HELMHOLTZ
Open Science

Helmholtz Open Science Briefing

2nd Helmholtz Open Science Practice Forum on Research Data Management

Report

Imprint

The online version of this publication can be found at:
https://doi.org/10.48440/os.helmholtz.058

Authors

Nina Leonie Weisweiler, Roland Bertelmann, Wolfgang zu Castell, Hannes Fuchs, Heike Görzig, Thomas Jejkal, Oliver Knodel, Markus Kubin, Christian Langenbach, Inga Patarčić, Ines Schmahl

Publisher

Helmholtz Open Science Office

Editors

Nina Leonie Weisweiler, Lea Maria Ferguson, Roland Bertelmann, Christoph Bruch, Heinz Pampel
Antonia C. Schrader, Paul Schultze-Motel

Contact

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

Version

December 15, 2022. Version 1.0

License

Content

HELMHOLTZ

Open Science

Abstract

To share best practices and to foster the research data management (RDM) community within Helmholtz, the Helmholtz Open Science Office hosted its first "Helmholtz Open Science Practice Forum Research Data Management" virtually in February 2022. A follow-up event on October 20, 2022 has taken up and continued this theme. The following aspects were highlighted through presentations with ample time for discussion in the forum:

- Thinking and linking data, text, and research software together
- Data Stewards, Data Librarians, Research Data Managers, Data Curators… – Their profiles and roles in Helmholtz
- Data Management Plans – DMPs as Living Documents
- Monitoring data publications

HELMHOLTZ

Open Science

Report

Introduction

The Centers of the Helmholtz Association conduct data-intensive research: Complex scientific and technical infrastructures, such as particle accelerators, satellites, and research ships and aircraft, produce a large volume of digital research data. Since the Helmholtz Association's commitment to the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities"[1], the Association has been committed to open access to research data. The Helmholtz Centers hold a high level of expertise in the field of research data management (RDM).

In 2016, the Helmholtz Association adopted the position paper "Making information resources more usable"[2] on the handling of research data. Research data policies are being developed and implemented at the Centers to give a concrete shape to this directional decision. To support the Centers in formulating their policies, the Helmholtz Association's General Assembly adopted practical recommendations on research data policies[3]. Since the publication of these recommendations, twelve Centers have already developed their own guidelines for handling digital research data. Since 2020, the Helmholtz Open Science Office, together with the "Task Group for the Implementation of the Guidelines on Research Data"[4] of the Helmholtz Open Science Working Group, has produced an annual internal report on the handling of research data and the status of the development of the implementation of research data policies at the Centers.

RDM is a lived practice at Helmholtz. This is evidenced by various initiatives and projects at the Centers as well as the Helmholtz-wide platforms in the field of Information and Data Science (Helmholtz Incubator)[5]. In addition, the Centers operate more than 100 digital repositories that store and curate valuable digital research data[6]. To promote collaboration in the field of research data, Helmholtz is involved in both national and international networks: Numerous consortia of the National Research Data Infrastructure (NFDI)[7] are implemented with substantial Helmholtz participation, and Helmholtz Centers are actively involved in the committees and projects of the European Open Science Cloud (EOSC)[8].

The Helmholtz Open Science Office works as a facilitator for dialogue and innovation around open science topics within the Helmholtz Association and beyond. To share best practices and to foster the RDM community within Helmholtz, the Office hosted its first "Helmholtz Open Science Practice Forum Research Data Management" on February 3, 2022. The high number of participants and lively discussions demonstrated a strong need for cross-domain community building in RDM within Helmholtz. A report documents the event (in German)[9].

---

[1] https://os.helmholtz.de/en/open-science-in-helmholtz/berlin-declaration/ [Accessed on: 02/12/2022]
[2] Mitgliederversammlung der Helmholtz-Gemeinschaft (Ed.) (2016): Making information resources more usable: A position paper on the management of research data in the Helmholtz Association, Potsdam: Helmholtz Open Science Office, https://doi.org/10.48440/os.helmholtz.026
[3] https://os.helmholtz.de/en/open-research-data/research-data-policies/ [Accessed on: 02/12/2022]
[4] https://os.helmholtz.de/en/open-science-in-helmholtz/working-group-open-science/task-group-research-data-policies/ [Accessed on: 02/12/2022]
[5] https://www.helmholtz.de/en/research/challenges/information-data-science/ [Accessed on: 02/12/2022]
[6] https://os.helmholtz.de/en/open-research-data/research-data-repositories/ [Accessed on: 02/12/2022]
[7] https://www.nfdi.de/ [Accessed on: 02/12/2022]
[8] https://www.eosc.eu/ [Accessed on: 02/12/2022]
[9] Weisweiler, N. L., Bertelmann, R., Bumberger, J., Elger, K., Fiedler, M., Fuhrmann, P., Knodler, O., Krahl, R., Özkan, Ö., Rhiem, F., Schmahl, I., Servan, S., Upmeier, A., Wedlich-Zachodin, K. (2022): Helmholtz Open Science

In this first practice forum on research data in February 2022, various exemplary approaches to organizing RDM at the Helmholtz Centers were presented. In addition, the focus was placed on concrete service offerings regarding RDM. Furthermore, networking activities with external actors, e.g., in the context of NFDI or EOSC, were highlighted. The second forum that took place in October 2022 revisited the topic of RDM, taking a closer look at four specific areas:

- Thinking and linking data, text and research software together
- Data Stewards, Data Librarians, Research Data Managers, Data Curators… – Their profiles and roles in Helmholtz
- Data Management Plans – DMPs as Living Documents

Each thematic part of the forum was followed by an open question and discussion slot. The present report looks into the second forum and documents key findings and discussion points.

Summary of the Forum

Part 1: Thinking and linking data, text and research software together

Wolfgang Graf zu Castell-Rüdenhausen, CIO of the GFZ German Research Center for Geosciences, head of its Department Geoinformation, and chair of the Helmholtz Open Science Working Group held the forum's first presentation. He demonstrated that the reusability and reproducibility of scientific results can be improved through open science practices, in particular by considering all steps in the research cycle and all forms of research outputs, such as research data and software.

The second talk by Oliver Knodel, head of the Data Management and HPC group at Helmholtz-Zentrum Dresden-Rossendorf (HZDR), described how software projects such as HELIPORT and HERMES, as well as interoperable repositories and PIDs enable an interlinked ecosystem of research data and software at HZDR.

Concluding the first part of the forum, Thomas Jejkal, IT specialist at the Karlsruhe Institute of Technology (KIT) and coordinator of the FAIR Data Commons unit of the Helmholtz Metadata Collaboration (HMC), presented FAIR Digital Objects as "a soft way of introducing FAIRness" within Helmholtz and beyond.

Part 2: Data Stewards, Data Librarians, Research Data Managers, Data Curators… – Their profiles and roles in Helmholtz

The RDM group leader in the IT department of the Helmholtz-Zentrum Berlin für Materialien und Energie (HZB), Heike Görzig, opened the second thematic part of the forum. She described the past and current development of the RDM Group, highlighting its organizational structure, the tools and services it offers, and the (initial) challenges it has faced.

Ines Schmahl, research data specialist in the Information Services department of the Central Library of Forschungszentrum Jülich (FZJ), continued the session by elaborating how a combination of centralized and decentralized roles and responsibilities in RDM helps to meet the heterogeneous needs at FZJ. She emphasized that such roles require a holistic view of RDM within the framework of open science.

Briefing. Helmholtz Open Science Praxisforum Forschungsdatenmanagement: Report. https://doi.org/10.48440/os.helmholtz.044

Part 3: Data Management Plans – DMPs as Living Document

To kick off the third themed session, Christian Langenbach, responsible for research data management at the German Aerospace Center (DLR), posed the question of whether data management plans (DMPs) have a 'demonic or angelic' effect on researchers' everyday work. He concluded that the creation of and adherence to DMPs have many positive effects on the work of researchers and are therefore highly recommended.

Inga Patarčič from the Research Data Management Services Unit at the Max Delbrück Center for Molecular Medicine (MDC) introduced the DMP tool "FAIR Wizard" and explained why and how it is implemented to support researchers at MDC with data management.

Closing the DMP theme, Hannes Fuchs from the eScience Centre at GFZ German Research Centre for Geosciences discussed the application of the MOSES DMP Tool at GFZ and clarified why DMPs should function as interactive and networked "living documents" rather than static entities.

Part 4: Monitoring of data publications

In the last part of the forum, Markus Kubin from Helmholtz-Zentrum Berlin (HZB) spoke in his role as FAIR data specialist in the Hub Matter team of the Helmholtz Metadata Collaboration (HMC). He introduced a recently developed dashboard approach to monitor research data publications within the HMC Hub Matter and to assess their conformance with the FAIR principles.

Concluding Remarks

The 2nd Helmholtz Open Science Practice Forum on Research Data Management provided an opportunity to further discuss the topics identified as particularly relevant in the first event: Interoperability and meaningful linking of tools and infrastructures so that research data can be processed smoothly in workflows, the roles and responsibilities of data stewards and comparable personnel in Helmholtz, and the conception of data, text, and software publications as interrelated entities.

The vivid exchange and high number of participants again demonstrated the need for continued discussion and sharing of best practices for RDM in Helmholtz: 107 people from all 18 Helmholtz Centers participated in the forum.

Appendix

Program of the event from October 10, 2022

Time	Agenda	Speaker
10:00 - 10:15	Introduction	Roland Bertelmann Helmholtz Open Science Office
10:15 - 10:30	Towards an Integrated Digital Research Ecosystem	Wolfgang zu Castell GFZ
10:30 - 11:45	Towards a Seamlessly Interlinked Research Data and Software Ecosystem at HZDR	Oliver Knodel HZDR
10:45 - 11:00	FAIR Digital Objects - A gentle way of introducing FAIRness	Thomas Jejkal KIT / HMC FAIR Data Commons
11:00 - 11:15	Questions & discussion	Moderation: Open Science Office
11:15 - 11:20	Break	
11:20 - 11:30	RDM structures @ HZB	Heike Görzig HZB
11:30 - 12:40	Roles in RDM at the Forschungszentrum Jülich – From RDM to Open Science	Ines Schmahl FZ Jülich
12:40 - 12:00	Questions & discussion	Moderation: Open Science Office
12:00 - 13:00	Lunch Break	
13:00 - 13:10	Data Management Plans – Angel or Devil for the Researcher Work	Christian Langenbach DLR
13:10 - 13:20	FAIR Wizard at the MDC	Inga Patarčič MDC
13:20 - 13:30	Data Management Plans – DMPs as Living Documents	Hannes Fuchs GFZ
13:30 - 13:45	Questions & discussion	Moderation: Open Science Office
13:45 - 14:00	Monitoring Data Publications: A Dashboard Approach in HMC Hub Matter	Markus Kubin HZB / HMC
14:00 - 14:30	Wrap-up and final discussion	Roland Bertelmann Helmholtz Open Science Office

Presentations

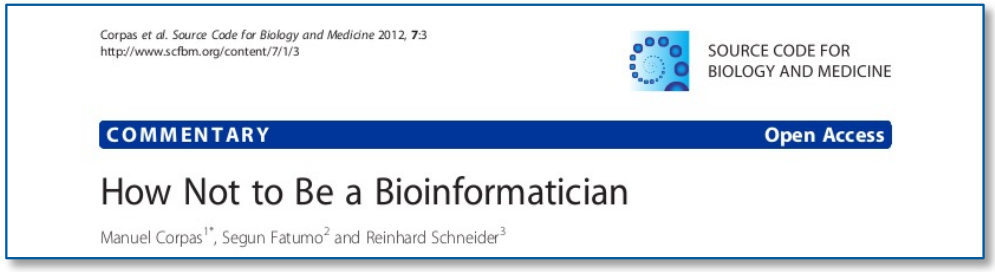
1. Wolfgang zu Castell: Towards an Integrated Digital Research Ecosystem
2. Oliver Knodel: Towards a Seamlessly Interlinked Research Data and Software Ecosystem at HZDR
3. Thomas Jejkal: FAIR Digital Objects – A gentle way of introducing FAIRness
4. Heike Görzig: RDM structures @ HZB
5. Ines Schmahl: Roles in RDM at the Forschungszentrum Jülich – From RDM to Open Science
6. Christian Langenbach: Data Management Plans – Angel or Devil for the Researcher Work
7. Inga Patarčič: FAIR Wizard at the MDC
8. Hannes Fuchs: Data Management Plans – DMPs as Living Documents
9. Markus Kubin: Monitoring Data Publications: A Dashboard Approach in HMC Hub Matter

Towards an Integrated Digital Research Ecosystem

2nd Practice Forum Research Data Management

Wolfgang zu Castell (Department Geoinformation)

From a talk at Open Access Days 2014 …



1. Be **open source** without being open. … Ensure that your code is not portable, it only works in outdated operating systems and assume only you will use your application. **Take for granted that everyone will be able to understand it.**
2. Never **maintain your databases**, web services or any information that you may provide at any time. Provide unstable data, unstable models and unstable services. Your ultimate goal in data curation should be to **propagate as many errors as possible** from one database to another, while still making sure that they sound realistic.
3. Do not ever **share your results** and do not reuse. Never discuss your results before your submission has been accepted in a lost conference proceeding. Consider that the work others are doing is probably a waste of time. Ignore whatever new algorithms and methods your colleagues have developed in the last two decades.

Highlightening not from the paper.

