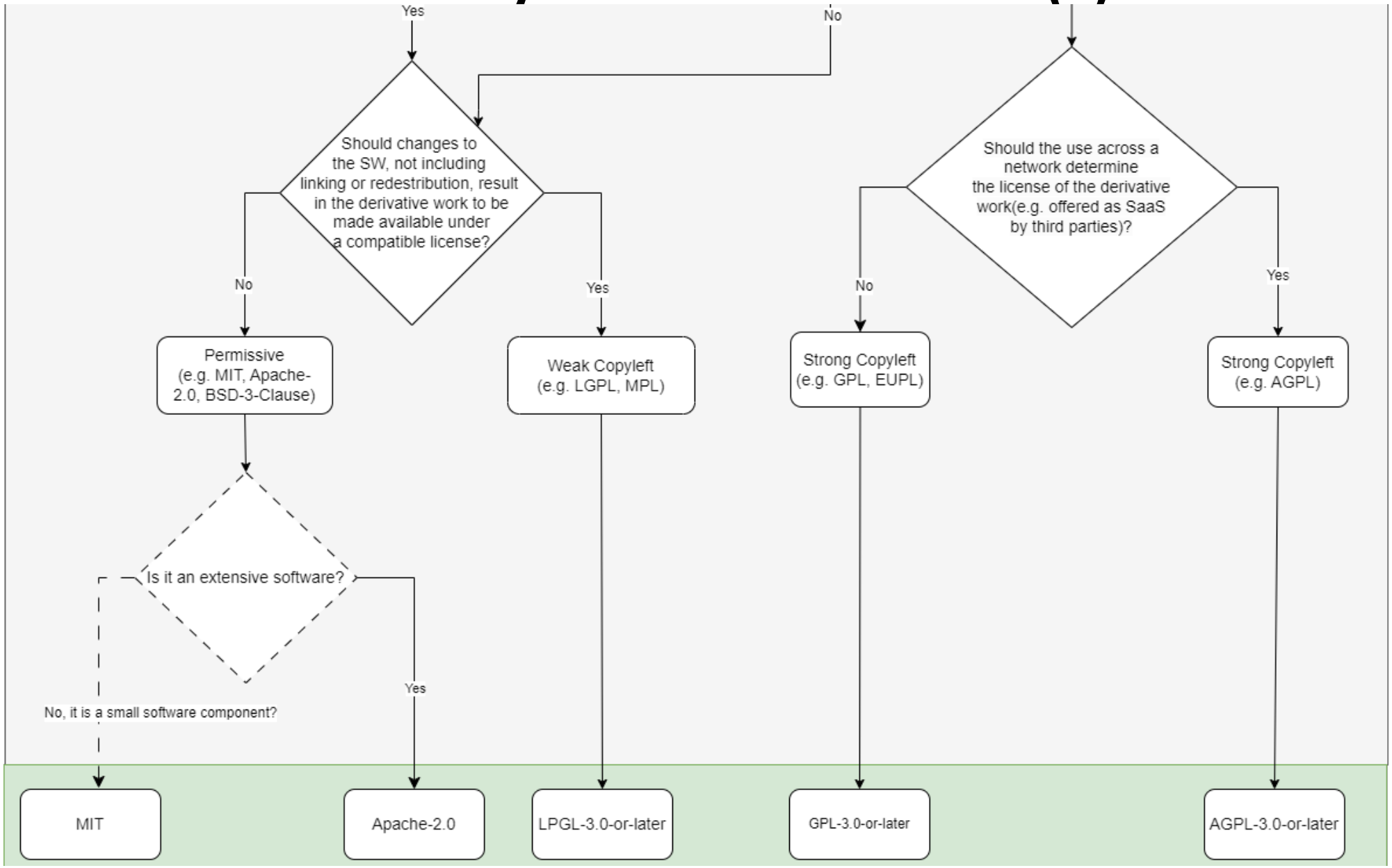
Strong vs. Weak Copyleft Licenses

Copyleft limits the choice of licenses, determines which license needs to be used, and demands that derivative work is to be made available under a specific license. Copyleft limits reusability because license compatibilities need to be taken into account.

Strong copyleft demands that all components are put under a specific license without exceptions, whereas weak copyleft defines exceptions to this rule.

Example of preferred choice of weak copyleft over strong copyleft:
Development of a library (e.g. includable code snippet, linkable module or library, or a loadable plugin) which can be reused easily; the work which uses the library can be put under a different license than the library itself.

HZDR Software Policy – OS License Choice (2)



Thank you for your attention !

m.bussmann@hzdr.de
c.hueser@hzdr.de
t.huste@hzdr.de
u.konrad@hzdr.de
m.moravcikova@hzdr.de
b.wolf@hzdr.de
n.wagner@hzdr.de and many more….

Special thanks for the work of the teams at DLR, FZJ, GFZ, the Helmholtz Open Science Office and the software working group!

6.5 From here to there – a long journey to an institute's software policy

From Here to There - a long journey to an institute‘s software policy

Bernadette Fritzsch

Stefan Hendricks/AWI

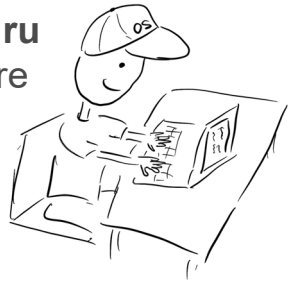
DOI 10.5281/zenodo.10605162

HELMHOLTZ

People behind Research Software

The Reproducibility Guru

- Learns lots of software tools in order to make his research reproducible



The Researcher

- Needs analysis scripts (or other software skills) for her research
- Learns what she needs

The Geek

- Writes software as part of her research project
- Would like to code more, but needs to think about her career and write papers

The Software Person

- Is hired to work on software for a research project

The Go-to Person in case if problems

- Knows how to solve all kinds of computer problems
- Is hired to work on other things but is kind enough to help because he likes it

Research Software in Helmholtz

Embedded in Open Science

2016
Task Group FoSW,
Workshop

2017
„Empfehlungen zur Implementierung von Leit- und Richtlinien zum Umgang mit wissenschaftlicher Software “
https://doi.org/10.2312/os.helmholtz.008

2019
„Model Policy on Sustainable Software at the Helmholtz Centers “
https://doi.org/10.2312/os.helmholtz.007 (German),
https://doi.org/10.48440/os.helmholtz.041 (English).

2021
„Checklist to Support the Helmholtz Centers in Implementing Policies on Sustainable Research Software “
https://doi.org/10.48440/os.helmholtz.031 (German)
https://doi.org/10.48440/os.helmholtz.038 (English)

Research Software Policy at AWI (1)

2017
„Empfehlungen zur Implementierung von Leit- und Richtlinien zum Umgang mit wissenschaftlicher Software “
https://doi.org/10.2312/os.helmholtz.008

2019
„Model Policy on Sustainable Software at the Helmholtz Centers “
https://doi.org/10.2312/os.helmholtz.007 (German),
https://doi.org/10.48440/os.helmholtz.041 (English).

2021
„Checklist to Support the Helmholtz Centers in Implementing Policies on Sustainable Research Software “
https://doi.org/10.48440/os.helmholtz.031 (German)
https://doi.org/10.48440/os.helmholtz.038 (English)

2016
Task Group FoSW, Workshop

2020
Mandate from IT Board

2020
Writing Team

2021
First version of AWI Research SW Policy presented at ED

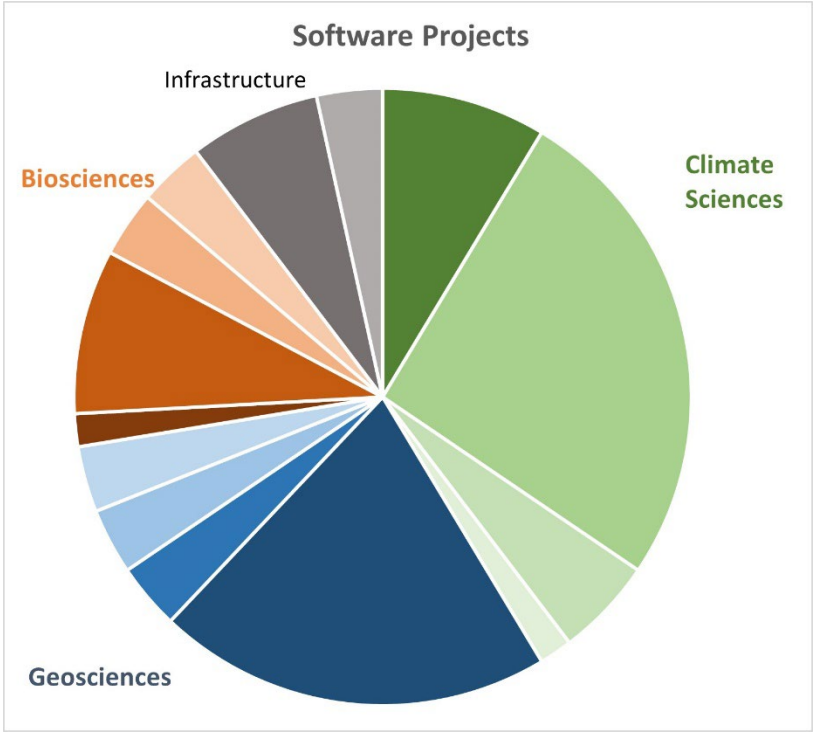
2021
Survey „Research Software at AWI“

Research Software at AWI - Survey

Survey on Research SW at AWI

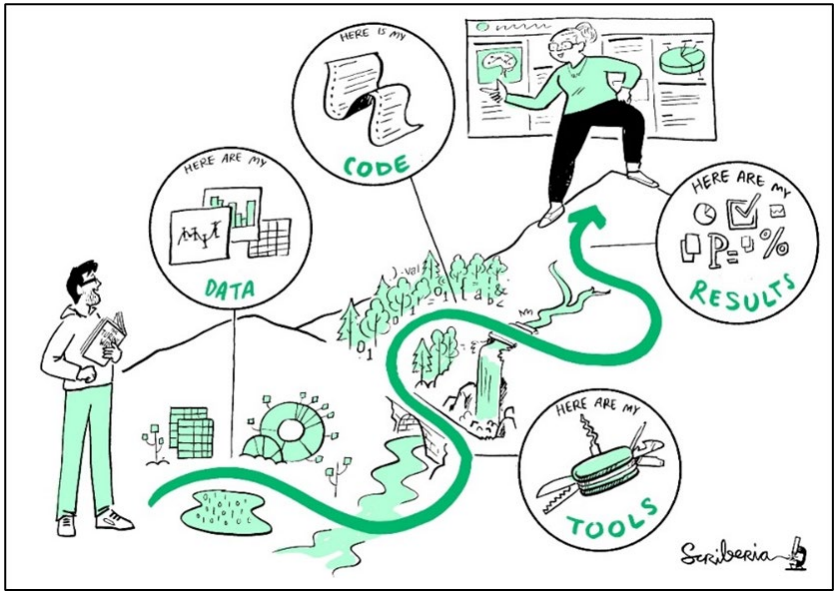
- 112 responses
- different quality of input
- 69 Research Software projects mentioned;
  55 of them further specified (name, aim, …)

➔Research Software is developed in all divisions

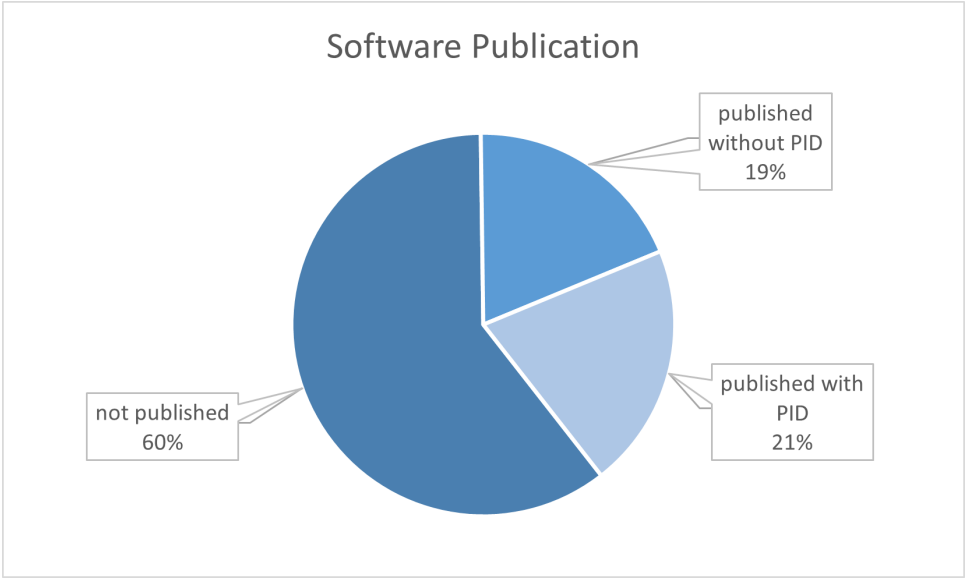


Survey results – Software publications

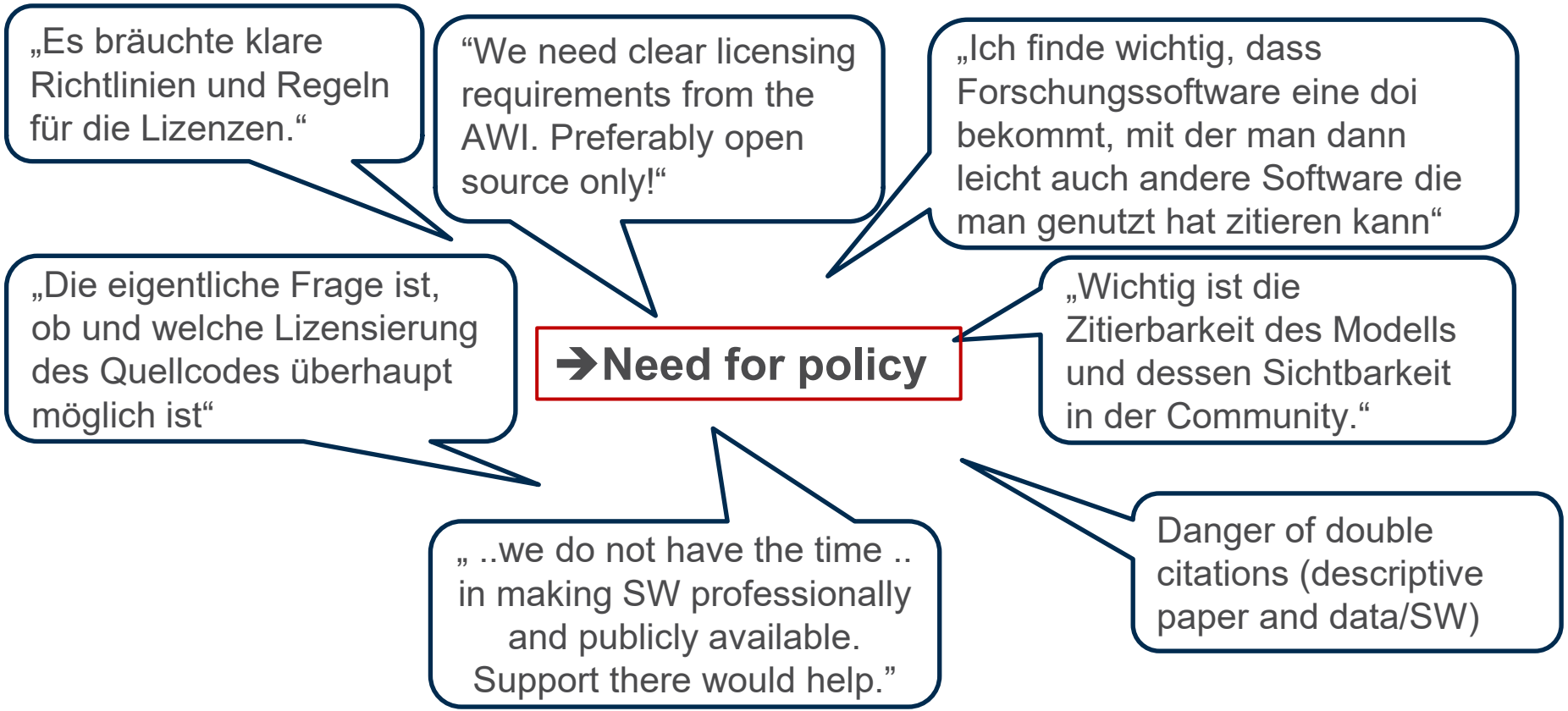
- Most software not published yet
- Software publication not well known
- Changes now with collecting the new indicators



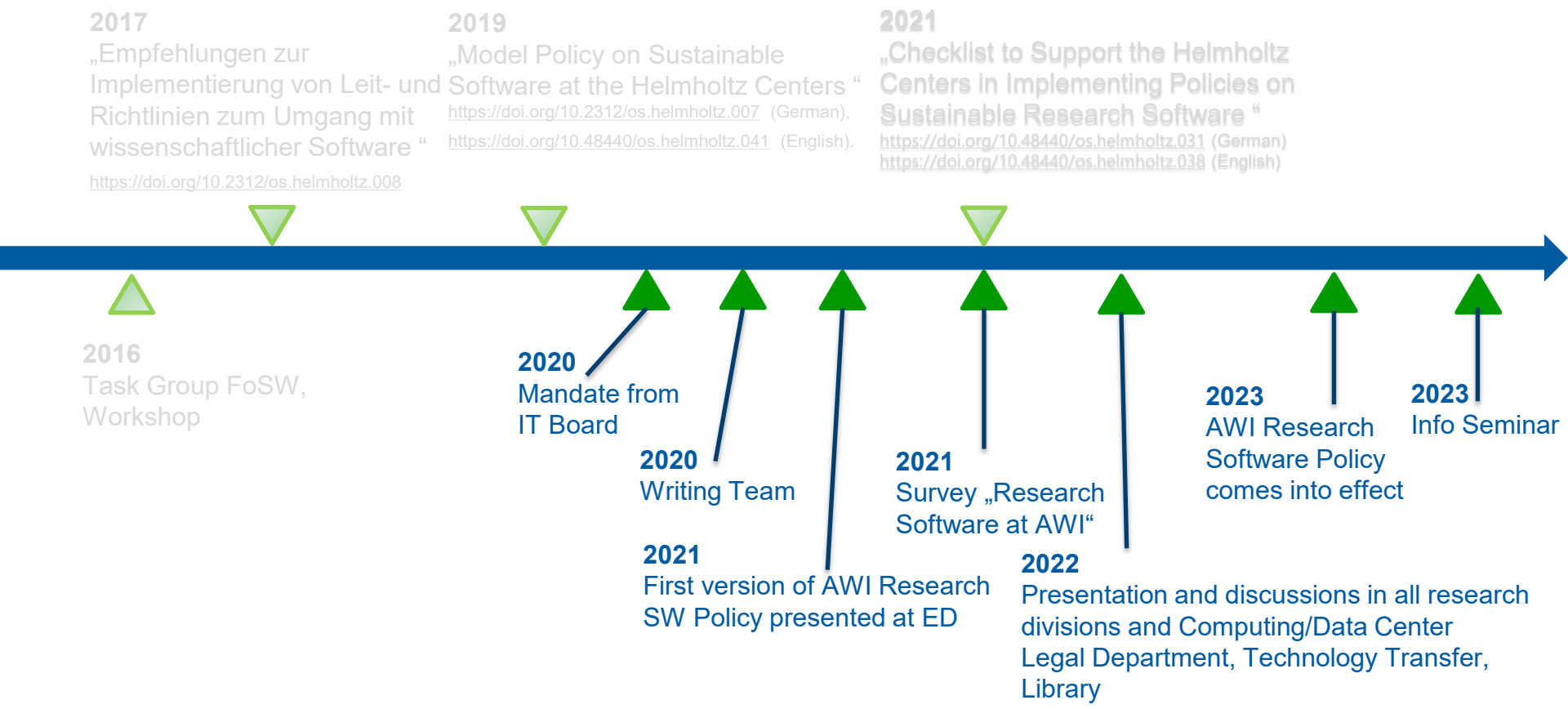
The Turing Way Community, & Scriberia. (2019). Illustrations from the Turing Way book dashes. Zenodo. https://doi.org/10.5281/zenodo.3332808



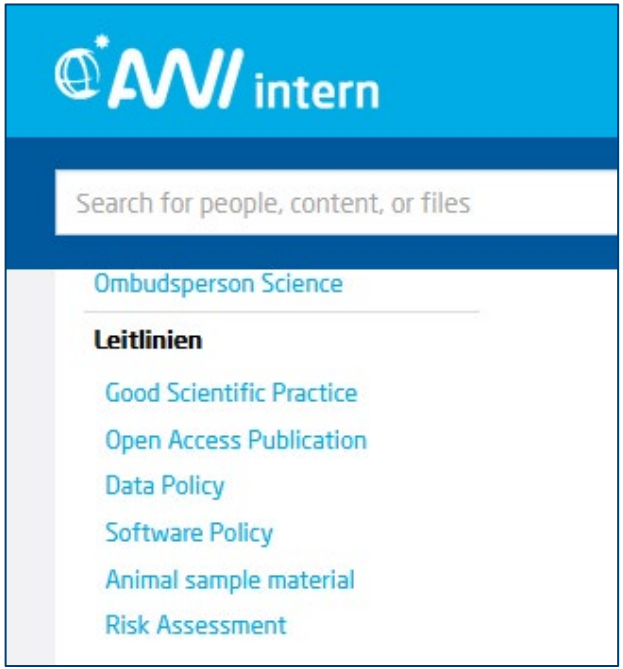
Survey results – free text



Research Software Policy at AWI (1)



Yet another Policy?



Remarks

- Policies and guidelines can support improving software quality
- Many stakeholders involved
  - Policy vs guideline
- Try not to write a „perfect“ policy
- Focus on most relevant aspects

Caspar David Friedrich, Public domain, via Wikimedia Commons

6.6 EOSC-EVERSE: Paving the way towards a European Virtual Institute for Research Software Excellence

EOSC-EVERSE

Paving the way towards a European Virtual Institute for Research Software Excellence

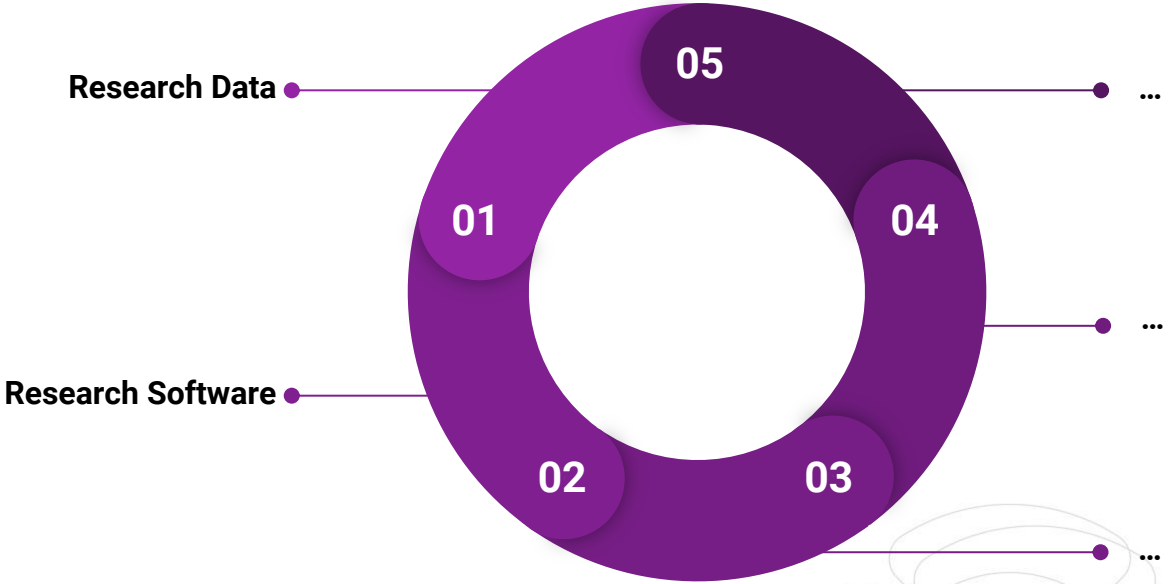
Funded by the European Union

This project has received funding from the European Union's Horizon Europe Programme under GA 101129744 — EVERSE — HORIZON-INFRA-2023-EOSC-01-02

Research Software: The backbone of research

- Software is ubiquitous in Research.
- Communities have created numerous software applications that are essential to progress in their fields.
- The reliability of these applications, how effectively they can be reused and their long-term sustainability, are critical aspects for future progress.
- For example:
  - software used in healthcare must be trustworthy to ensure the safety and well-being of patients.
  - reliable software ensures accurate weather forecasts that we all rely on for planning.
- This necessary trust in research/community software longevity requires a transparent display of good engineering and clear organisational processes that enable continuity.
- Ensure research software curation, quality, preservation and adoption of best practices tailored to developers at all levels.

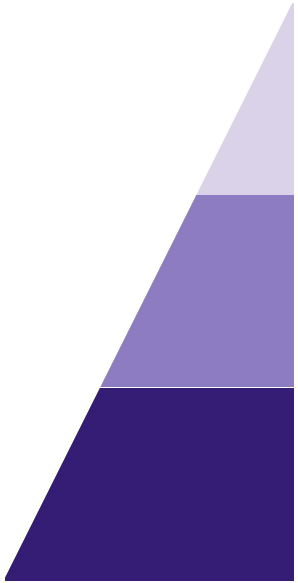
Making Research Software a first class citizen for the scientific endeavours



In a given programming language design, a first-class citizen[a] is an entity which supports all the operations generally available to other entities. These operations typically include being passed as an argument, returned from a function, and assigned to a variable.[1]

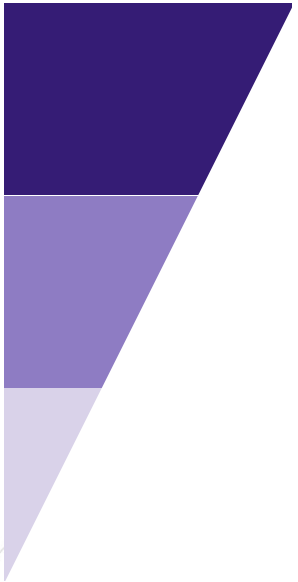
Wikipedia (2023/11/12). https://en.wikipedia.org/wiki/First-class_citizen

Not all software has the same level of importance …



abundance

1. **Research software infrastructure**
   It involves research software that captures more broadly accepted and used ideas, methods and models for use in research, and warrants close researcher involvement in their development.
2. **Prototype tools**
   It refers to research software that demonstrates a new idea, method or model for use by others outside the project within which it originated, often as a substantive intellectual contribution in its own right and often in the form of a proof of concept.
3. **Analysis code**
   It includes research software that captures computational research processes and methodology, and often occurs in the context of simulation, data generation, preparation, analysis and visualisation.



Importance

Foundational Software

REFERENCE TO THE 3-TIERS RESEARCH SOFTWARE MODEL

… but it is ubiquitous across data-intensive scientific domains

ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC
https://zenodo.org/record/367508

EVERSE

Paving the way towards a **E**uropean **V**irtual Institut**e** for **R**esearch **S**oftware **E**xcellence

EVERSE aims to create a framework for research software and code excellence, collaboratively designed and championed by the research communities, in pursuit of building a European network of Research Software Quality and setting the foundations of a future Virtual Institute for Research Software Excellence

ensure research software curation, quality, preservation and adoption of best practices, by the Communities, for the Communities, build on collaboration with the five EOSC Science Clusters

adopt a three-tier model for research software, i.e., analysis code, prototype tools and research software infrastructure, which captures the varying complexity of research software and its development, and can be used as a basis for research software excellence

credit and recognition for both developers and software are essential components of our strategy to promote sustainable software practices

- **Start**: 1 March 2024 (36 months)

Partners, associates, and affiliated entities

Consortium:
15 full beneficiaries, 2 associated partners & 1 affiliated entities across 10 countries

Objectives (1/2)

Objective #1

Build a collaborative, community-led structure for **evaluating**, **verifying**, and **improving the quality** of **research software** and code, by actively involving researchers, software developers, and other stakeholders in the research community.

Objective #2

Leverage existing tools and resources to support the evaluation, verification and improvement of research software and code quality, based on **existing practices and standards** across **research communities** represented by the five EOSC Science Clusters.

Objectives (2/2)

Objective #3

Establish a **sustainable** and **collaborative ecosystem of stakeholders** across the research communities associated with the five EOSC Science Clusters to ensure research software and code quality assurance and support the advancement of **reliable and reproducible research**.

Objective #4

Provide a **framework** that will ensure appropriate **recognition**, **reward**, and **career development** for researchers and **RSEs** who implement research software and code quality assurance practices and policies

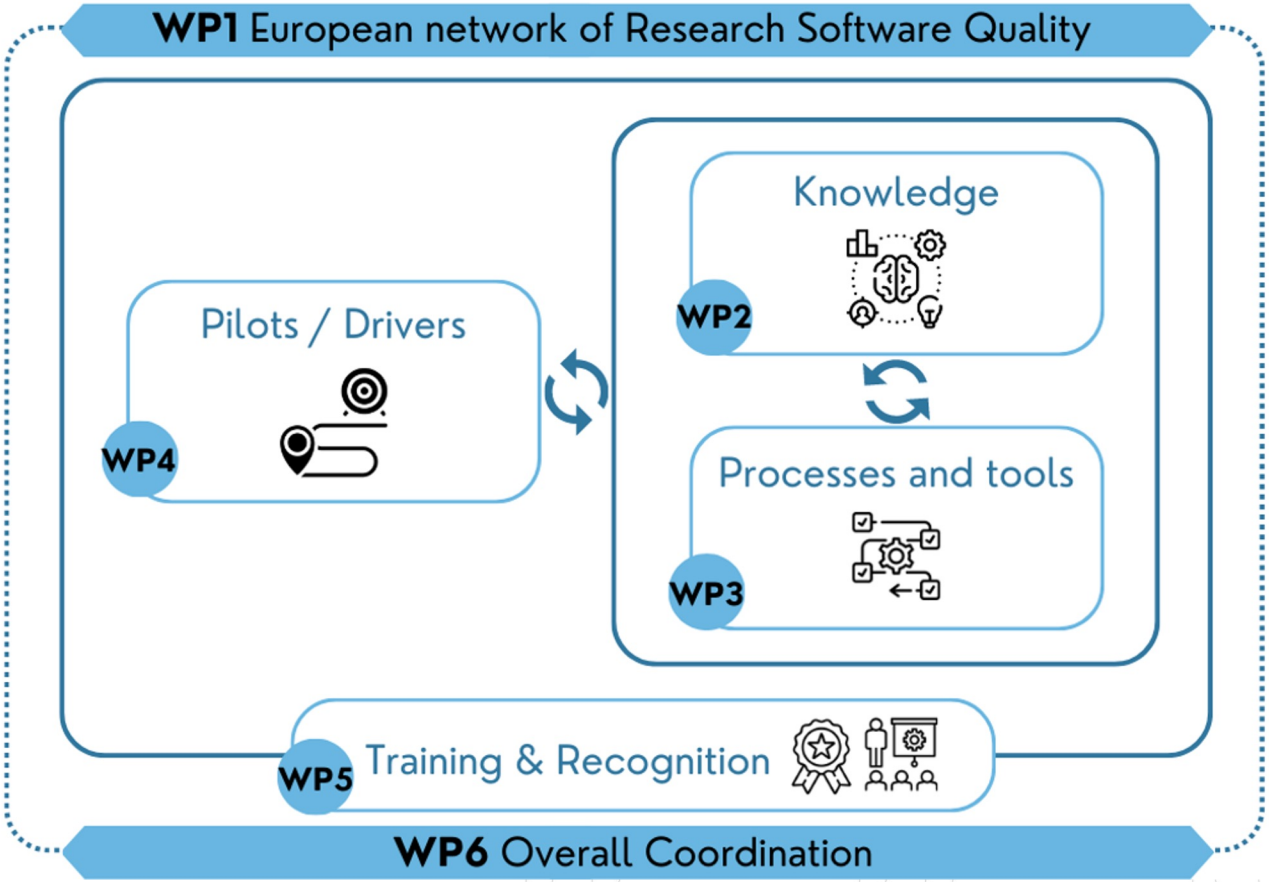
Expected Outcomes and Impact (1/3)

- A framework of **community curation** is established and promoted that ensures **quality of software and code** across the **different disciplines**.
- **Infrastructure, tools and services** are deployed that allow researchers to properly develop, describe with proper metadata, version, archive, share and reuse research software.
- The notion of **software quality** is defined in the context of EOSC and builds upon established practices by the FAIR and other communities.

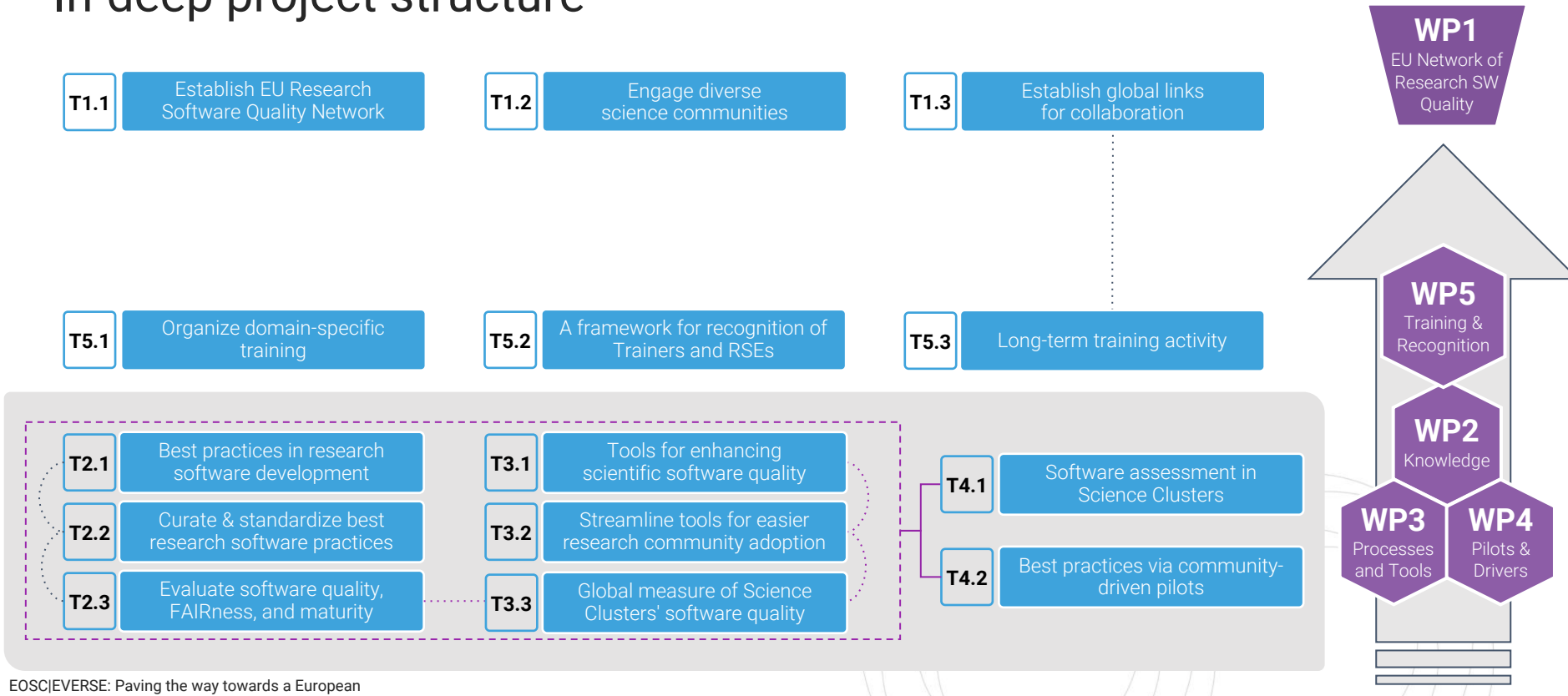
Expected Outcomes and Impact (2/2)

- Baseline **quality indicators** of “minimum quality” defined for the different types of digital objects targeted (software, code, etc), taking into account the concept of “**fit for purpose**”.
- The quality of research software (technical and organisational) improved, in general (e.g. software for data analysis) and in particular for software used in the services offered through EOSC.
- Software is developed in a **sustainable way and its reuse is maximized**.

Project Overview



In-deep project structure



Pilots & Drivers

Environmental Sciences: Integration of Science Cluster ENVRI through ENVRI-HUB

- Integrate EVERSE framework into the ENVRI-HUB Knowledgebase and Virtual Research Environment
- Apply to the development of the Essential Climate Variable computing program and cloud workflows

Life Sciences: Integration of Science Cluster EOSC-Life through ELIXIR

- Make RO-Crate actionable by incorporating the five safes concept into WfExS for secure and federated workflow orchestration
- Use of community-led standards for materialising research software packaged using container technologies and mobilising encrypted data whenever needed

Astronomy and particle physics: Integration of Science Cluster ESCAPE through the Dark Matter Test Science Project

- ML for scientific data compression (standalone code, python)
- A Common Tracking Software
- Choose an ATLAS trigger algorithm as an option for the collaboration

Proton and neutron science: Integration of Science Cluster PaNOSC through LEAPS/LENS
Transition software to high performance computing (HPC) and heterogeneous computing architectures

Social sciences: Integration of Science Cluster SSHOC
Develop a multilanguage textual analysis pipeline of tools that use a combination of open source tools and own code to create an integrated SotA tool capable of deploying locally or as a service

Project Ambition

Collaborative effort to produce and improve high-quality research software, code and other digital artifacts

- Knowledge Hub (RSQkit) containing:
  - Curated best practices.
  - Tools and processes to support the best practices and to assess software and code quality.
  - Guidelines and policies.
  - Catalogue of training materials and courses.
- Additional elements contributing to enhance quality.
- Examples for increasing researchers and RSEs recognition.
- Focus capacity building action through training to increase reach.

European Virtual Institute for Research Software Excellence

eosc | EVERSE

Thank you!

contact: everse-contact@lists.certh.gr

+info: everse.software/

Funded by the European Union

This project has received funding from the European Union's Horizon Europe Programme under GA 101129744 — EVERSE — HORIZON-INFRA-2023-EOSC-01-02

6.7 News from HIFIS

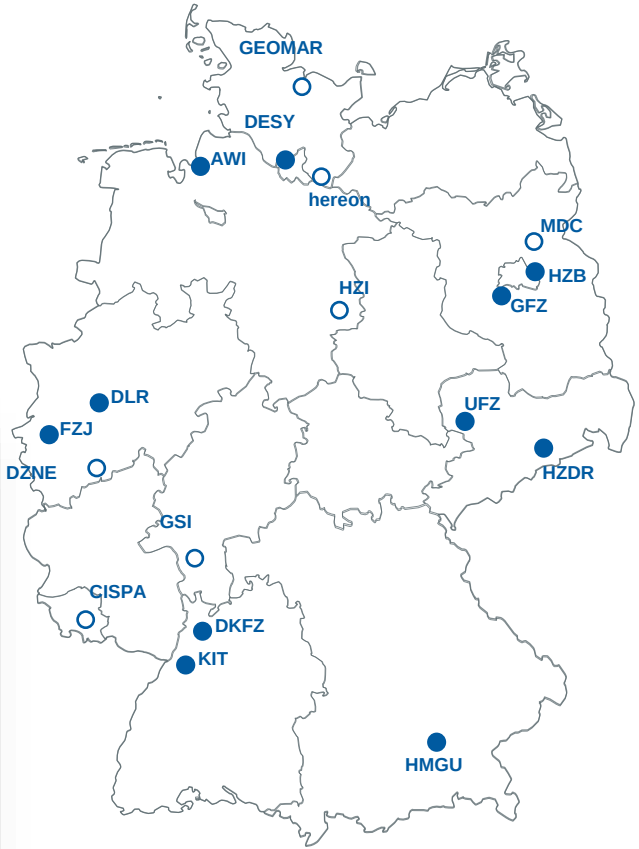
News from HIFIS

Tobias Huste
HZDR

Helmholtz Open Science Forum:
Research Software
2024-02-05

The HIFIS Team

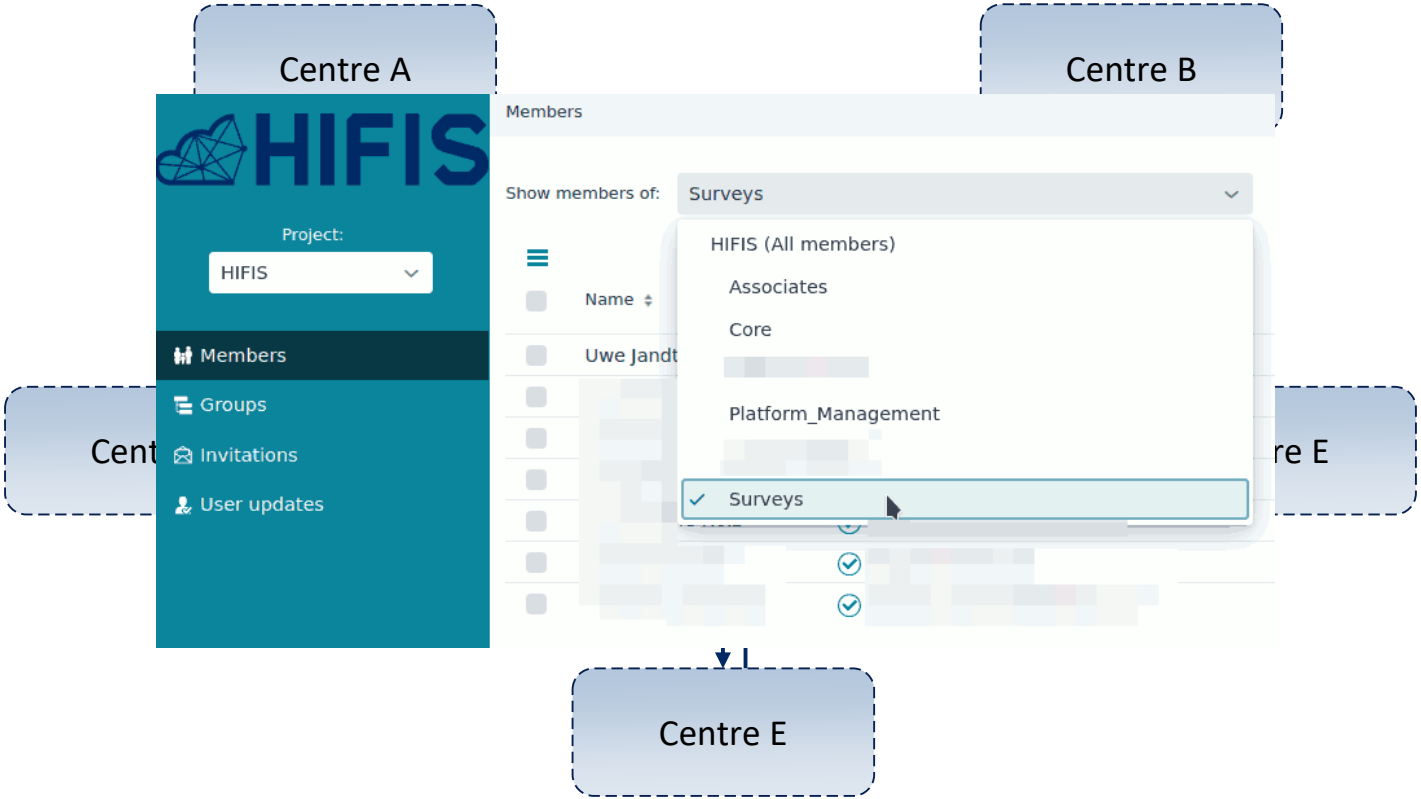
- 11 Centres for all-Helmholtz & Partners
- 3 Organisational Clusters: Backbone, Cloud, Software



HIFIS

HELMHOLTZ

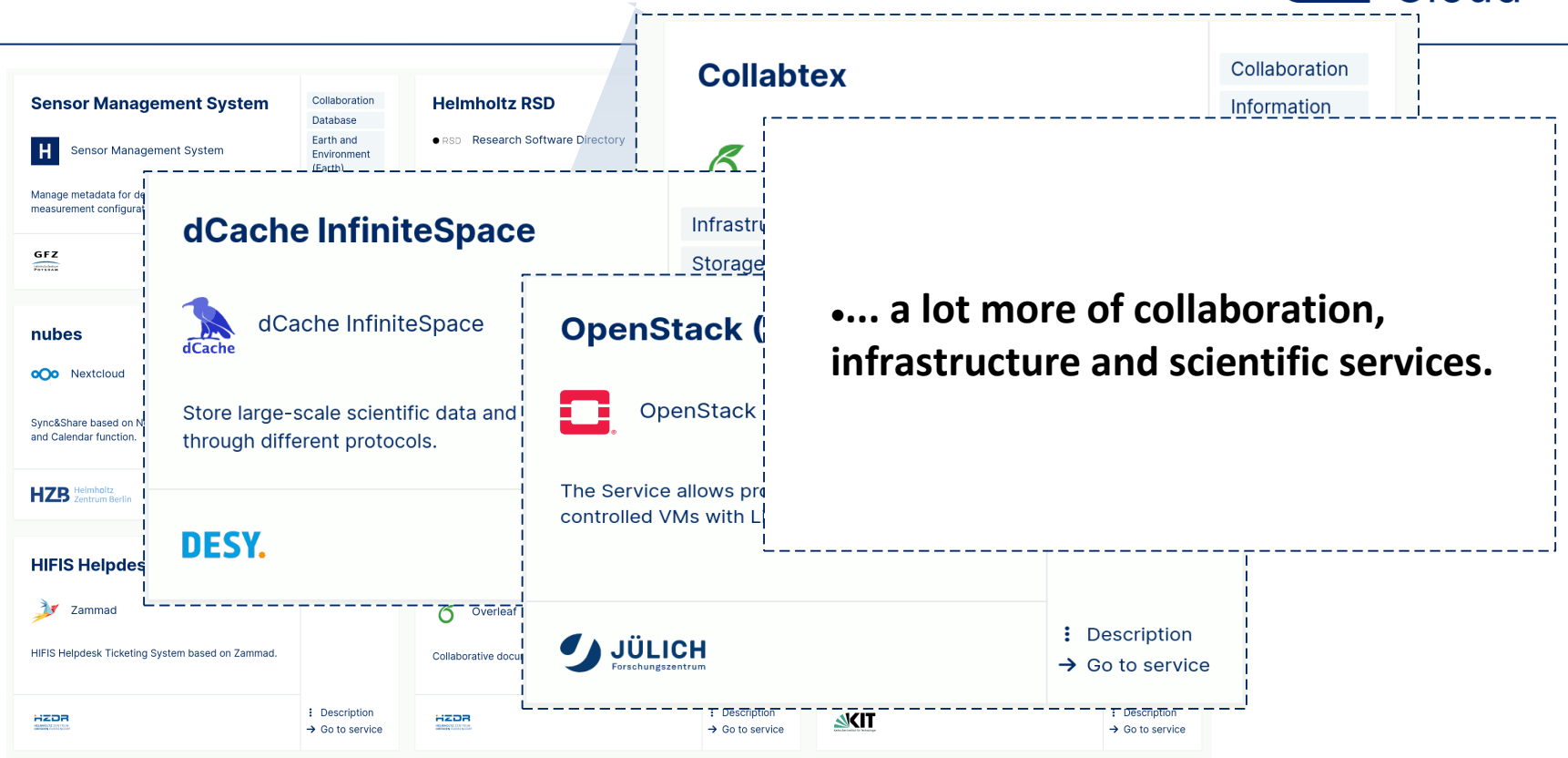
Helmholtz ID: Unified User & Group Management



AARC Blueprint

Infrastructure & Cloud Services

HELMHOLTZ Cloud



HIFIS

➢ https://helmholtz.cloud

HELMHOLTZ

Recap: What HIFIS Software is all about?

Education & Training	Community	Consulting	Technology
Courses, material and **workshops** for getting you started or boosting your software engineering practice.	• Build and foster communities to support the **cultural change** when dealing with research software. • Maintain a SW Directory	**Contact points** for researchers for questions and problems in the context of RSE.	Provide a sustainable, well integrated and easy to use **technology infrastructure** for research software development.

HIFIS

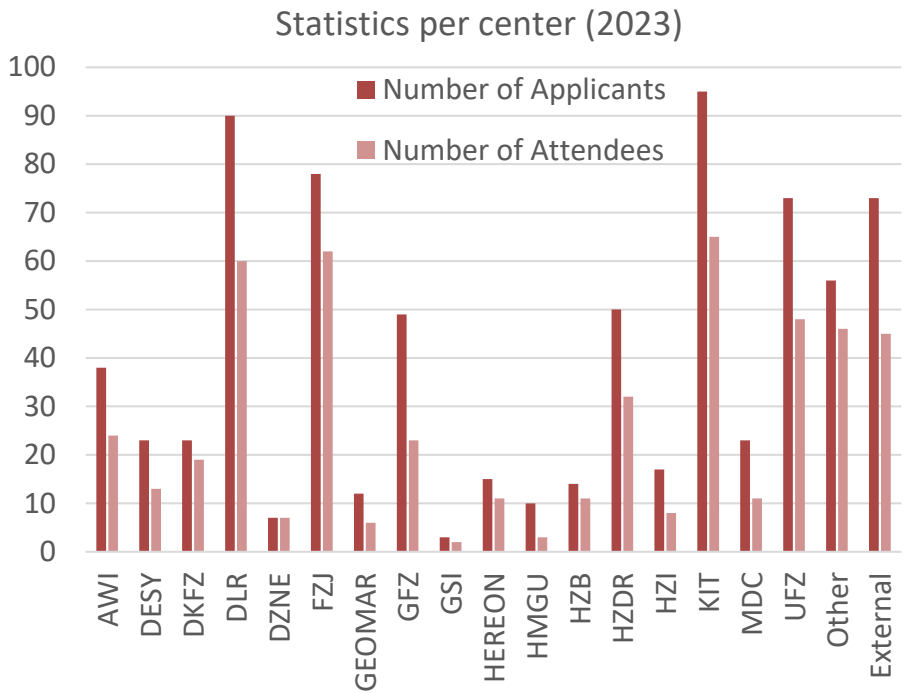
Education & Training

Education & Training

Statistics 2023

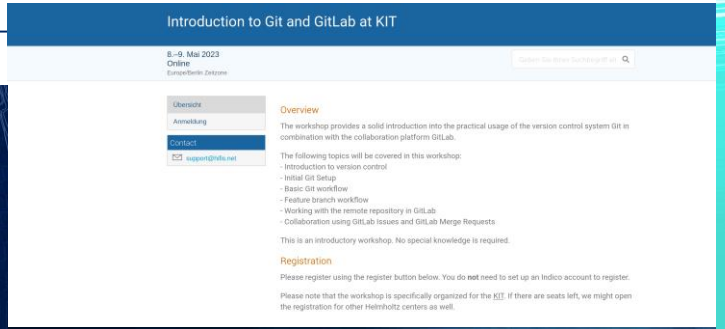
Summary of 2023 (Total):

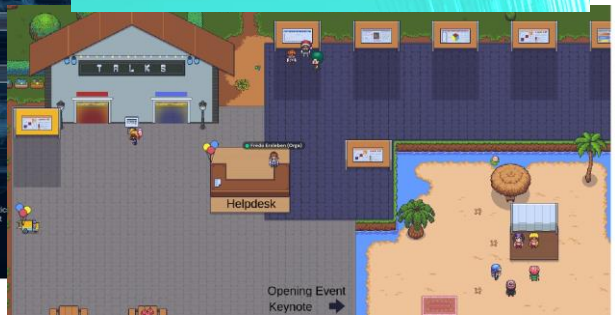
- 29 (112) workshops
- 496 (1.900) attendees
- No-show rate ~18% (20%)
- 352 (1.186) hours of instruction duration

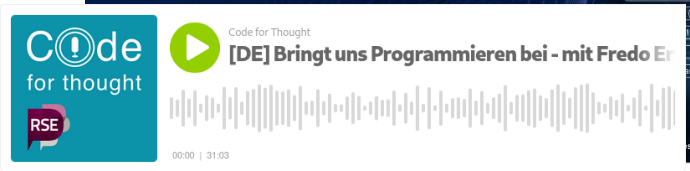


Education & Training

Events 2023 – A Selection







HIFIS

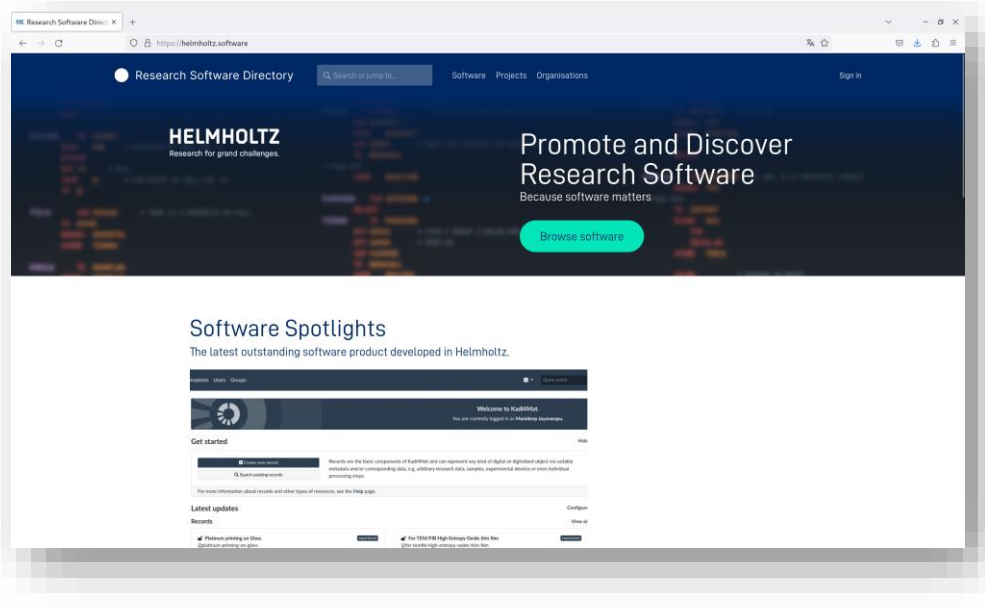
Community

Community

Helmholtz Research Software Directory

- Built on top and in close cooperation with the Netherlands eScience center
- **203 software packages** from 17 out of 18 Helmholtz centers

→ Advertize https://helmholtz.software



Community

Helmholtz Research Software Directory

Build recognition and **Cross-linking** of communities

- Helmholtz Software Forum
  - Joint exchange format driven by HIFIS and the Helmholtz Open Science Office
  - So far 3 events with more than 100 participants each time organized
- Contribute to community activities like deRSE conferences or SORSE
- Contributions at similar activities of MPG, FhG, NFDI, EOSC AG...

HELMHOLTZ Open Science

SORSE

de RSE

Community

Helmholtz Incubator Software Award

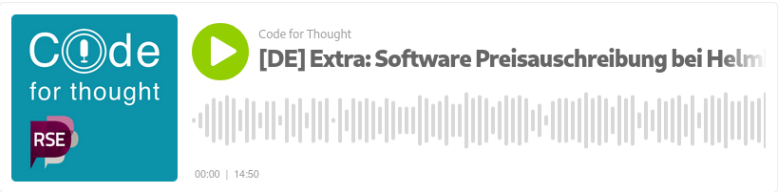
- Promote the development of professional and high-quality research software
- Recognize the commitment to software as the basis of modern Data Science
- 41 submissions from 16 Helmholtz centers
- Three categories
  - Scientific Originality
  - Sustainability
  - Newcomer

HIFIS

HiDA HELMHOLTZ Information & Data Science Academy

HIRSE_PS

HELMHOLTZ Open Science



HIFIS

Consulting

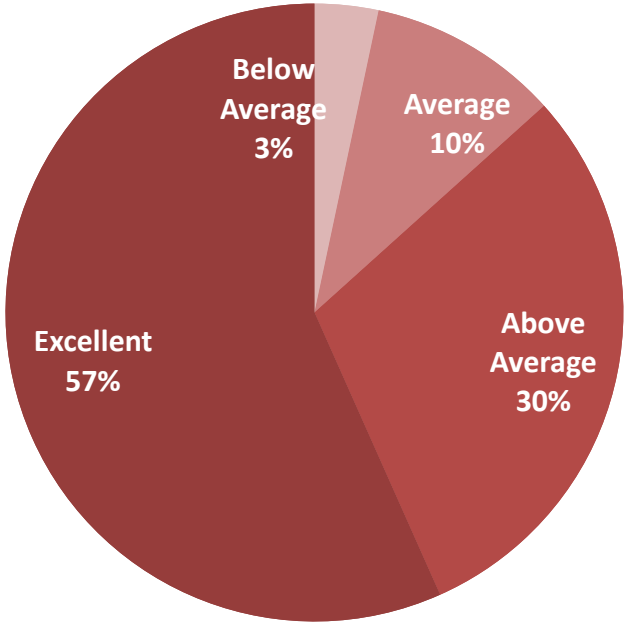
Consulting

- Free-of-charge software consulting for research groups within Helmholtz
- Possible topics include, but are not limited to licensing and Open Source, setting up new projects, code migrations etc.
- Material collection:
  - Consulting Handbook: https://hifis.net/consulting-handbook/
  - Awesome List RSE: https://github.com/hifis-net/awesome-rse

Great idea and a great support especially since there is no other person programming in my research group. Very happy that you came up with this!

For us, it would be perfect to have such a consulting service over a longer period of time, e.g. for 6-12 months with regular meetings.

Impact of the consultation on your project or work



HIFIS

Technology

Technology

Supporting the whole software development lifecycle

Dependabot

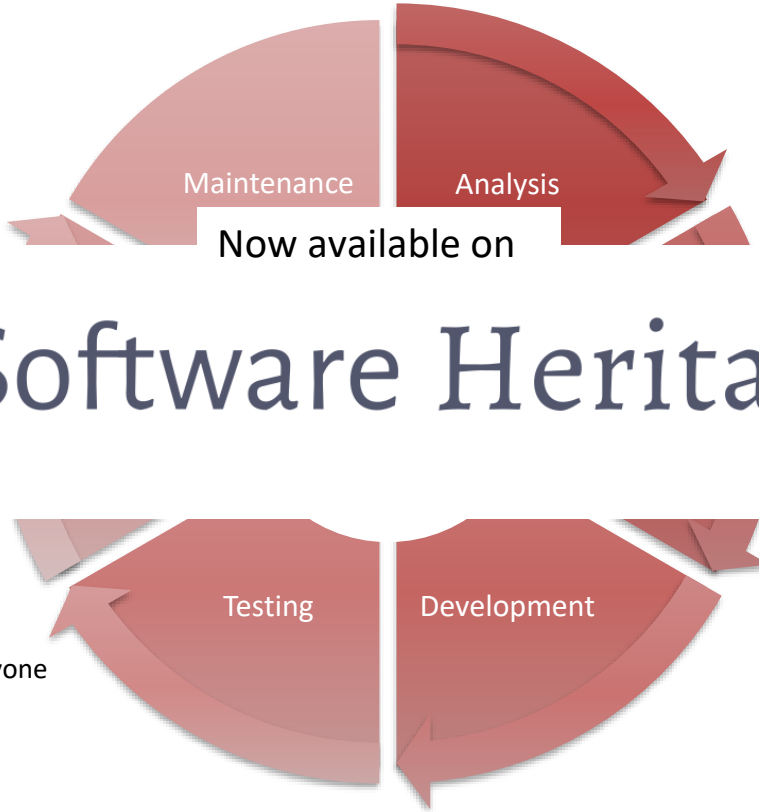
Security: Automate dependency updates

Deployed and made available with

ANSIBLE

https://github.com/hifis-net

Continuous Integration and Deployment available for everyone by default



Now available on

Software Heritage

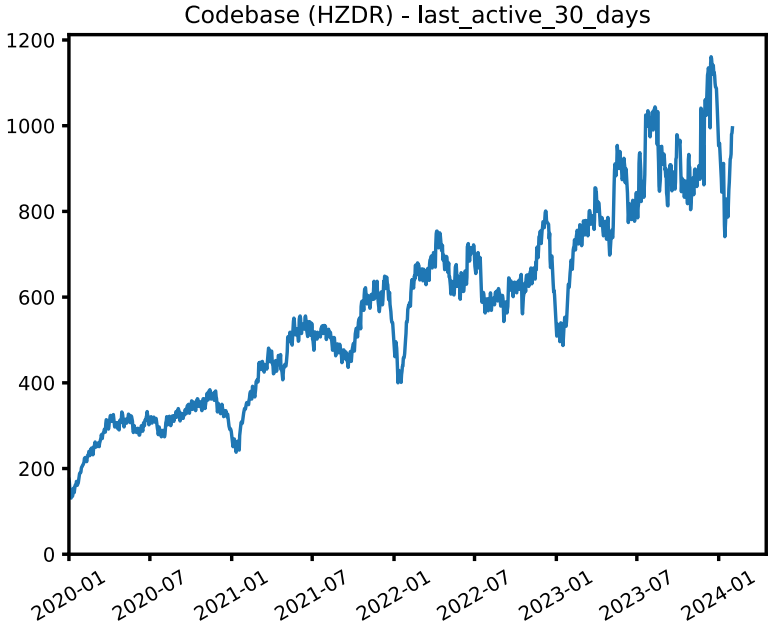
Kroki

Diagram creation

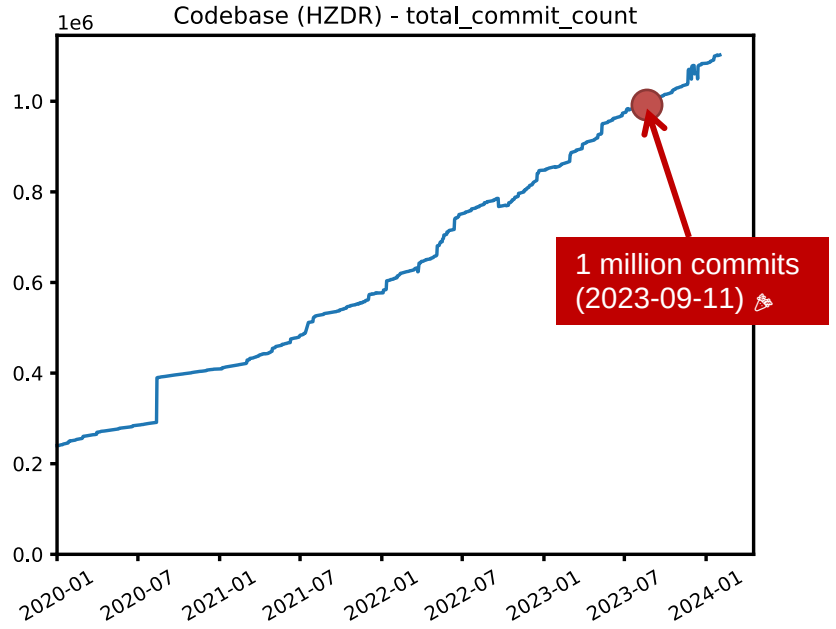
Mattermost

Team communication

Monthly Project Activity



Total Number of Commits



HIFIS

112
Workshops

1.900
Attendees

HIFIS supports Research Software Engineering with

- **Education & Training**,
- **Community**,
- **Consulting** and
- **Technology** Services.

203
RSD Software Packages

>1.000.000
Commits in Codebase

1.161
Peak number of monthly active projects

4.944
Peak number of monthly active Mattermost users

>440.000
CI Jobs Processed in 2023

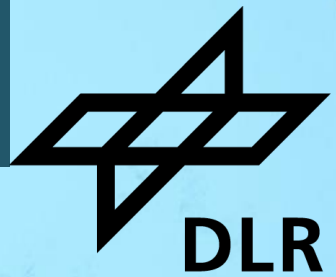
Community
Statistics

6.8 Establishing Software Citation Practices at DLR

ESTABLISHING SOFTWARE CITATION PRACTICES AT DLR

05.02.2024, Helmholtz Open Science Forum: Research Software, UFZ, Leipzig

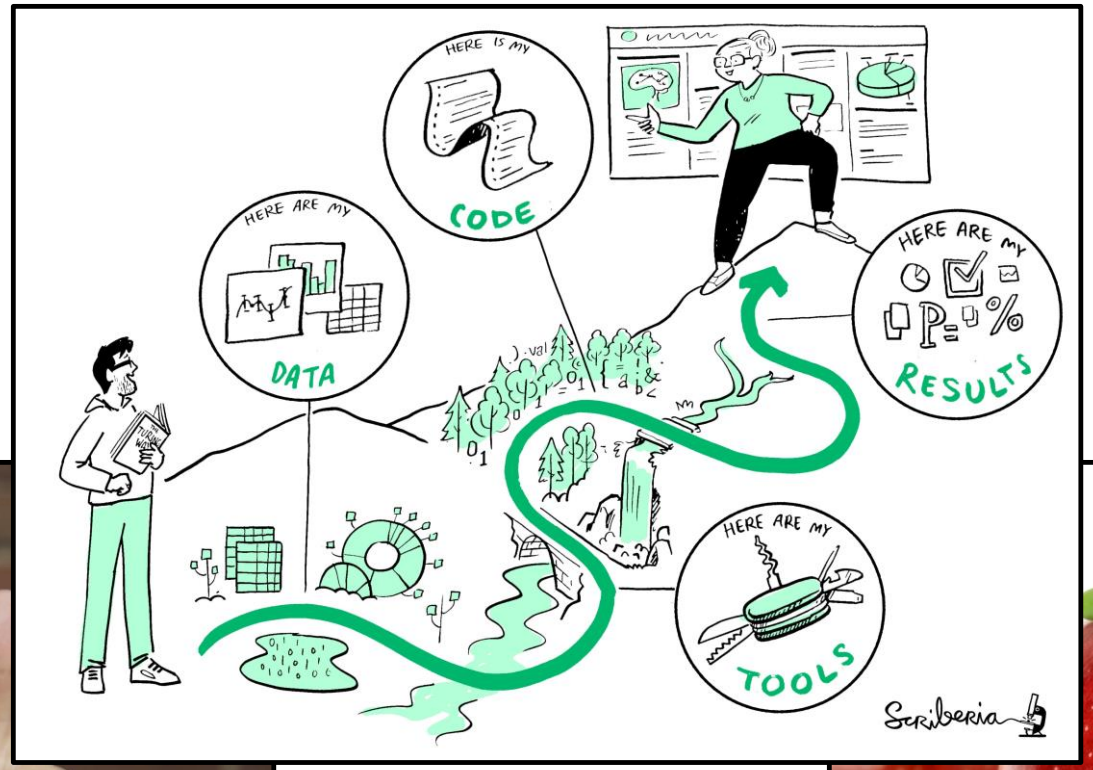
Tobias Schlauch <Tobias.Schlauch@DLR.de>
Institute for Software Technology
German Aerospace Center (DLR)
http://www.dlr.de/sc



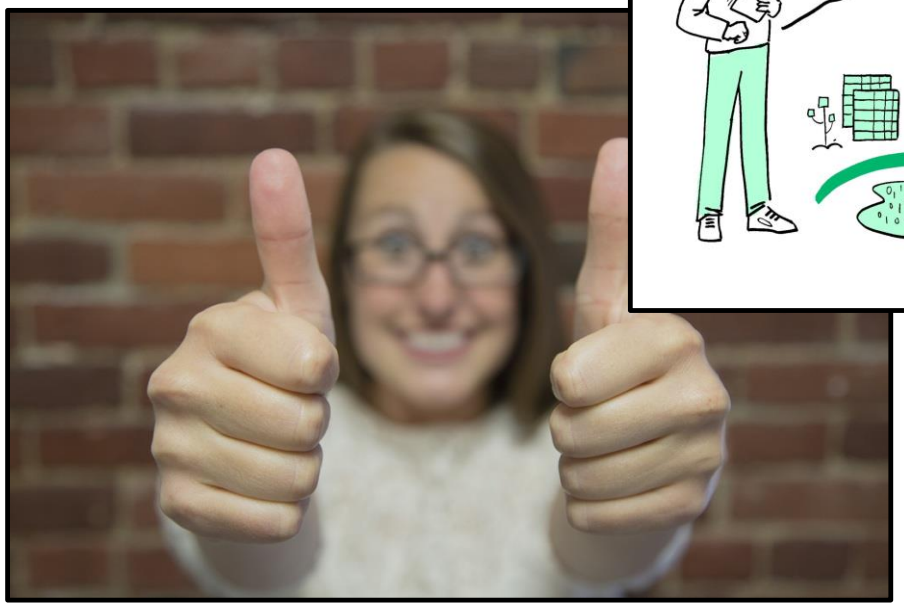
Plans concerning the Update of the DLR Software Policy

- **What policies / recommendations are already available at DLR?**
  - Framework Directive Software Engineering
    (initial version: 2008, last major update: 2016)
  - DLR Software Engineering Guidelines (initial version: 2016)
  - DLR Open Source Brochure (initial version: 2013, last major update: 2022)
- **Current approach:**
  - Close implementation of the Model Policy on Sustainable Software at the Helmholtz Centers
  - Map existing policies
- **Main improvement areas:**
  - Establish a more binding open source policy
  - Update DLR SE Guidelines
  - Enable software citation

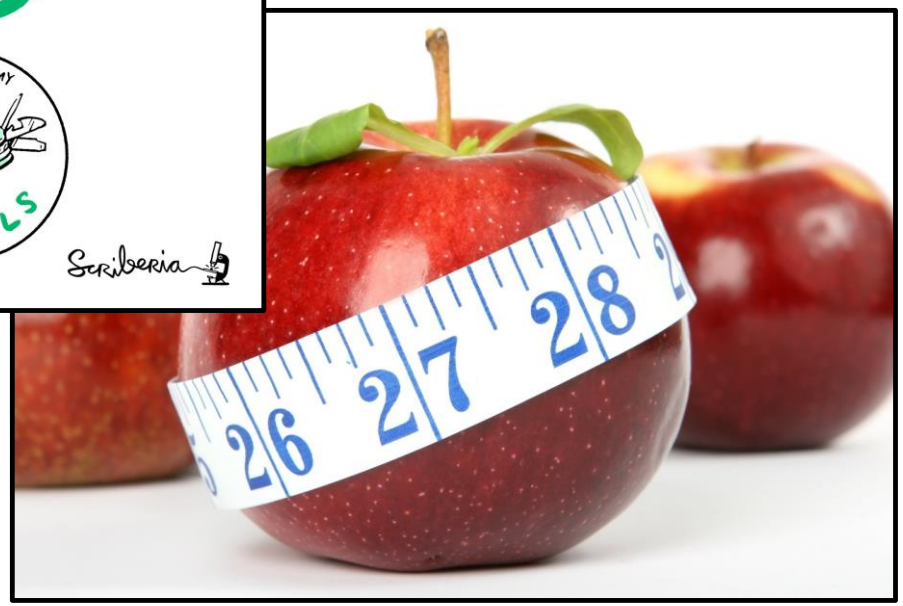
Why is Software Citation important?



Reproducibility



Give Credit



KPIs

How to Cite Software?
Example

“The Jupyter notebook containing the analysis details has been published separately [11].”

References:

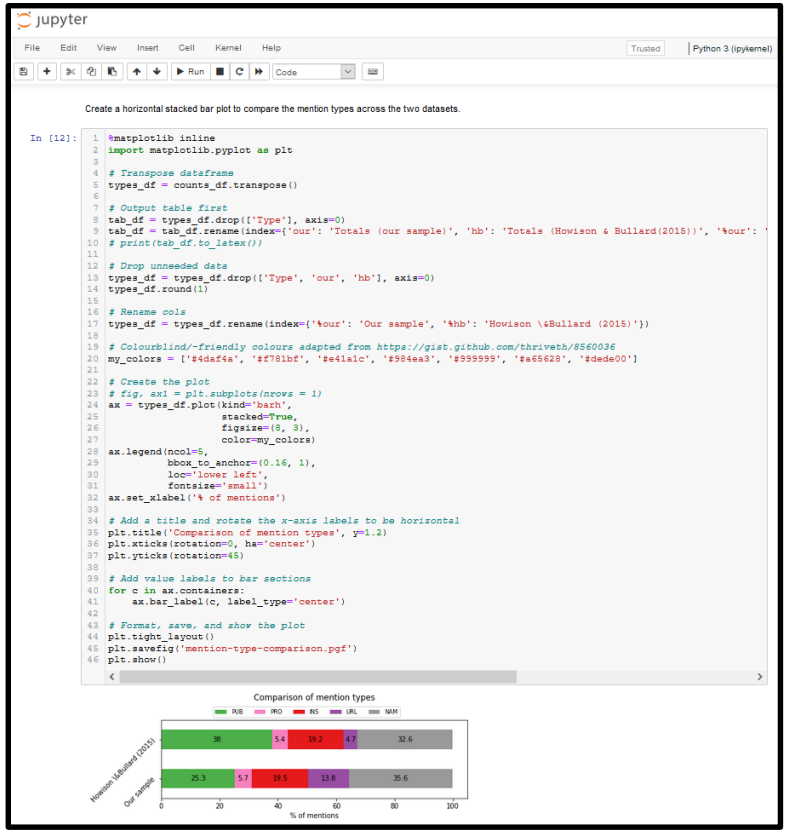
[11] Schlauch, Tobias, & Haupt, Carina. (2019).
Analysis of the DLR Knowledge Exchange Workshop Series on Software Engineering (Version 1.2.0).
Zenodo. https://doi.org/10.5281/zenodo.3403991

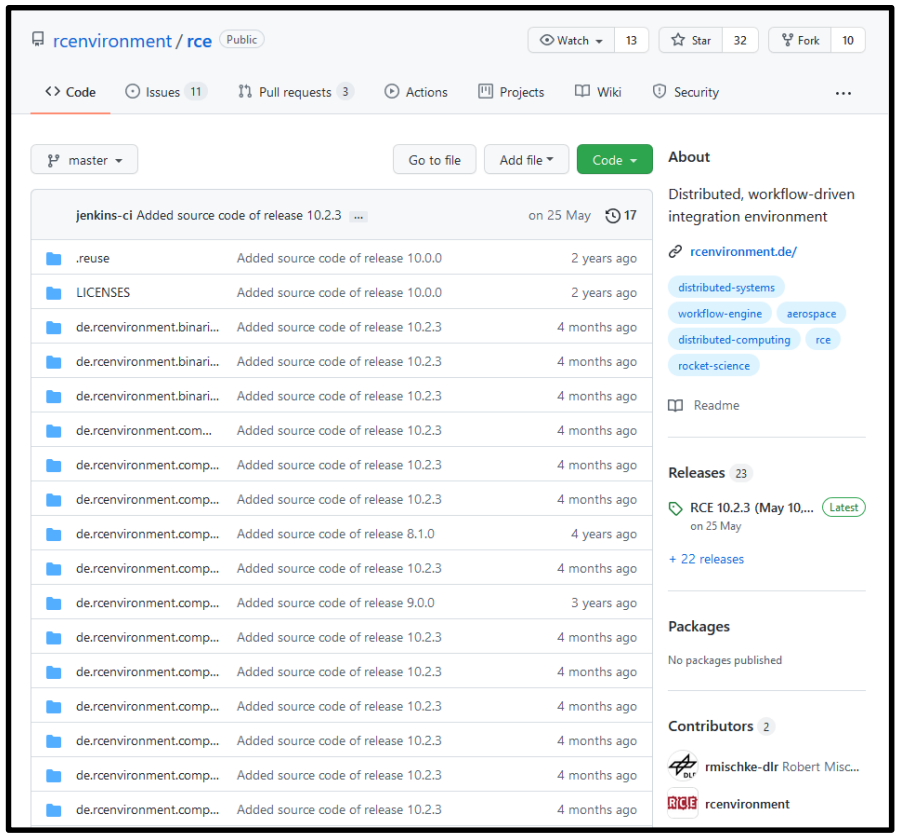
- **Authors**
- **Software name**
- **Software itself**
- **Source code version**
- **Exact version**
- **Publication date**
- **Persistent identifier (PID)**

- **See also:** Software Citation Principles

How to Cite Software?

Sometimes Software Citation is not easy … ☹





METADATA: Name? Authors? Version? Publisher? Publication date?

Make your Software Citable!

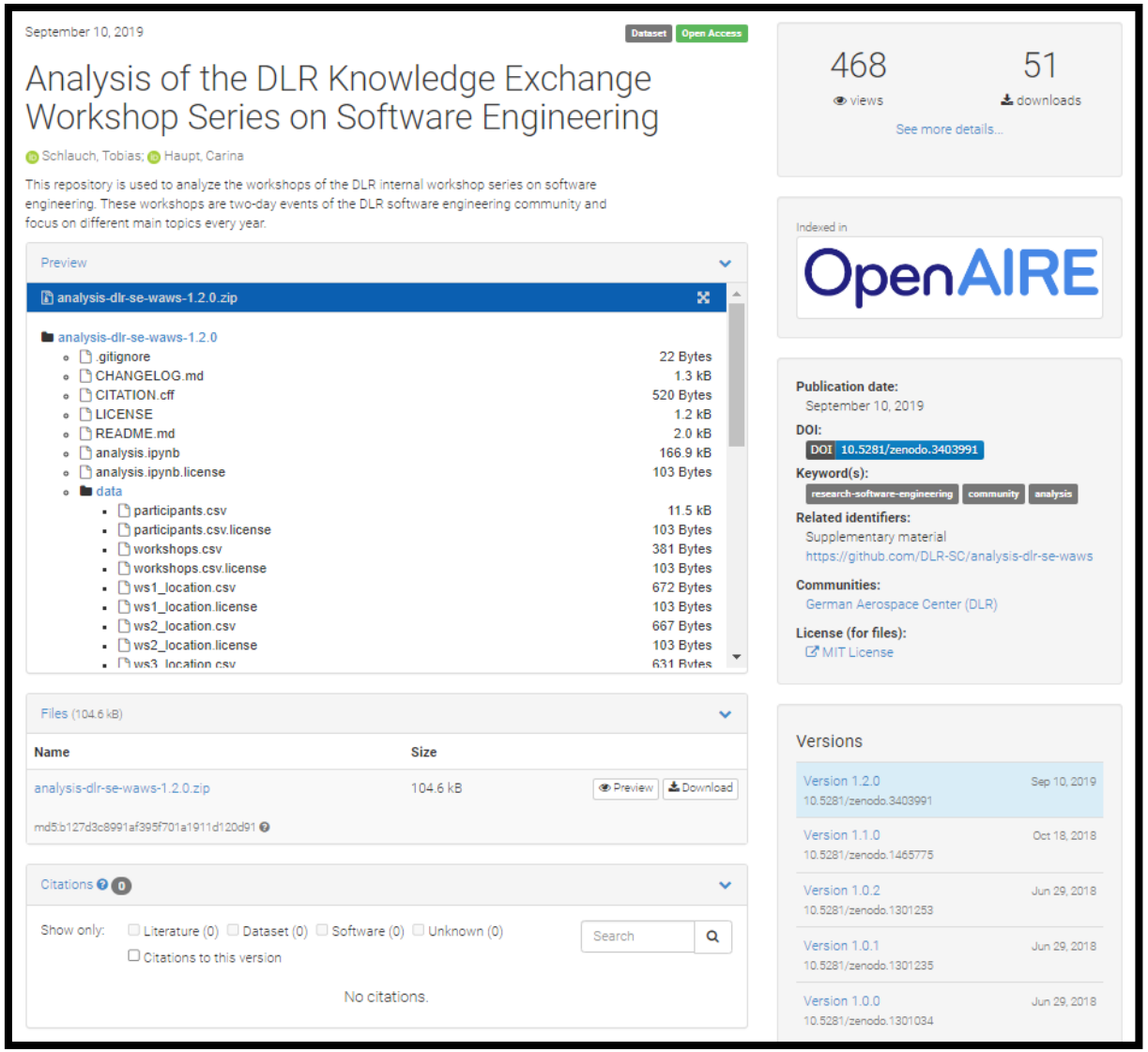
Step 1: Provide Citation Metadata for your Software

```yaml
cff-version: 1.2.0
message: If you use this software, please cite it using these metadata.
title: My Research Software
abstract: This is my awesome research software. It does many things.
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: "https://orcid.org/0000-0003-4925-7248"
version: 0.11.2
date-released: "2021-07-18"
identifiers:
  - description: This is the collection of archived snapshots of all versions of My Research Software
    type: doi
    value: "10.5281/zenodo.123456"
  - description: This is the archived snapshot of version 0.11.2 of My Research Software
    type: doi
    value: "10.5281/zenodo.123457"
license: Apache-2.0
repository-code: "https://github.com/citation-file-format/my-research-software"
```

+ references, preferred citation, experimental data set support

METADATA: In a `CITATION.cff` file in the code repository!