An analysis during the pandemic …

EDITORIALS

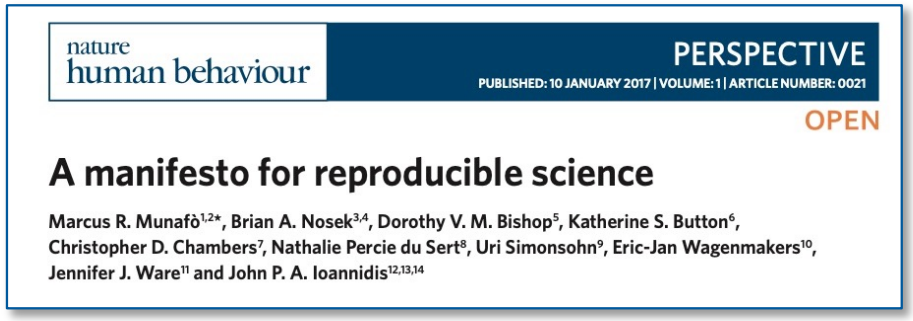
Check for updates

Waste in covid-19 research

A deluge of poor quality research is sabotaging an effective evidence based response

Paul P Glasziou *professor of evidence based medicine*, Sharon Sanders *assistant professor*, Tammy Hoffmann *professor of clinical epidemiology*

Institute for Evidence Based Healthcare, Bond University, Australia



nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

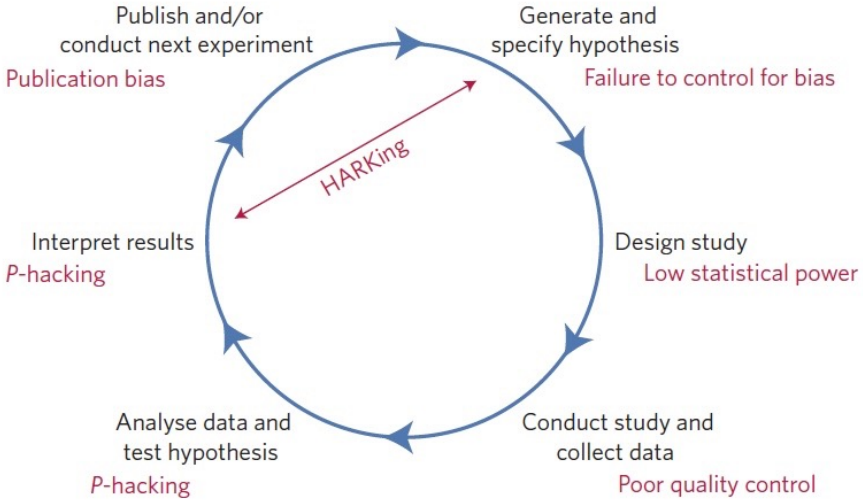
Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

observation

- waste due to lack of scientific scrutiny
- existing research infrastructures to enable collaboration and communication are extremely limited
- coordination has been lacking

conclusion

open research increased poor quality research

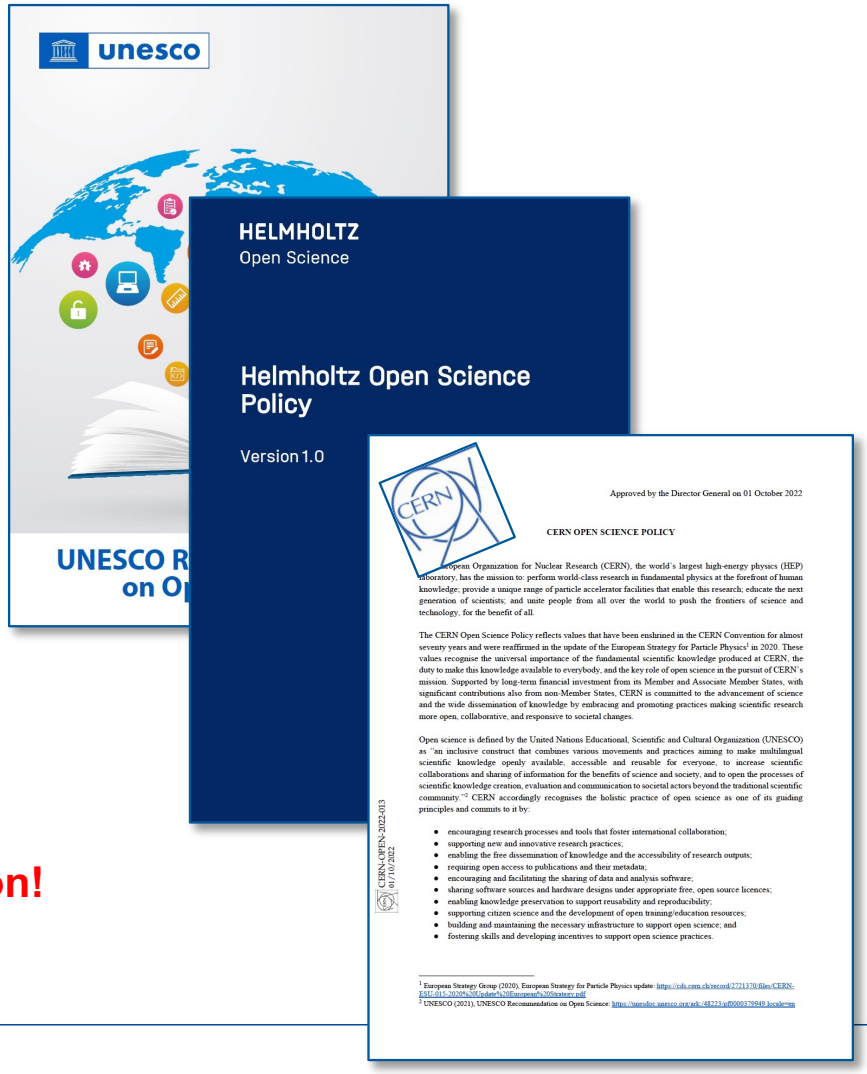


Munafo et al., Nat Hum Beh 2017

Recommendations …

Munafo et al. suggest (among others) …

- protecting against cognitive biases
- encouraging collaboration and team science
- improving methodological training
- encouraging transparency and open science (open data, open materials, open software …)
- rewarding open and reproducible practices



Policies are helpful but we need to put them into action!

„We need to talk about reproducibility“

Crick, Hall and Ishtiaq cite studies, finding that **50% of published studies** (including top-tier journals) **cannot be repeated** by an industrial lab.


Journal of open research software

Crick, T, et al. 2017 Reproducibility in Research: Systems, Infrastructure, Culture. *Journal of Open Research Software*, 5: 32. DOI: https://doi.org/10.5334/jors.73

ISSUES IN RESEARCH SOFTWARE

Reproducibility in Research: Systems, Infrastructure, Culture

Tom Crick[1], Benjamin A. Hall[2] and Samin Ishtiaq[3]

[1]Cardiff Metropolitan University, UK

[2]University of Cambridge, UK

[3]Microsoft Research Cambridge, UK

Corresponding author: Tom Crick (tcrick@cardiffmet.ac.uk)

conclusion

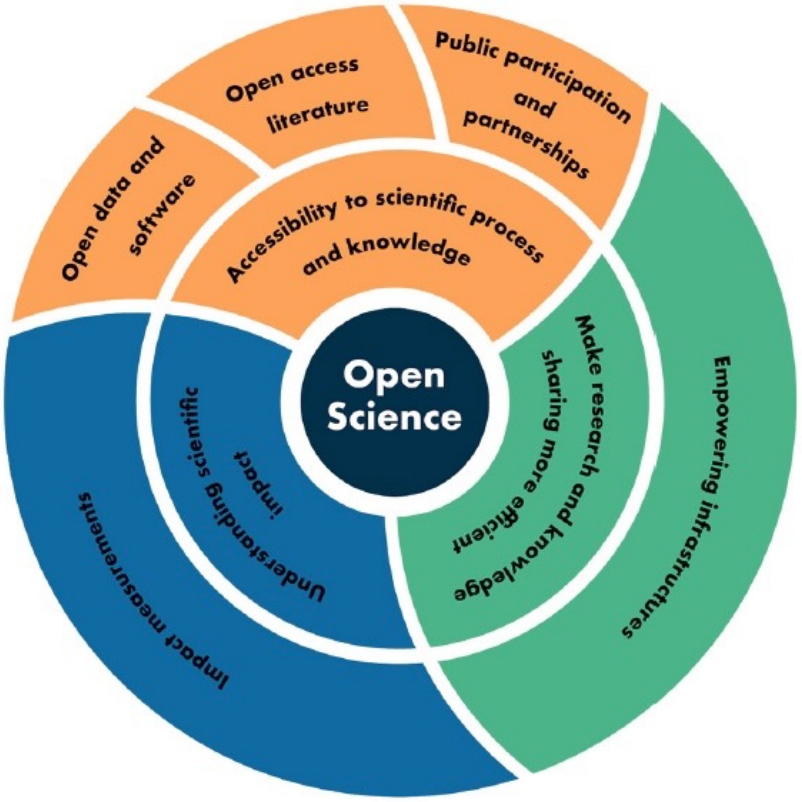
- just publishing linked data is not enough
- „set the code free“
- welcome Web 2.0 technologies (use and share workflows, provide web services, use cloud …)

there is also substantial potential[1]

In 2009 Tim Gowers (mathematician, Fields medalist) posted an unsolved problem in a blog. Within little more than a month, 27 people had made 800 comments leading to the solution of the problem.

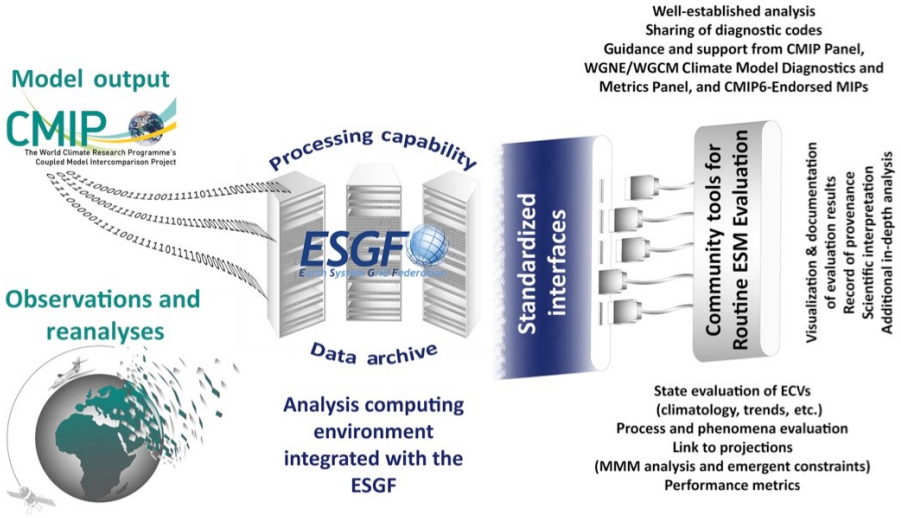
[1] example cited from: The Royal Society: Science as an open enterprise, June 2912

There are guiding examples …



Ramachandran, Bugbee & Murphy, Earth Space Sci 2021

Coupled Model Intercomparison Project (CIMP)



Eyring et al., Earth Syst Dynam 2016

Geosci. Model Dev., 11, 3659–3680, 2018
https://doi.org/10.5194/gmd-11-3659-2018

Geoscientific Model Development
Open Access
EGU

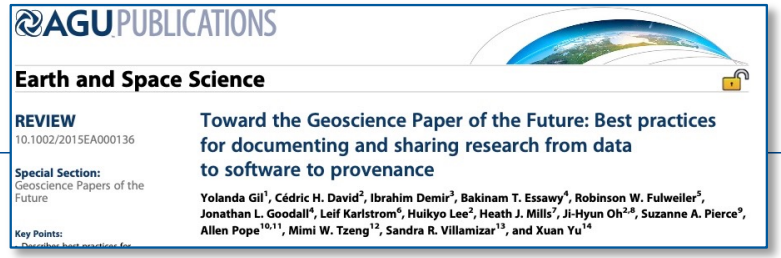
Requirements for a global data infrastructure in support of CMIP6

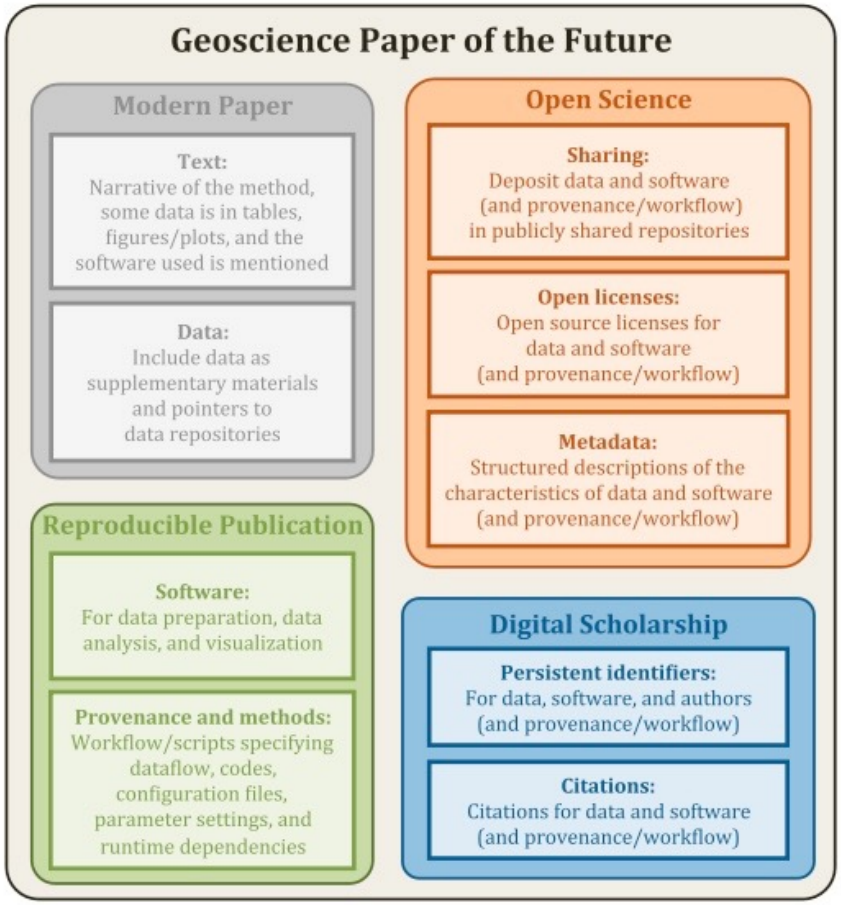
Venkatramani Balaji[1,2], Karl E. Taylor[3], Martin Juckes[4], Bryan N. Lawrence[5,4], Paul J. Durack[3], Michael Lautenschlager[6], Chris Blanton[7,2], Luca Cinquini[8], Sébastien Denvil[9], Mark Elkington[10], Francesca Guglielmo[9], Eric Guilyardi[9,4], David Hassell[4], Slava Kharin[11], Stefan Kindermann[6], Sergey Nikonov[1,2], Aparna Radhakrishnan[7,2], Martina Stockhause[6], Tobias Weigel[6], and Dean Williams[3]

Towards new ways of sharing scientific results …

1. *Make data reusable* through publication in a **public repository**, with documentation (**metadata**), a clear **license** specifying conditions of use, and citable using a **unique and persistent identifier**.
2. *Make software reusable* through publication in a **public repository**, with documentation, a **license** for reuse, and citable with a **unique and persistent identifier**. This includes modeling software as well as all software for data (re)formatting, conversions, filtering, analysis, and visualization.
3. *Document the computational provenance of results* by explicitly describing the series of computations and their outcome in a high-level workflow diagram, a formal workflow, or a **computational provenance record**, possibly stored in a **shared repository** and citable with a **unique and persistent identifier**.