Make your Software Citable!
Step 2: Publish your Software in a Publication Repository



Zenodo an example of a public publication repository:

- Storage of standardized metadata
- Archived artifacts
- DOIs as persistent identifiers
  - Individual DOIs for every software release
  - A conceptual DOI for the software itself (points to the latest release)

Make your Software Citable!
Practical Considerations

- **Where to start?**
  - Do you write a research paper? Please cite all relevant software and make it citable!
  - Citable Software != Open Source Software
- **Citation metadata:**
  - Manage citation metadata inside your source code repository in a `CITATION.cff` file (Citation File Format)
- **Publication repositories:**
  - Open Source Software: Zenodo
  - Closed Source Software: Organizational publication repository
- **(Automated) publication approaches:**
  - Custom solution
  - GitHub Zenodo integration
  - HERMES ("**HE**lmholtz **R**ich **ME**tadata **S**oftware publication)

Enabling Software Citation at DLR

- Provide practical guidelines for ***publishing citable software****:*
  - Recommend default technologies cornerstones:
    - Citation metadata: ***CITATION.cff***
    - Publication automation tool: ***HERMES***
    - Publication repository: ***DLR-internal publication repository***
  - Propose concrete publication process / provide concrete how-tos
- Establish ***DLR-specific publication repository*** (InvenioRDM)
  - Start with a pilot phase with interested early adopters
  - Start “fleshing out” practical processes (e.g., publication, curation, etc.)
- Make software citation ***a required part*** of DLR`s software policy

Discussion

- How do you address / do you plan to address the topic ***software citation*** in your research software policy?
- What technologies/standards do you use / plan to use?
- Do you have established / plan to establish a ***publication repository suitable for software deposits*** at your research center?
- How do you organize / plan to organize the ***practical processes*** (e.g., publication, curation)?

Copyright and License Information

Thank you!

What are your Questions?

Email: Tobias.Schlauch@dlr.de
Mastodon: https://norden.social/@schlauch
HIFIS Mattermost: @schlauch

6.9 Helmholtz Platform for Research Software Engineering – Preparatory Study (HiRSE_PS)

HELMHOLTZ

SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

Helmholtz Platform for Research Software Engineering – Preparatory Study (HiRSE_PS)

René Caspart

Karlsruhe Institute of Technology

Together with Markus Diesmann, Stefan Blügel, Robert Speck (Jülich), Achim Streit, Markus Götz (KIT), Johannes Reuther (HZB), Christian Cyron, Regine Willumeit-Römer & Daniel Höche (Hereon)

HIRSE_PS

JÜLICH Forschungszentrum

KIT Karlsruher Institut für Technologie

HZB Helmholtz Zentrum Berlin

Helmholtz-Zentrum hereon

Research Software Engineering (RSE) – Why?

HIRSE_PS

(Open Source) Software =

- **Key component of scientific work**
- **Software ≈ data ≈ devices**
- **Software = research infrastructure**
- **Valuable assets**

in all research fields of Helmholtz

Sources: https://www.software.ac.uk/about, https://www.helmholtz.de/en/research/

Innovation pool project HiRSE_PS

https://www.helmholtz-hirse.de

HIRSE_PS

- **HiRSE_PS** can only provide a first impetus of a RF-wide, much larger activity
  - Focus on software as an infrastructure (open, reliable, sustainable, reproduceable)
  - Testbed for structural RSE support within a research field (toward HiRSE)
  - Human-centric view: enable RSEs to work best and together on their codes/project
- **3 pillars** in 2 work packages
  - Specific **Community Software Infrastructure (CSI) groups** for already existing/established Community Codes Teams
    - NEST and FLEUR (FZJ), HeAT (FZJ/KIT/DLR/Intel), PFFRG (HZB), 4C (Hereon)
  - Central **support and consulting unit** for RSE (FZJ, KIT)
  - Open **HiRSE Seminar** for a regular exchange of information

WP 1: CSI groups (PI: Markus Diesmann)

Community Software Infrastructure Groups

HIRSE_PS

- **Goals of the WP**
  - Establish five CSI Gruppen (topical width across all 3 programs in the RF-Information and diversity of already established and young codes)
  - Fostering sustainability and long-term stability of specific codes
  - Supporting the community
  - Generates insights about the requirements for WP2
- **Structure and location of CSI-groups**
  - Expert knowledge from a scientific domain
  - High community trust through in-person-responsibilities in the domain-institute
- **Tasks of the CSI-groups**
  - Coordination of the development, e.g. doing code reviews, generating releases, monitoring of Cx-technologies
  - Taking over hard and longer-lasting development tasks (e.g. refactoring of existing codes)
  - Organization of trainings and Hackathons

WP 2: Consulting & Networking (PI: Achim Streit)

Goals and structure

HIRSE_PS

- **Goals:**
  - Establishing the technological basis for RSE
  - Supporting established CSI groups and codes in Cx environment usage and software engineering
  - Taking young codes by the hand introducing modern RSE practices
- **3 sub-WPs**
  1. Framework for CI/CT/CD
  2. Support & consulting
  3. Community building & networking

- **Usage of modern Supercomputing infrastructures**
  - E.g., JUWELS (FZJ-JSC) or HoreKa (KIT-SCC)
  - Future Technologies Partition at KIT-SCC
    for CI on different HPC resources and architectures
  - Cloud resources via OpenStack in the Helmholtz Data Federation

openstack.

Supporting RSEs

Cx Infrastructure and services

HIRSE_PS

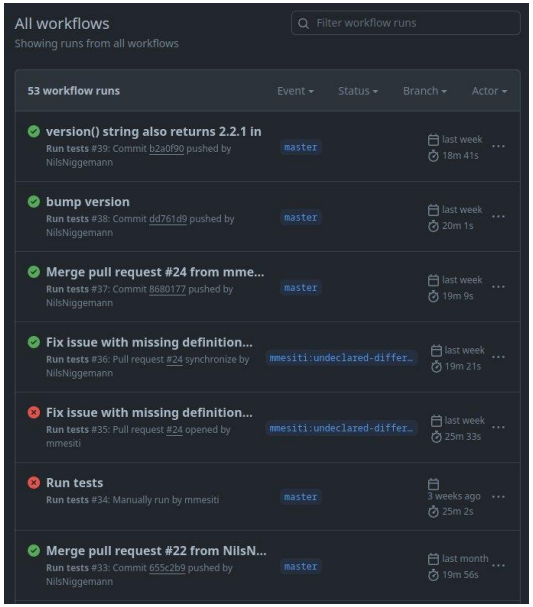
- Providing and enabling possibilities for Cx (=CI/CT/CD/CB) on HPC
  - HPC systems at KIT (HoreKa and Future Technologies Partition) and FZJ (Jacamar on JUWELS)
  - Wide range of available HPC architectures
  - Easily integrable into GitLab projects for users of the systems
- Enabling usage with GitHub repositories
  - Github2lab Action for using GitLab CI from GitHub repositories

Supporting RSEs

Consulting CSI Groups

HIRSE_PS

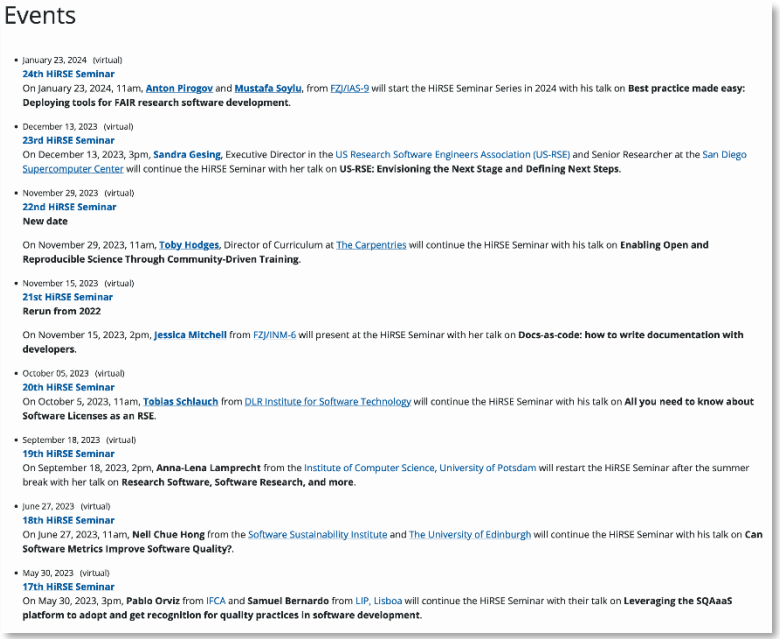
- Consulting on the PMFRG project (presented in the last forum)
  - Improved **parallelization** and HPC specific optimization
  - Improvements to the code and **RSE methods**
- Work on the project over a 3 months period
- Main output
  - incremental MPI parallelization of the PMFRG code
  - Now able to scale up to multiple nodes
- Additional outputs:
  - **CI** with GitHub Actions setup for PMFRG.jl
  - **Regression tests** added to identify possible optimizations
  - Using Julia package extension mechanism -> ideal for keeping the project owners close to the code



Update on the HiRSE Seminar Series

HIRSE_PS

- HiRSE Seminar started in April 2022
- Now at 24 talks, much more planned
- About 20-70 participants per talk
- Significantly broadened the outreach over time
  - National: HIFIS, HGF OS Office, de-RSE
  - International: UK RSE, US RSE, bssw.io
- Feedback form: 4.6/5 stars rating, very good feedback
- Did first reruns of seminars due to feedback
- YouTube Channel + Zenodo Community

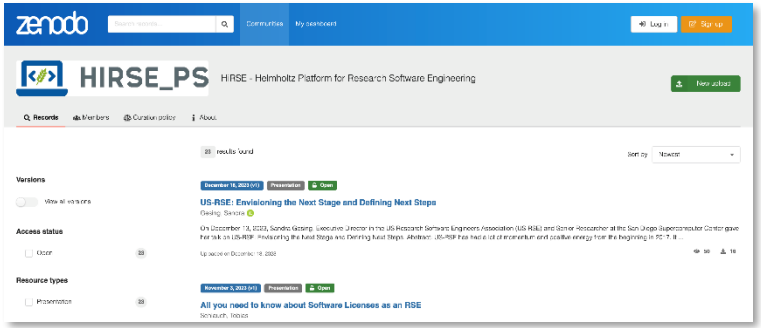


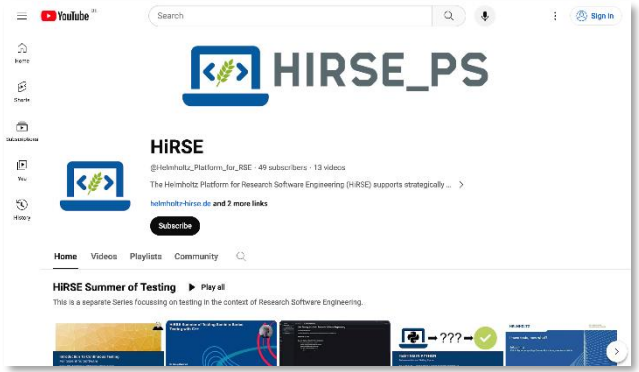
Events

- January 23, 2024 (virtual)
  24th HiRSE Seminar
  On January 23, 2024, 11am, **Anton Pirogov** and **Mustafa Soylu**, from FZJ/IAS-9 will start the HiRSE Seminar Series in 2024 with his talk on **Best practice made easy: Deploying tools for FAIR research software development.**
- December 13, 2023 (virtual)
  23rd HiRSE Seminar
  On December 13, 2023, 3pm, **Sandra Gesing**, Executive Director in the US Research Software Engineers Association (US-RSE) and Senior Researcher at the San Diego Supercomputer Center will continue the HiRSE Seminar with her talk on **US-RSE: Envisioning the Next Stage and Defining Next Steps.**
- November 29, 2023 (virtual)
  22nd HiRSE Seminar
  New date
  On November 29, 2023, 11am, **Toby Hodges**, Director of Curriculum at The Carpentries will continue the HiRSE Seminar with his talk on **Enabling Open and Reproducible Science Through Community-Driven Training.**
- November 15, 2023 (virtual)
  21st HiRSE Seminar
  Rerun from 2022
  On November 15, 2023, 2pm, **Jessica Mitchell** from FZJ/INM-6 will present at the HiRSE Seminar with her talk on **Docs-as-code: how to write documentation with developers.**
- October 05, 2023 (virtual)
  20th HiRSE Seminar
  On October 5, 2023, 11am, **Tobias Schlauch** from DLR Institute for Software Technology will continue the HiRSE Seminar with his talk on **All you need to know about Software Licenses as an RSE.**
- September 18, 2023 (virtual)
  19th HiRSE Seminar
  On September 18, 2023, 2pm, **Anna-Lena Lamprecht** from the Institute of Computer Science, University of Potsdam will restart the HiRSE Seminar after the summer break with her talk on **Research Software, Software Research, and more.**
- June 27, 2023 (virtual)
  18th HiRSE Seminar
  On June 27, 2023, 11am, **Nell Chue Hong** from the Software Sustainability Institute and The University of Edinburgh will continue the HiRSE Seminar with his talk on **Can Software Metrics Improve Software Quality?.**
- May 30, 2023 (virtual)
  17th HiRSE Seminar
  On May 30, 2023, 3pm, **Pablo Orviz** from IFCA and **Samuel Bernardo** from LIP, Lisboa will continue the HiRSE Seminar with their talk on **Leveraging the SQAaaS platform to adopt and get recognition for quality practices in software development.**

HiRSE Seminar Series - extended

- Popular request I: provide slides and record talks, done via Zenodo and YouTube ☑
- Popular request II: seminar(s) on testing, done via HiRSE Summer of Testing ☑
- Also: good practice series, for now with talks on documentation, licenses, citation
- 49 subscribers on YouTube, >250 views for (some) talks, >650 views of (some) slides
- Notable recurring attendee: DFG
- Ideas for new topics/speaker? Let us know!





Events

- Organized HiRSE_PS CB Hackathon
  - Jülich Supercomputing Centre, Oct 17-18, 2023
  - 31 participants, who are eager to continue the Cx Hackatons…
- Active in the GI SIG on RSE:
  - Founded in Berlin, June 8-9, 2023
  - HiRSE_PS part of the steering committee and WGs
- Co-organized un-deRSE23:
  - Dornburger Schlösser Jena, Sep 26-28, 2023
  - First (and last?) German RSE unconference
- Co-organizing deRSE24:
  - Julius-Maximilians-Universität Würzburg, Mar 5-7, 2024

RSE Summer School 2024

HIRSE_PS

Announcement:

First ever German
Summer School on Research Software Engineering
hosted at KIT, Sep 23-27, 2024.

Facts (so far):

- Awesome lineup of trainers also from HiRSE_PS, e.g. Jessica Mitchell, Tobias Schlauch, the JuRSE team, KIT
- Broad range of topics, e.g. Cx, documentation, collaborative coding and more
- Aiming at a low attendance fee, up to 70 participants possible
- Interfacing with communities outside Helmholtz e.g. NHR
- Interested? More information soon.

Summary and Outlook

HIRSE_PS

- **Continuing and extending the HiRSE Seminar Series**
  - HiRSE Tutorials: short (2-3h) tutorials on a specific topic
  - International outreach and speakers
- **More collaboration and outreach**
  - GI special interest group RSE
  - Close-by universities and local partners
  - RSE Summer School 2024
- **Integration of more groups**
  - Simulation and Data Labs (SDLs) within Helmholtz Information
  - Support for individual RSEs within the centers
- **More visibility of RSE and RSEs within RF and HGF**

JÜLICH Forschungszentrum

KIT Karlsruher Institut für Technologie

HZB Helmholtz Zentrum Berlin

Helmholtz-Zentrum hereon

6.10 Helmholtz Research Software Directory: Status report and roadmap for 2024

HIFIS HELMHOLTZ FEDERATED IT SERVICES

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Helmholtz Research Software Directory
Status report and roadmap for 2024

Christian Meeßen[1]

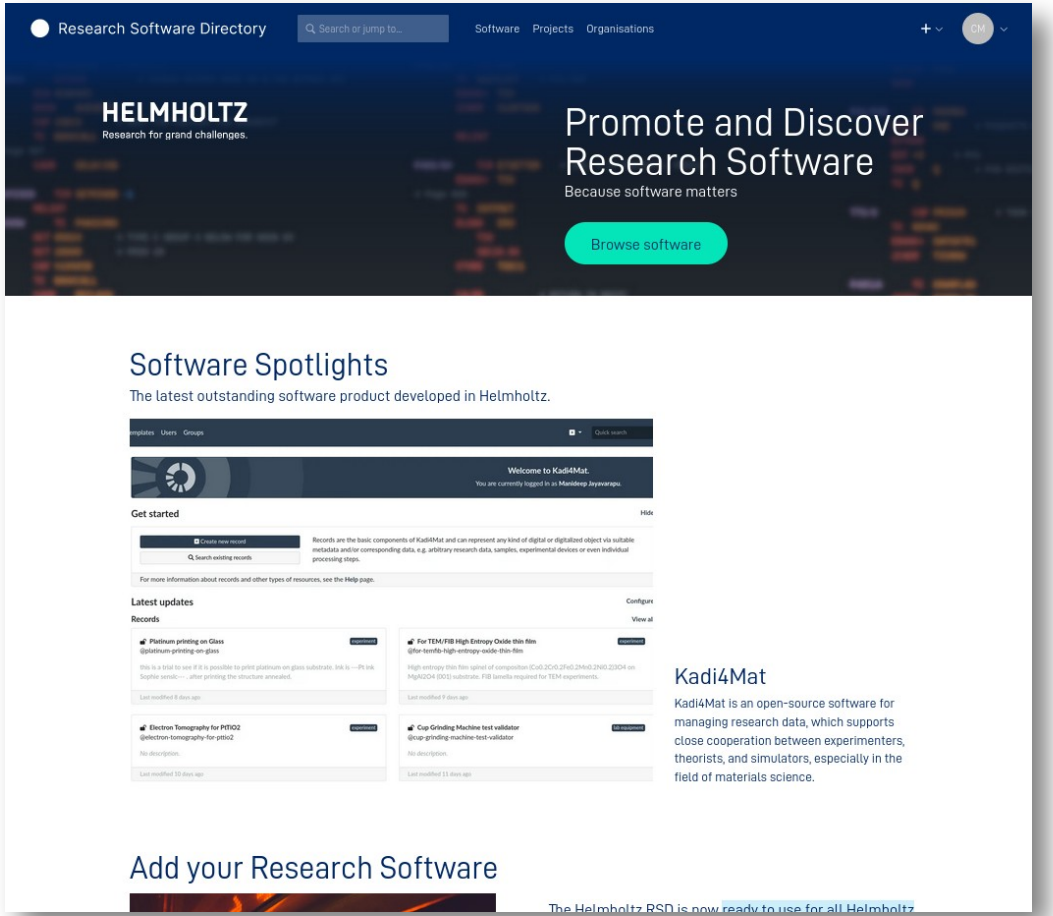
Helmholtz Open Science Forum Research Software Policies

Leipzig
06. Februar 2024

Contributions
Felix Mühlbauer, Tobias Huste, Norman Ziegner, Martin Hammitzsch, Uwe Konrad

1) Helmholtz Zentrum Potsdam, Deutsches GeoForschungsZentrum Potsdam GFZ

The Helmholtz Research Software Directory



- HIFIS Software Community project
- Forked from NLeSC RSD
- Pilot: July 2022
- Official launch: February 2023
- Helmholtz Fork
  - Corporate Design
  - Software Spotlights
  - Research Fields (WIP)
  - Helmholtz ID Authentication
  - EUPL-1.2
- Helmholtz Cloud Service
- https://helmholtz.software

The Helmholtz RSD in numbers

17 +4
Helmholtz centres with software contributions

6
Helmholtz centres are maintained

128 +93
Partner organisations

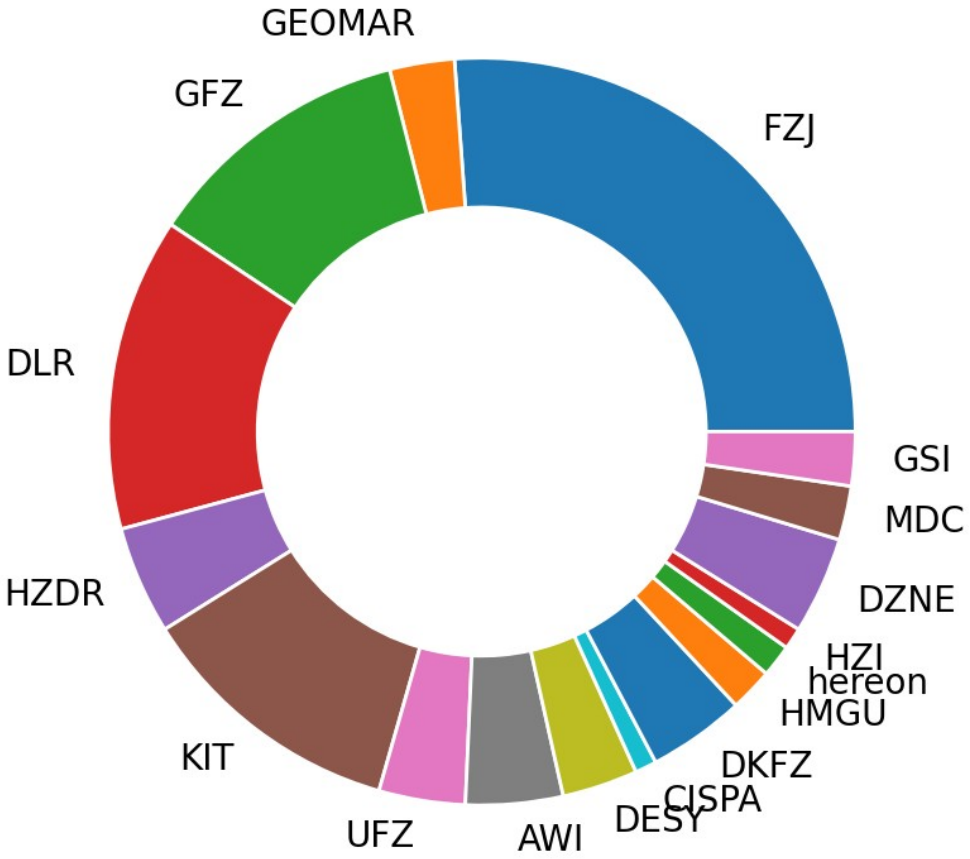
1152 +936
Software contributors

203 +128
Software entries

6541 +6137
Software mentions

Data as of 05. February 2023, compared to 15. May 2023.

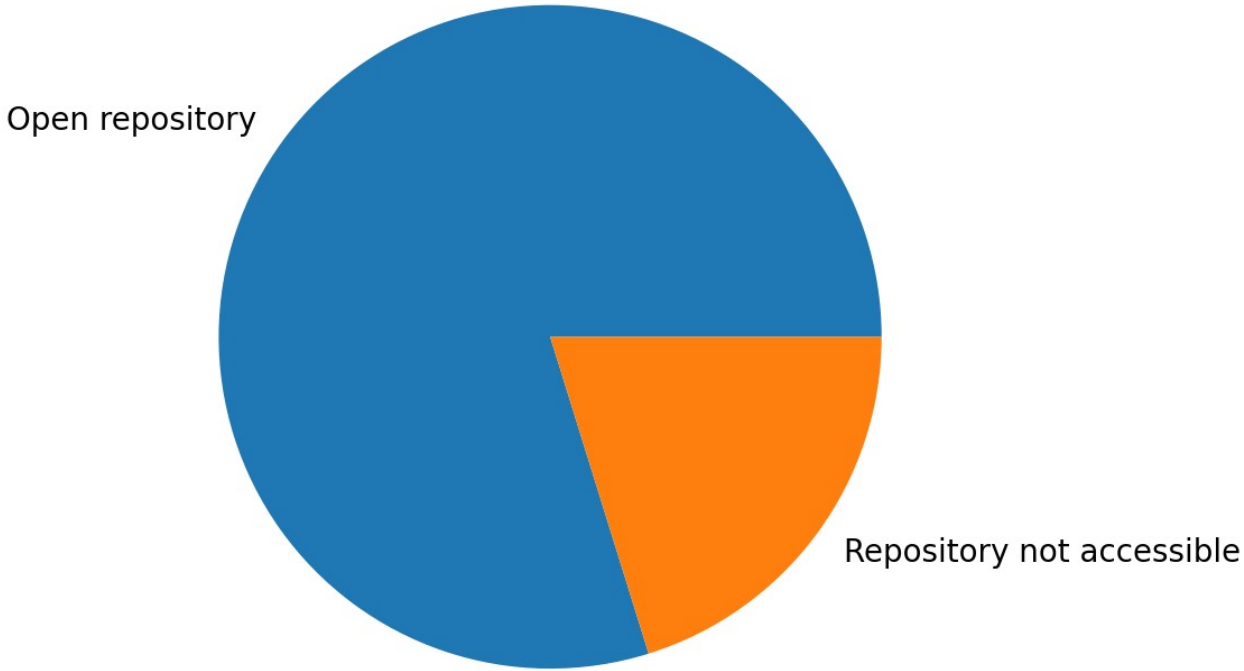
Software count by research centre



Total: 214

Data as of 5. February 2023

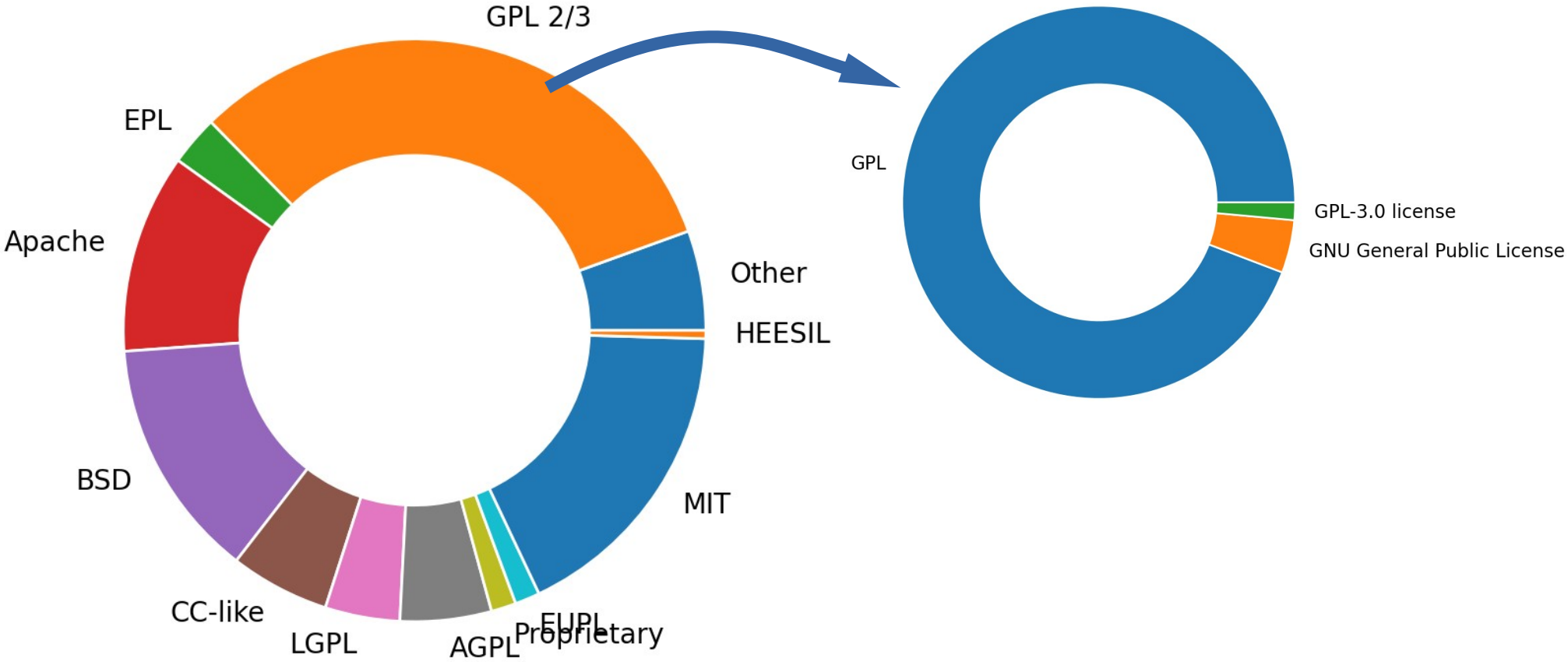
Repository accessibility of available software entries



Data as of 5. February 2023

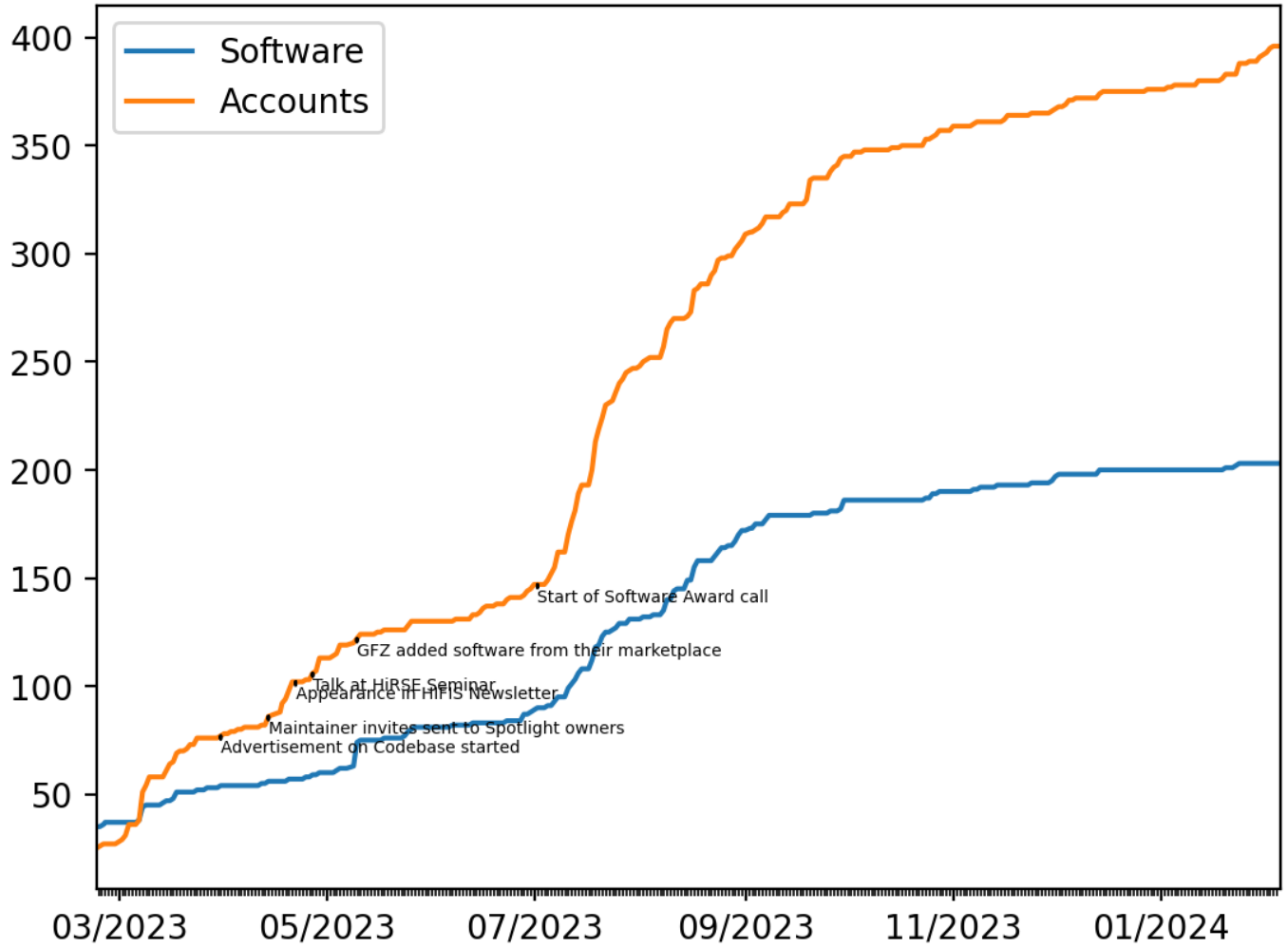
Data as of 5. February 2023, n=217 licenses

Repository accessibility of available software entries



Other: (we are currently working towards an open source license), Academic Public License, Artistic-2.0, EMODNET-1.0, GEBCO-1.0, ICES-1.0, MPL-2.0, Non-commercial, https://messy-interface.org/licence/conditions/, https://www.desy.de/~xraypac/license.html, https://xm.cfel.de/research/scientific_software/xcascade_3d/, scientific

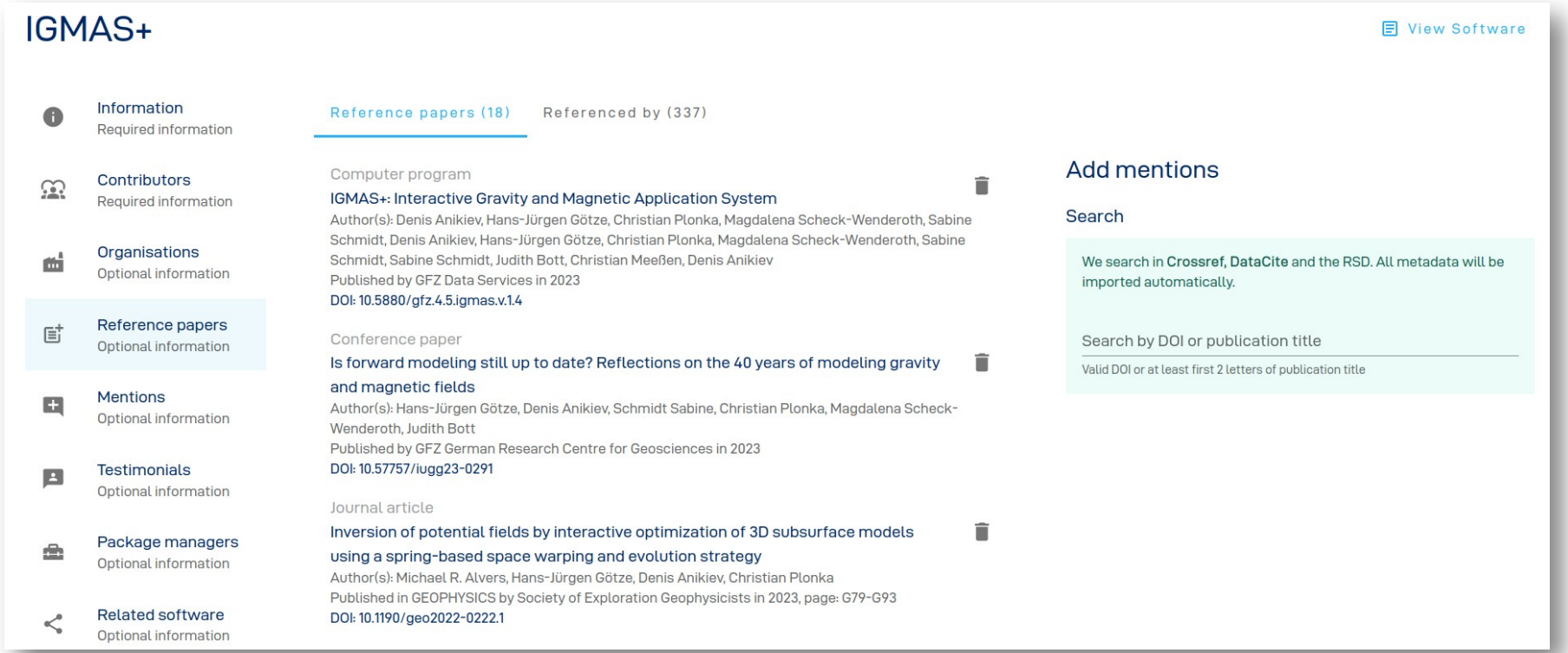
History of Software entries and Accounts



New features in the Helmholtz RSD

New: Automated mention scraping

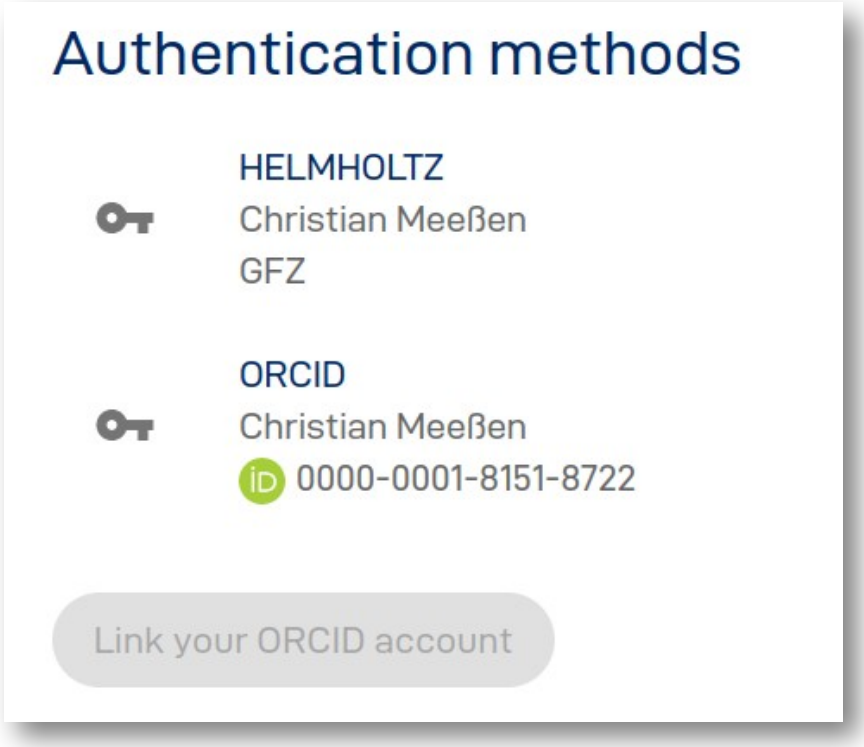
Scrape citations of **Reference papers** from OpenAlex. Scraped citations appear in the mentions section.



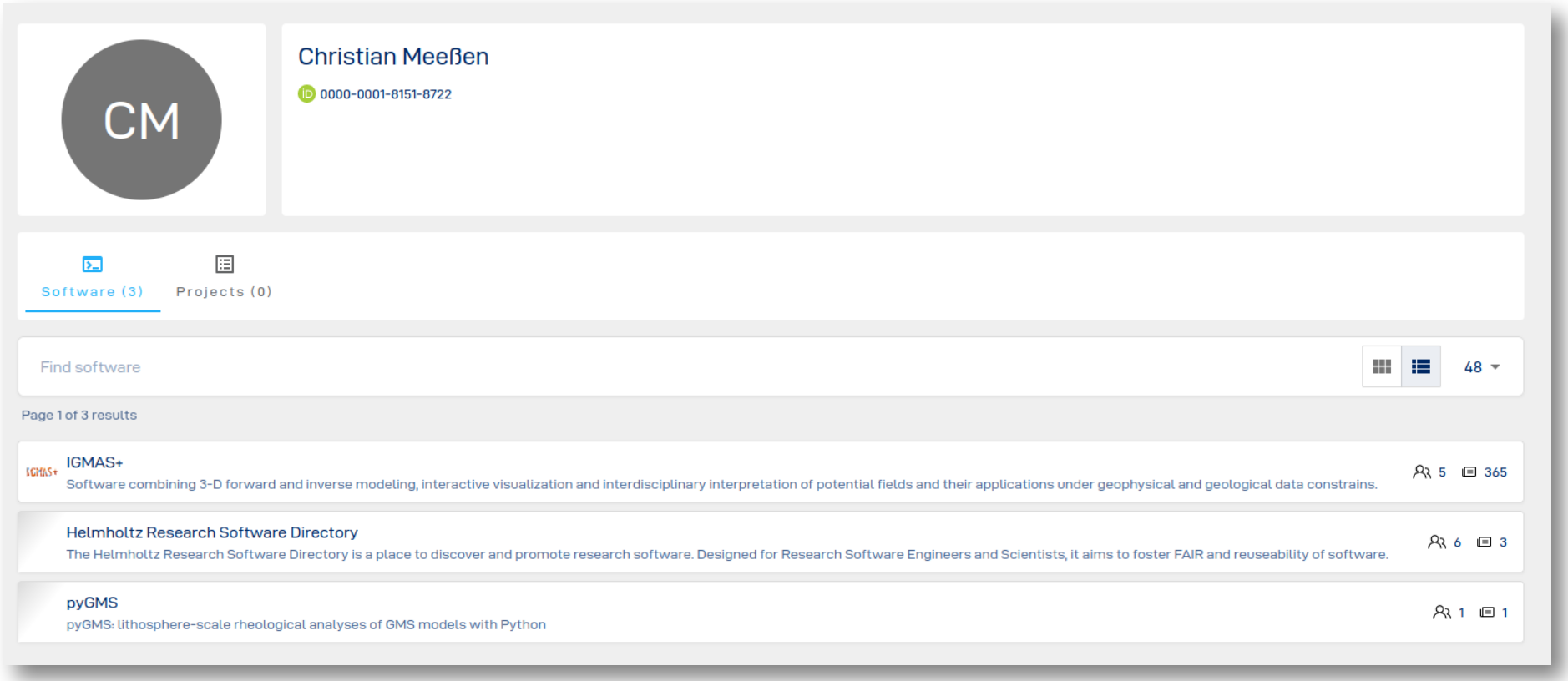
Automated mention scraping

New Community feature: personal profile pages

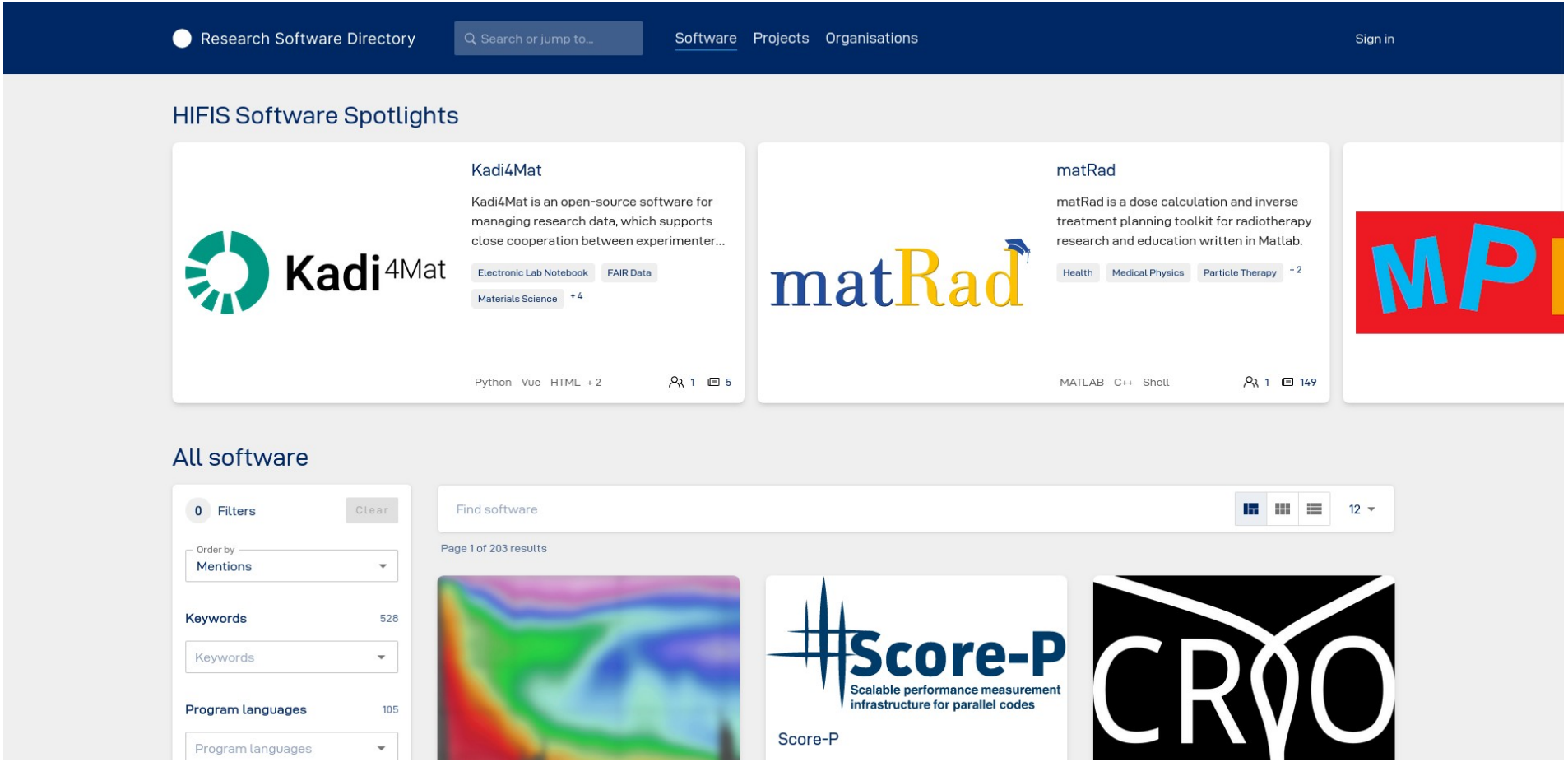
- Users can couple their ORCID to their RSD account
- Personal profile page shows all software and projects a user is related to



Community features: personal profile pages



New: HIFIS Spotlights

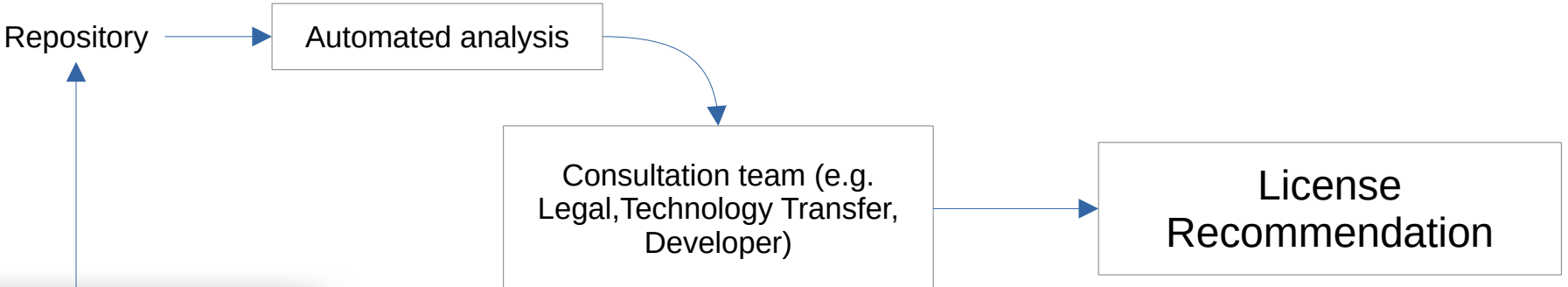


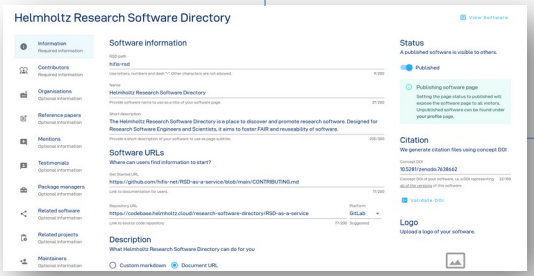
1st international RSD Developer Meetup

- Overall goals for 2024
  - Organise community (devs and users)
  - Steering committee
  - Support for communities in the RSD
- Joined NLeSC & HIFIS Roadmap
  - Jul/Aug: Standardised API (probably CodeMeta)
  - Sept/Oct: Communities
  - Nov/Dec: License consultation
- Additional HIFIS Goals
  - Python package for automated push (FZJ + GFZ)
  - Software indicators (prototype)

January 2024

License consultation





Manually entered metadata

- Consultation at the centres with centre-specific workflow
- Alternatively: standardised approach via HIFIS Software Consulting

Software Indicators

- Helmholtz Open Science Task Group Indikatorik: White papper on Research Software Quality indactors
- “rs-eval” by Jonas Rimatzki
- Goal for 2024: prototypical implementation of an indicator in RSD with semi-automatic assessment using the HGF OS TG SW questionnaire

Thank you for your attention

For **questions**, **requests** or **contributions** contact me:

Mattermost HIFIS Channel

or

Our HIFIS Software communtiy bi-weekly
(see Mattermost channel for announcements)

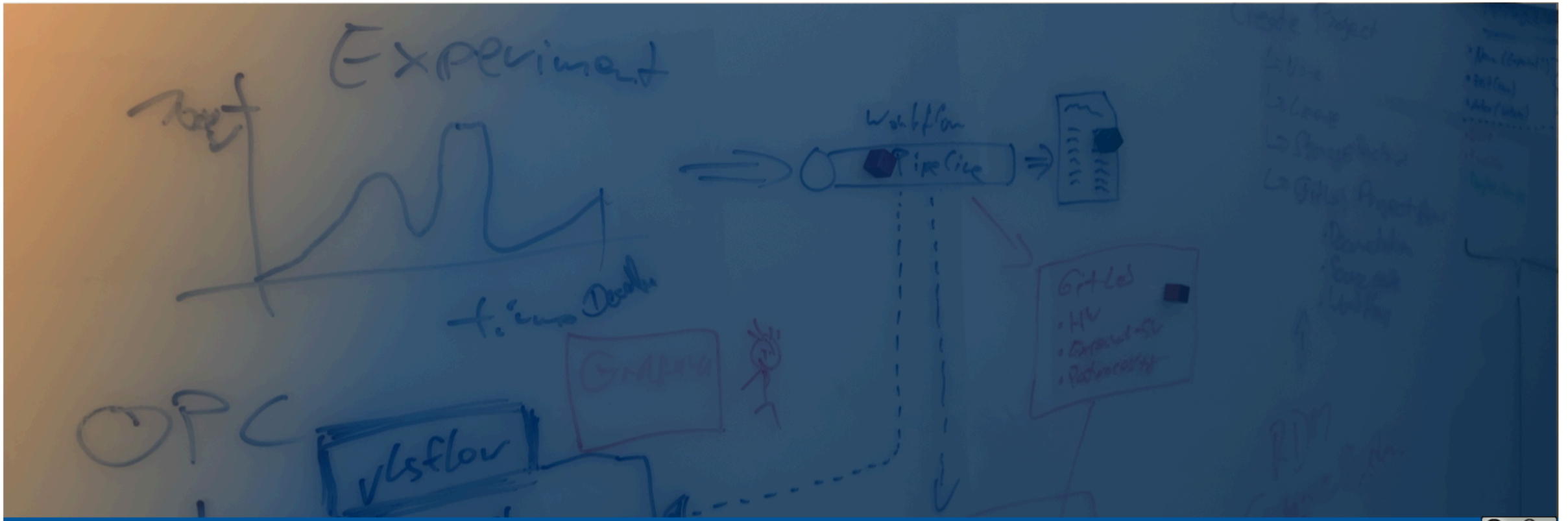
or

support@hifis.net

6.11 Open Research Project Guidance System: HELIPORT

HELIPORT

JÜLICH FORSCHUNGSZENTRUM
HI JENA Helmholtz Institute Jena
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF



Open Research Project Guidance System: HELIPORT

Helmholtz Open Science Forum: Research Software // February 6th, 2024

Tobias Huste, Oliver Knodel, Martin Voigt, Robert Ufer, David Pape, Mani Lokamani, Jeffrey Kelling, Stefan E. Müller, Thomas Gruber, Guido Juckeland, Alexander Kessler, Chien-Li Lee , Joachim Hein, Bernd Schuller // contact: o.knodel@hzdr.de

DRESDEN concept

Our Research Facility and our Large Scale Research Infrastructures

The Helmholtz-Zentrum Dresden - Rossendorf

— Employees approx. 1,470. Thereof 670 scientists.

— HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Research Fields

— Energy, Health and **Matter**.

ELBE – Center for High-Power Radiation Sources

— Electron accelerator, free-electron lasers & THz source.

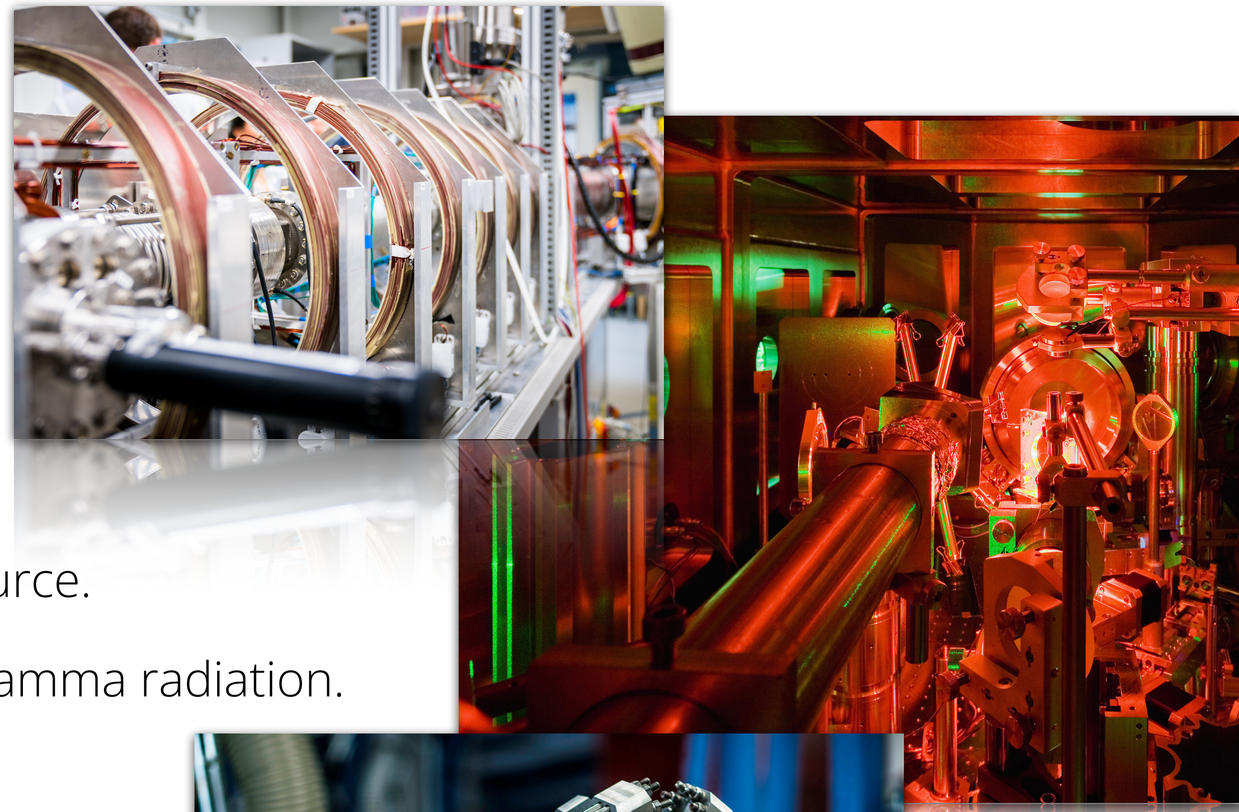
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

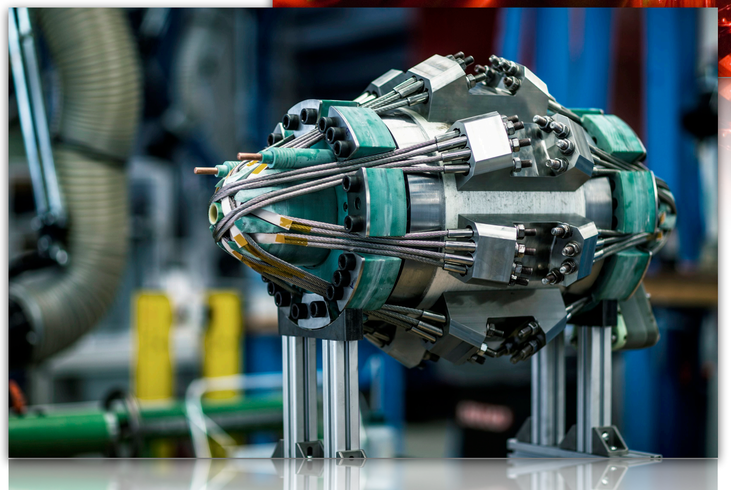
Dresden High Magnetic Field Laboratory (HLD)

— Europe’s highest pulsed magnetic fields.

Ion Beam Center (IBC)

— Nanoscale surface analysis and modification.

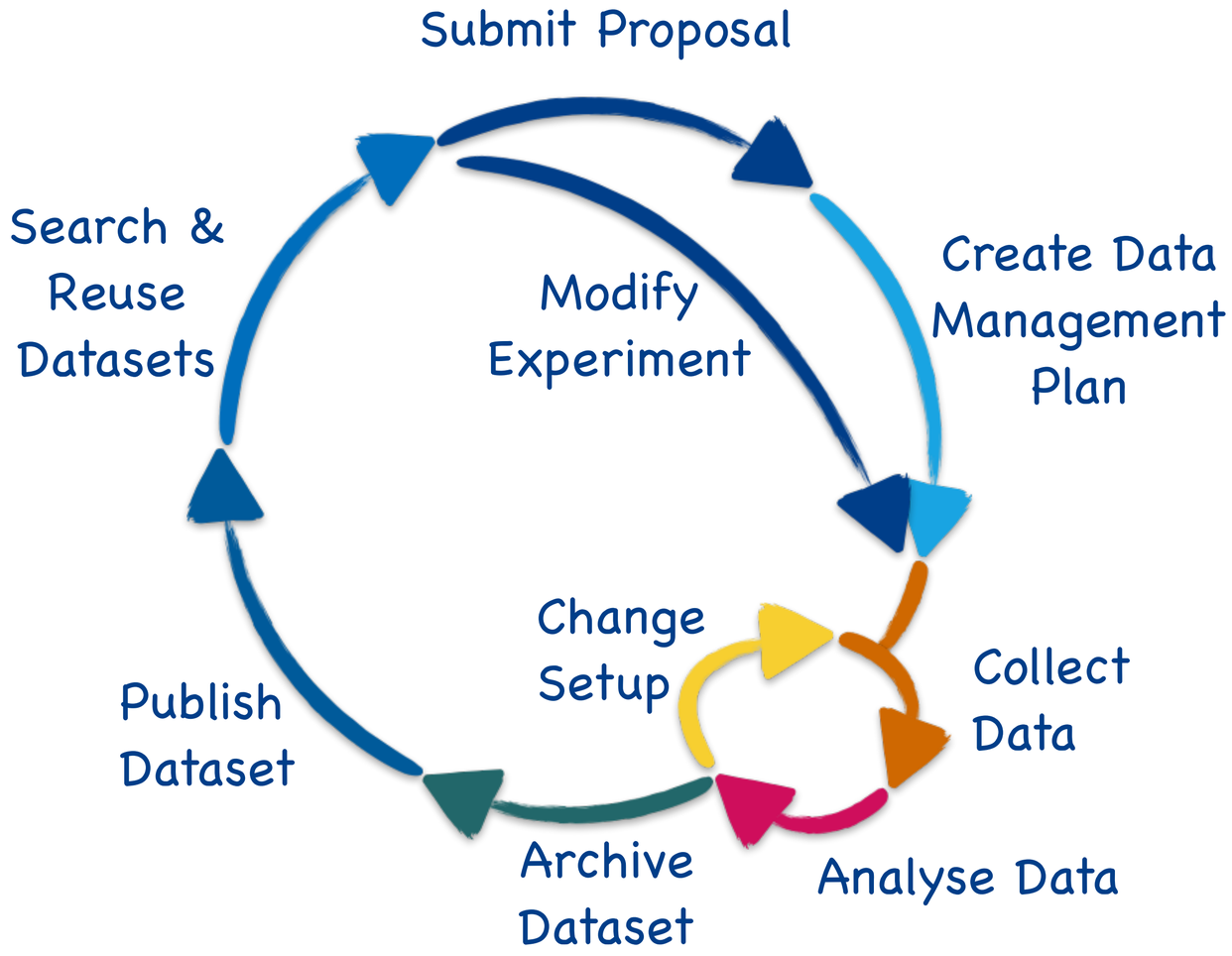




HZDR

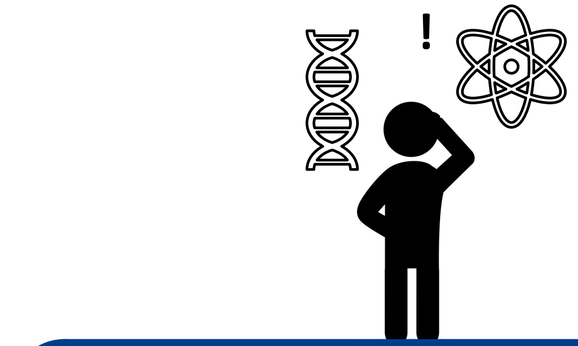
Our Challenge: An End-to-End Digital Data Lifecycle

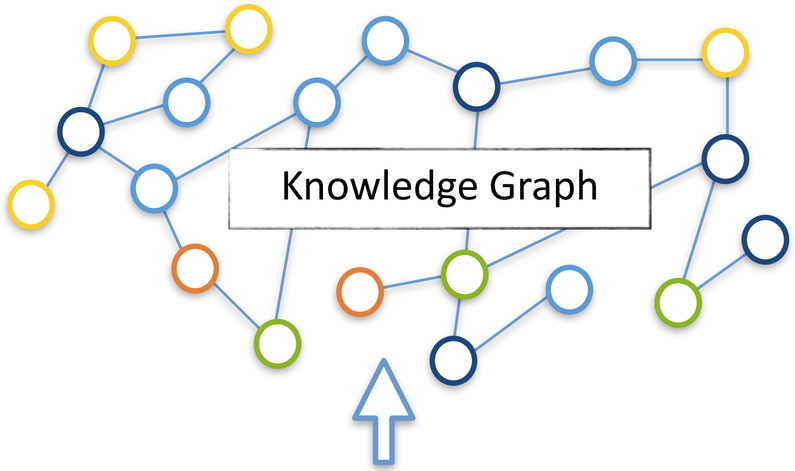
- We support many steps of our different research experiment (matter, energy and health) with tools:
  - electronic lab books,
  - interactive analysis,
  - publication of datasets,
  - scientific workflow management,
  - Handle generation and management.
- A uniform and smooth access to and between all services and systems in our ecosystem is necessary.
- The documentation of all these linked resources is essential to create a comprehensible and FAIR data lifecycle.

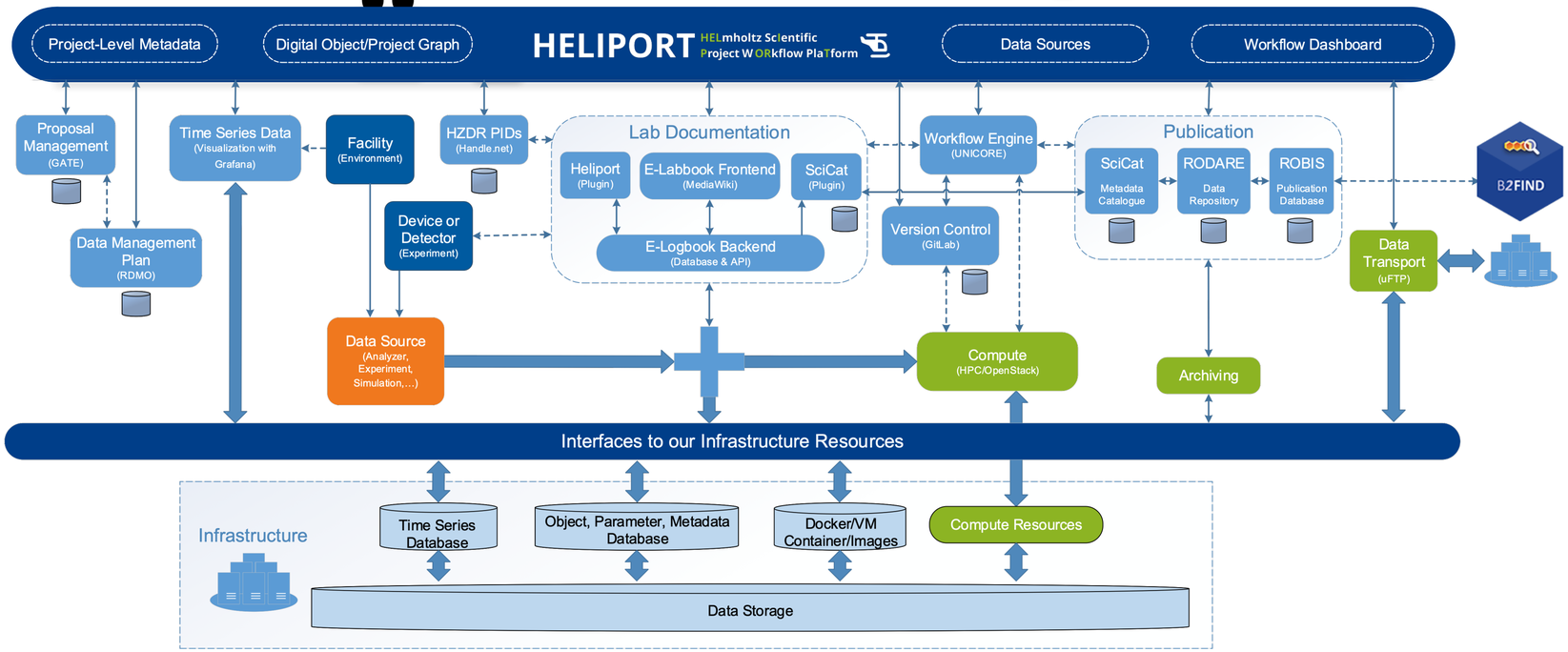


Our Observations and Experiences

- Our IT infrastructures can support various experiments, but they are complex…
- Scientists often don't know which services are available and how to use them.
- An overarching system guiding our scientists (and visitors) through the lifecycle of their research project is essential.
- In the future we can provide an overall Helmholtz-wide knowledge graph!







DRESDEN concept HZDR

HELIPORT

HELmholtz Sc**I**entific
Project W**OR**kflow Pla**T**form

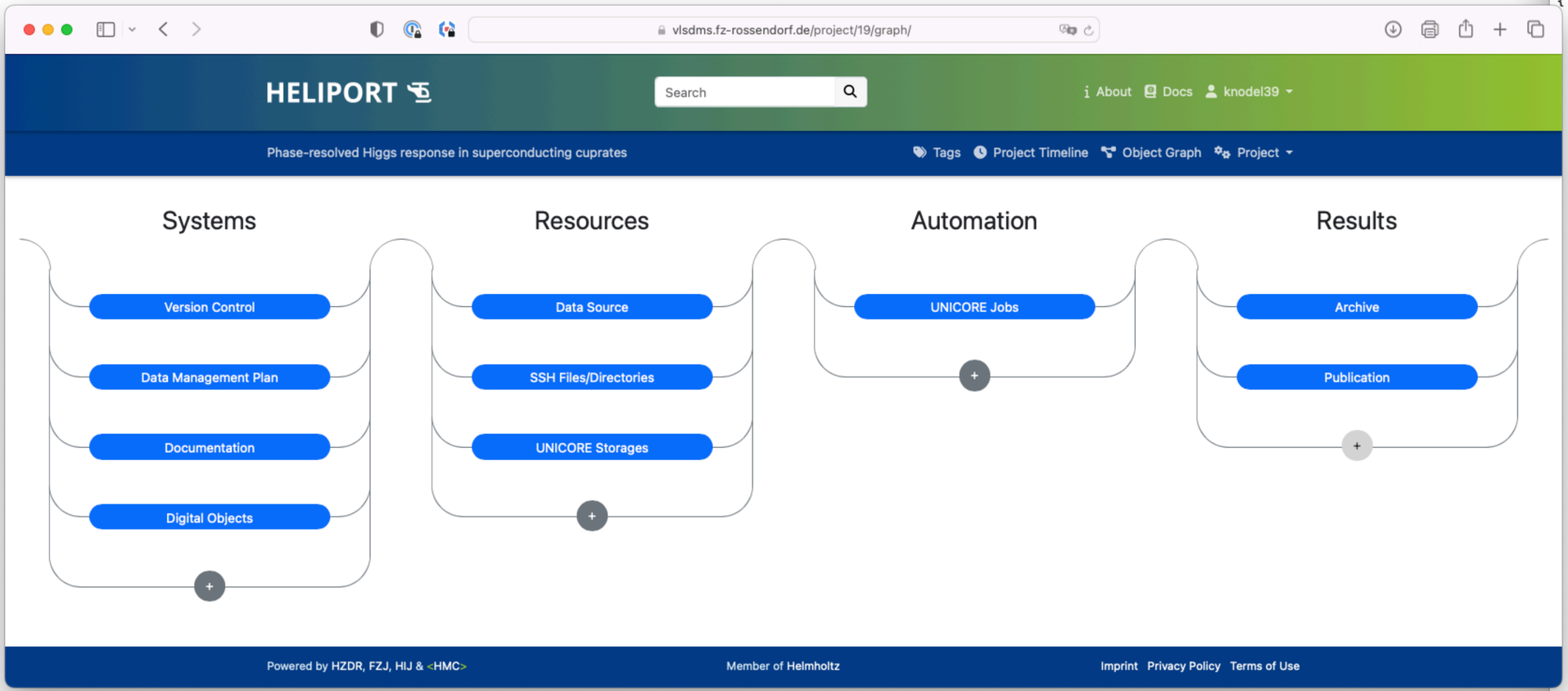
> The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR and comprehensible** project description.

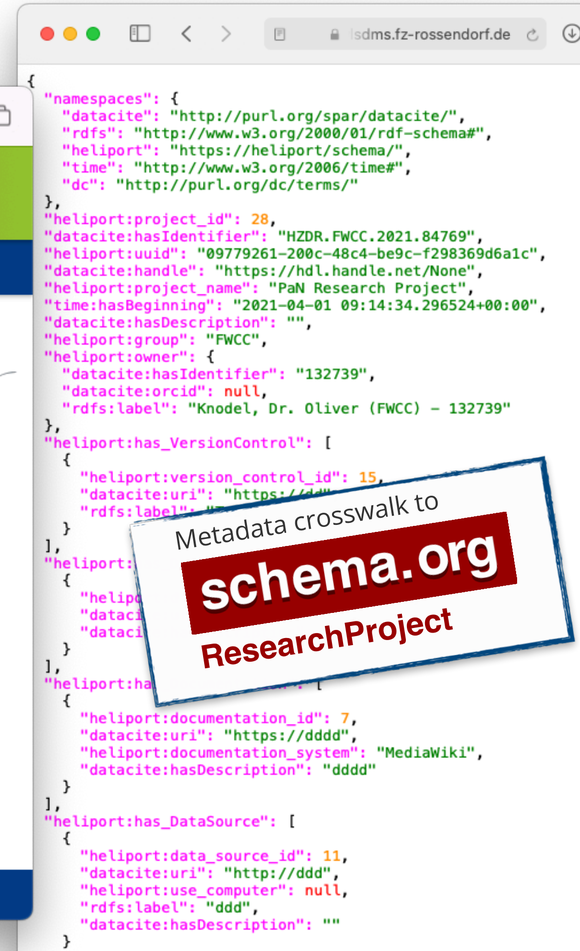
Project Members:

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
HI JENA Helmholtz Institute Jena
JÜLICH FORSCHUNGSZENTRUM

Funded by:

<HMC> HELMHOLTZ Metadata Collaboration





The Motivation to Develop HELIPORT

— HELIPORT was originally intended to provide only the **proposal's metadata**, to allow the assignment of resources.

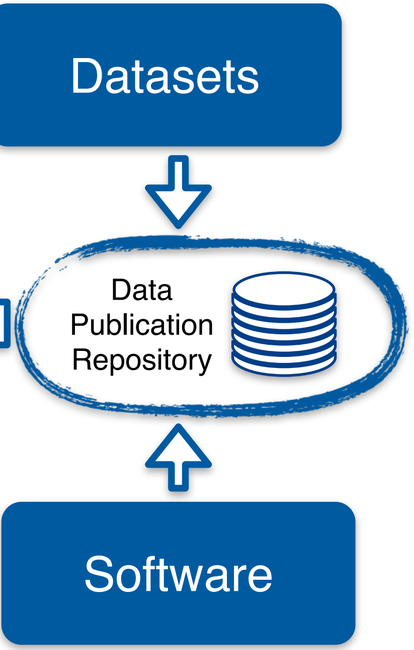
— Over time, we realised that HELIPORT can also answer our scientists' most important questions, such as:

What are the necessary steps towards a full comprehensible and FAIR research experiment ensuring data provenance?

How we can reduce computations and save energy?

How can we **automate recurring processes** and keep track of status and data products?

Which datasets or software can be **published** (and how)?



How can we bring **new team members** or external scientists into our project lifecycle and associated services/tools?

Where are data, software and how can I gain **access** to both of them?

HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

HELIPORT Features

- Entry point for experiments and scientific projects
- User and group authorisation/management
- Overview of systems and services involved in an experiment
- Provision of metadata from proposal systems (e.g. GATE)
- Registration of and access to internal file systems
- Automated transfer of metadata between involved systems/services
- Background data publication of datasets (e.g. Zenodo, Rodare)
- Integration of reproducible computational workflows
- HPC cluster access (slurm, UNICORE)
- Digital object and handle management with graph visualisation
- Timeline representing changes
- HELIPORT REST API
- Authentication via Helmholtz ID

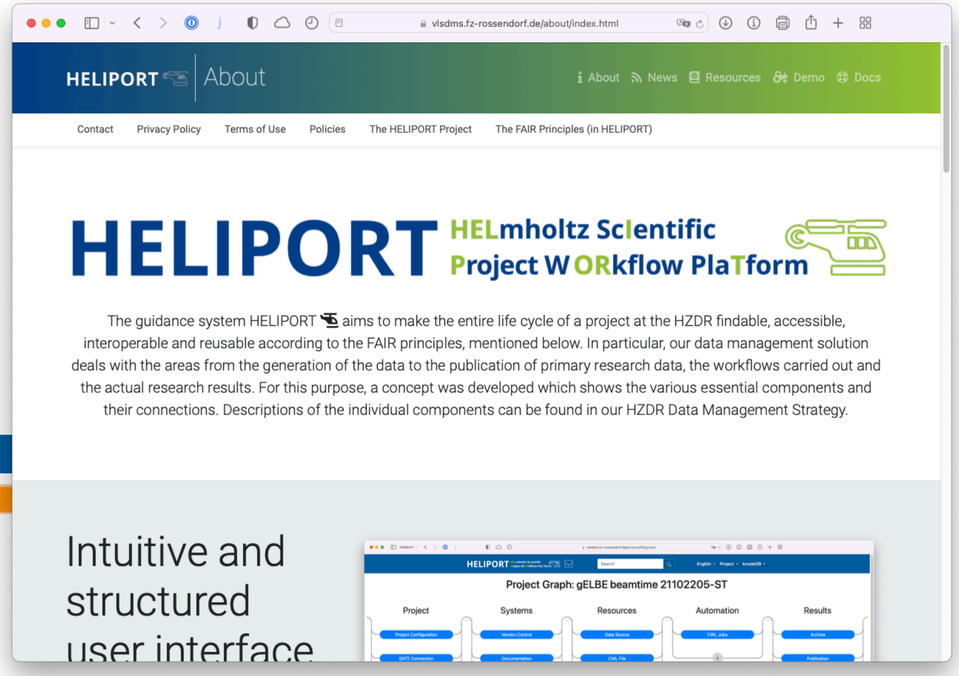
HELMHOLTZ ID

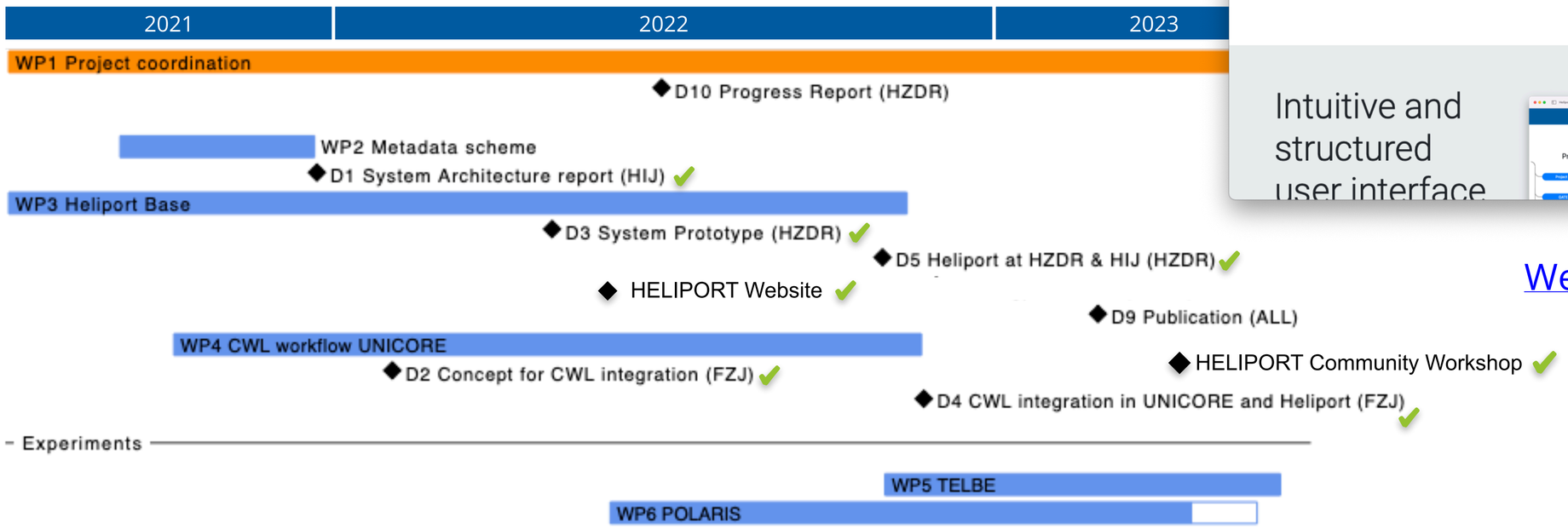
HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

Timescale for the HMC-Project HELIPORT

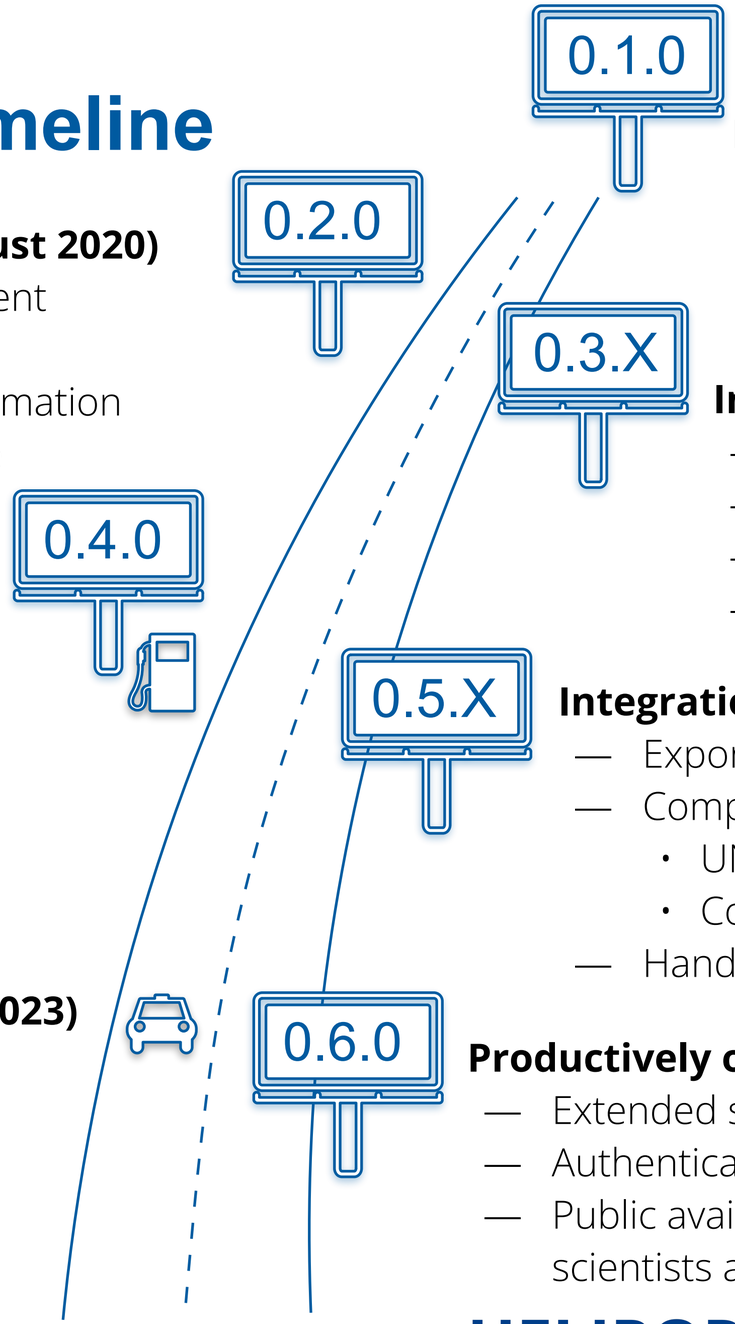
- The deliverables and our prototype are available on our website.
- We are in contact with different Helmholtz centers, universities and European partners and build a HELIPORT community.
- Overview of work packages and milestones:





[Website: heliport.hzdr.de](heliport.hzdr.de)

Heliport (Project) Timeline



Initial Version (June 2020)
- DMS Projects and proposal information from the **HZDR GATE proposal** database
- Webinterface with user authentication (LDAP)

First Draft: Project Plan (August 2020)
- Project and user management
- Configurable stages
- **REST API** for proposal information
- CWL visualization prototype

Improved Project Plan (December 2020)
- Configurable stages and modules
- Infrastructure and database updates
- Daily proposal database update
- Advanced logging and monitoring

Modular Structure (July 2021)
- Official start of the HMC founded Heliport project:

  HMC | HELMHOLTZ METADATA COLLABORATION
- Redesign to provide modular and highly configurable system

Integration of various Apps and Features
- Export for (different) metadata schemas
- Computational/scientific workflow execution
  - UNICORE support ()
  - Computing job management and monitoring
- Handle management with public landing pages

HELIPORT Community Workshop (July 2023)

GFZ Helmholtz-Zentrum Potsdam · HZDR Helmholtz Zentrum Dresden Rossendorf · Jülich Forschungszentrum · GSI · HI JENA

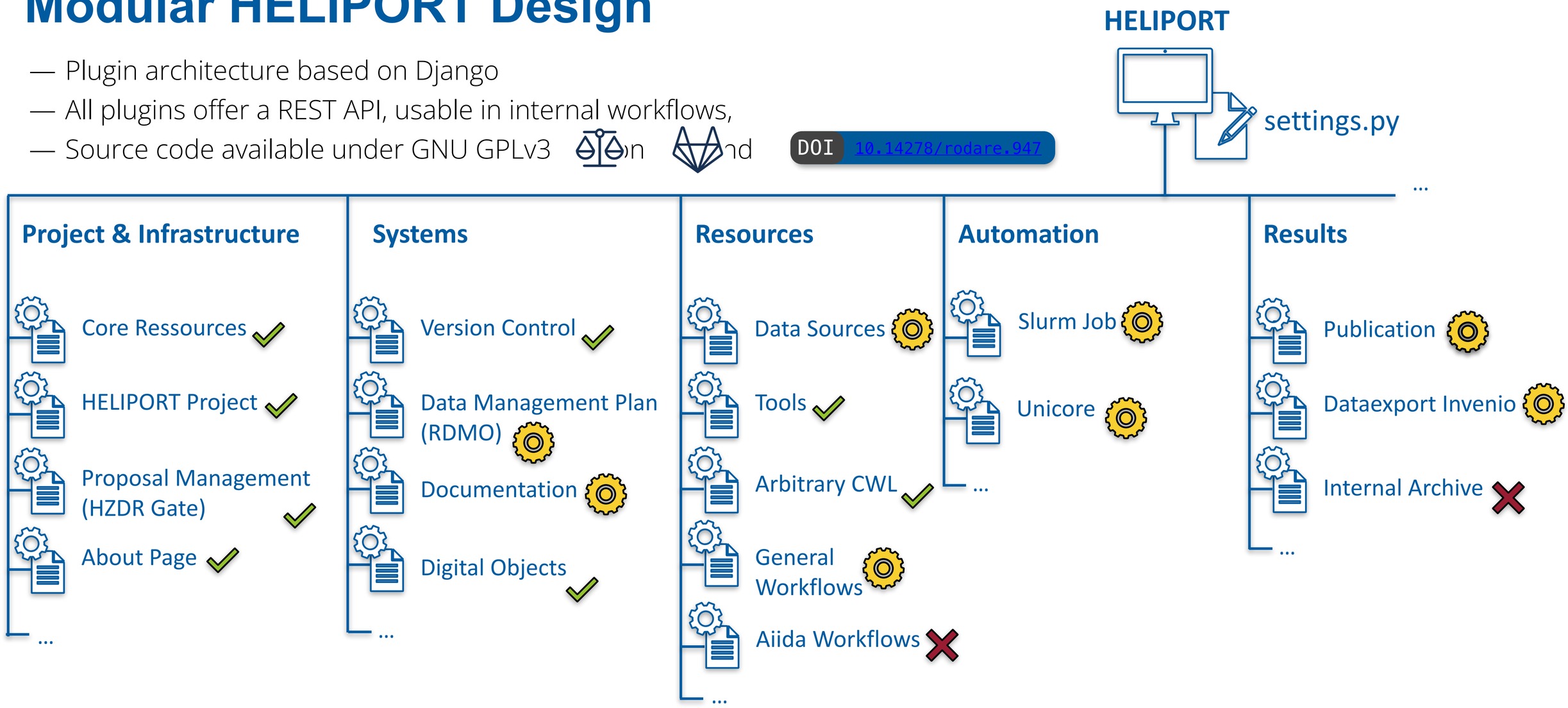
Productively operating HELIPORT for different RIs
- Extended support for a proposal system (GATE)
- Authentication with OpenID Connect (Helmholtz ID)
- Public available HELIPORT instance for remote/visiting scientists at HZDR

HELIPORT · Jülich Forschungszentrum · HI JENA · HZDR

Modular HELIPORT Design

— Plugin architecture based on Django
— All plugins offer a REST API, usable in internal workflows,
— Source code available under GNU GPLv3 on GitLab

DOI 10.14278/rodare.947

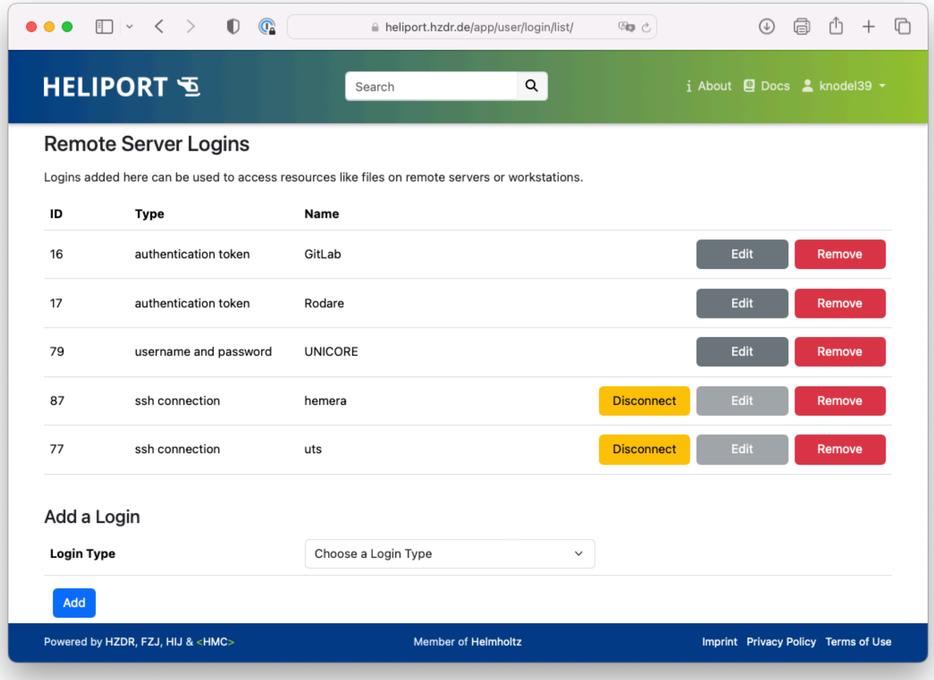


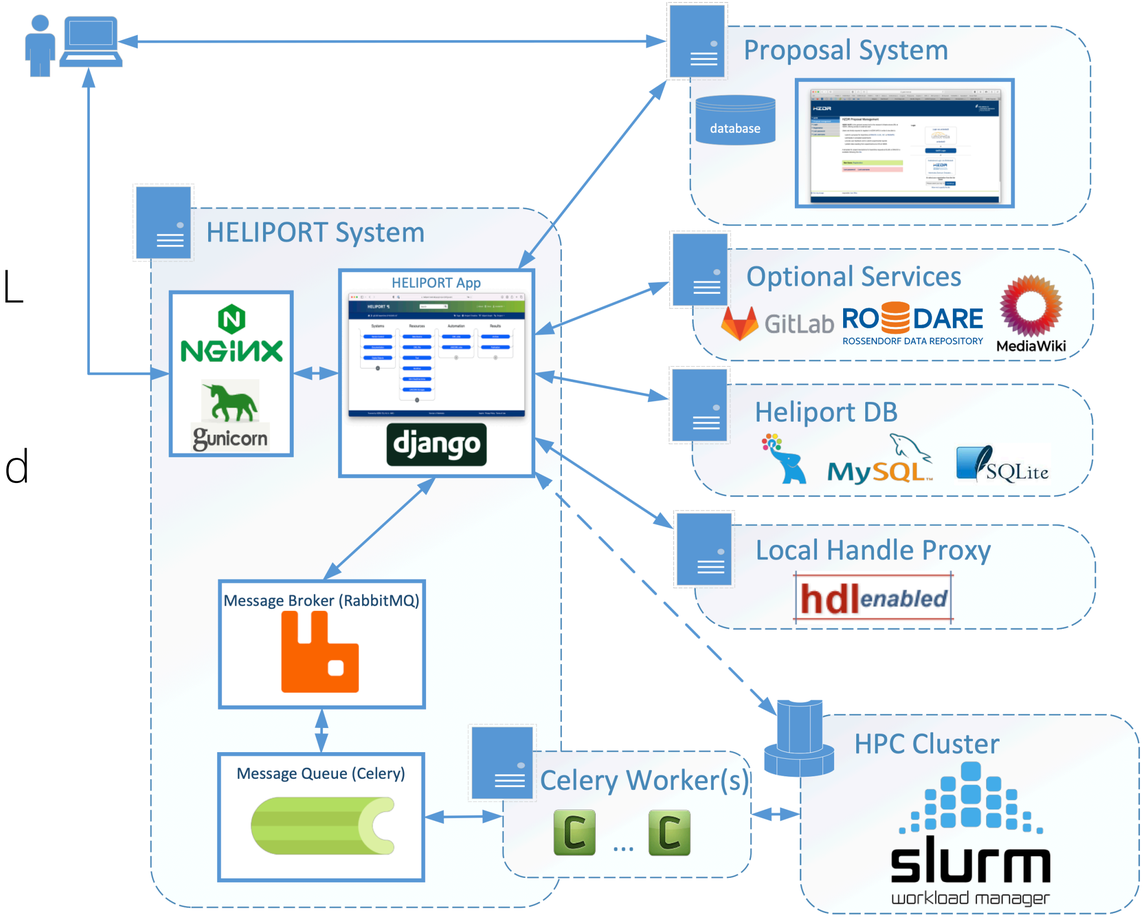
HELIPORT
JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

HELIPORT Infrastructure

— The HELIPORT web app is based on Django:
  - Heliport communicates with various system through REST APIs,
  - The project-level metadata is stored in a PostgreSQL database and can be exported in various metadata schemes.

— Computational workflows are managed in HELIPORT and executed on HPC clusters using slurm or UNICORE.



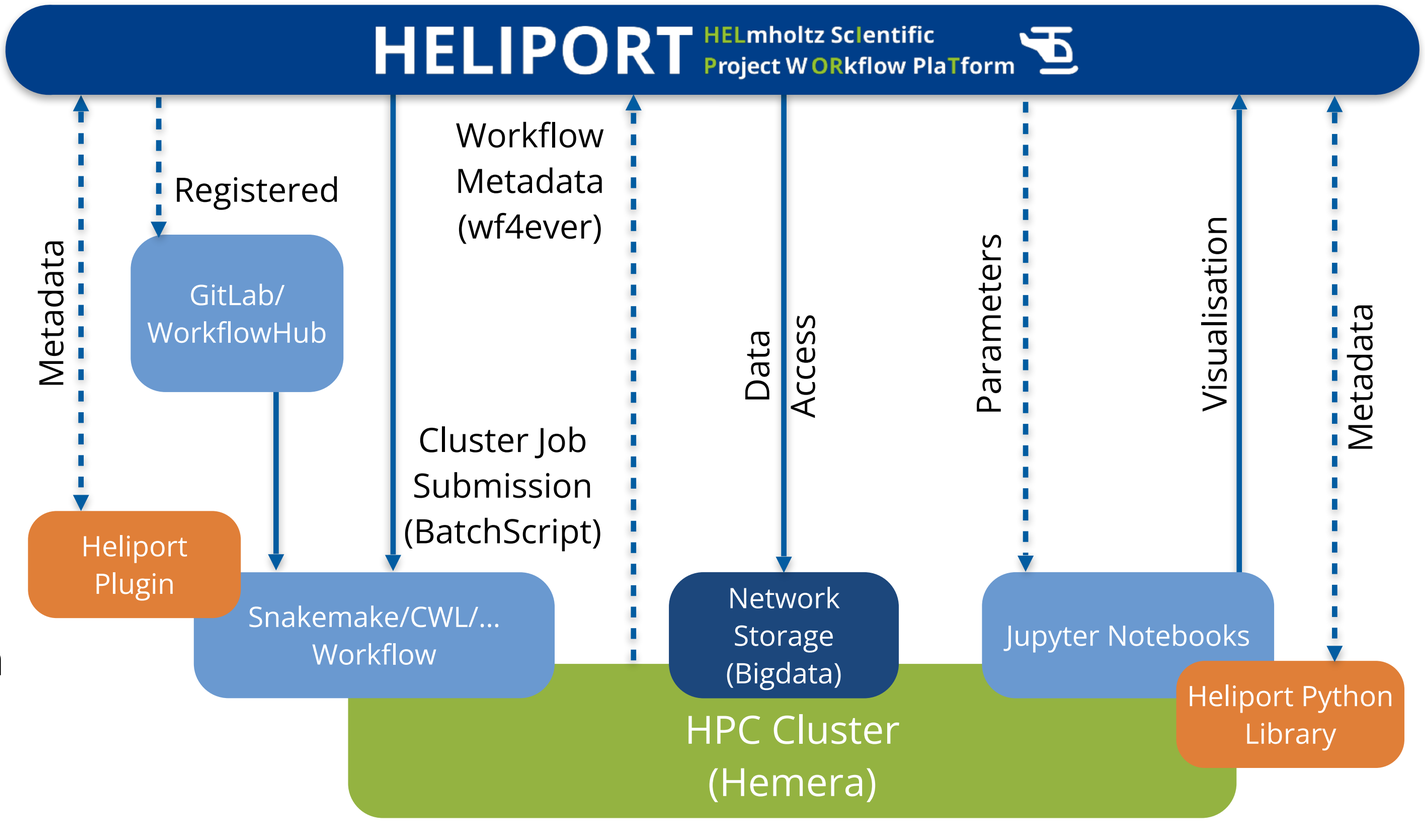


HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

Workflow Architecture (in development)

- HELIPORT offers an infrastructure which permits the integration of various workflow languages and access modes to HPC infrastructures.
- The infrastructure keeps track of and collects the metadata and enables access to all resources involved.
- Next steps:
  - Python library sending workflow information directly to HELIPORT,
  - Provision of provenance information from Jupyter notebooks,
  - Use case: **PIConGPU**

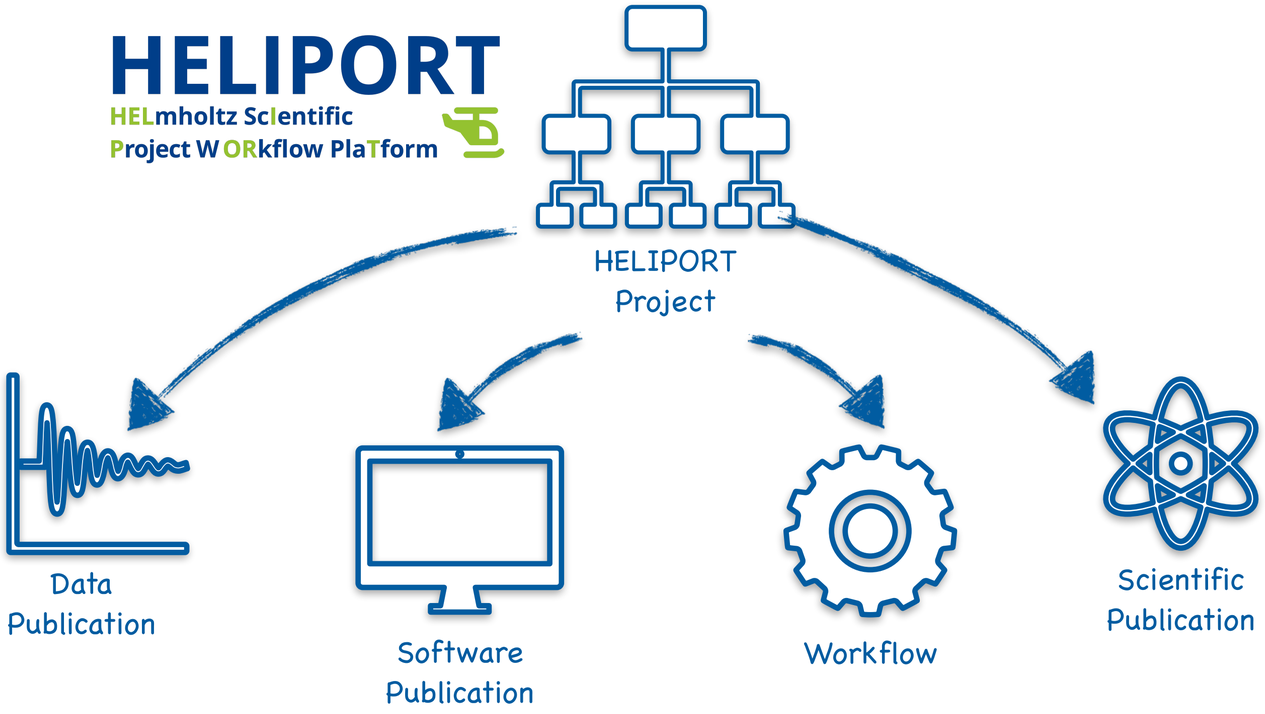


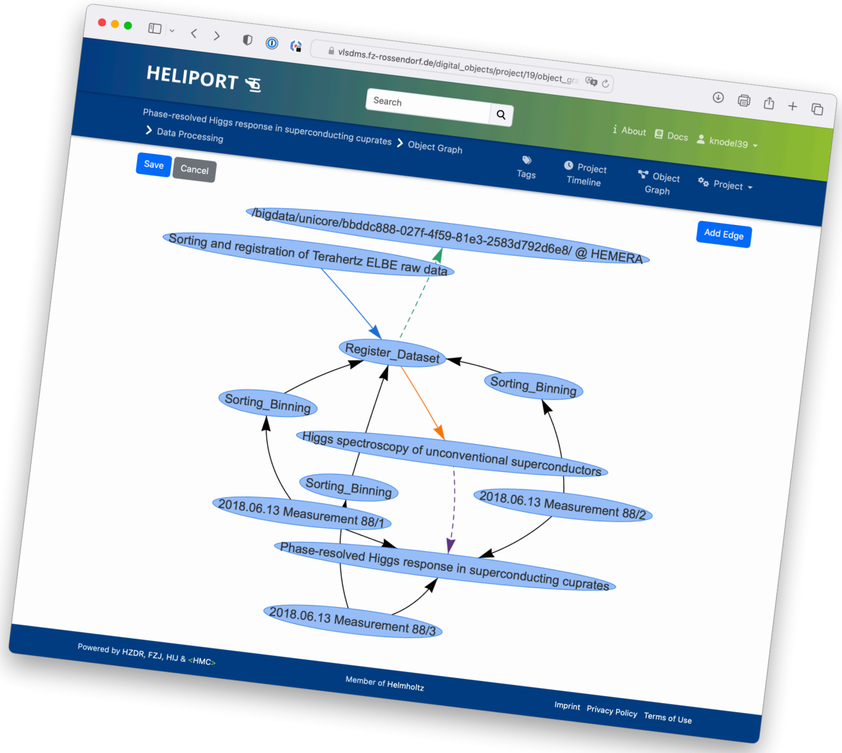
Conclusions

— HELIPORT describes and collects all metadata from all services and systems involved in an scientific experiment.

— Such an approach is desirable and leads us to a fully **FAIR** and **comprehensible** research project.

— The computational workflows are essential to keep track of everything what happened during the experiment.

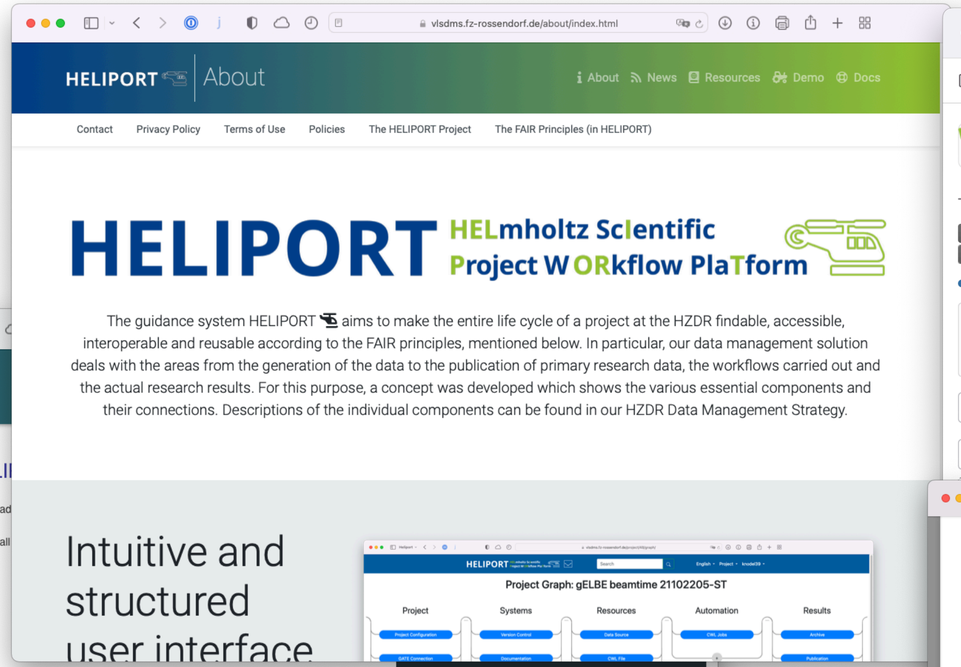


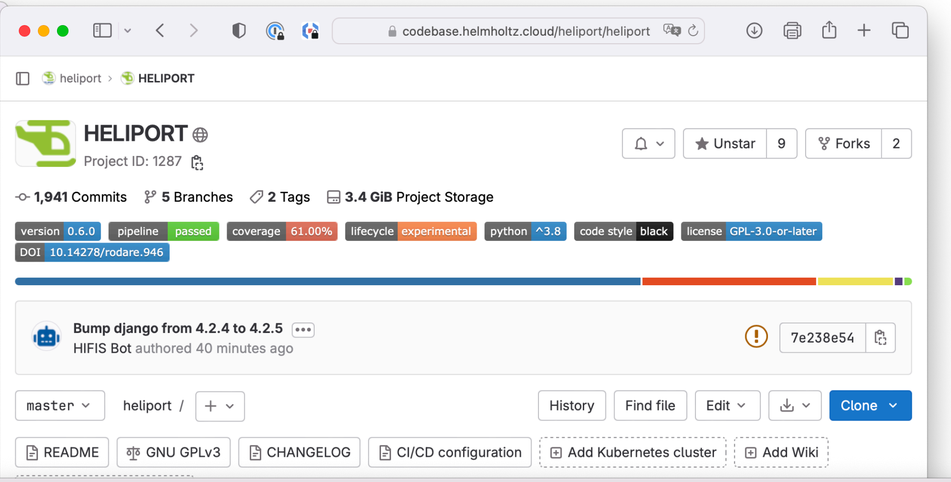


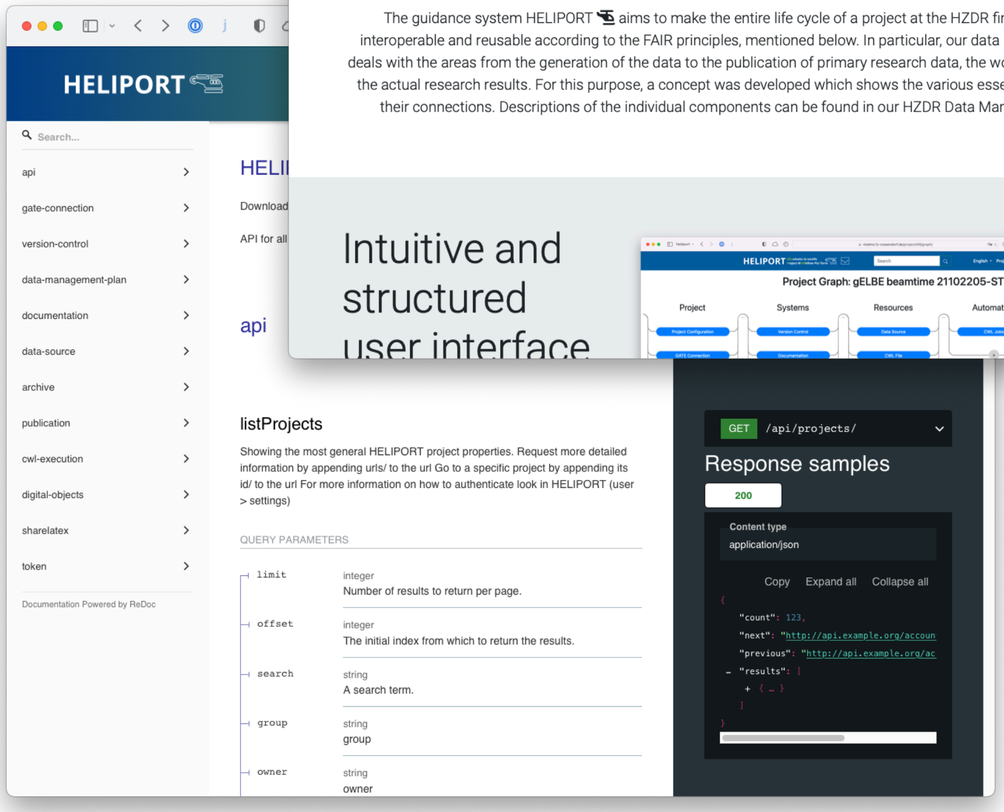
Resources

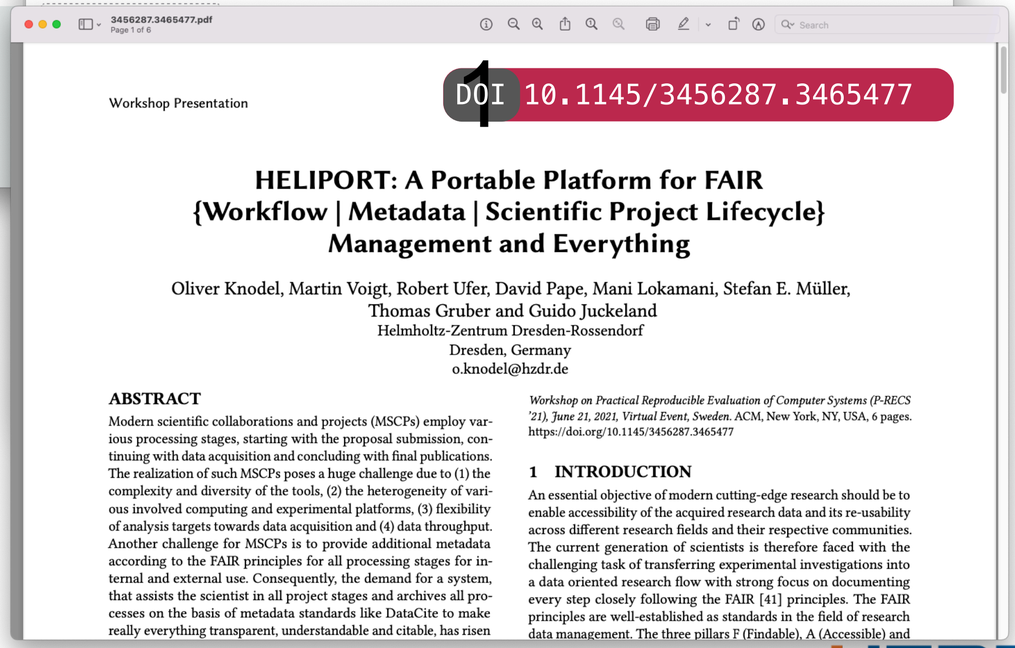
Website: heliport.hzdr.de

Repository: codebase.helmholtz.cloud/heliport









API Doc: heliport.hzdr.de/redoc/

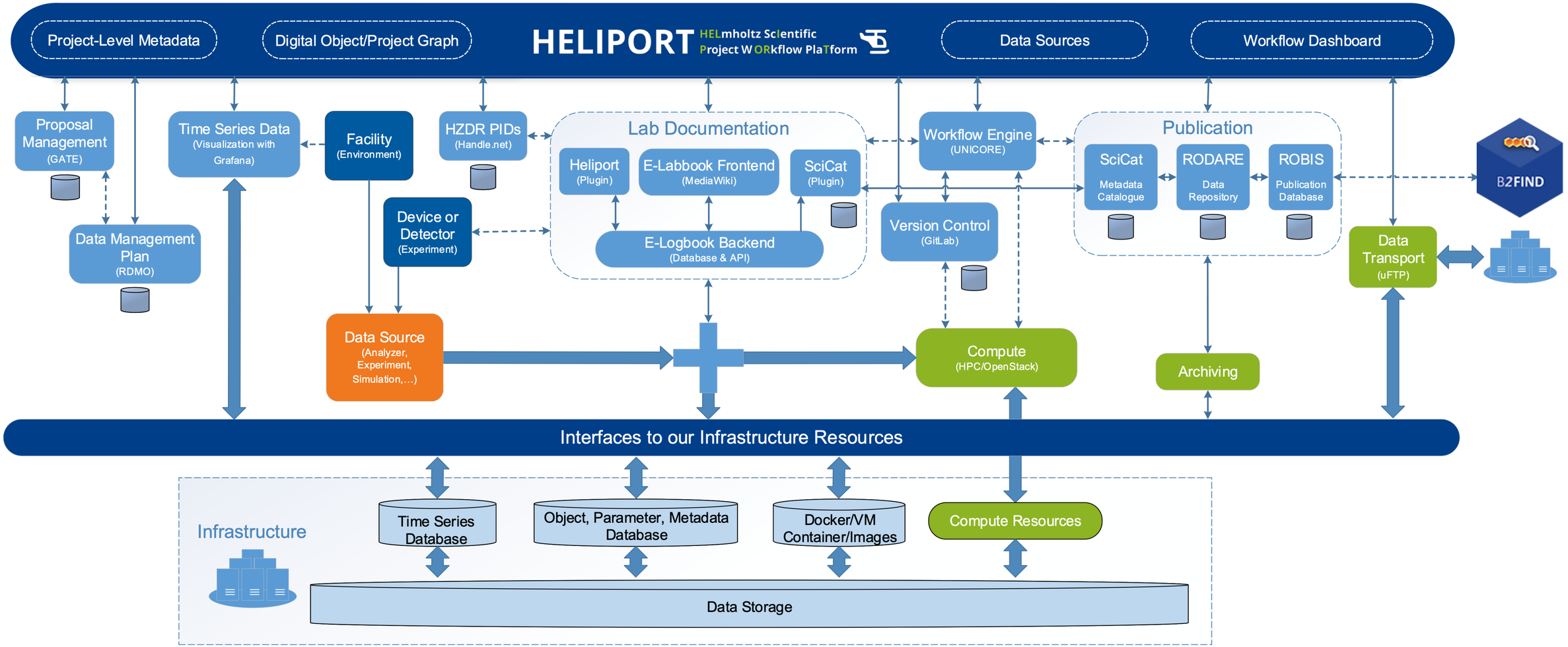
HELIPORT

JÜLICH FORSCHUNGSZENTRUM
HI JENA Helmholtz Institute Jena
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Appendix