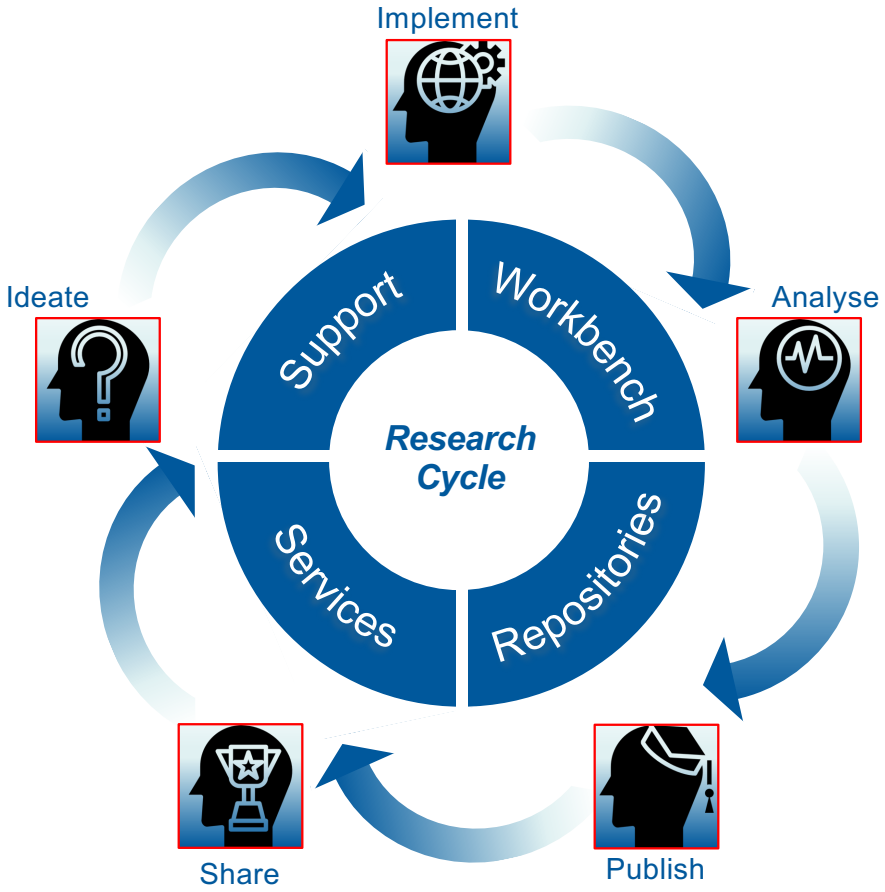
Highlightening not from the paper.





Gil et al., Earth Space Sci 2016

Starting point is the research cycle …

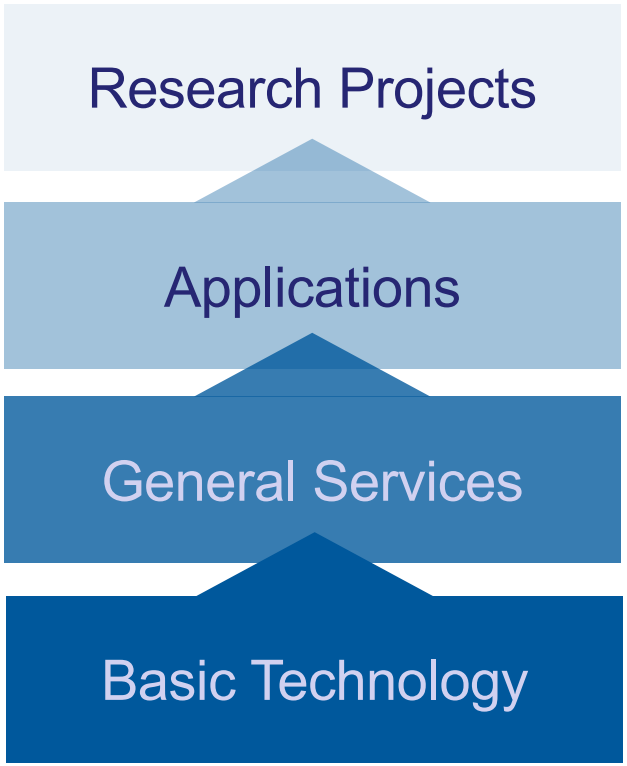


Almost all artefacts of scientific work today are digital.

The goal is to build a digital ecosystem which allows to move through the steps of the cycle in arbitrary ways.

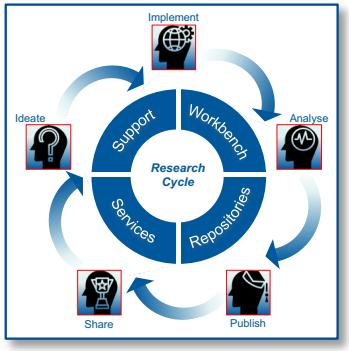
Therefore, we must integrate services on all levels of the technological stack.

Technological Stack



Our vision: a Digital Research Ecosystem …

DataHub
Research Field Earth and Environment



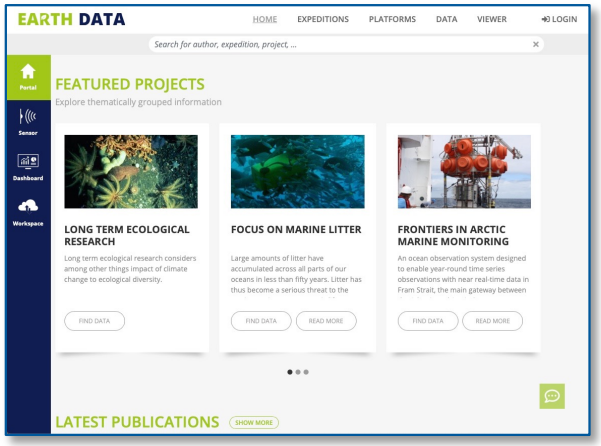
The **cloud** herby abstracts users from the explicit instantiation of technology.

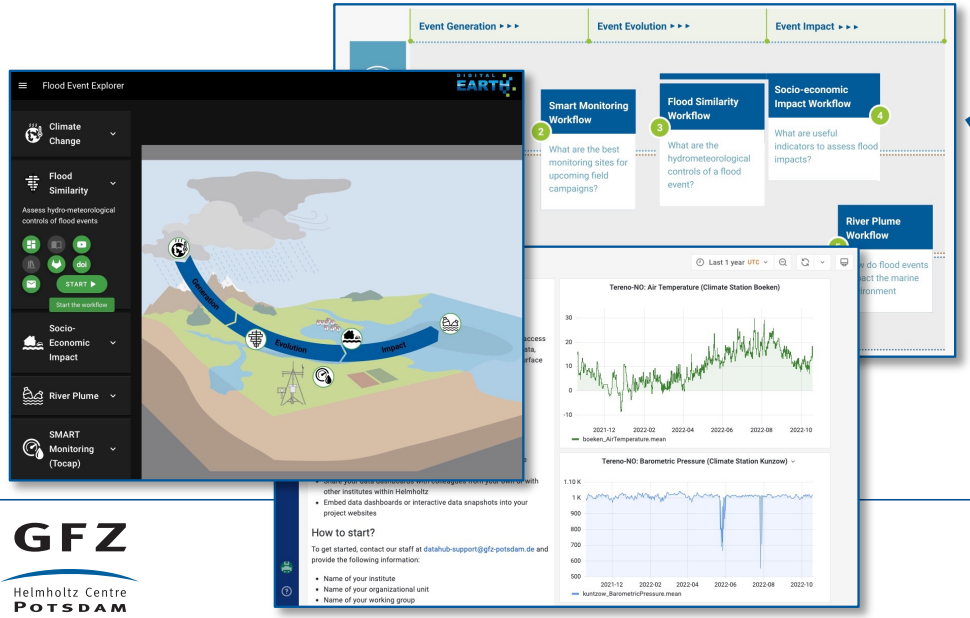
GFZ Helmholtz Centre Potsdam

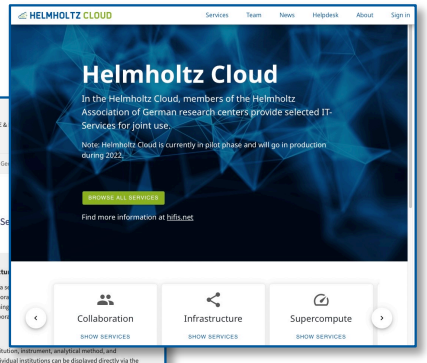
HELMHOLTZ

Digitalization in the Research Field Earth & Environment

DataHub
Research Field Earth and Environment







GFZ Helmholtz Centre Potsdam

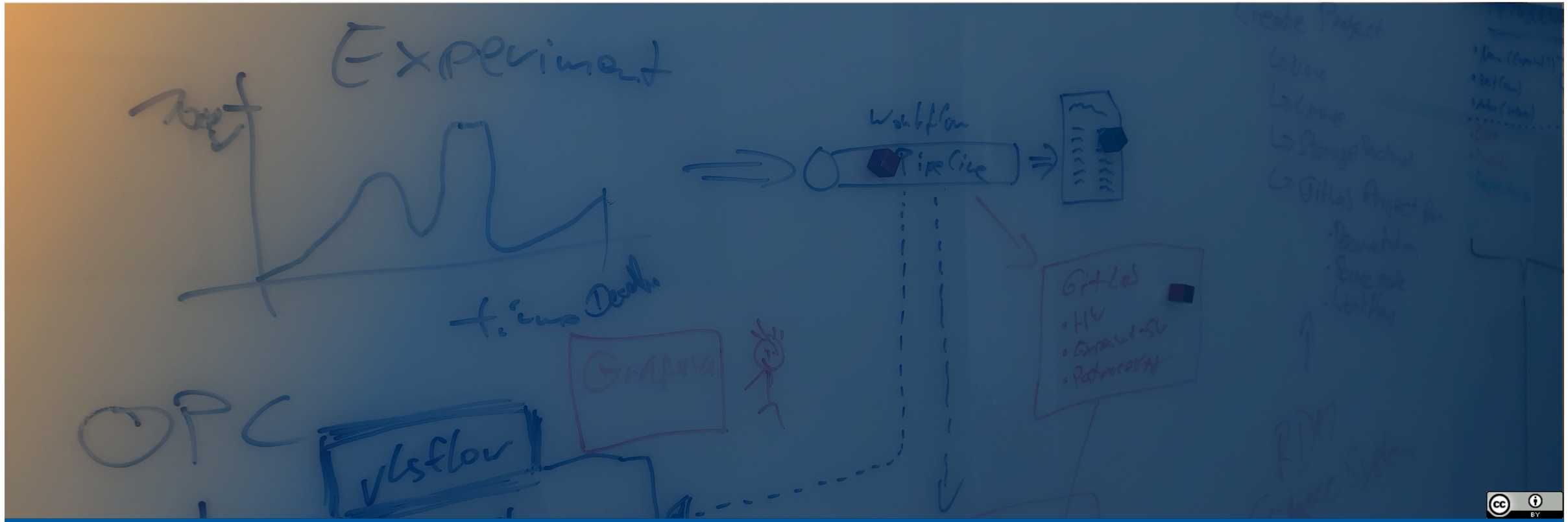
HELMHOLTZ

There are many (parallel) initiatives …

… let‘s start to bring them together!

Thank you for your attention!

HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF



Towards a Seamlessly Interlinked Research Data and Software Ecosystem at HZDR

2nd Practice Forum Research Data Management, October 20, 2022

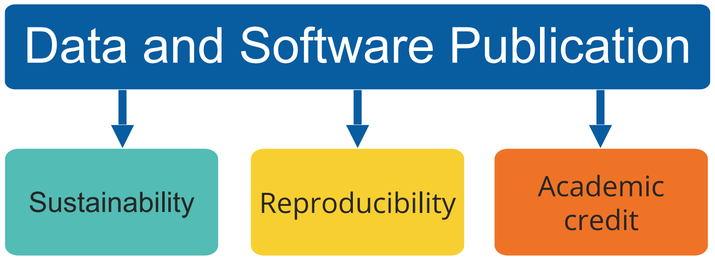
Oliver Knodel // contact: o.knodel@hzdr.de

DRESDEN concept

Motivation — Software, Data and Everything in Between

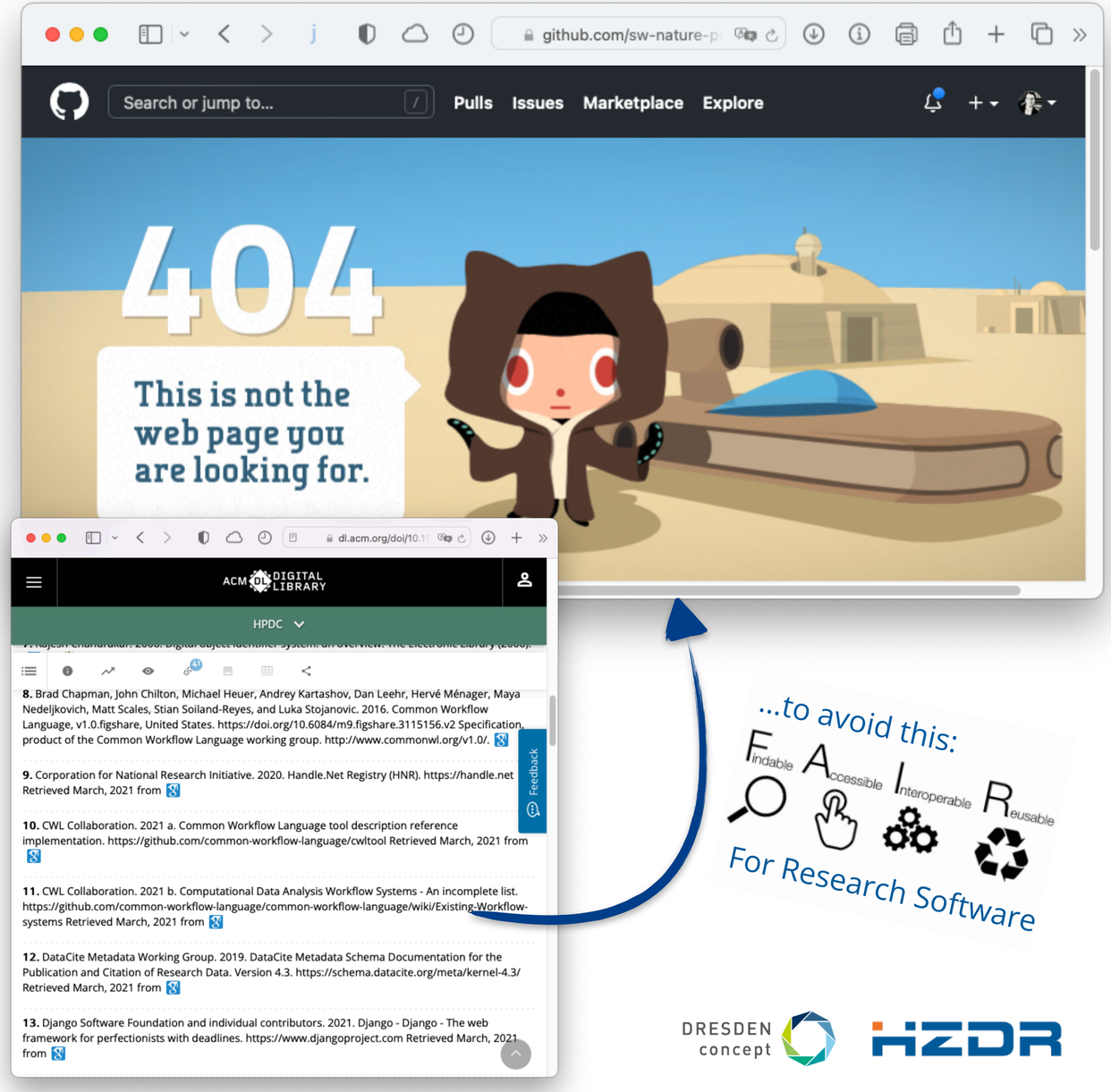
— Data and software are an important result of a scientific experiment.

— Scientific publication, research software and data must receive the same academic credit:



— FAIR principles also exist for research software and should be taken into account [1].

— In addition to the publication itself a seamless interlinking between all available data products is also necessary to improve findability.



[1] Barker, M., Chue Hong, N.P., Katz, D.S. *et al.* Introducing the FAIR Principles for research software. *Sci Data* **9**, 622 (2022). doi.org/10.1038/s41597-022-01710-x

DRESDEN concept

HZDR

The Git Repository is not a Publication…

Software Repository

— Software is typically available (not *published)* in version control systems with open or restricted access:

— We need workflows or methods to publish software and data to ensure long-term *availability* and to meet the FAIR principles.

Software Publication

— Software must be cited in a similar way to scientific publications.

— Common data repositories (e.g. institutional, domain-specific, Zenodo) support typically the publication type *software*.

— With an additional *software publication* we can cite specific versions of a software including rich metadata:

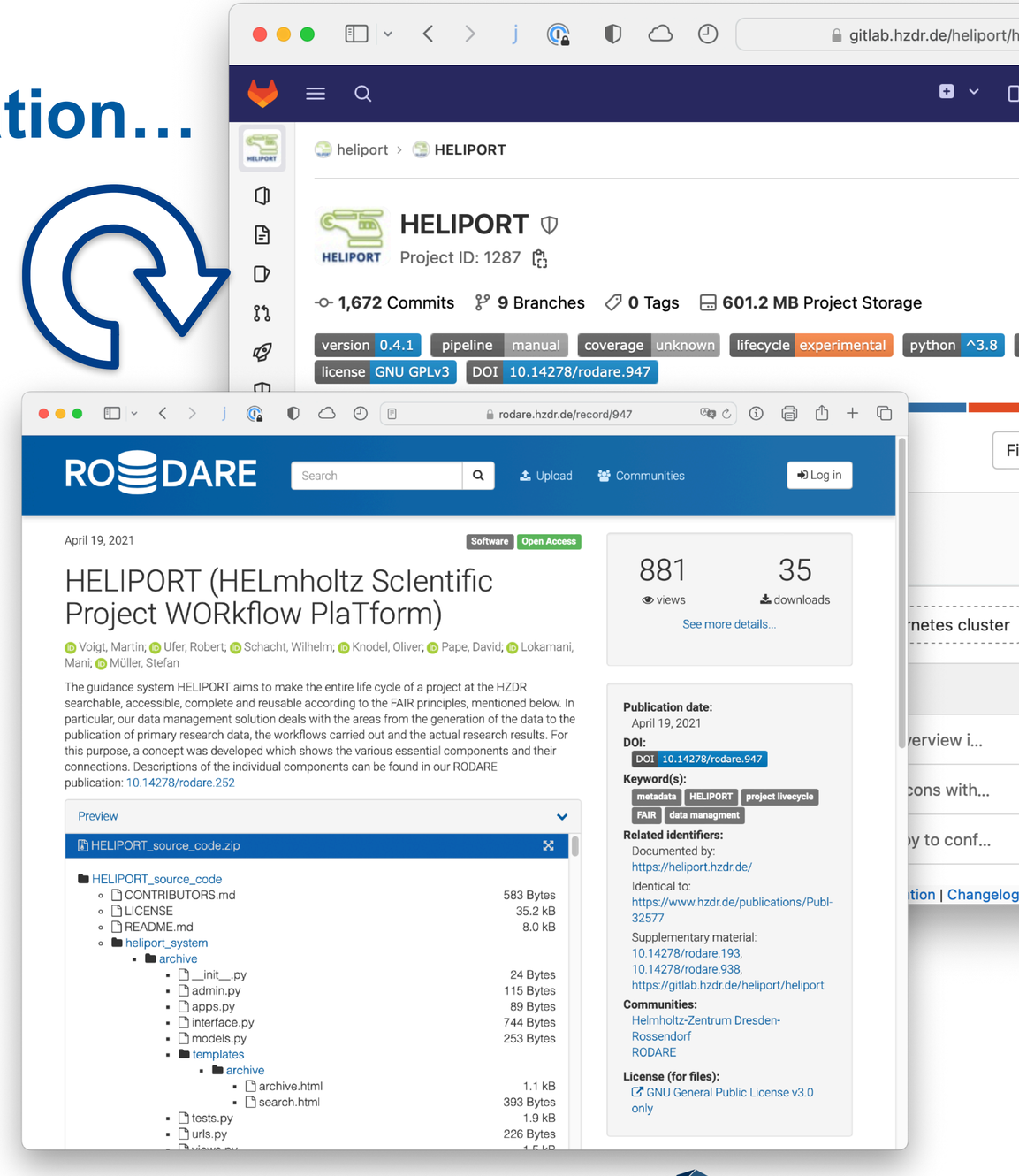
- Title, authors (including ORCIDs), Abstract, license, …
- Related Identifiers to link additional resources.
- Typically it would not be practical to link *all* scientific datasets produced or analyzed by the software to the software publication.

Repository Structure

— Software and date mixen in one repository,

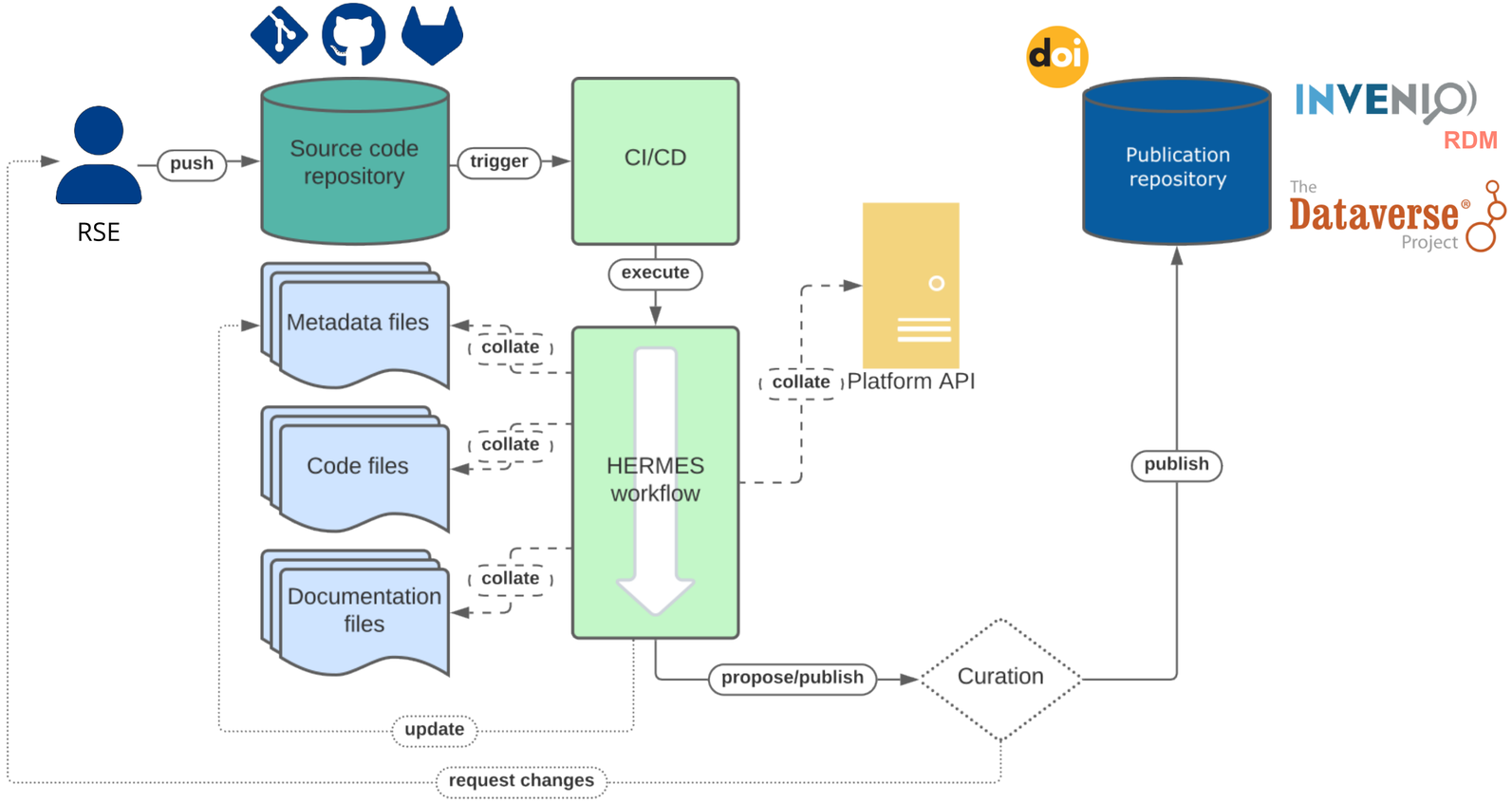
— Separation between data and software repositories,

— Something in between…



The HERMES Project: Automated Software Publication Workflow

- A simple and transparent software publication workflow for open and closed access software can be a platform for an understandable science.
- The metadata harvesting is essential to create a findable software publication.



HERMES

project.software-metadata.pub

github.com/hermes-hmc

team@software-metadata.pub

HMC | HELMHOLTZ METADATA COLLABORATION

DLR JÜLICH Forschungszentrum HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

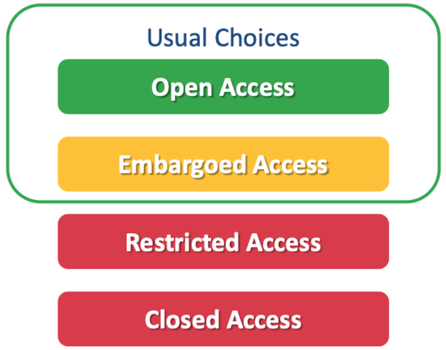
- 07/2021 – 06/2023
- Aim: Support RSEs in automatedly publishing their software with rich metadata

DRESDEN concept HZDR

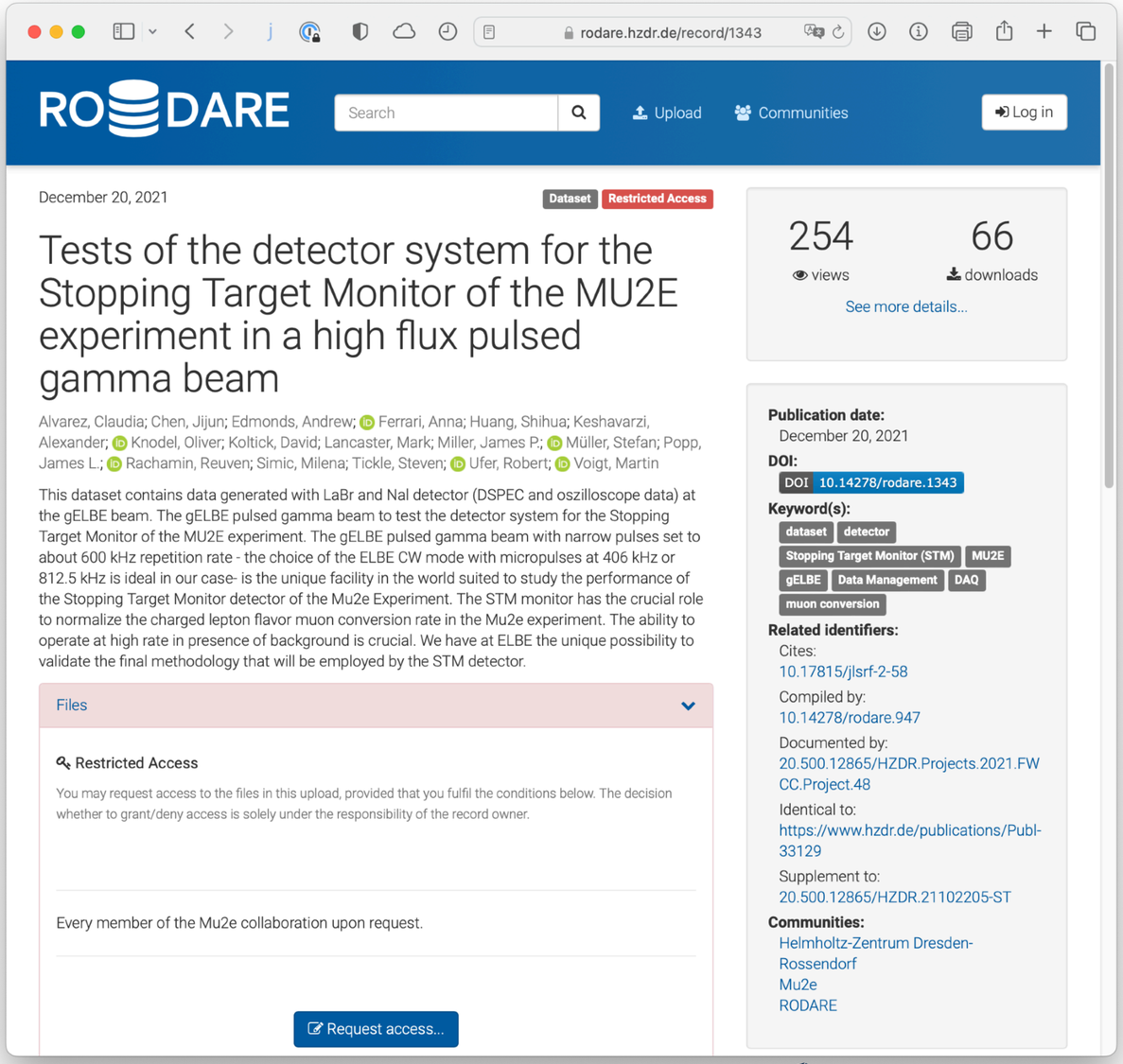
Data (and Software) Publication Repositories

- The data can be published in the same or a different repository as the software (possibly there is a domain-specific repository for the scientific data).
- A dataset (at least the metadata) should always be published to provide sustainable scientific evidence.
- The data itself can also be published under restricted access.*

*Nevertheless it fulfills the FAIR principles, because the steps to access the data are documented.