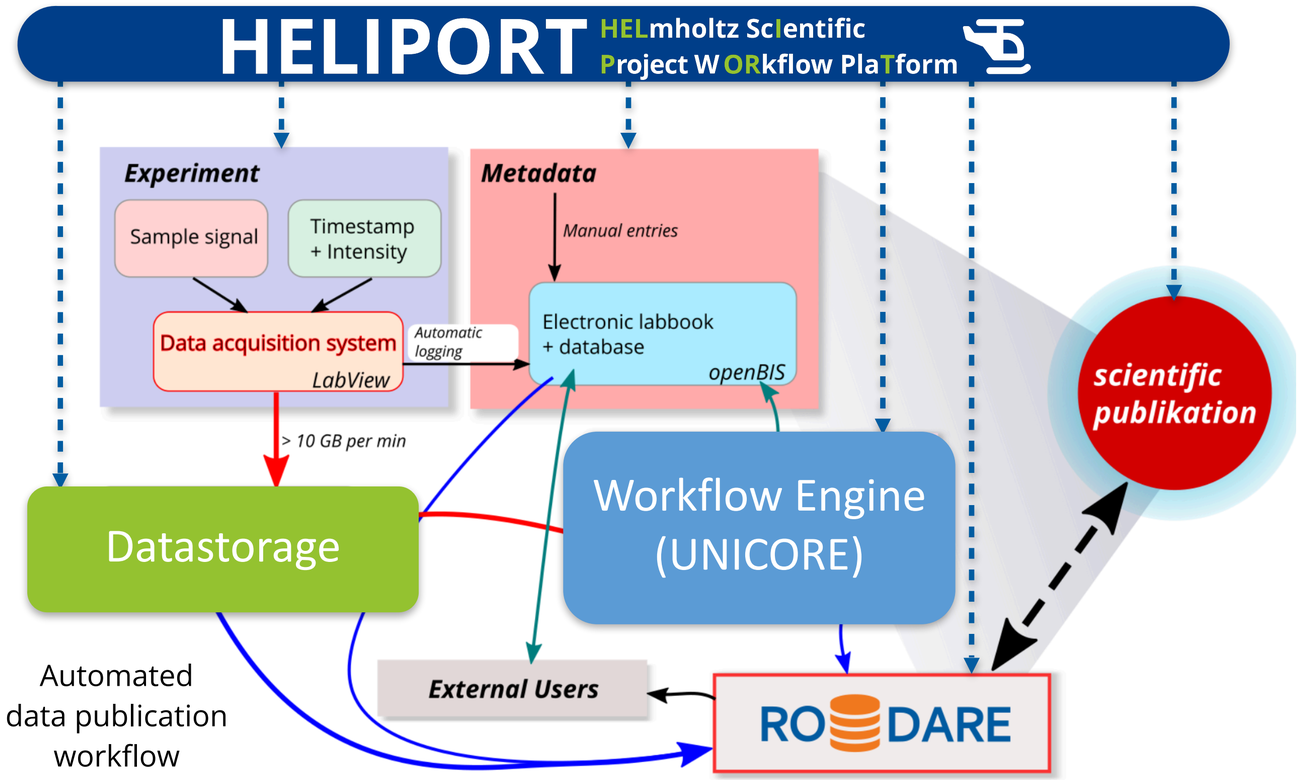
DRESDEN concept

HZDR Research (Infrastructure) Landscape



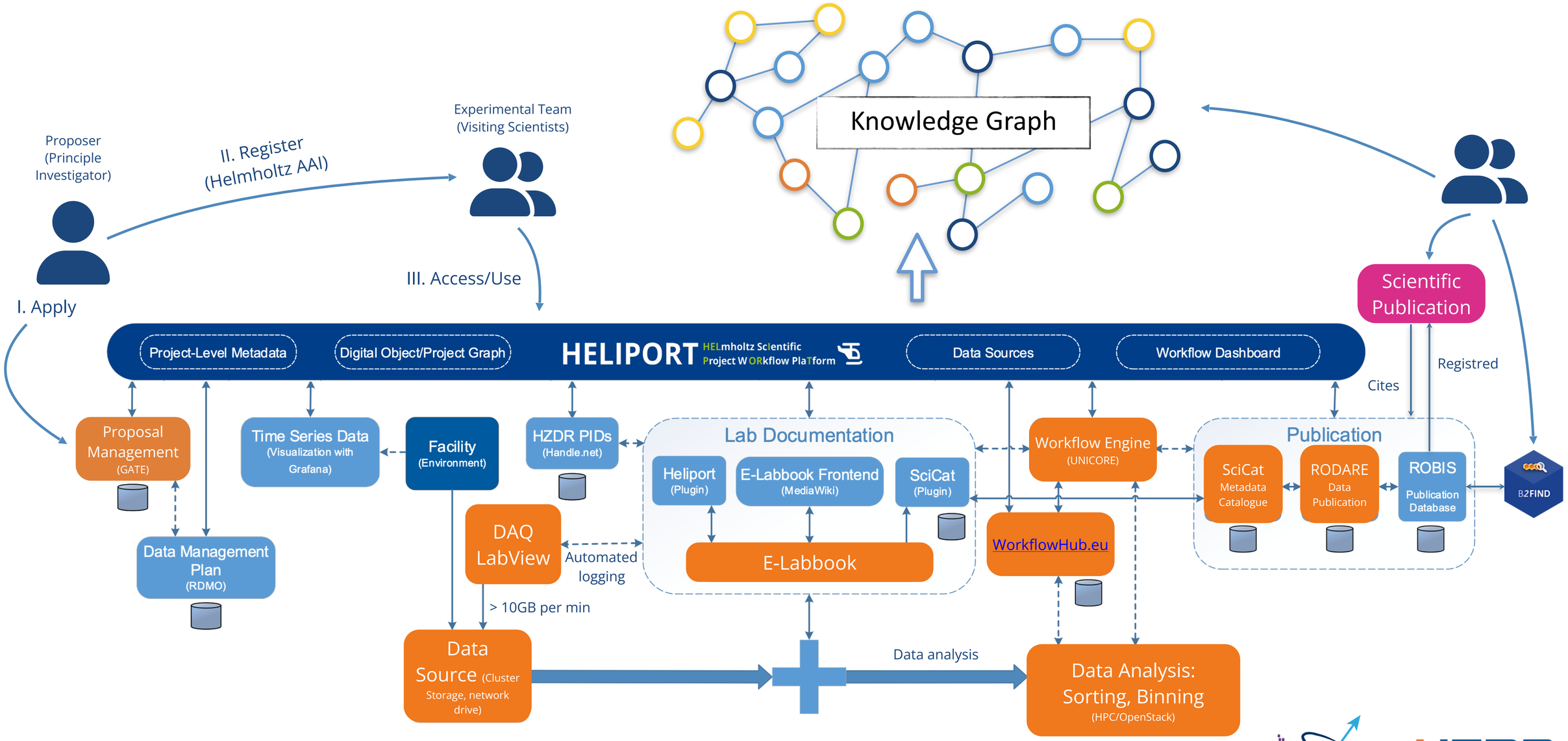
TELBE Data Flow

- Terahertz facility at the ELBE center for High-Power Radiation Sources.
- In the future HELIPORT guides (external) scientists through the complete experiment.
- Submission of data analysis Jobs from LabView to UNICORE with visualisation in HELIPORT



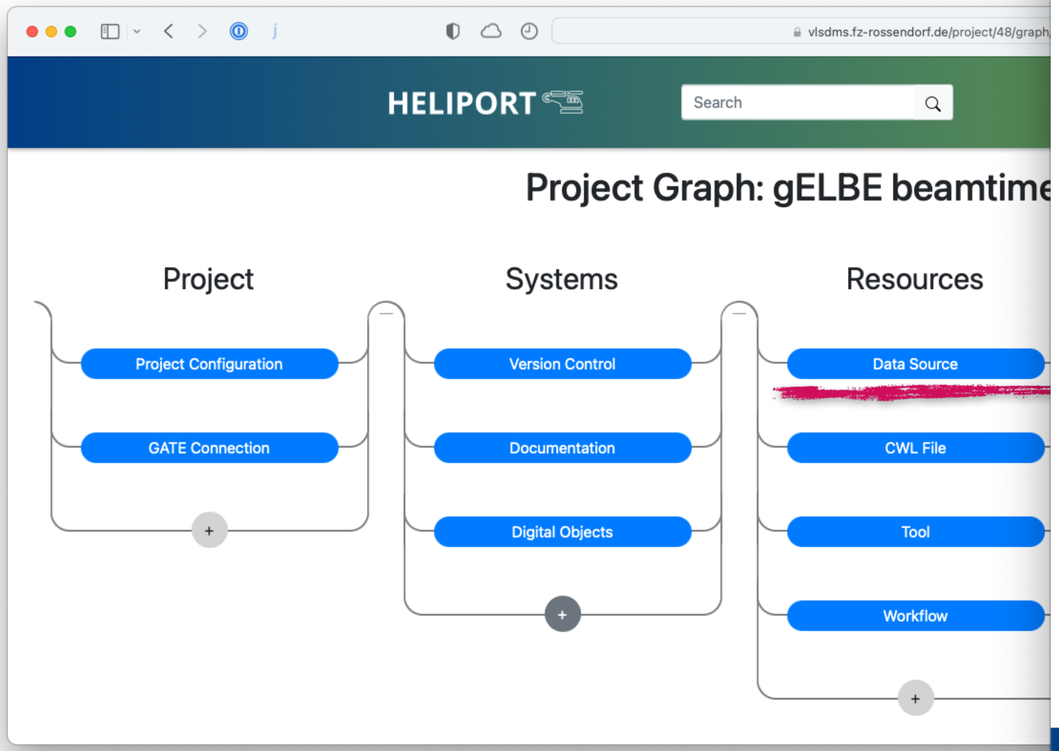
Deinert, Jan-Christophi. (2021, November 18). TELBE data analysis workflow and the PaN training platform UX — Booklet of presentations from the PaN EOSC Symposium 2021. Zenodo. https://doi.org/10.5281/zenodo.5636331

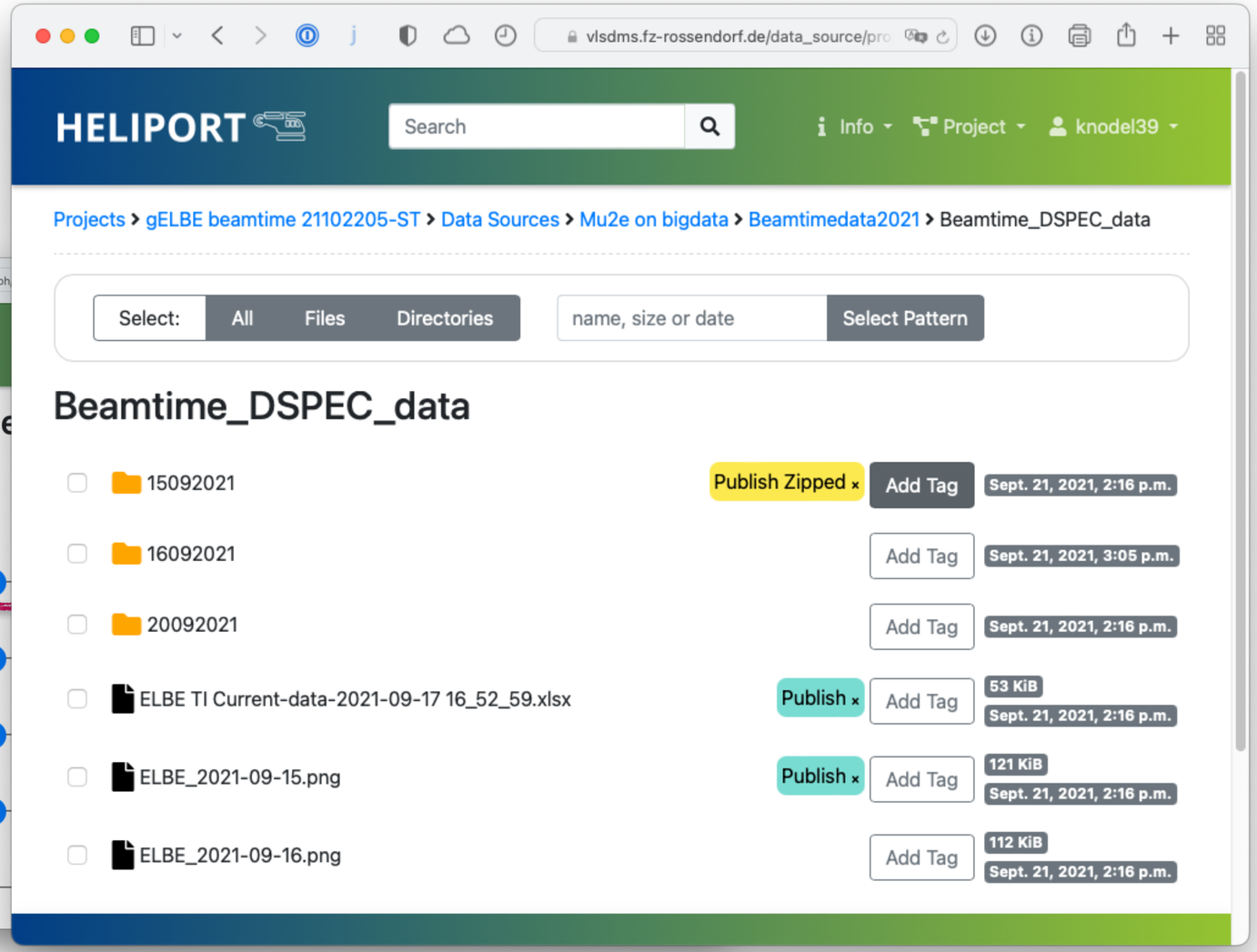
Mapping of the TELBE Resources to HELIPORT



Data Sources can be Registered…

…and files (selected for publication) can be transferred directly to the data publication.



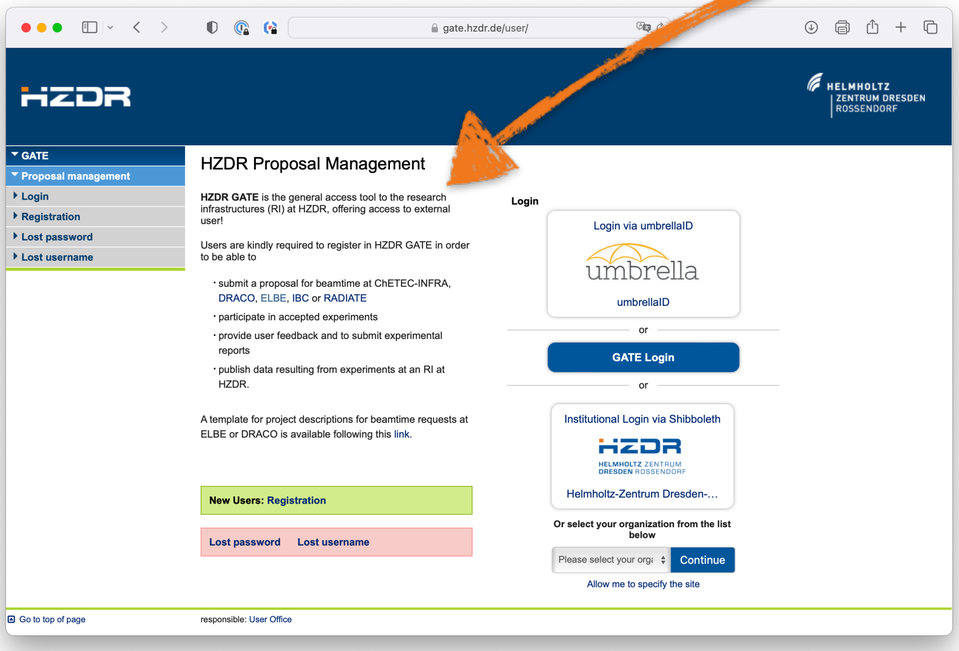


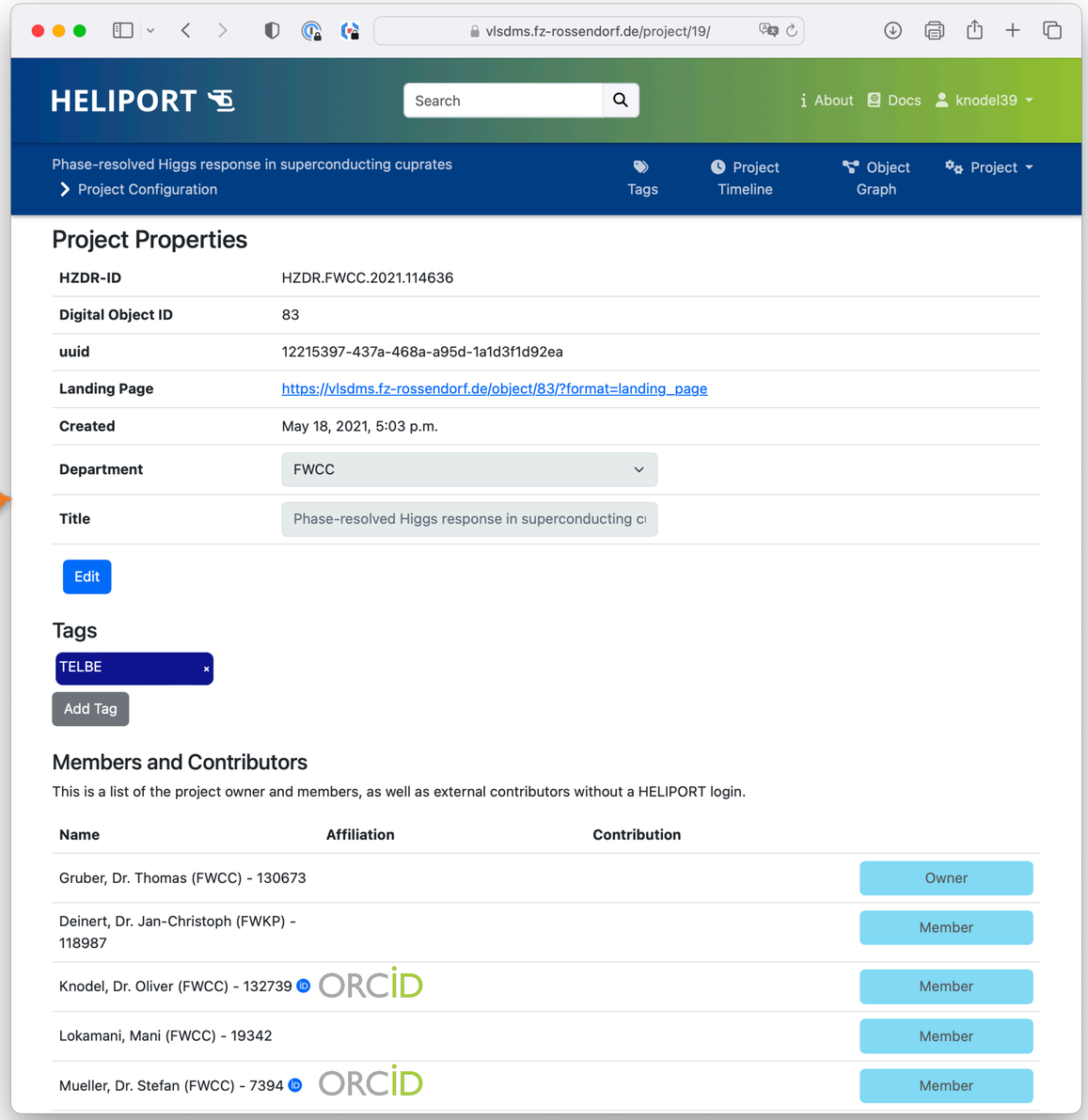
HELIPORT JÜLICH HI JENA HZDR
FORSCHUNGSZENTRUM

I. Proposal Submission

Automated transfer of project metadata from the proposal system (GATE) into HELIPORT:

- Title, Authors, Description,
- Beamtime schedule,
- Large-scale facility used,
- Scientific method (PaNET)

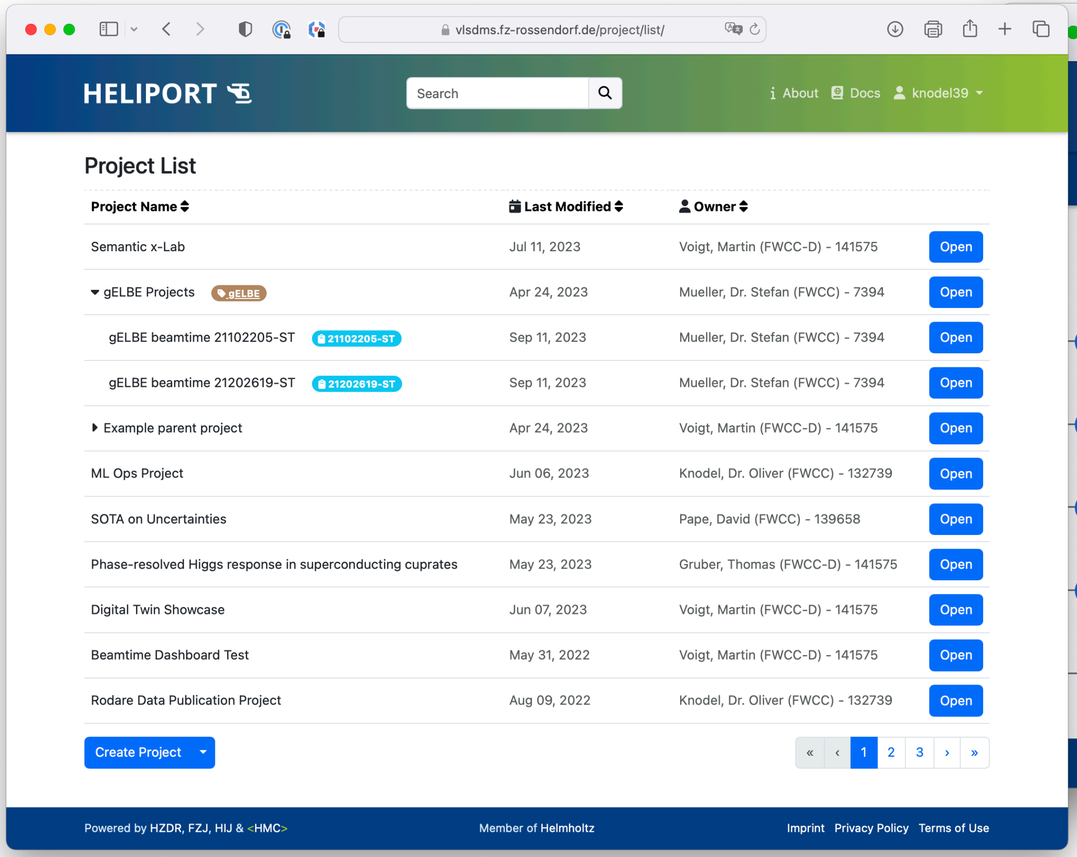


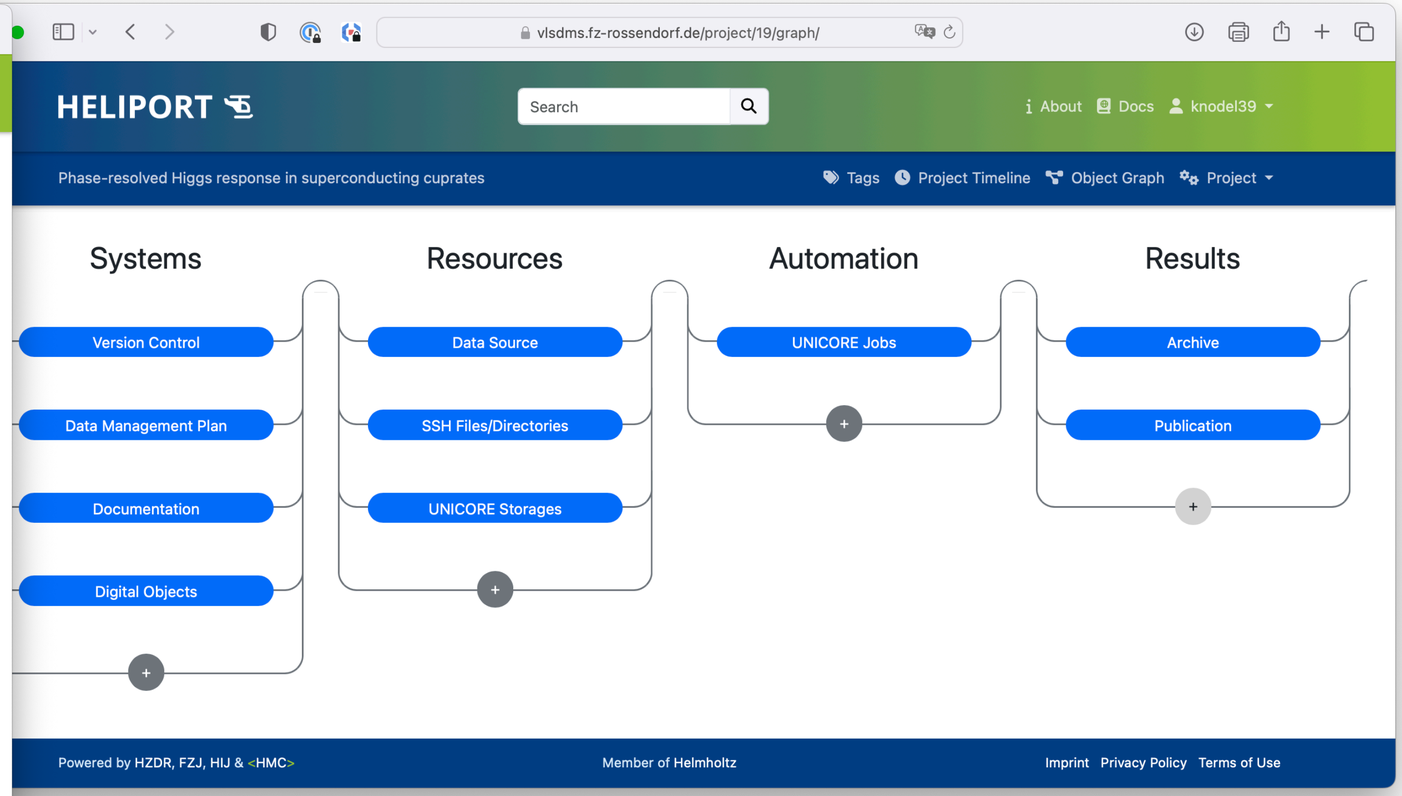


II. Project List and Dashboard

— Typically, a beam line scientist is the owner of a HELIPORT project and the proposer has the role of the manager and can add additional project members.

— Tags and sub-projects including inheritance are possible in the project list.

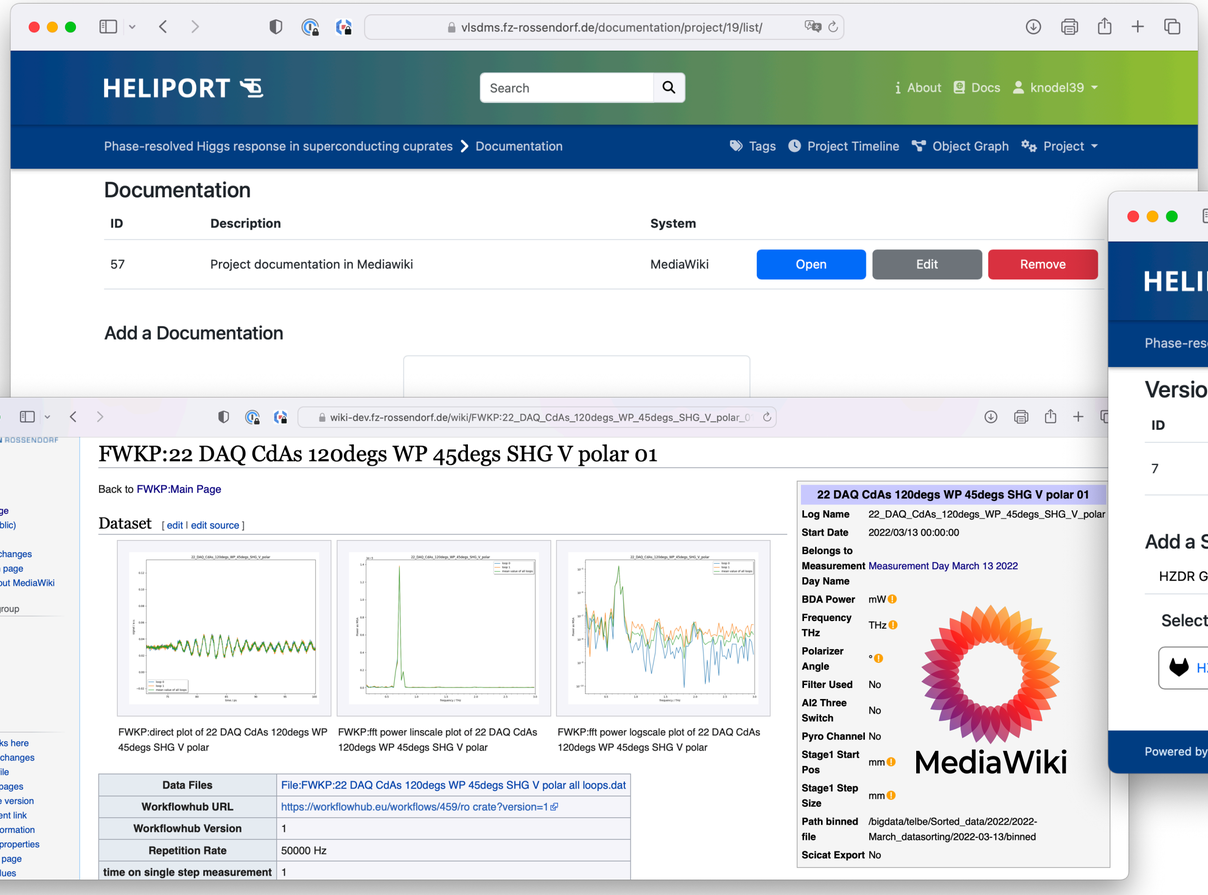


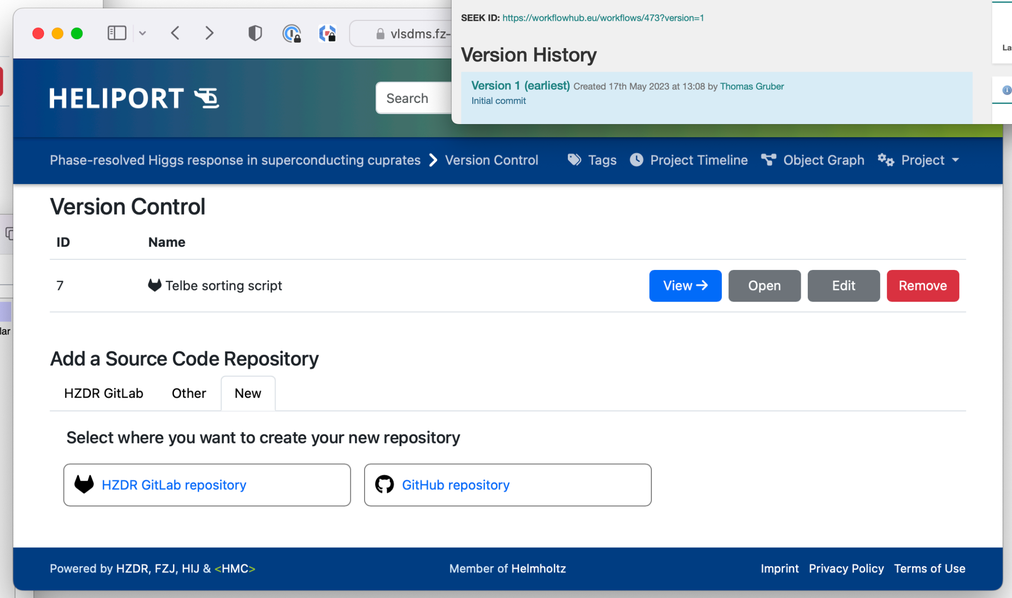


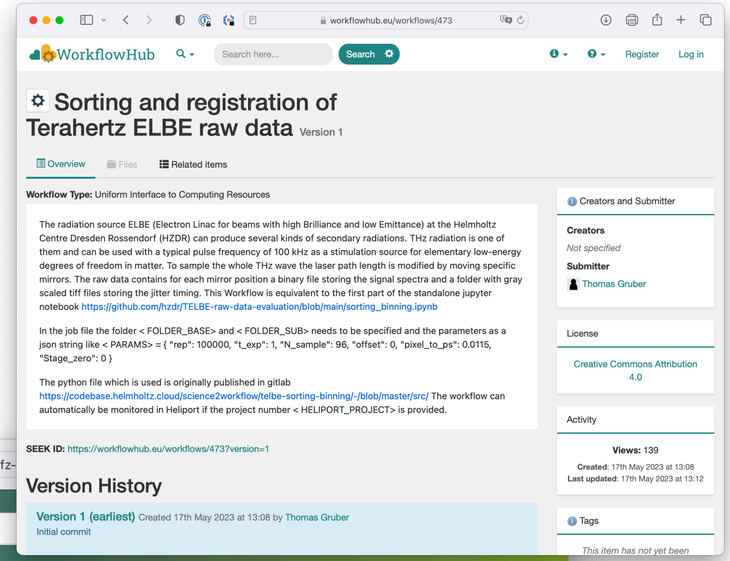
III. Resources: Documentation and Repositories

The documentation section is typically used to refer to all internal and external systems or services used:

- E-Labbook (Mediawiki),
- GitLab, Github, Workflowhub, …

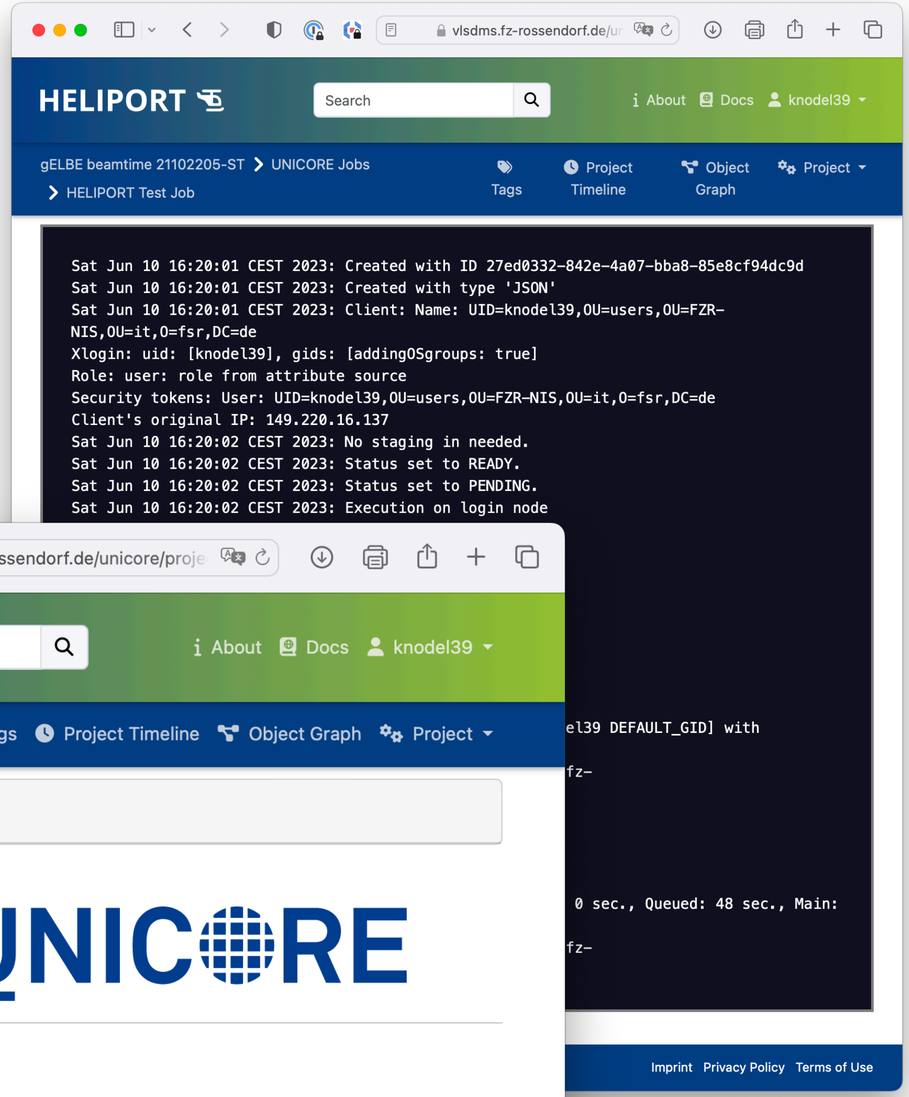


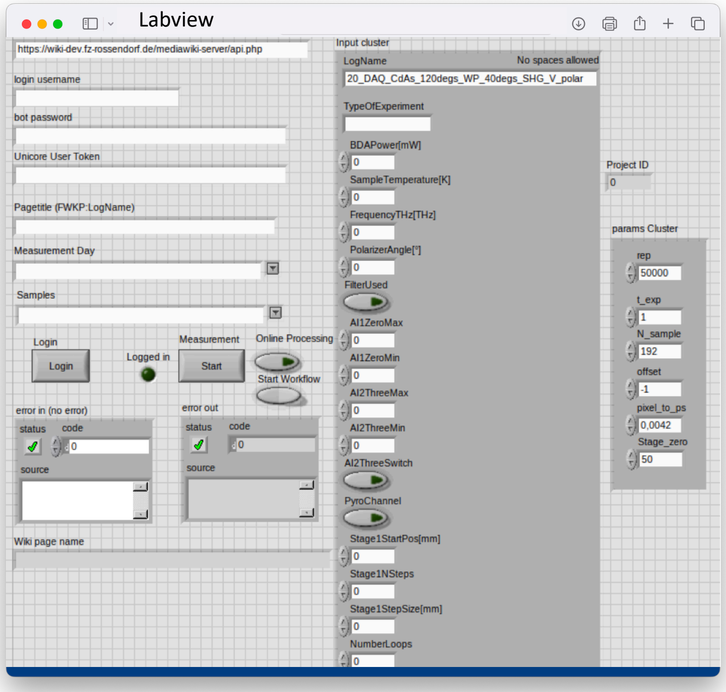


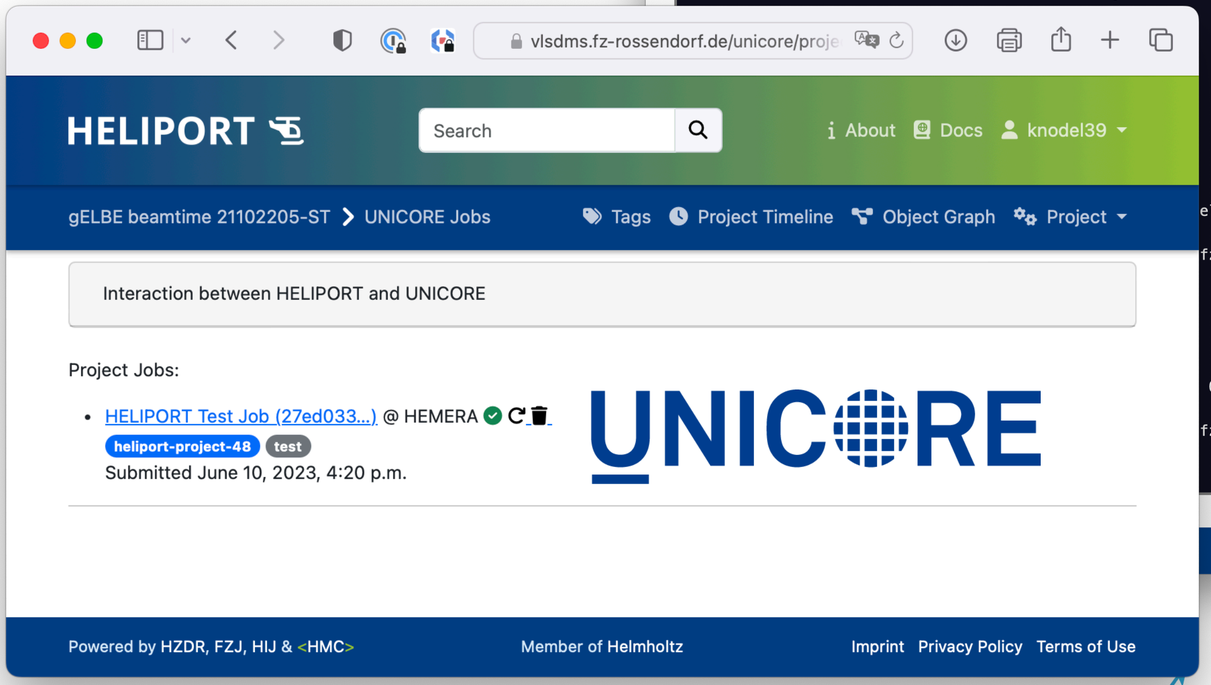


IV. Detector Control and Workflows

- The HELIPORT REST-API enables the transfer of metadata between HELIPORT and external systems (e.g. detector control in LabView).
- The integrated workflow management system (e.g. UNICORE) provides metadata for the provenance information required by HELIPORT.
- Workflows (on our HPC cluster) can be accessed by any project member directly in the HELIPORT web frontend.