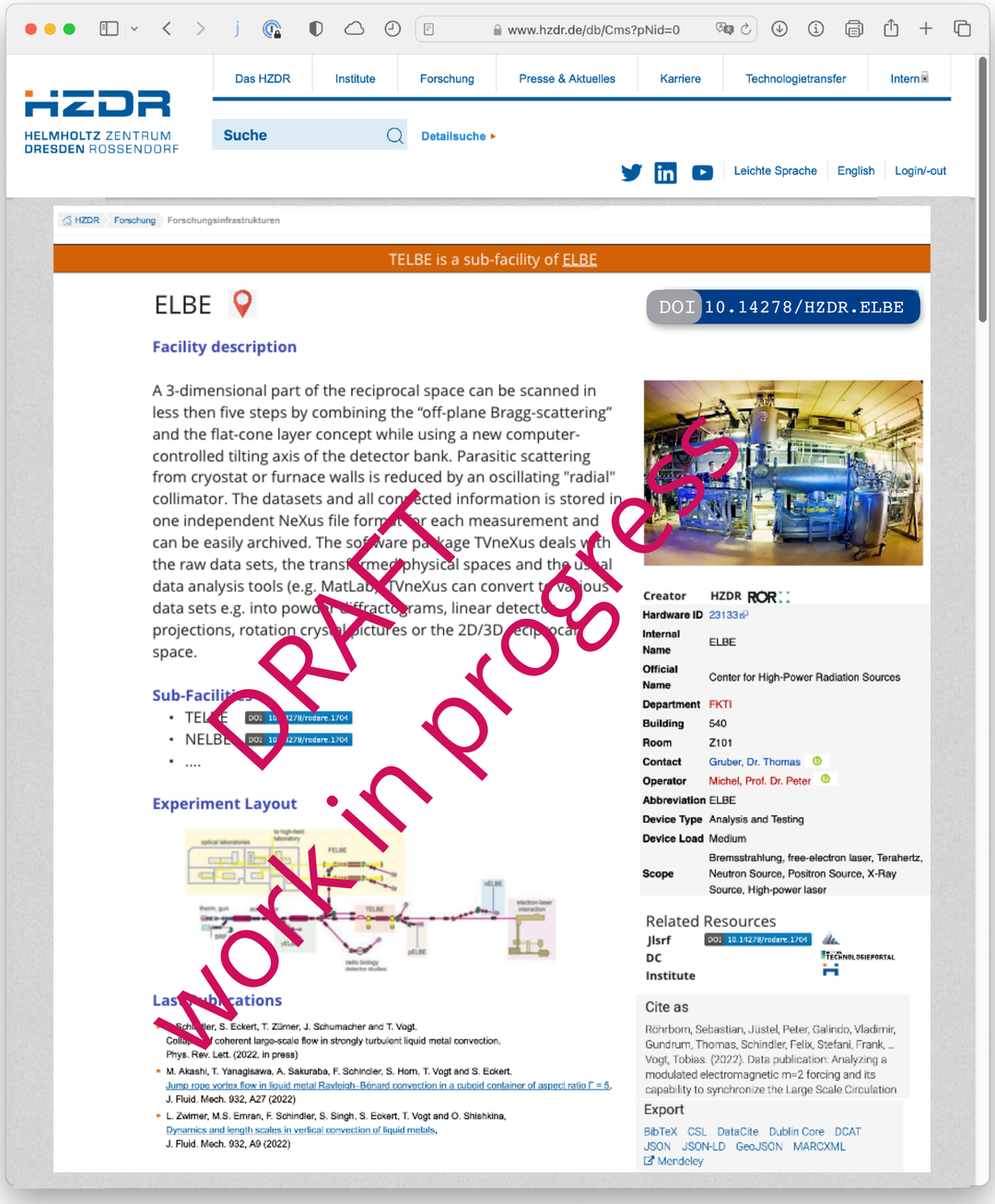
- The data publication should reference:
  - The software repository used to create or analyse the dataset.
  - The scientific publication based on the datasets.
  - The instrument or facility where the data was generated…



DRESDEN concept HZDR

Instrument DOIs and Landing Page

- For data publications we have the field *related identifiers*, were we can refer research facilities and instruments.
- Therefore, we plan to assign DOIs to instruments and provide DataCite records [2] and additional metadata on public landing pages.
- Components of the landing pages:
  - Mandatory: DOI, name, description, contacts, scope, location, ROR, device type.
  - Optional: Image, layout, sub-facilities, additional resources (JLSRF publication, internal website, …) and the latest publications.
  - Citation export to BibTex, JSON, …

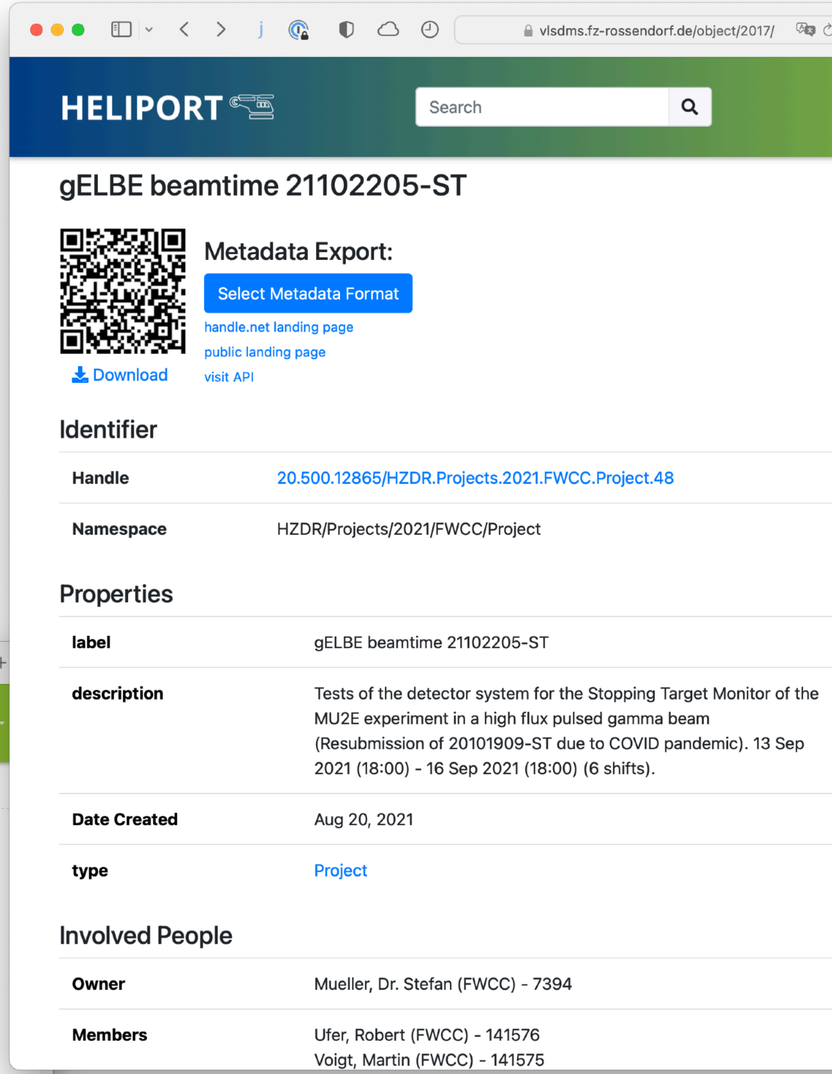


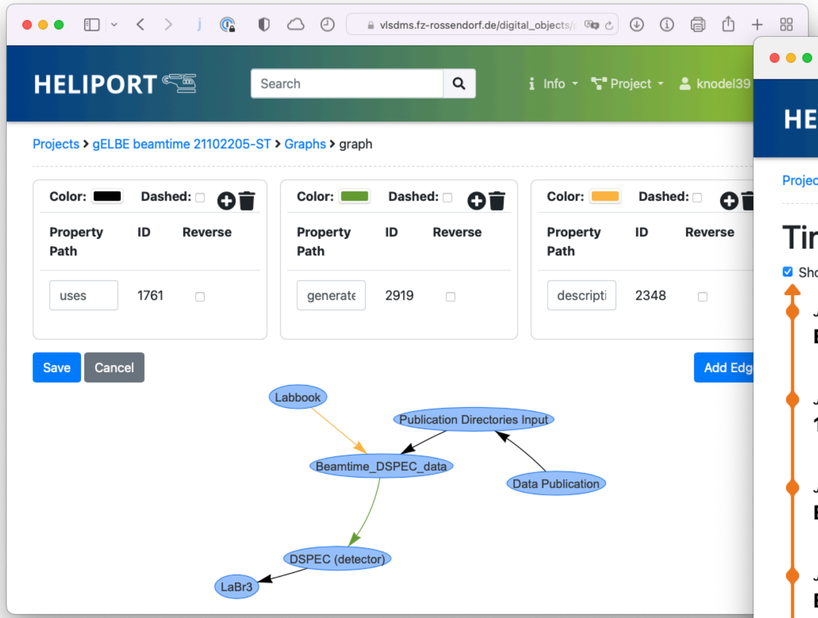
[2] Bunakov, Vasily, Krahl, Rolf, Matthews, Brian, Vizcaino, Noeland Vukolov, Andrey, "Advanced infrastructure for PIDs in Photon and Neutron RIs", ExPaNDS project deliverable D2.5, Zenodo, Mar. 2022. doi: 10.5281/zenodo.5905351.

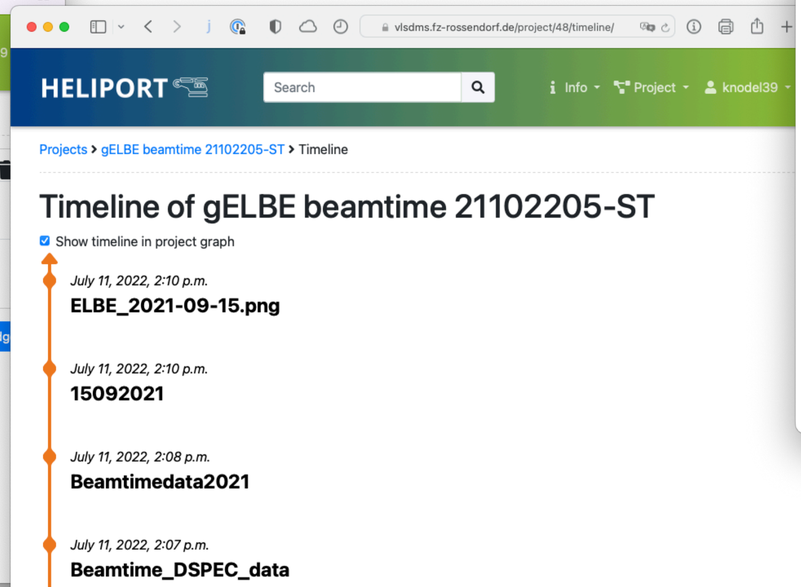
DRESDEN concept HZDR

Digital Objects and Handles Enable Long-term Sustainability

- At the HZDR, we use DOIs for resources containing the whole set of bibliographic metadata:
  - Scientific articles,
  - Published datasets and software,
  - Instruments.
- Other identifiers included in the metadata are ORCID and ROR available in our internal databases for almost every scientist at HZDR.
- Further digital objects can request PIDs from our Handle hdl enabled server.
- The digital objects in our ecosystem can be correlated with each to create a comprehensible experiment providing data provenance.







Linking Data, Text and Research Software Together

Software:

I. HERMES can extract the metadata provided by Github or GitLab.

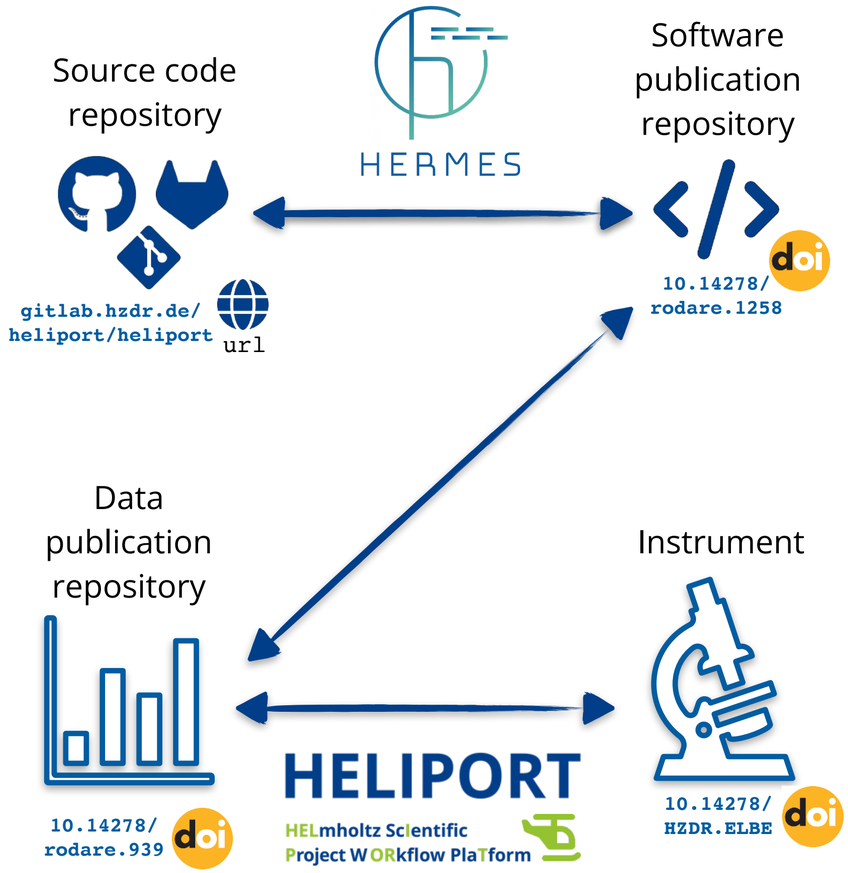
II. A software release can trigger a pipeline that initialises a publication with DOI based on the available (and third-party) metadata.

III. In a subsequent step, the DOI is added to the Readme file in the Git repository and the cross-linking is completed.

Data:

I. After data collection or processing, a pipeline can start collecting metadata from a proposal system or other related services.