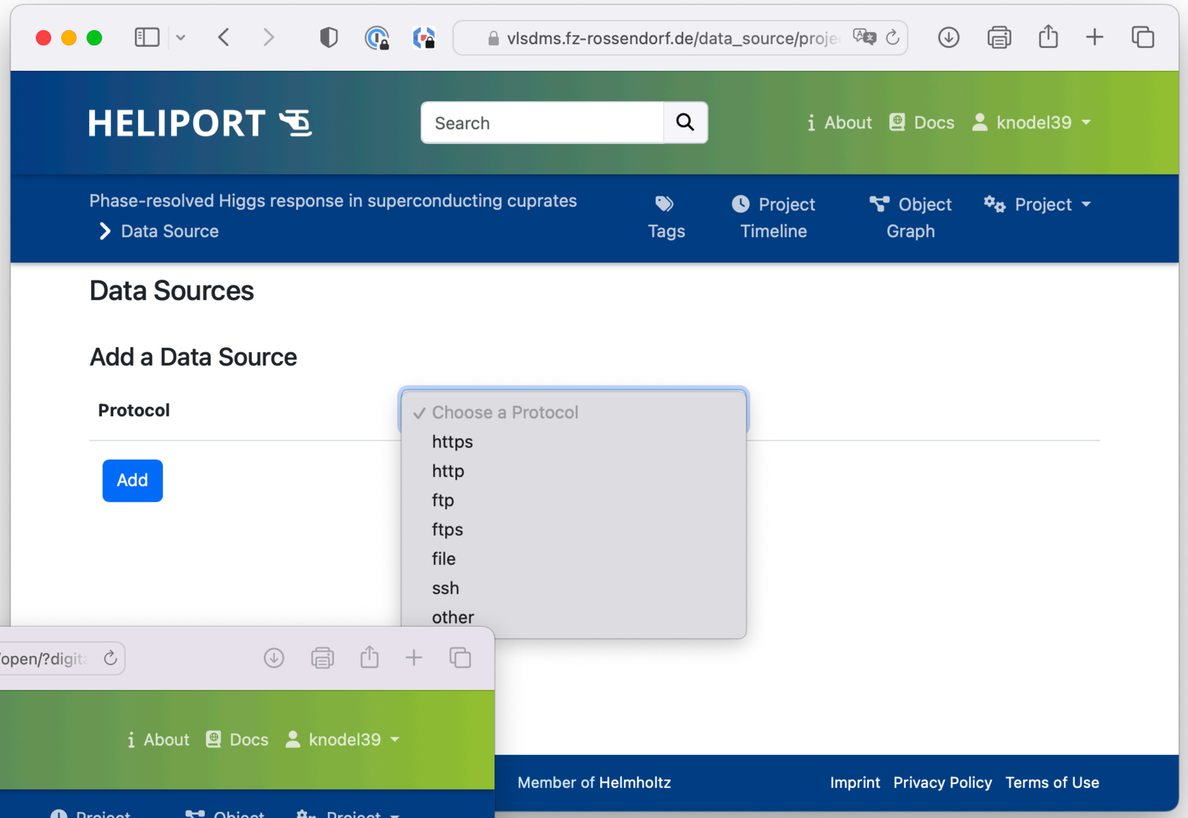


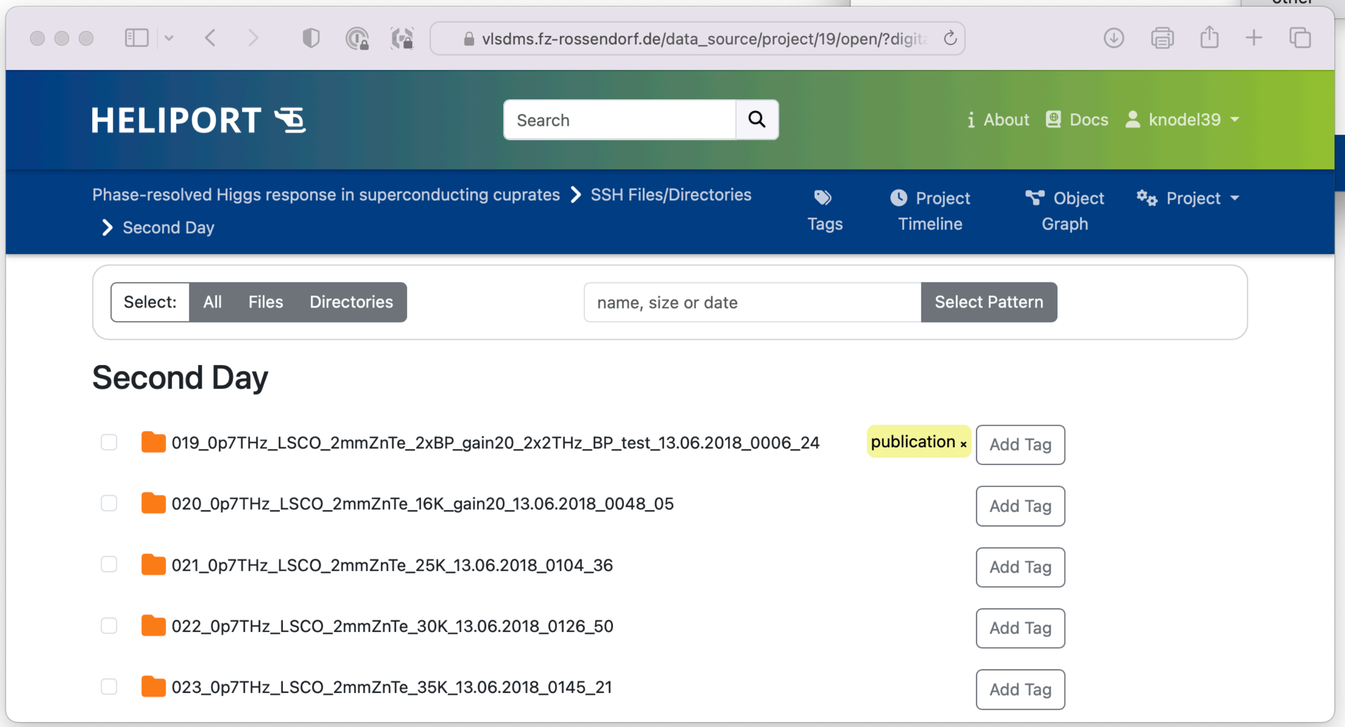
V. Data Sources

— Folders and Files in our internal filesystem can be registered in HELIPORT as data source.

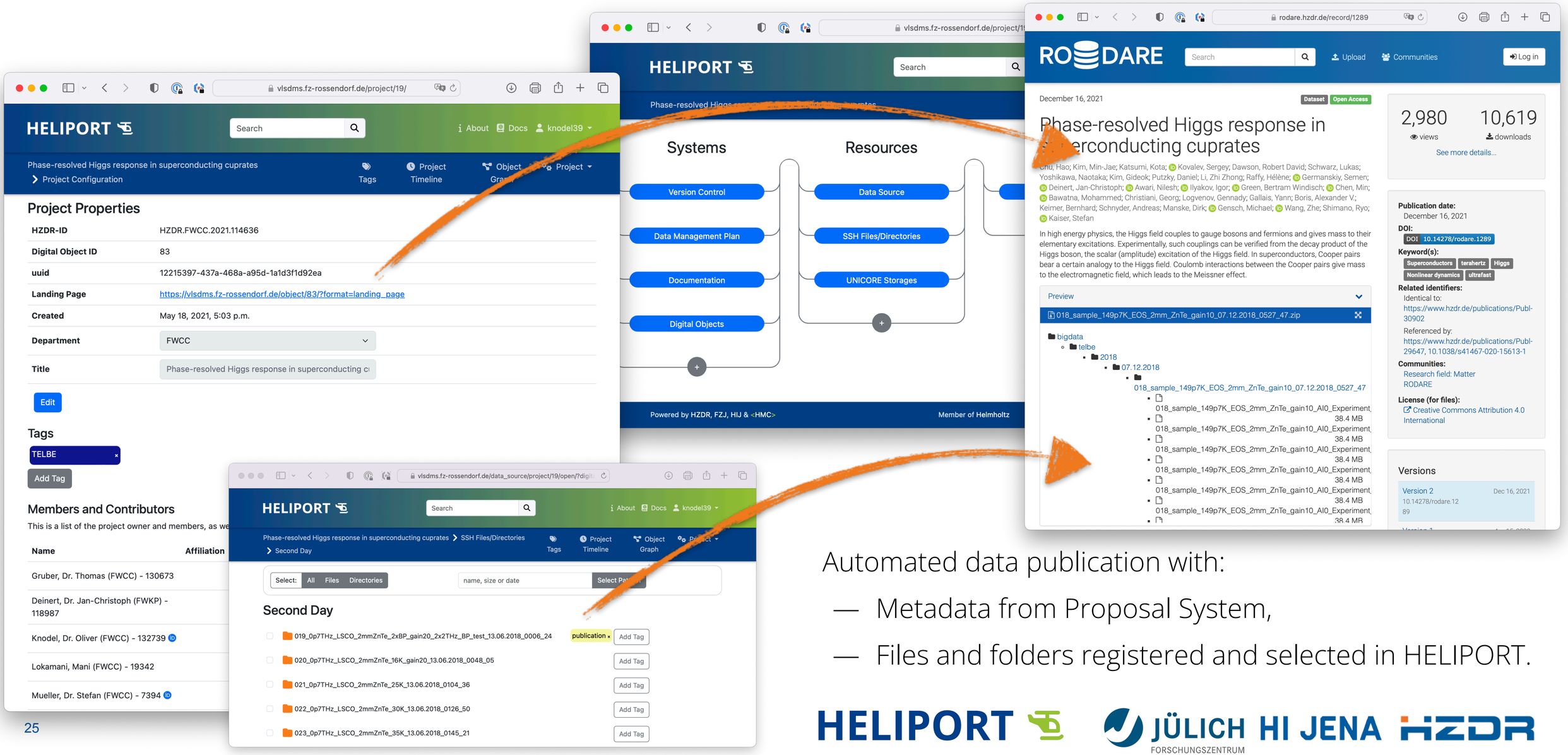
— Each member of a HELIPORT project has access to the files and folders.

— The provenance of the data sets generated from an experiment is entirely comprehensible.





VI. Integration in an Overall Data Publication Workflow



Automated data publication with:

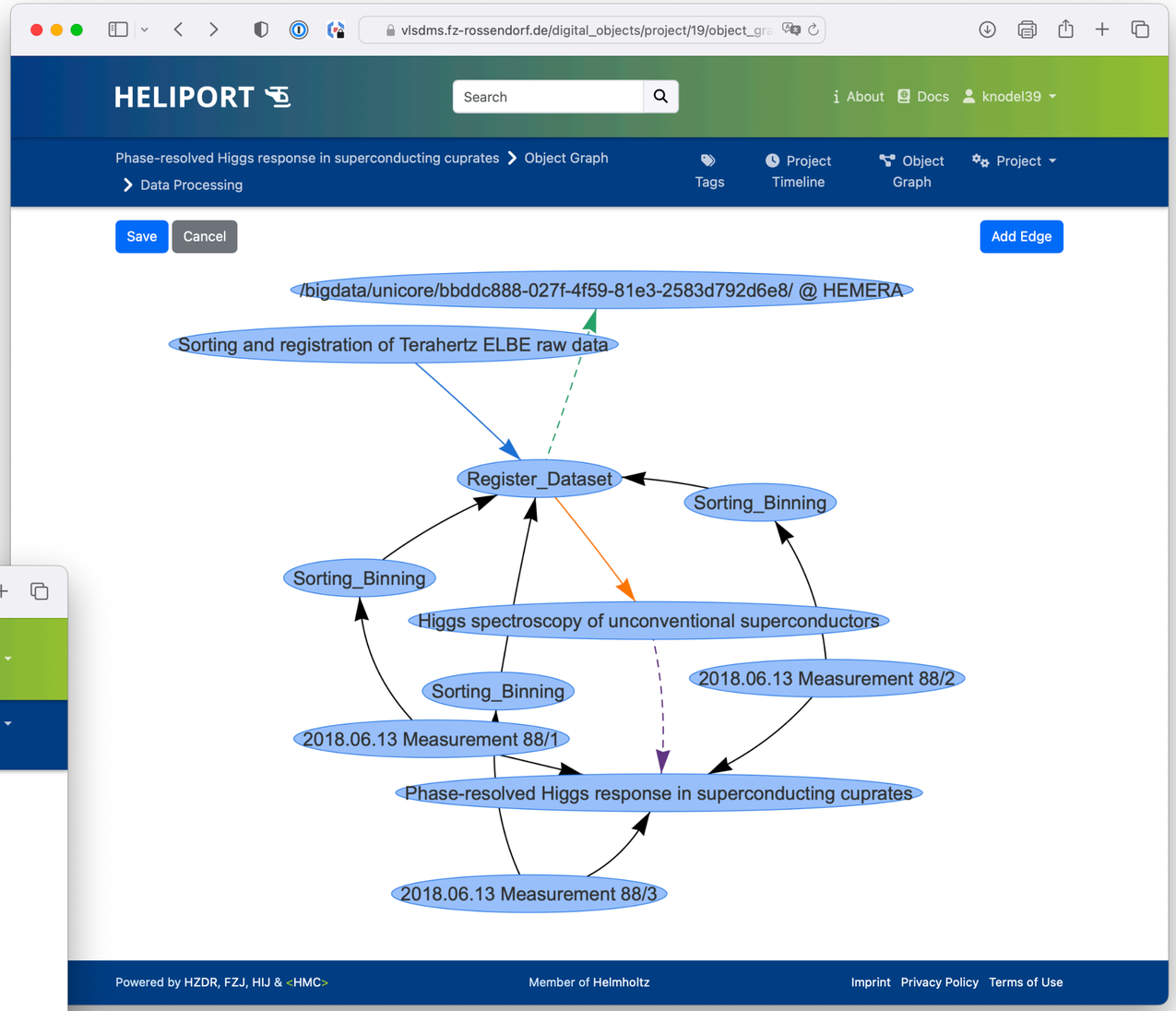
- Metadata from Proposal System,
- Files and folders registered and selected in HELIPORT.

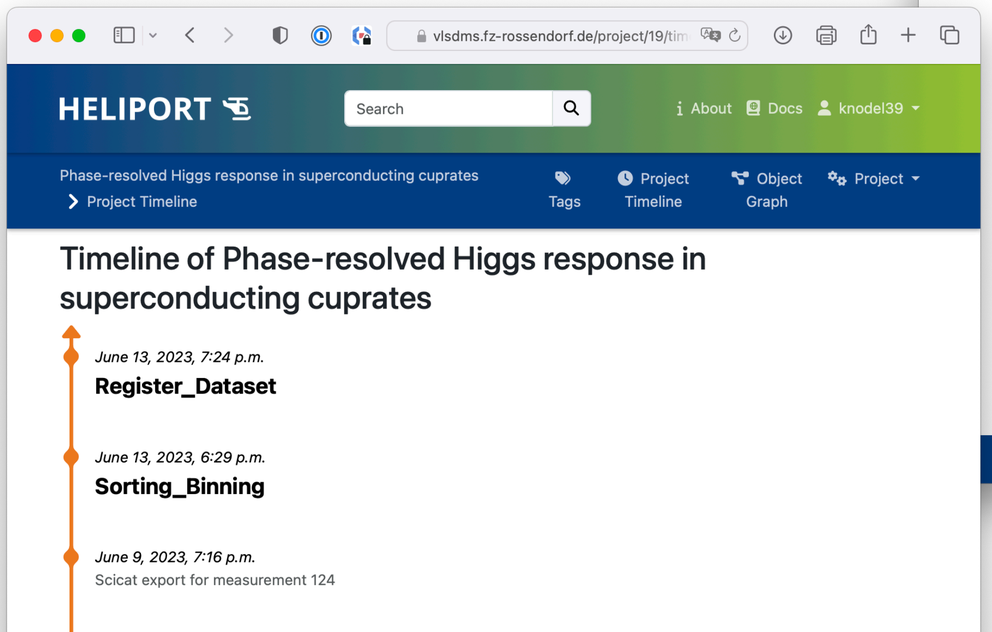
HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

VII. Relations Between Digital Objects and

- Relations between digital objects are visualized to provide a top-level view on the project with dependencies.
- The relationships between simulation (surrogate model) and experiment can also be demonstrated.
- The versioning of an experiment is an essential extension, and first approaches via a timeline are being evaluated.



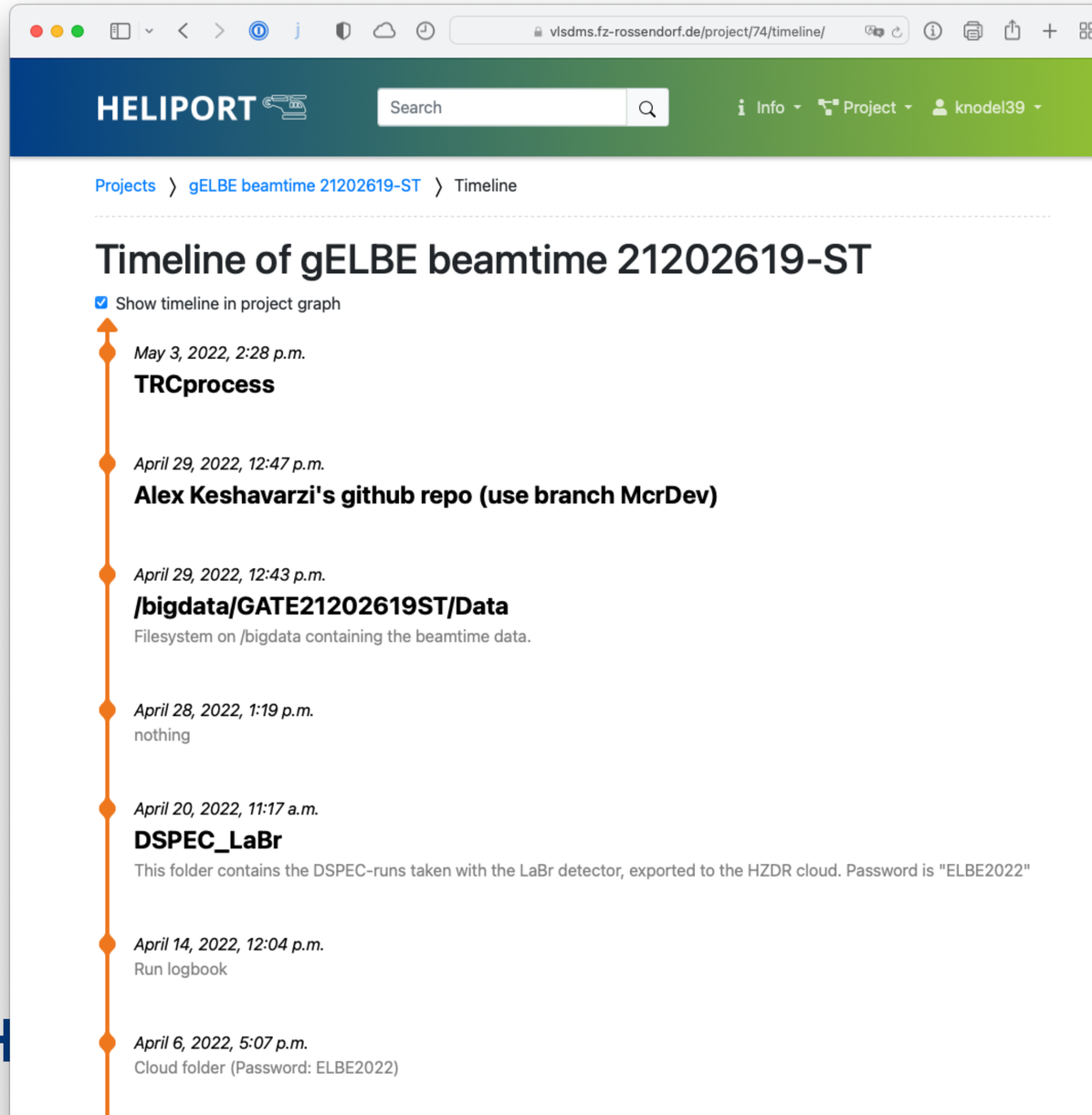


HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

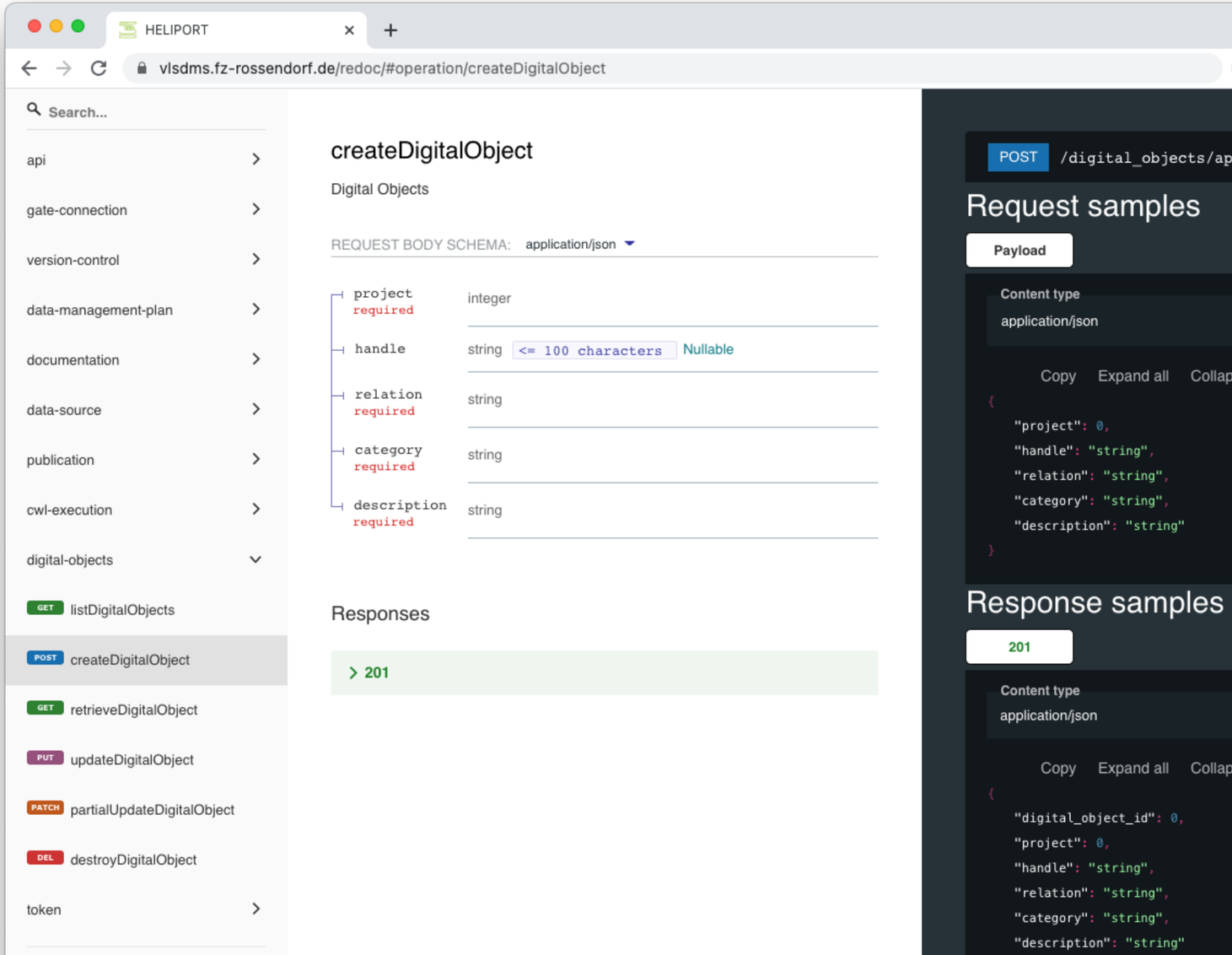
Data provenance and Comprehensibility

- For many systems and services we still have to develop necessary plug-ins for the integration into Heliport.
- The versioning of an experiment lifecycle is unavoidable and we are still discussing how we can present the feature in our web frontend:
  - A Git project with all metadata to restore a lifecycle,
  - Or an implementation direct in Heliport?
- Inheritance of projects,
- Different views based on roles (owner, beam line scientist, data curator, …)



Heliport REST API

- The API provides access to our full Heliport infrastructure:
  - Proposal access (GATE),
  - Handle management,
  - CWL execution and monitoring,
  - Project metadata export,
  - Digital Object and
  - Lifecycle management.
- API documentation (ReDOC) available.
- Essential to integrate the Heliport Infrastructure in Experiments.
- Everything can be documented with less user interaction.



6.12 Building Scalable Time Series Data Infrastructures: State-of-the-Art Solutions for Accessible and Interoperable Environmental Data

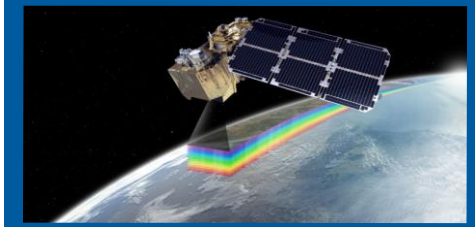
UFZ HELMHOLTZ Centre for Environmental Research

RDM Research Data Management

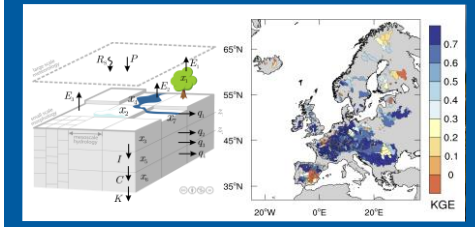
DataHub Research Field Earth & Environment

TERENO TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

eLTER

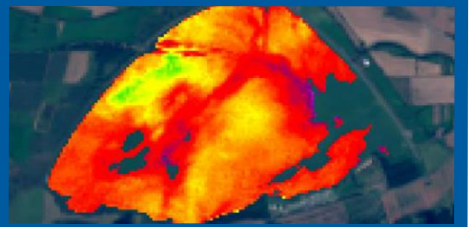










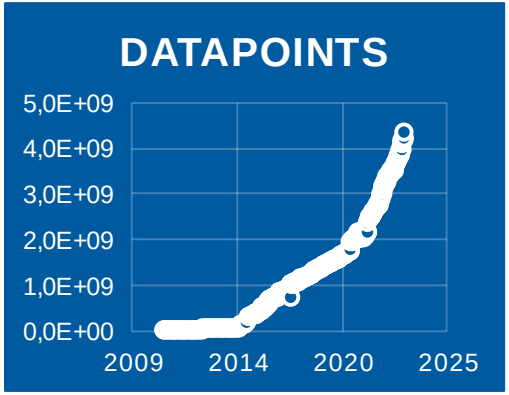


Building Scalable Time Series Data Infrastructures: State-of-the-Art Solutions for Accessible and Interoperable Environmental Data

D. Schäfer, M. Abbrent, F. Gransee, J. Hemmen, T. Kuhnert, L. Nendel, B. Palm, M. Schaldach, C. Schulz, M. Schrön, T. Schnicke, and J. Bumberger

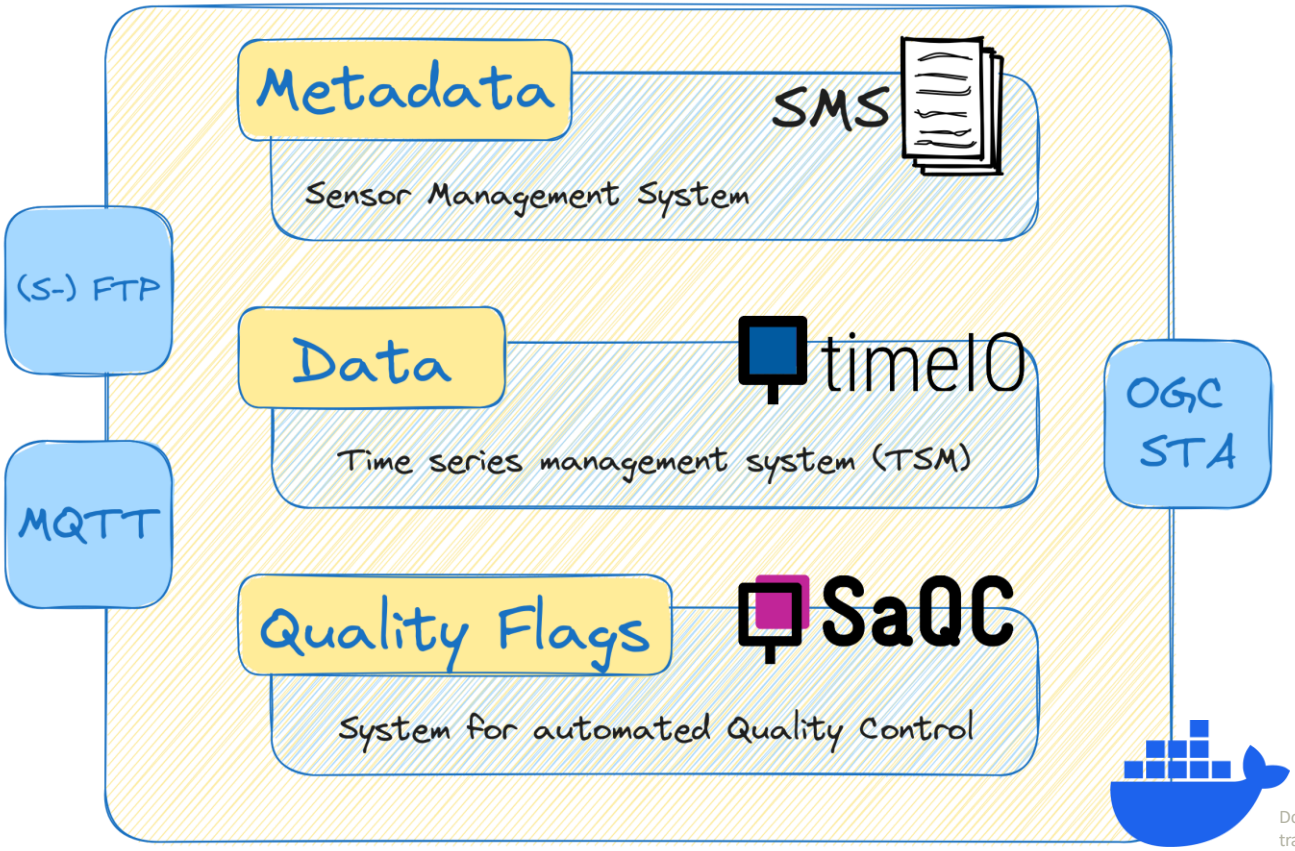
Environmental Earth Science needs FAIR Sensor Data

- Assessment of environmental conditions and changes
- Building and validating stable models
- Data science approaches
- Stakeholder interaction



Digital Ecosystem for FAIR Time Series Data

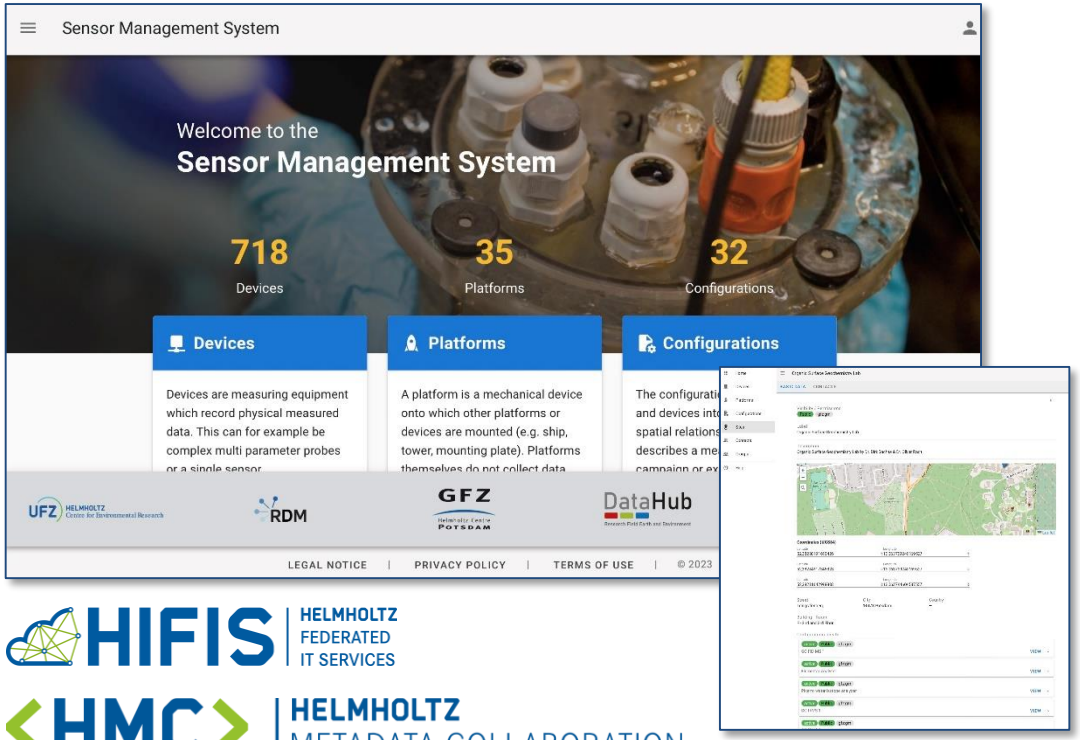
From databases to data management



- Universal
- Extendable
- Scalable
- Accessible
- Transferable
- Equipped
- Open Source

Sensor Management System (SMS)

Findable and accessible data sources



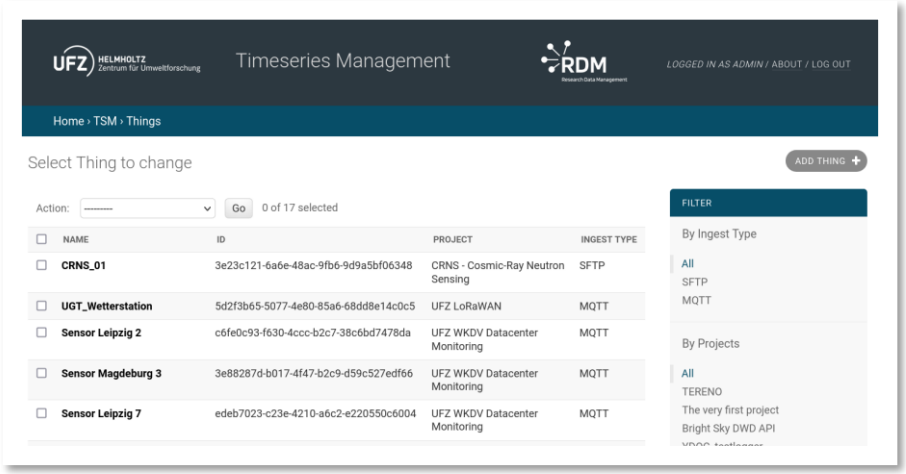
- Register and manage sensors and measuring parameters
- Use a community driven CV
- Plan and manage complex measuring setups and changes over time
- Provision metadata, via standard interfaces following international standards (OGC SensorML)

https://codebase.helmholtz.cloud/hub-terra/sms

time.IO

Integrated timeseries data management

- Distribute/link data to the other components
- Transfer and store timeseries data
- Visualize data streams using Grafana
- Provision data and metadata via STA
- Integrate seamlessly into existing IT infrastructures.



https://codebase.helmholtz.cloud/ufz-tsm/

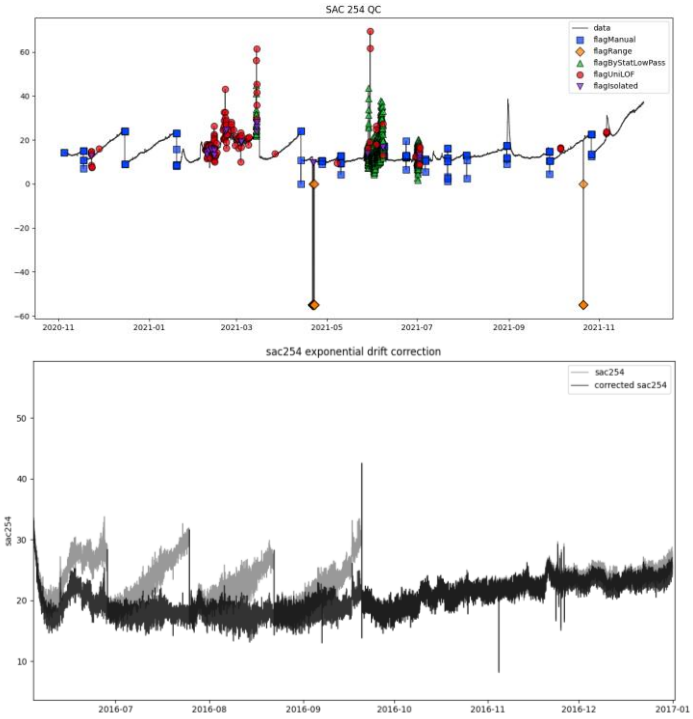
Schäfer, David, Abbrent, Martin, Gransee, Florian, Hemmen, Joost, Nendel, Luca, Palm, Bert, Schaldach, Maximilian, Schulz, Christian, Schnicke, Thomas, & Bumberger, Jan. (2023). timeIO - A fully integrated and comprehensive timeseries management system (0.1). Zenodo. **https://doi.org/10.5281/zenodo.8354840**

Open Geospatial Consortium. MQTT Grafana SaQC docker Compose

System for automated Quality Control

SaQC - Quality control of timeseries data

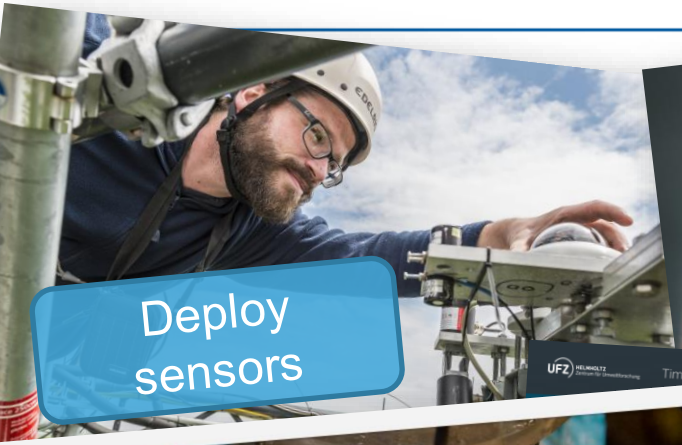
- Analyze, annotate and process data
- Enrich metadata from end to end
- Use predefined or custom quality annotation schemes
- Interact with one of several user interfaces, a Python API, text based configuration or a web application
- Get from PyPI, conda-forge and Gitlab



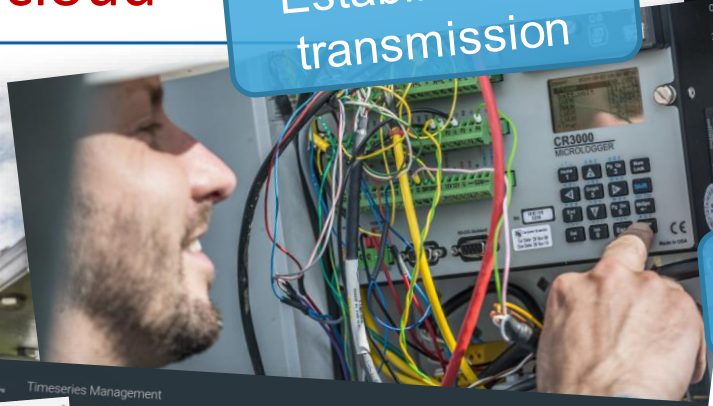
Lennart Schmidt, David Schäfer, Juliane Geller, Peter Lünenschloss, Bert Palm, Karsten Rinke, Corinna Rebmann, Michael Rode, Jan Bumberger, System for automated Quality Control (SaQC) to enable traceable and reproducible data streams in environmental science, Environmental Modelling & Software, 2023, 105809, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2023.105809.