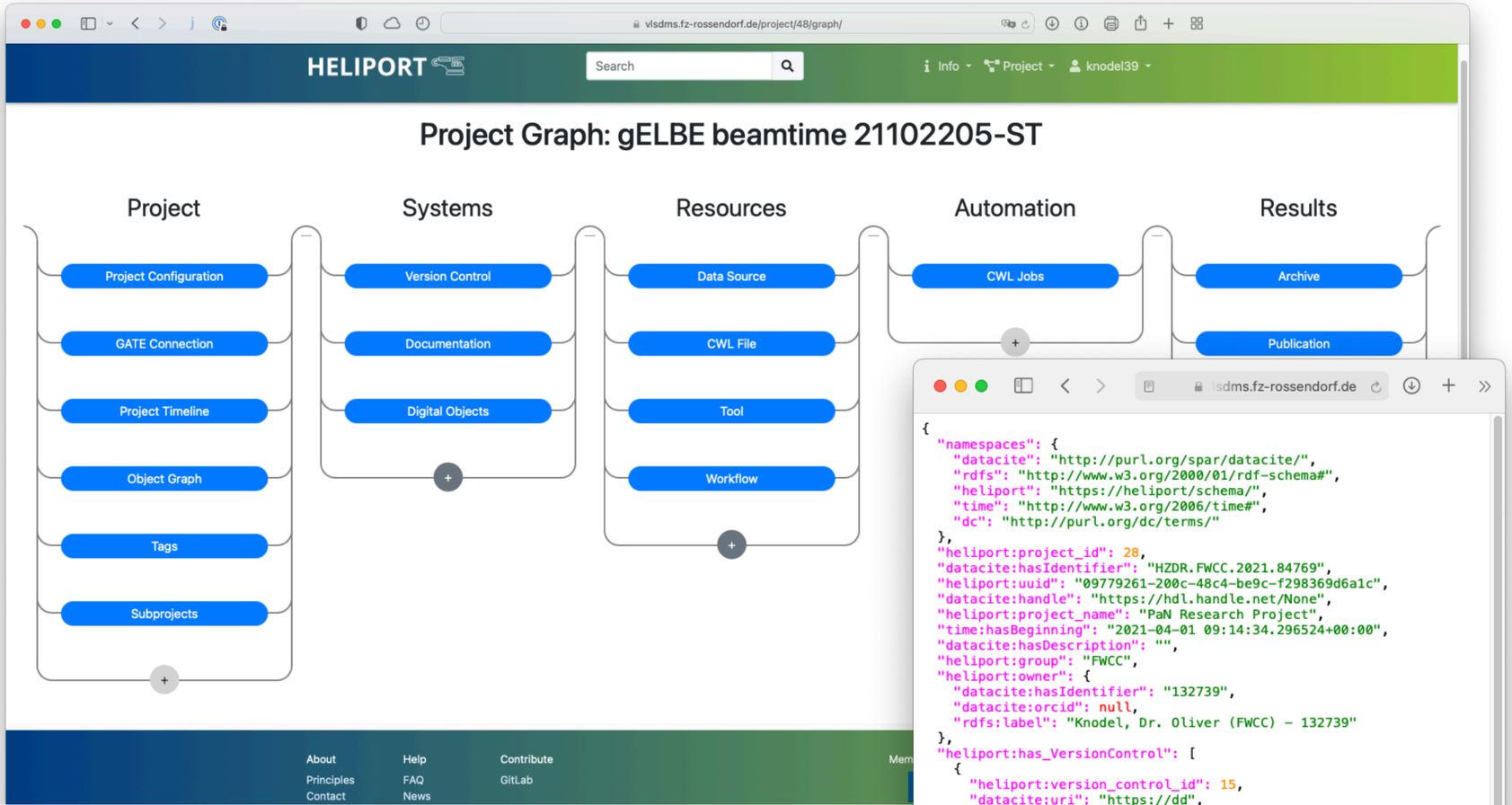
II. The metadata and information from a computational workflow can be used to create a data publication with references to a specific software version (DOI) and the instrument where the data was taken.



Overview of the Project Resources from a Higher Level

> 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.

HELIPORT
HELmholtz Scientific Project WORkflow PlaTform



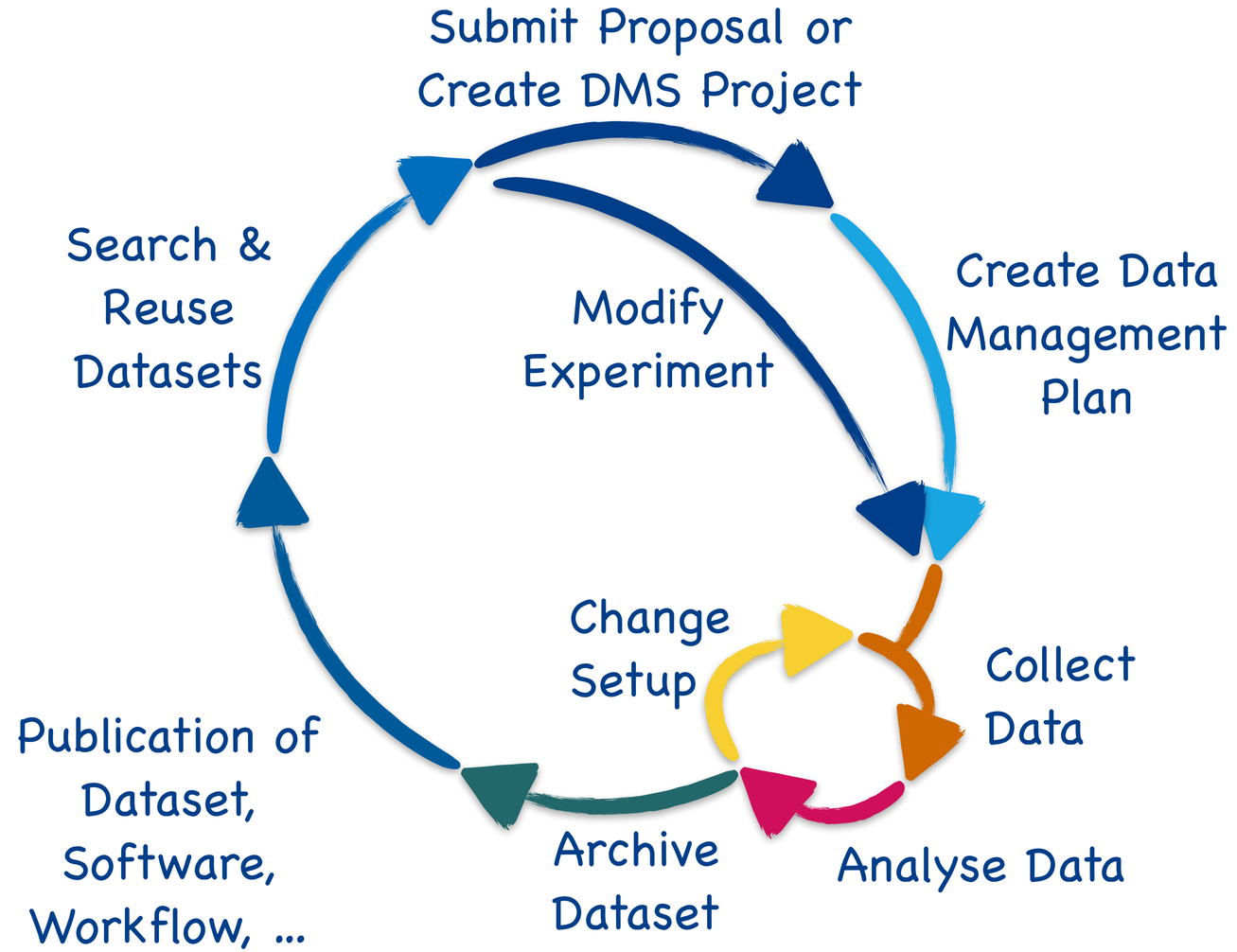
heliport.hzdr.de
codebase.helmholtz.de/heliport
heliport@hzdr.de

HMC | HELMHOLTZ METADATA COLLABORATION
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
HI JENA Helmholtz Institute Jena
JÜLICH FORSCHUNGSZENTRUM

- 07/2021 – 06/2023
- Aim: Collect every system, service or digital product of a research project in an uniform metadata package.

DRESDEN concept
HZDR

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

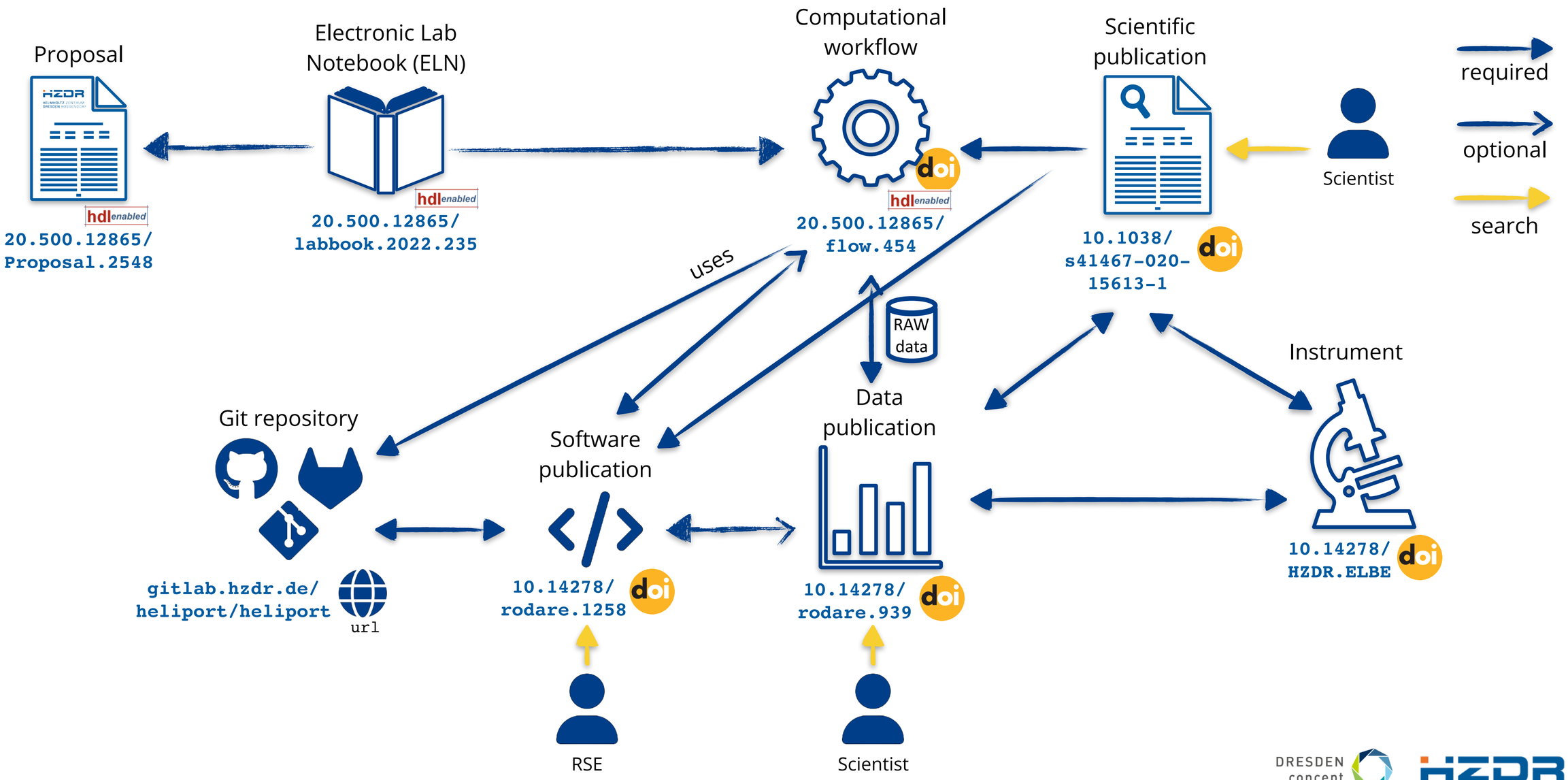


HELIPORT
HELmholtz Scientific Project WORkflow PlaTform

HERMES

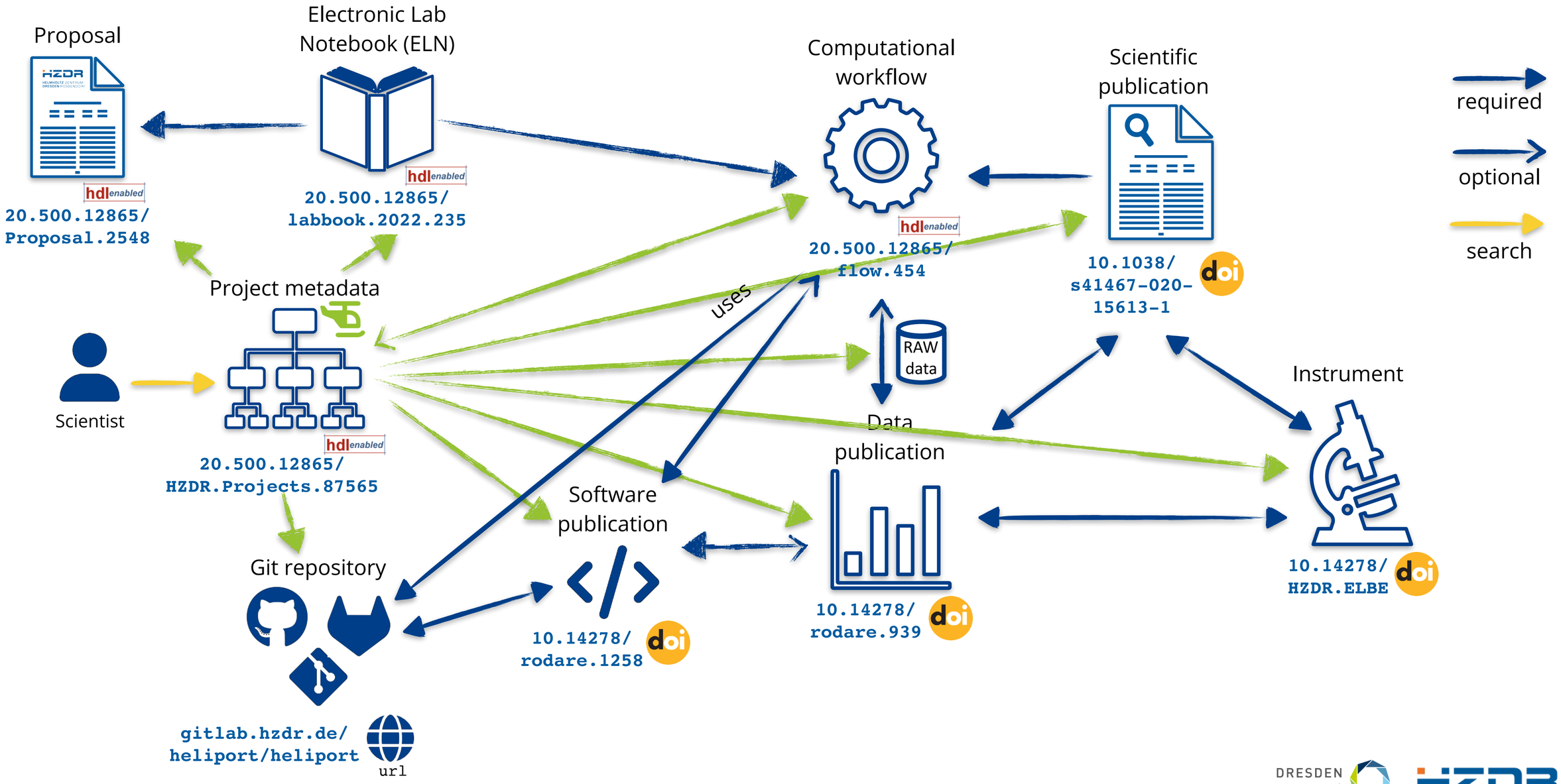
DRESDEN concept
HZDR

Top-Level View of the Interlinked Digital Objects of an Experiment at the HZDR



Including HELIPORT

Top-Level View of the Interlinked Digital Objects of an Experiment at the HZDR

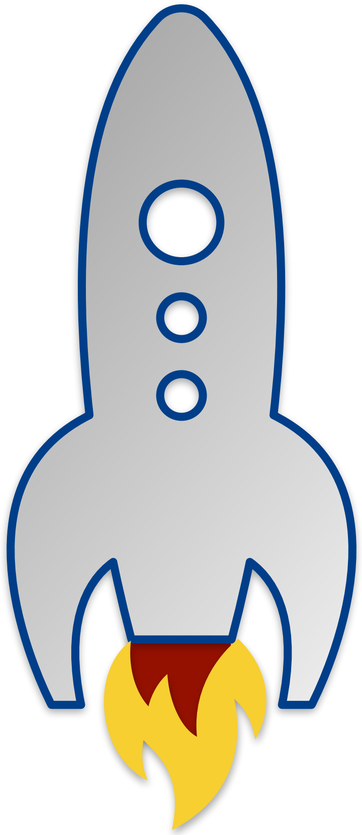


DRESDEN concept HZDR

Conclusions and Outlook

Conclusion:

- For an interlinked ecosystem, it is necessary to consider different entry points for the provision of metadata.
- The cross-linking of the services and systems is unavoidable to enable comprehensible science.
- ➡ Automated pipelines and workflows are the key to exchange metadata and support scientists and RSEs.



Status:

- We provide DOIs for software and data (instrument DOIs are work in progress),
- Handles can be created for all types of digital objects.
- ➡ With HELIPORT and HERMES, we develop systems that automate the exchange of metadata between internal and external systems and services.

HELIPORT
HELmholtz Scientific Project WORkflow PlaTform

RODARE
ROSSENDORF DATA REPOSITORY

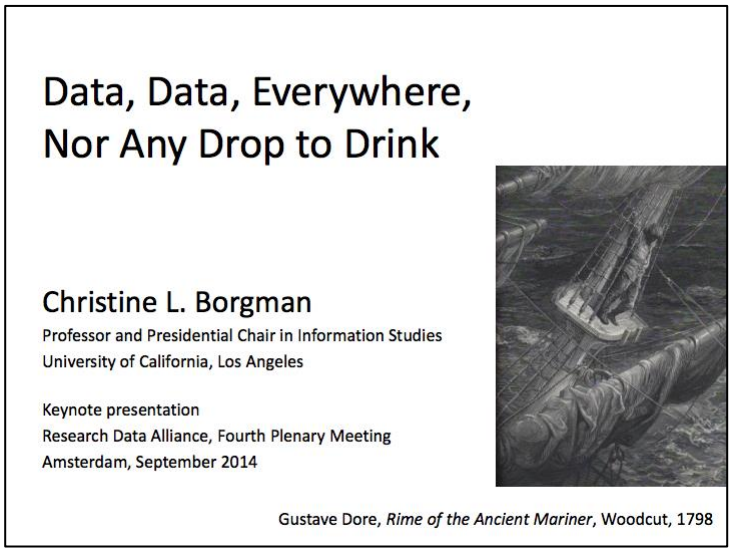
HERMES

FAIR Digital Objects
A gentle way of introducing FAIRness

Thomas Jejkal, et al.

Steinbuch Centre for Computing (SCC)

Motivation



- 481 entries for Germany (https://www.re3data.org/search?query=&countries%5B%5D=DEU, 2022)
  - **F**: DOI (218), hdl (37), URN (22), PURL (9), none (159)
  - **A**: REST (61), OAI-PMH (58), SOAP (11), SPARQL (6), FTP (27)
  - **I**: DataCite (92), DC (78) ISO 19115 (34), DDI (31), Custom (18)
  - **R**: License (huge majority), Provenance/Versioning (169), Quality management (275)

- What is inside?
- How many of these systems may a researcher access?
- How many of these systems are still actively maintained?

  → Repository software: 122 other, 188 unknown

Overarching commonality to make content available to researchers.

FAIR Digital Objects in the International Context

- Lots of **standardization and conceptual work** ongoing
- **Different implementation options** under discussion
  - PID-based, Linked Data-based
- Some **prototypical/demonstrator-like implementations** available
- 1st International FDO Conference (26.10. – 28.10.)

thomas.jejkal@kit.edu
Steinbuch Centre for Computing

FAIR Digital Objects for the Helmholtz Association

- Evaluate FAIR DOs as potential top-level commonality across all research fields
- Focus on PID-based implementation
- […]

21.T11981/6ab464ed-978b-4996-876f-f68ea913a308

Key	Value
KernelInformationProfile	HelmholtzKIPMultipleTypes
digitalObjectType	ScanningElectronMicroscopy
digitalObjectLocation	https://b2share.eudat.eu/api/files/[...]

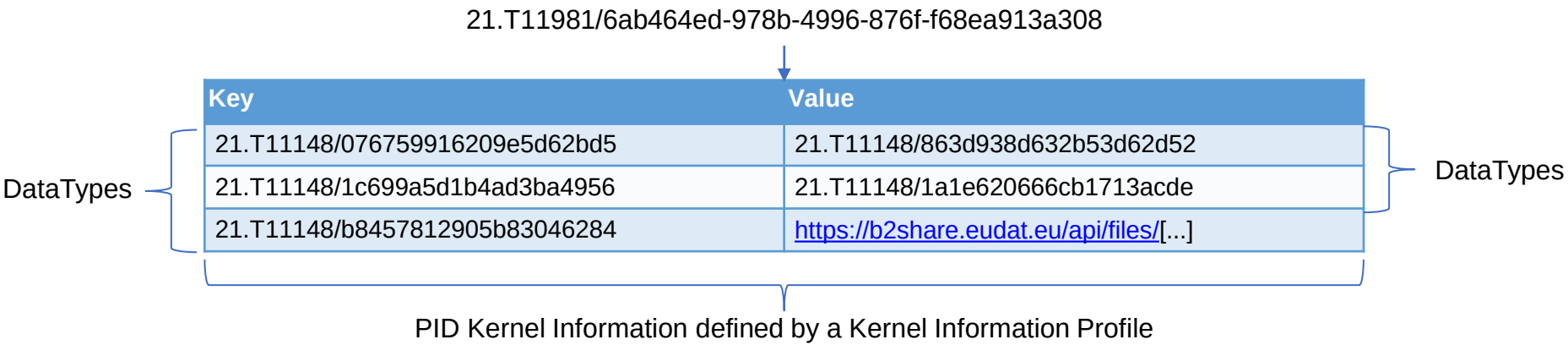
PID Kernel Information

thomas.jejkal@kit.edu

Steinbuch Centre for Computing

FAIR Digital Objects for the Helmholtz Association

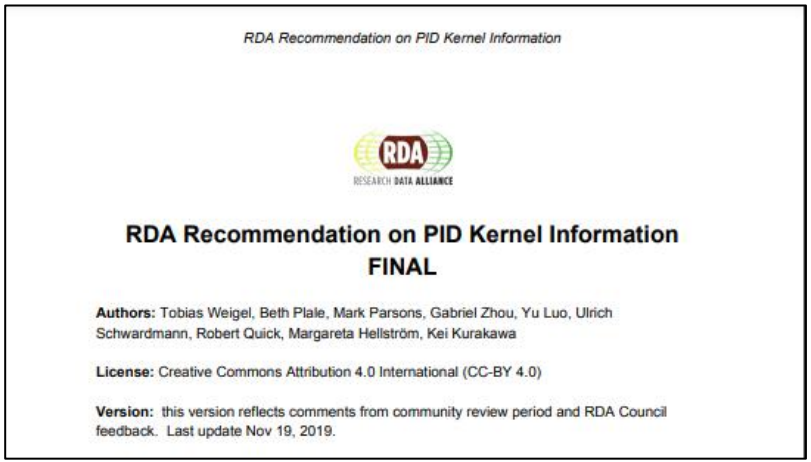
- Evaluate FAIR DOs as potential top-level commonality across all research fields
- Focus on PID-based implementation
- […]



thomas.jejkal@kit.edu
Steinbuch Centre for Computing

FAIR Digital Objects for the Helmholtz Association

- Evaluate FAIR DOs as potential top-level commonality across all research fields
- Focus on PID-based implementation
- Agree on common properties every Helmholtz FAIR DO must follow
- […]



RDA Recommendation on PID Kernel Information

RDA
RESEARCH DATA ALLIANCE

RDA Recommendation on PID Kernel Information
FINAL

Authors: Tobias Weigel, Beth Plale, Mark Parsons, Gabriel Zhou, Yu Luo, Ulrich Schwardmann, Robert Quick, Margareta Hellström, Kei Kurakawa

License: Creative Commons Attribution 4.0 International (CC-BY 4.0)

Version: this version reflects comments from community review period and RDA Council feedback. Last update Nov 19, 2019.

DOI: 10.15497/rda00031

- **Goal**: Facilitate fast decision making by machines
- **RDA Draft Kernel Information Profile** defined
- Contains **15 basic attributes**, e.g., location, creation date
- Mostly **administrative information**

thomas.jejkal@kit.edu

Steinbuch Centre for Computing

The Helmholtz Kernel Information Profile

- **Extension** of Draft KIP **by contextual and relational attributes** agreed on between representatives from all research fields
- **Goal**: **Increase immediate (scientific) benefit** of using FAIR DOs
- **Compatible** to RDA Recommendations
- **Basis for all FAIR DOs** created within the Helmholtz Association
- **Extensible** by additional attributes if required
- **Guidance document available**, publication soon

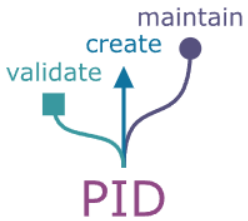
Additional Helmholtz KIP Attributes	Comment
digitalObjectLocation-AccessProtocol	Access information for digitalObjectLocation, e.g., protocol, protocol version, and client
underEmbargoUntil	Access restrictions probably apply before
license	Extracted from digitalObjectPolicy
checksum	Renamed from ‚etag‘ to be more specific
signature	Cryptographic signature of PID record
topic	Topic term from vocabulary for additional context
locationPreview	Optional preview for digitalObjectLocation
contact	Contact information, e.g., ORCiD or ROR
hasMetadata	PID pointing to a related FDO containing metadata
isMetadataFor	Inversion for hasMetadata
wasGeneratedBy	W3C PROV-DM element to refer to tool/agent used for generating the digital object
provenanceGraph	Optional PID of full provenance graph

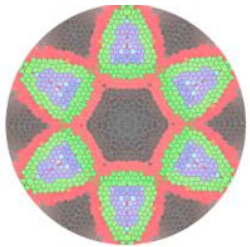
thomas.jejkal@kit.edu

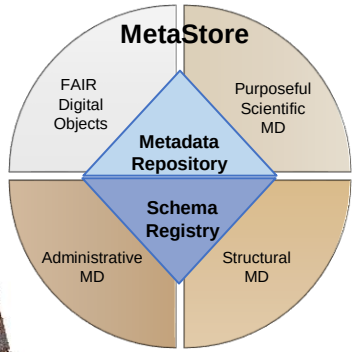
Steinbuch Centre for Computing

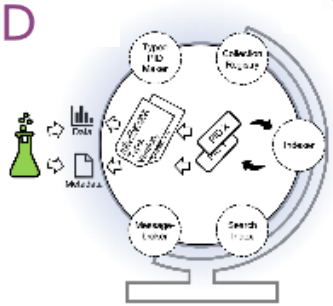
FAIR Data Commons – Tools and Services

- Evaluate FAIR DOs as potential top-level commonality across all research fields
- Focus on PID-based implementation
- Agree on common properties every Helmholtz FAIR DO must follow
- Work on filling gaps in existing landscape to realize FAIR DOs for the Helmholtz Association







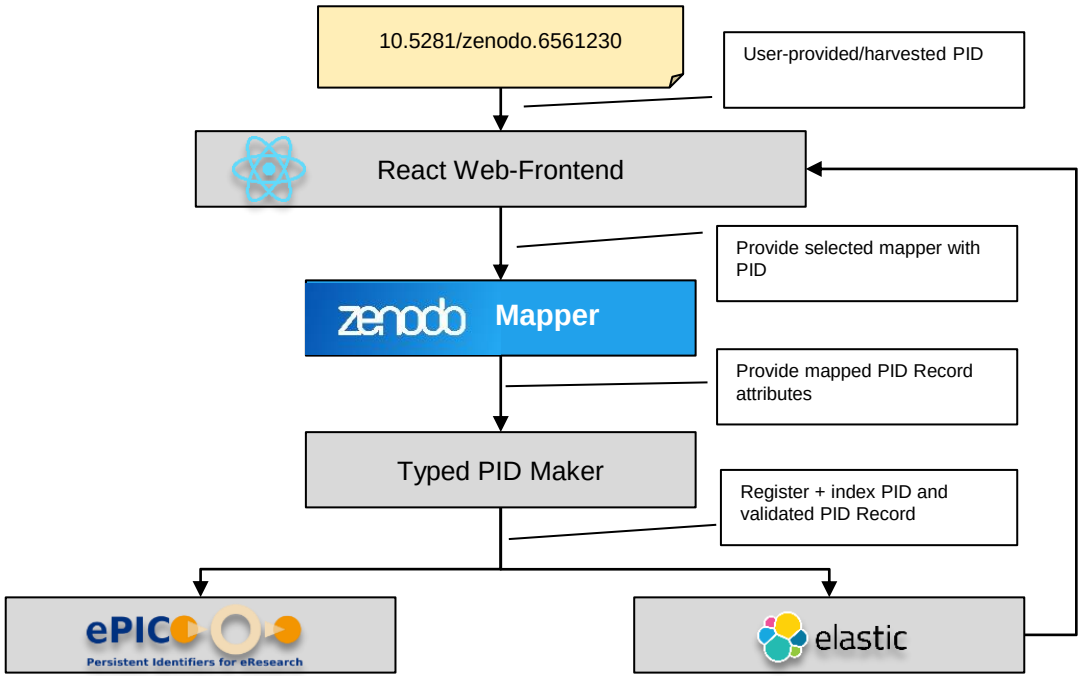


https://github.com/kit-data-manager

thomas.jejkal@kit.edu

Steinbuch Centre for Computing

A Helmholtz Kernel Information Profile Demonstrator



- Showcase implementation for **evaluating applicability** for existing repository (Zenodo)
- **Blueprint** for extension to additional repository platforms
- Basis for constantly **growing collection of FAIR DOs**