Customer/Data Journey

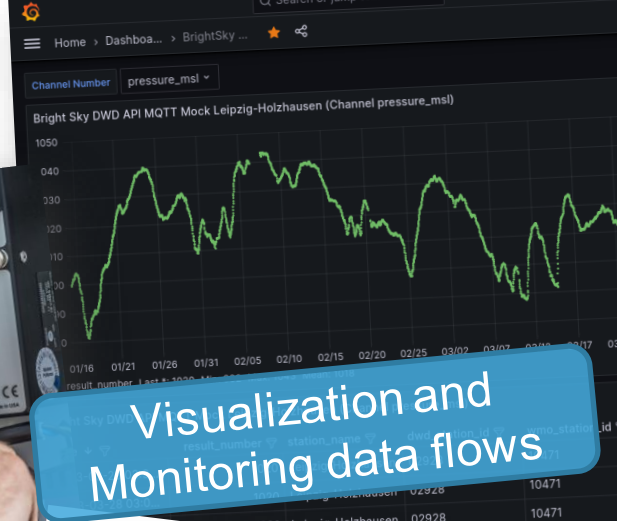
From the field to the cloud



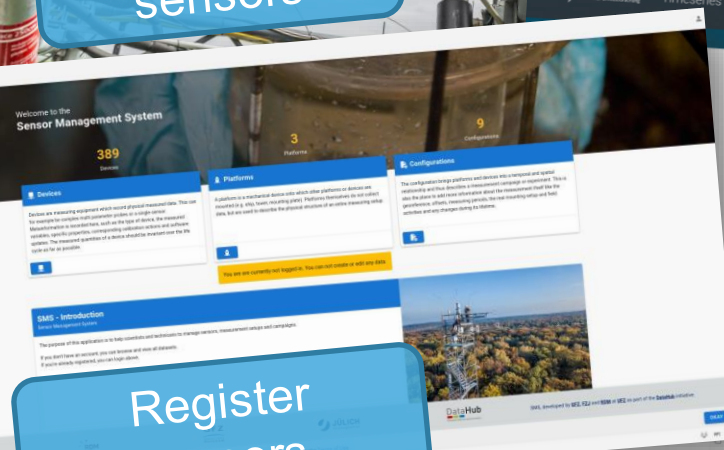
Deploy sensors



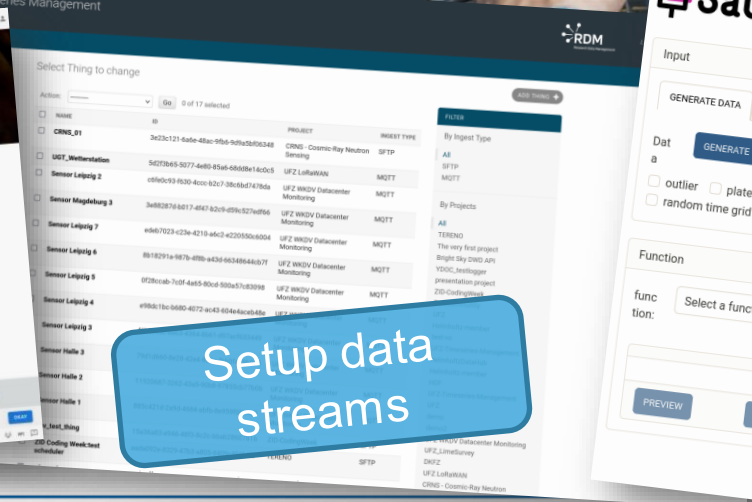
Establish data transmission



Visualization and Monitoring data flows



Register sensors

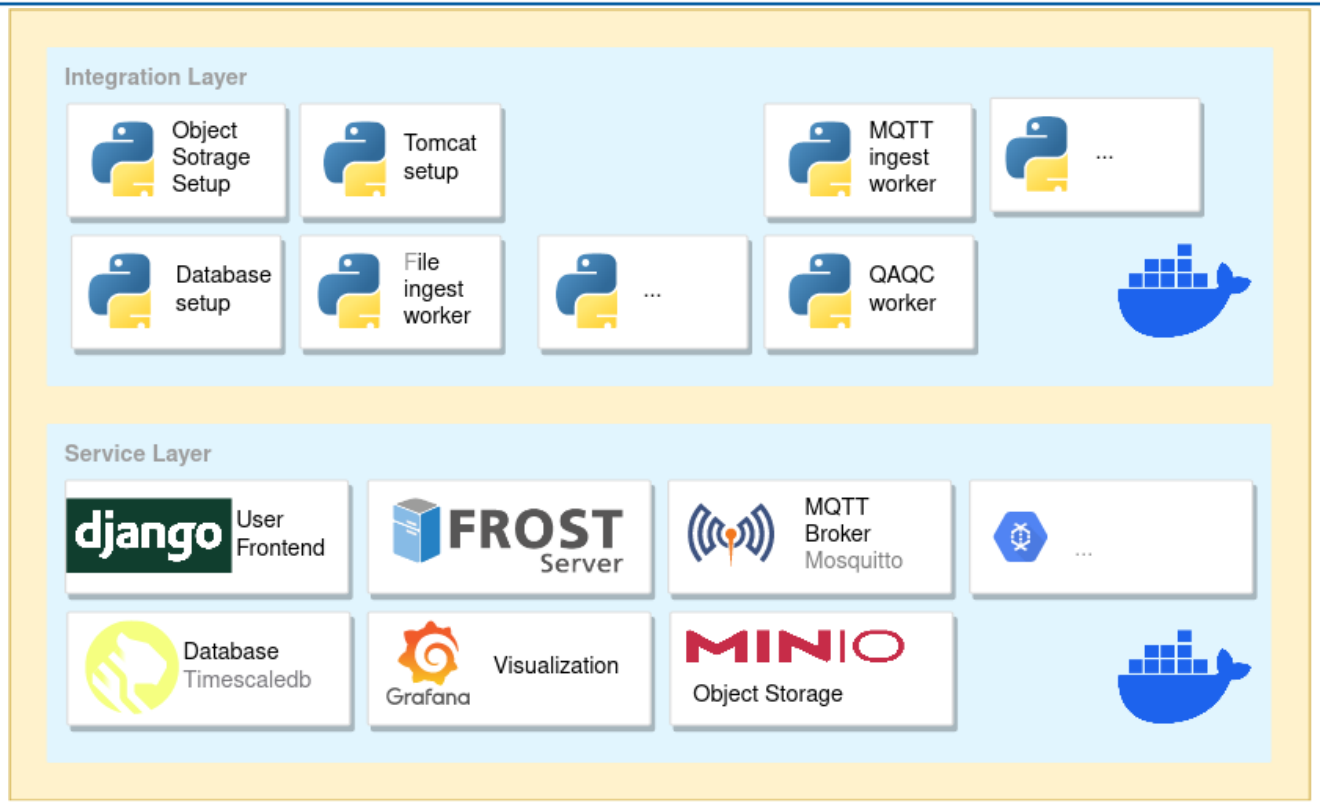


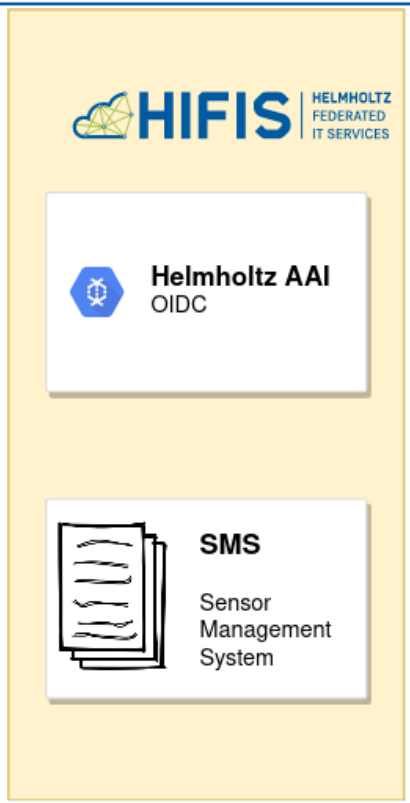
Setup data streams

Configure quality control

Pluggable Container-/Microservice Architecture

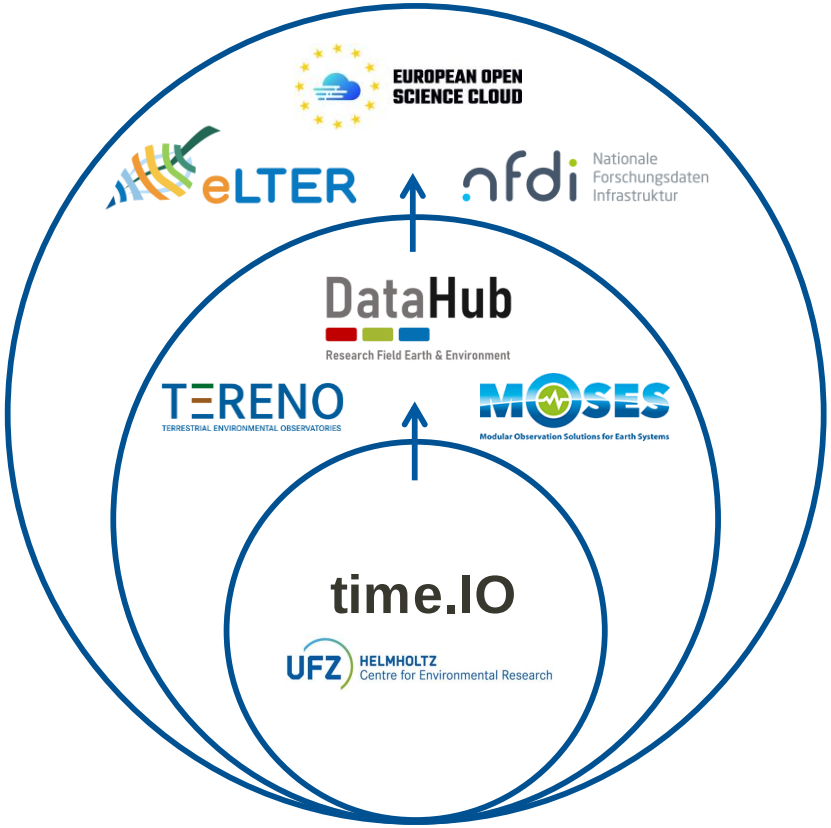
"docker compose up" approach





Integrated Timeseries Management System – time.IO

Embedding into federated higher level data infrastructures



UFZ

- First pilots running at the UFZ
- Production use and regular release cycle 2024
- Migration from a legacy system in 2024/25

General

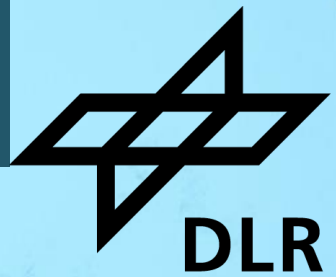
- Embedded into broader developments of the DataHub
- Contributing to efforts to harmonize metadata (HMC, OGC)
- Part of upcoming large scale infrastructures

6.13 Overview about the EU Cyber Resilience Act

OVERVIEW ABOUT THE EU CYBER RESILIENCE ACT

06.02.2024, Helmholtz Open Science Forum: Research Software, UFZ, Leipzig

Tobias Schlauch <Tobias.Schlauch@DLR.de>
Institute for Software Technology
German Aerospace Center (DLR)
http://www.dlr.de/sc



What is the Cyber Resilience Act all about?

- **What is the Cyber Resilience Act?**
  - Cyber Resilience Act is an upcoming EU legislation which establishes general cybersecurity rules for placing software and hardware products on the EU market.
  - It is part of different current EU legislations (e.g., Product Liability Act, AI Act) to ensure safety and security of products placed on the EU market.
  - In Future: CE mark on products = safe & secure

- **Important revisions:**
  - 15.09.2022: Initial proposal
  - 13.07.2023: EU Council position
  - 26.07.2023: EU Parliament position
  - 20.12.2023: Result of the trialogue negotiations

What products are in scope or out of scope of the CRA?

- **In scope:**
  - Hardware product (e.g., laptops, mobile phones, smart appliances, CPUs, …)
  - Software products (e.g., operating systems, word processing, games, mobile apps, software libraries, …)
  - Remote data processing components required by these products

- **Out of scope:**
  - Non-commercial products (e.g., hobby products)
  - Services, in particular standalone SaaS which covered by NIS2 (e.g., websites, purely web-based offerings, …)
  - Products which are covered by other regulations (e.g., cars, medical devices, certified aeronautical equipment, …)

Source: Presentation of Benjamin Bögel during FOSDEM 2024

Main elements of the law

- **General cybersecurity rules** for placing software and hardware products on the EU market
- **Obligations for manufacturers, distributors, and importers**
- Definition of **essential requirements across the whole lifecycle**
- **Harmonized cybersecurity standards** (to be established)
- Required **conformity assessments** (depending on risk)
- **Reporting** obligations
- **Market surveillance and enforcement**

Source / based on: Presentation of Benjamin Bögel during FOSDEM 2024

What was the problem for the Open Source community with the initial CRA draft?

"(10) In order not to hamper innovation or research, free and open-source software developed or supplied outside the course of a commercial activity should not be covered by this Regulation. …"

Source: Initial proposal of the CRA

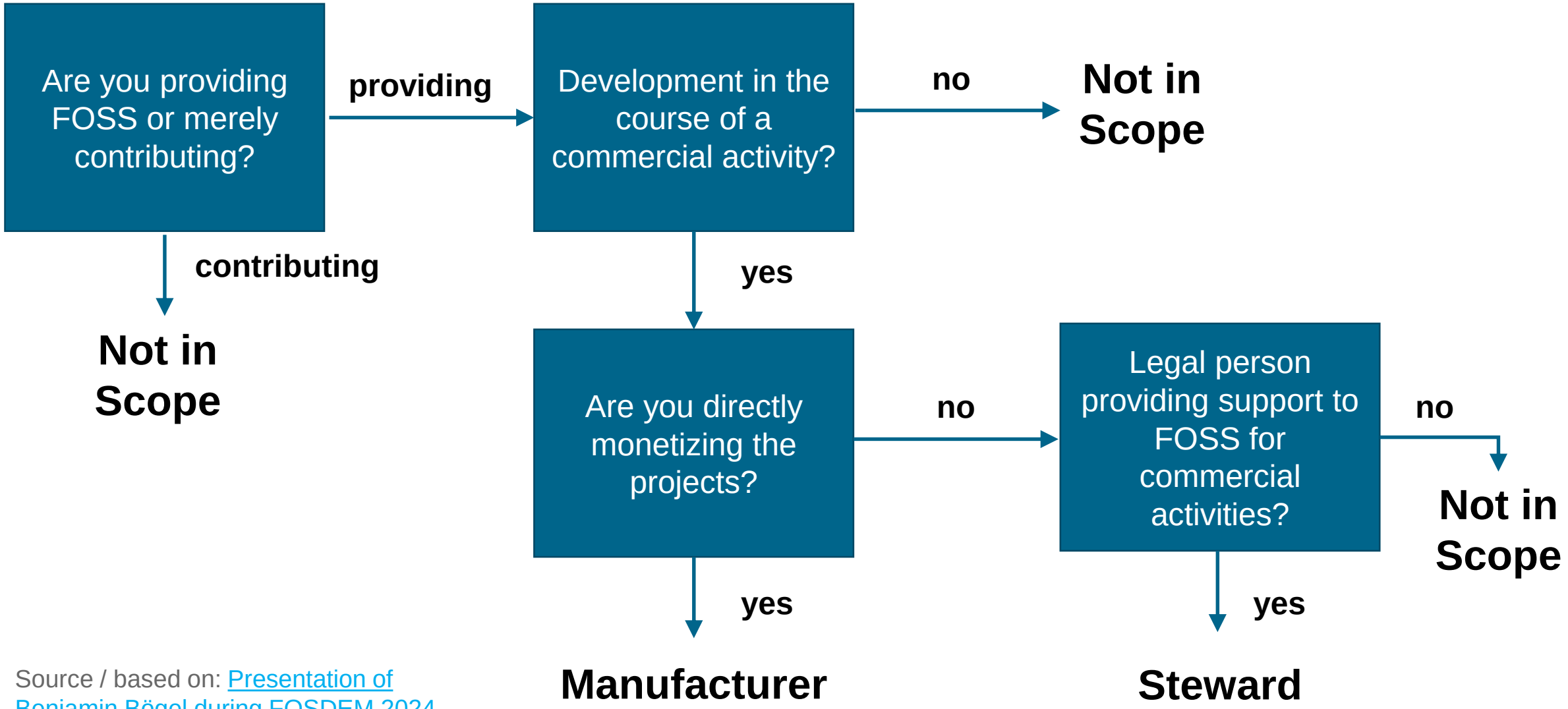
- Defined exception left too many uncertainties: What about donations, fundraising, selling merchandise? What about committers from a commercial company?
- Platforms such as GitHub or PyPI could be considered distributors
- For more information:
  - The ultimate list of reactions to the Cyber Resilience Act - Voices of Open Source
  - Update on the European Cyber Resilience Act (youtube.com)

- **Incalculable risks could have led to unavailability of important open-source projects in the EU!**

What did the Open Source community gain in the end?

- Much more differentiated consideration of the open-source specific development model!
  - Responsibility concerning the CRA is on the legal entity that monetizes the software / places the “final” product on the market.
  - Funding of the development is out of scope when considering the "commercial nature" of an open source project.
  - Exceptions for platforms such as GitHub and PyPi have been added. I.e., they are no longer considered as "distributors" in accordance to the CRA.
  - Introduction of the new role “open-source software steward” on which relaxed rules are imposed. This role has been primary introduced
  - And much more ... (at least 6 dedicated paragraphs + further additions on 4+ pages)
- **Open-source software is more or less out of scope of the strict CRA rules!**

Is your open-source project covered by the CRA?



Source / based on: [Presentation of Benjamin Bögel during FOSDEM 2024]

What is the expectable timeline of the next steps?

- Final text is currently prepared
- March/April 2024: Final vote in the EU Parliament
- Around mid 2024: CRA is officially published
- Around mid 2027: CRA is in full force

- After its official publication there are 36 months for:
  - Harmonization of the 40+ cybersecurity standards
  - Preparation for the application of the CRA rules

- **Standardization is an important undertaking – offers opportunities for finetuning but also new risks => engagement is important!**

What does this mean for Research Software Development?

- I am not sure but I think that in research we will have a mix of software which falls under the CRA and software that does not
- I think that two aspects are important:
  - Research needs to analyze the impact of the CRA (in connection with the revised Product Liability Directive) for typical scenarios in which software is made available (as open-source software or not) and is / might be monetized
  - Research should actively(!) engage in the standardization process to rule out possibly identified uncertainties and problems (as good as it is still possible)

Thank you!

What are your Questions?

Email: Tobias.Schlauch@dlr.de
Mastodon: https://norden.social/@schlauch
HIFIS Mattermost: @schlauch

Copyright and License Information

6.14 Open-Source Software Development with Industry



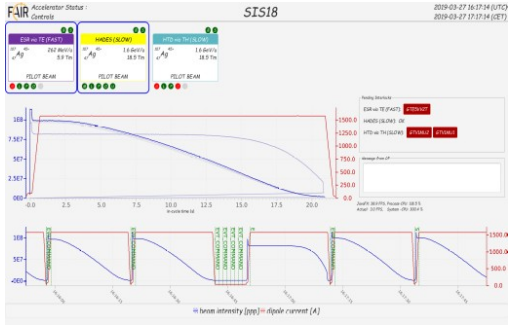
Open-Source Software Development with Industry

Alexander Krimm, GSI Helmholtz Centre for Heavy Ion Research GmbH
HELMHOLTZ Open Science Forum: Research Software, Feb 5 – 6, 2024 Leipzig

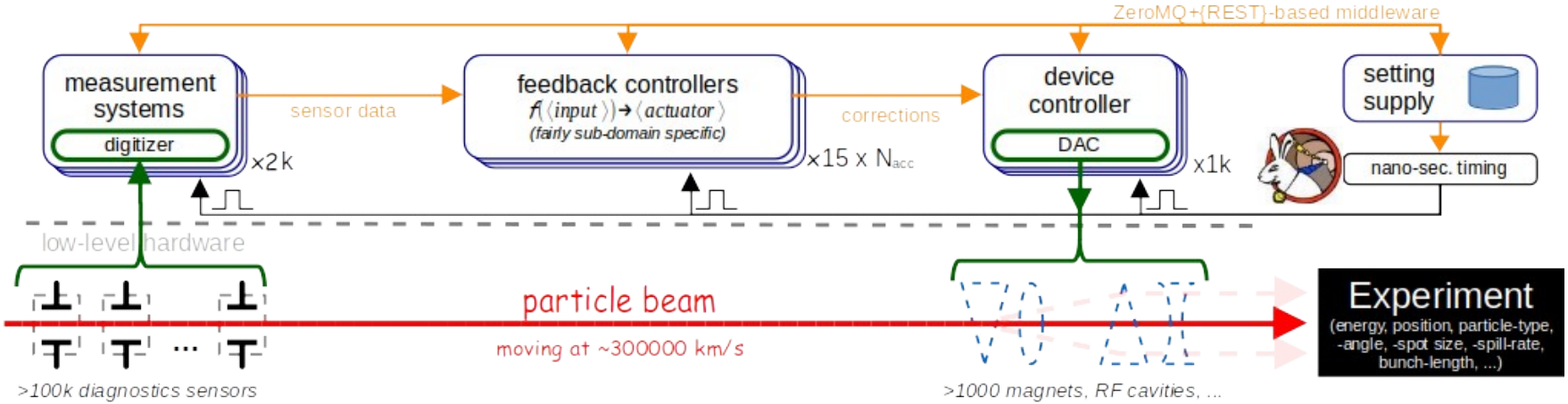
Who are we?

team working on system integration, beam-based diagnostic and feedback systems for the FAIR accelerator facility

- 100k signals from various data-acquisition sources & 100s of systems
- 20+ core services per machine for commissioning, first-line diagnostics and operation + co-use by experiments
- **60+ person-years of scheduled work → only possible with external partners**



=> huge demand for **sustainable** software development



Motivation

Why have external Partners at all?

- **lack of in-house resources:** freeing in-house team to focus on FAIR-accelerator-specific tasks.

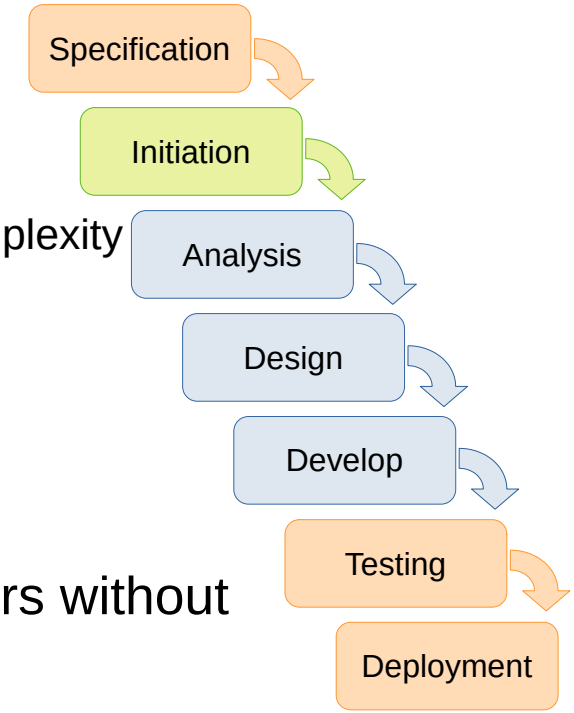
1) Mitigating Risks → ‘agile’ and ‘lean’ Framework Contract
   - general FAIR contract designed for established technologies but unsuitable for SW development
2) Software Quality and Efficiency mirrors Communication Structure (→ M. Conway’s Law)
   - **collaborations must be based on trust + open and transparent communication**
3) Public Documentation of Technologies & Outreach
   - fostering seamless collaboration & attract a diverse talent pool
   - break free from vendor lock-in, circumvent single service-provider problem

Creating Shared Value through creating public ‘clean’ and ‘lean’ Code-Base

- efficient response to 24h/7 operational needs and sustainable long-term maintenance
- FAIR as technology forge: develop technologies critical to FAIR
  & enable industry and general public in new emerging key technologies.

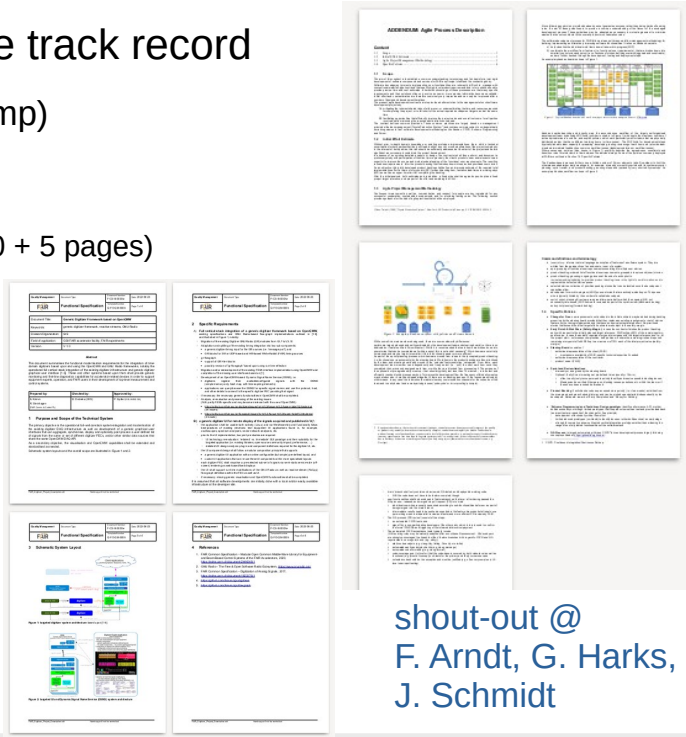
Common Problems

- “old contract”: ‘waterfall’ design process with few pros and mostly cons
  - strong tendency to over- or underspecification for software
    - overspecification time consuming for us and industry partners
    - design effort nearly identical to implementation effort
  - asymmetry between in-house and external understanding of project complexity
    - initial over- or underestimation of the project effort
  - slow feedback loops and expensive ECRs
  - “code-dump”: quality control and future maintainability
- closed in-house solutions: complicate onboarding
- hard to assess qualities and prior work of new industry partners without reliable public track record



New ‘agile’ and ‘lean’ Framework Contract (dt. “Rahmenvertrag”)

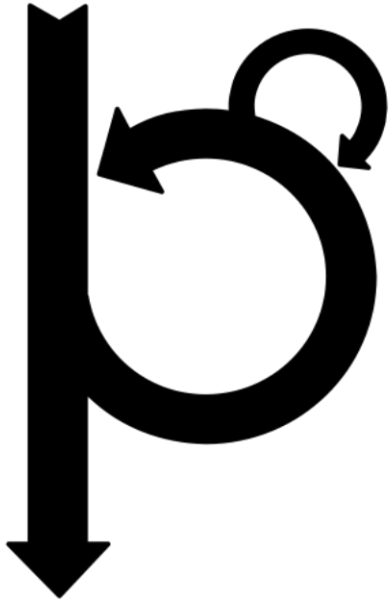
- enables fast and lightweight dispatch of projects to a pre-qualified pool of industry partners.
- pool selection process:
  - demonstrated skills and expertise proven by public open-source track record
    - quality of work & sustainability (actively maintained projects vs code-dump)
    - style of community engagement
- short-form legal contract + agile process description (10 + 5 pages)
  - small, re-usable, and describes the collaborative process
  - live iterations: ‘change for free’ policy
  - incentive for efficiency: ‘early finish’ & ‘risk share’
- project description (2-10 pages, by FAIR)
  - outlines scope of work that should be performed
  - details provided in referenced public resources



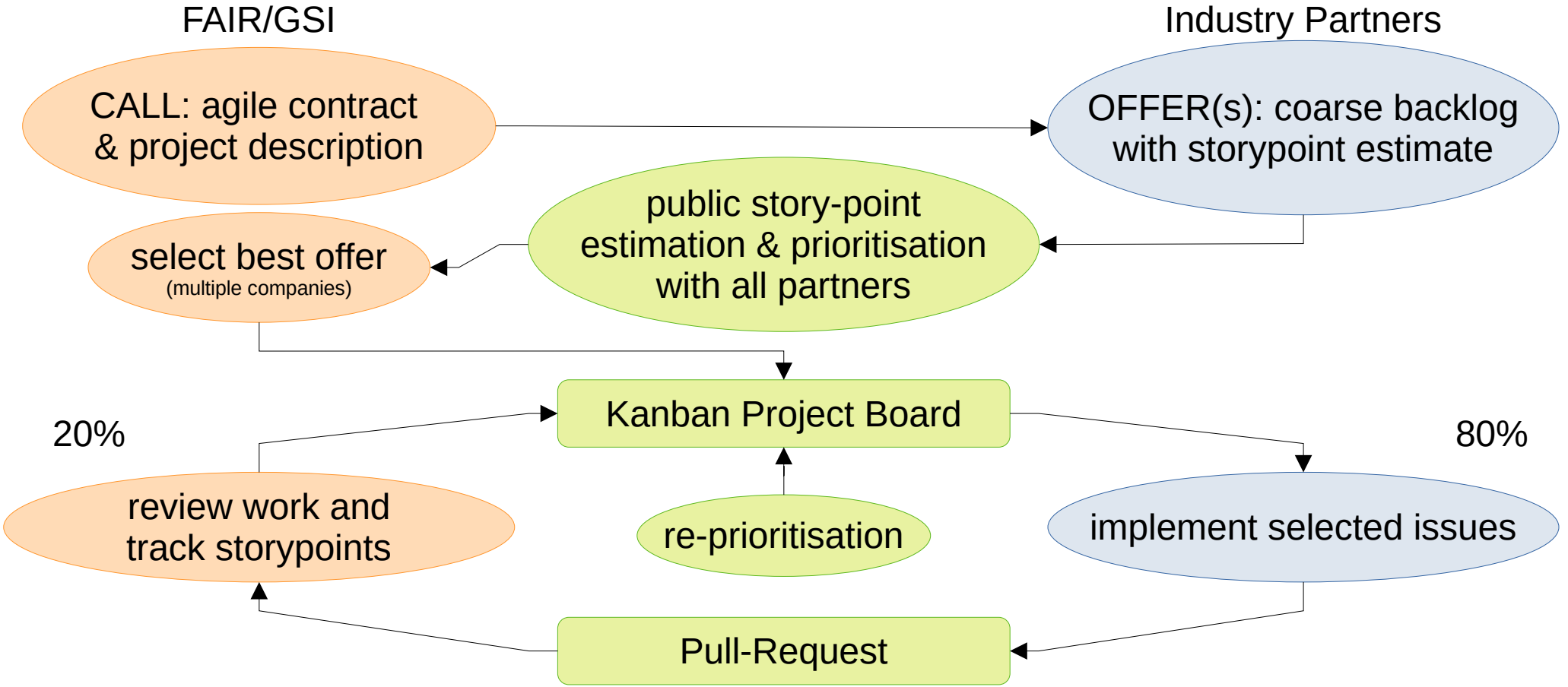
shout-out @
F. Arndt, G. Harks,
J. Schmidt

Co-Development and Continuous Iteration & Communication

- **function over form**
  - focus on design, function and intended use
  - ... rather than overspecified implementation details that are impossible to know at project start
- **prefer building upon or extending existing projects** rather than creating new ones
  - sharing of expertise and ideas
  - spread mainteinance on more shoulders
- **co-development**
  - lightweight specification + iterative process
  - allow re-priorisation and agree on a fair sharing of risks
- **shared code ownership**
  - early integration of individual functionalities
  - review process of small functional units
  - code and development process in the open as much as possible
    - for direct communication → document results publicly
- **collaboration must be based on trust and open communication**
  - communicate early on wrong estimates

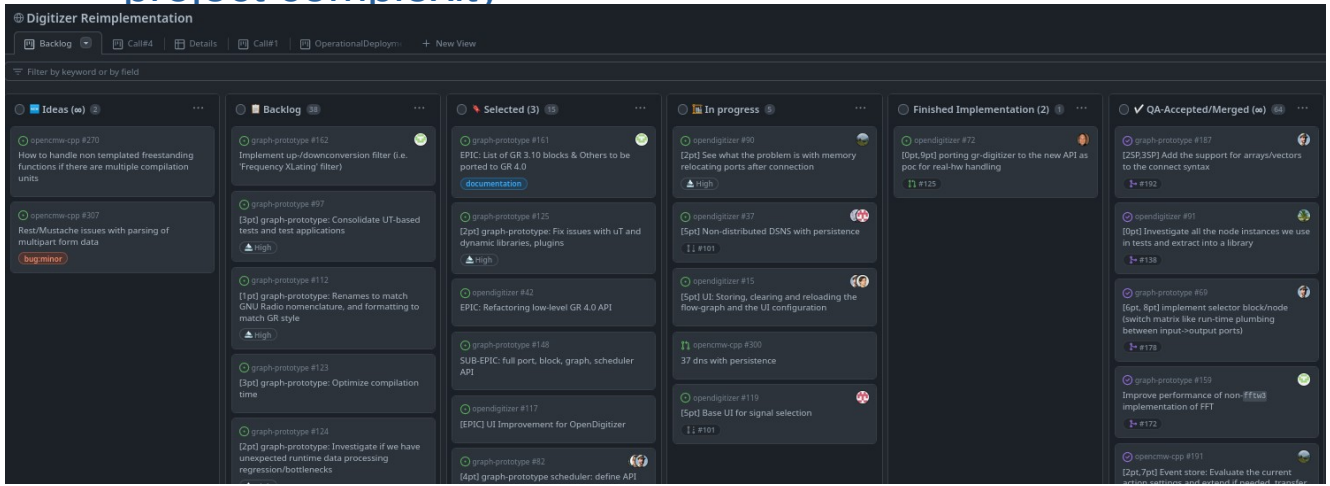


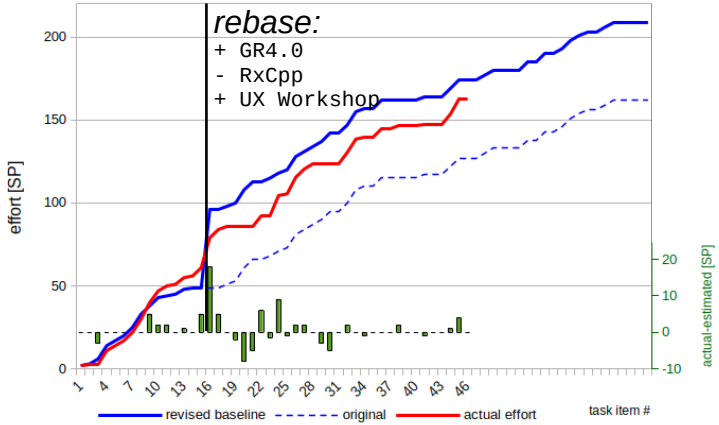
Joint Open Source Development - Overview

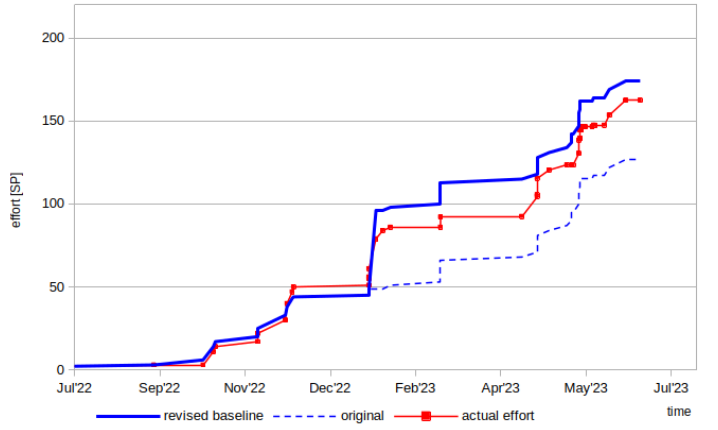


Backlog and Story-Point Estimation

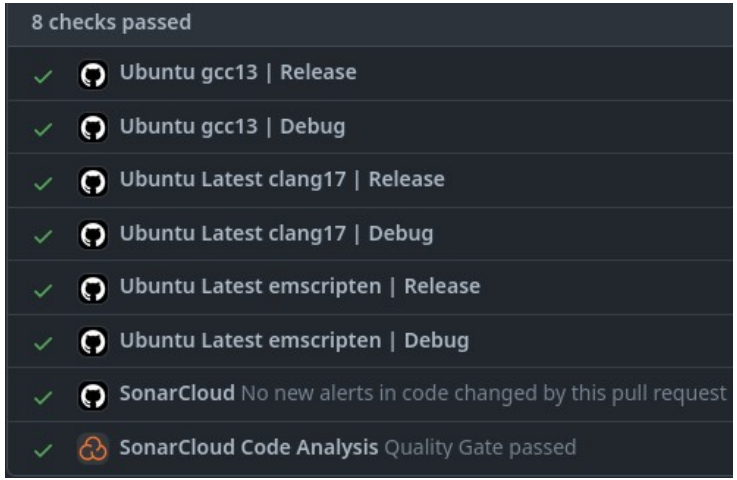
- story-point: one unit of work, approximately one workday
- [estimated, actual] tracking on kanban board issues
  - functional and financial records
  - but more important: continuous feedback and improvement on shared understanding of task and project complexity







- issues on the Kanban board have a predefined **deliverable** and story-point **estimate**. Once done,
  - reviewed by GSI/FAIR,
  - deliverable merged, and
  - story points get tracked (triggers payment)
- PR-**reviews** are assisted by CI:
  - must build without compiler warnings,
  - must pass static code analysis tools (aka. ‘linters’),
  - must pass unit tests and code coverage criteria,
  - automated code formatting, ...

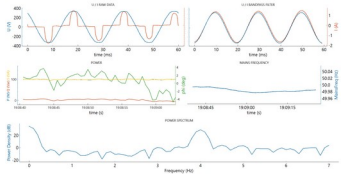


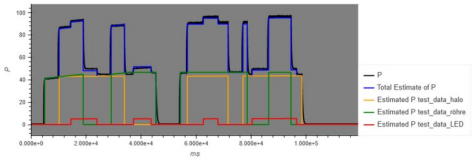
- for big differences between estimated and used story points, discuss causes and possible ways to get back on track

Example collaborations

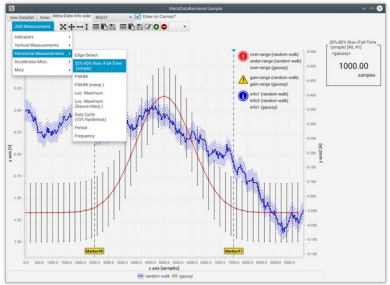
- digitizer framework + GNURadio + KDAB (3 x 180SP, 2019-ongoing)

- AI-based pulsed-power monitoring + Infoteam AG (2 x 90SP)





- chart-fx (Java charting library) + HEBI Robotics (20SP)



Summary

Why have external Partners at all?

- **lack of in-house resources:** freeing in-house team to focus on FAIR-accelerator-specific tasks.

1) Mitigating Risks → ‘agile’ and ‘lean’ Framework Contract
   - general FAIR contract designed for established technologies but unsuitable for SW development
2) Software Quality and Efficiency mirrors Communication Structure (→ M. Conway’s Law)
   - **collaborations must be based on trust + open and transparent communication**
3) Public Documentation of Technologies & Outreach
   - fostering seamless collaboration & attract a diverse talent pool
   - break free from vendor lock-in, circumvent single service-provider problem

Creating Shared Value through creating public ‘clean’ and ‘lean’ Code-Base

- efficient response to 24h/7 operational needs and sustainable long-term maintenance
- FAIR as technology forge: develop technologies critical to FAIR
  & enable industry and general public in new emerging key technologies.



Thanks for your Attention!

Do you use/develop/manage open-source projects?
What are your experiences and approaches?

Alexander Krimm, A.Krimm@gsi.de, GSI Helmholtz Centre for Heavy Ion Research GmbH
HELMHOLTZ Open Science Forum: Research Software, Feb 5 – 6, 2024 Leipzig

HELMHOLTZ

Open Science