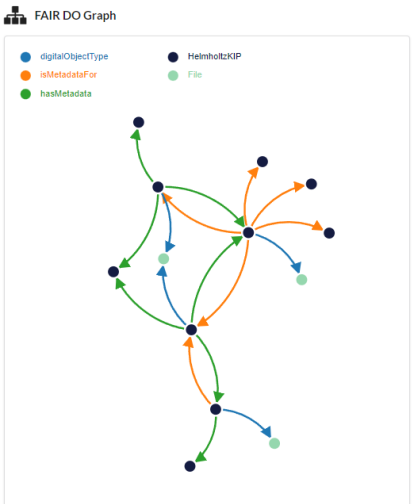
thomas.jejkal@kit.edu

Steinbuch Centre for Computing

Conclusions and Outlook

- Evaluated FAIR DOs as potential top-level commonality across all research fields
- Agreed on common properties every Helmholtz FAIR DO must follow
  - Extension of RDA Draft Kernel Information Profile by (mostly optional) contextual attributes
- Filled gaps in existing landscape to realize FAIR DOs for the Helmholtz Association
  - Implemented first version as demonstrator for mapping digital assets from Zenodo

- Building a growing collection of FAIR DOs
- Elaborate possibilities for further automation
- Integration into scientific application cases
- Identify and realize further applications based on FDOs



thomas.jejkal@kit.edu
Steinbuch Centre for Computing

HZB Helmholtz Zentrum Berlin

RDM STRUCTURES @ HZB

Heike Görzig, Michael Götte, Britta Höpfner, Tamara Husch, Rolf Krahl,
Marcus Lewerenz, Hector Perez Ponce

https://os.helmholtz.de/en/events/fora/2nd-practice-forum-research-data-management/

2nd Practice Forum Research Data Management

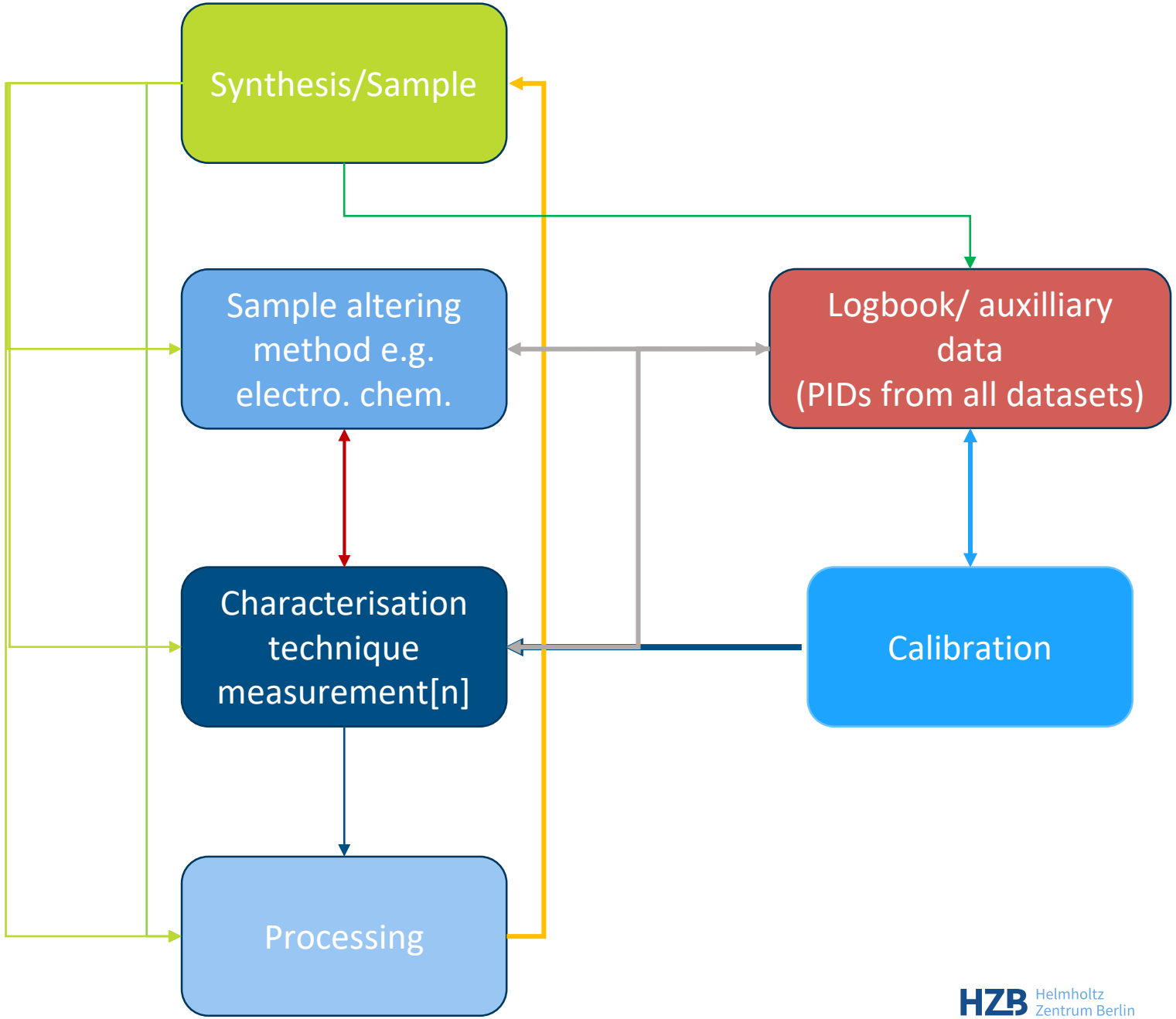
20.10.2022

THE HZB

Research Areas:
Energy, Information, Matter
Data from e.g.:

- device/instrument design and construction
- sample synthesis
- sample characterisation
- measurements at the BESSY II synchrotron radiation source
- data analysis
- long term solar cell, batteries etc. measurements
- simulations

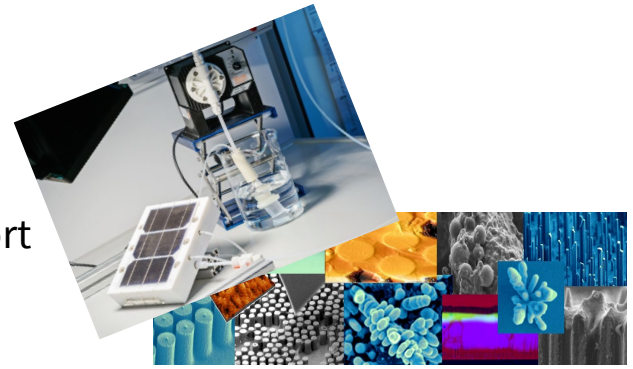
- Focus on energy materials
- User facility for any kind of sample



RDM group Tools and Services

Tools and Services

- Data repository / catalogue
- Repository ingest support
- (Meta)data standardisation coordination/support
- ELNs
- PIDs/DOIs
- Consulting

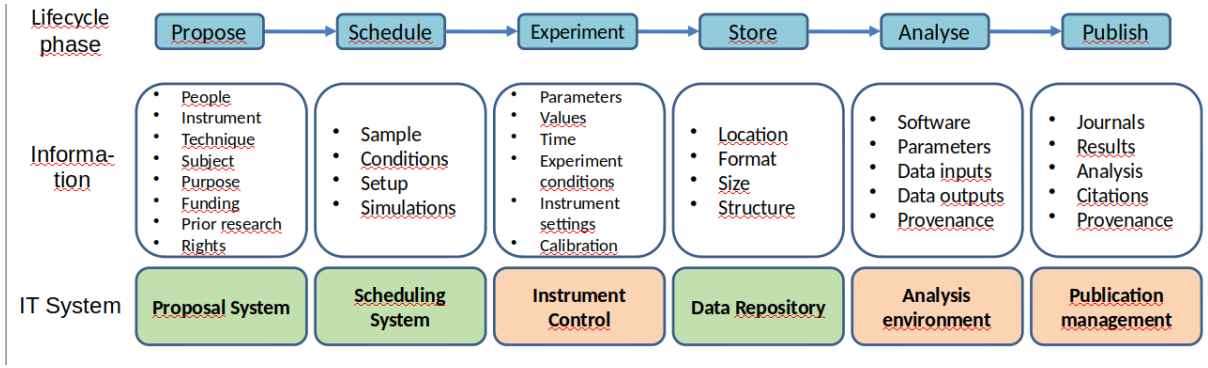


Service requests

- OpenAccess
- DMP support
- FAIR data
- Digital twins

Other ground work

- Discussions with groups outside the HZB about FAIR data requirements
- Organising scientists around same requirements
- Metadata modelling for
  - files
  - data catalogue
  - PIDs
- Organising storage
- Authentication & authorization
- Data policy



Source: https://doi.org/10.5281/zenodo.5636096

RDM group in IT department and support

Hector Perez Ponce (IT)
Writing files at the instruments (converter) and ingest support
Since 2022

Marcus Lewerenz (IT)
DAPHNE4NFDI (Repository / ELN)
Since 2022

Heike Görzig (IT)
Holding everything together / Metadata standards (Group leader RDM in IT)
Since 2017

Rolf Krahl (IT)
Repository and PIDs
Since 2013

Michael Götte (Energy)
Coordination research area Energy (Data steward)
Interface to scientist
Since 2022

NN (BESSY II/Matter)
Coordination research area Matter (Data steward)
Interface to scientist

(INITIAL) PROBLEMS

- Central RDM was not considered necessary as researchers took their data home for analysis
- In the past HZB did not provide much centralised IT infrastructure to support scientific workflows
- Scientist are used to help themselves and have low expectations on central IT services
- RDM is additional workload for scientists
- RDM team needs use cases and friendly / interested users to develop demonstrators
- Scientists need to see a benefit for their scientific work in order to dedicate time to RDM

Big gap between RDM team and scientists

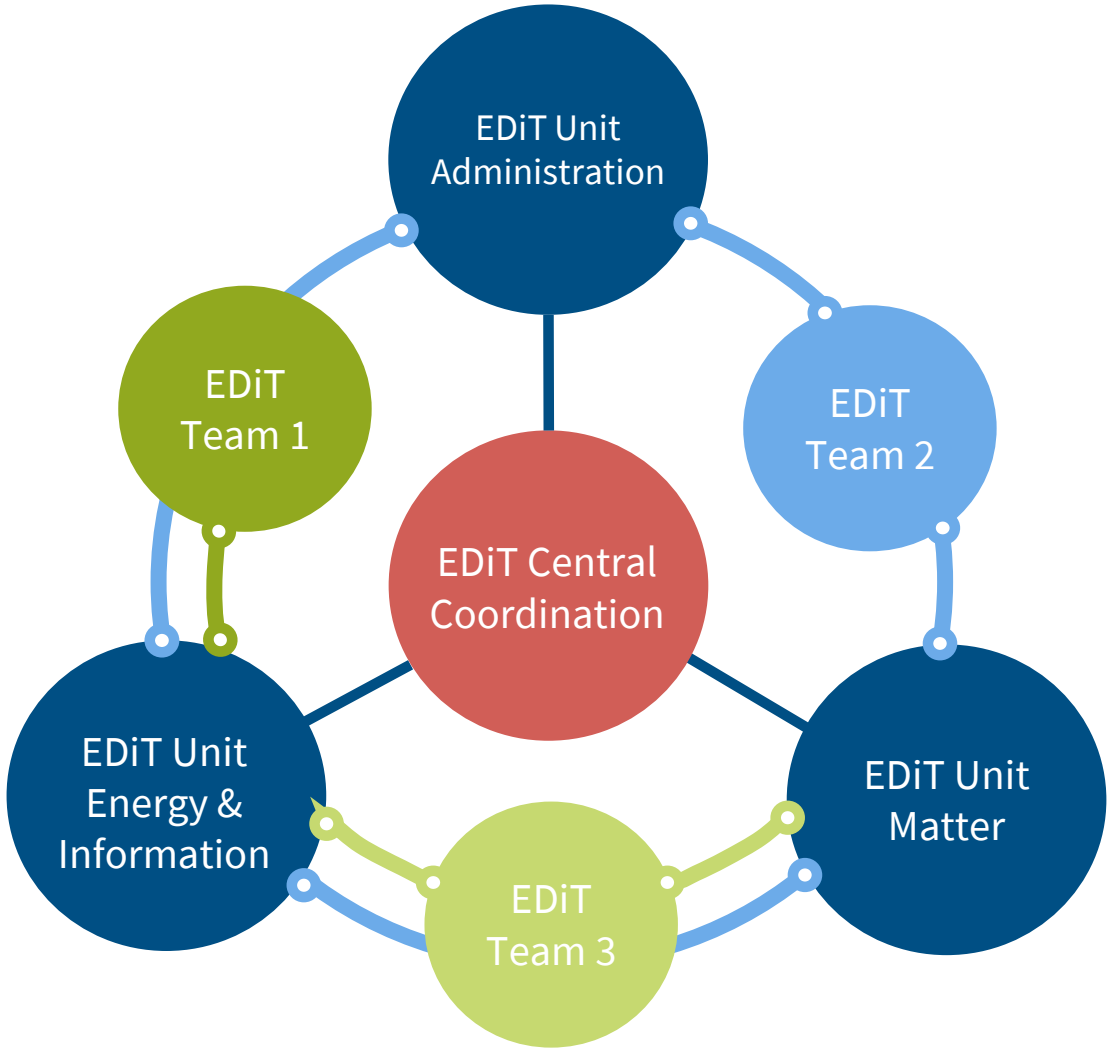
SINCE EARLY 2022
RDM STRUCTURES @ HZB TODAY

- **Initiative from board of directors**
- **New organisational structure:**

  RDM and scientists work in teams on RDM topics

  Support requirements can be formulated by the scientists AND the IT RDM team
- **Three new positions connecting scientists and IT**
- **Teams**

  ELN, Metadata standards, (DMPs?)



RDM group in IT department and support

Hector Perez Ponce (IT) Writing files at the instruments (converter) and ingest support Since 2022  **Marcus Lewerenz (IT)** DAPHNE4NFDI (Repository / ELN) Since 2022	**Heike Görzig (IT)** Holding everything together / Metadata standards (Group leader RDM in IT) Since 2017  **Rolf Krahl (IT)** Repository and PIDs Since 2013 	**Michael Götte (Energy)** Coordination research area Energy (Data steward) Interface to scientist Since 2022 **NN (BESSY II/Matter)** Coordination research area Matter (Data steward) Interface to scientist

EXAMPLES

- **RIXS (organising scientists around requirement)**
  - Standardisation for scientific technic
  - Organize scientists, support with approach and infrastructure
- **ICAT connect**
  - Convert instrument data into standardized format
  - Workflow for data ingest into repository
- **ELNs**
  - Provide basic ELN
  - Integration into infrastructure

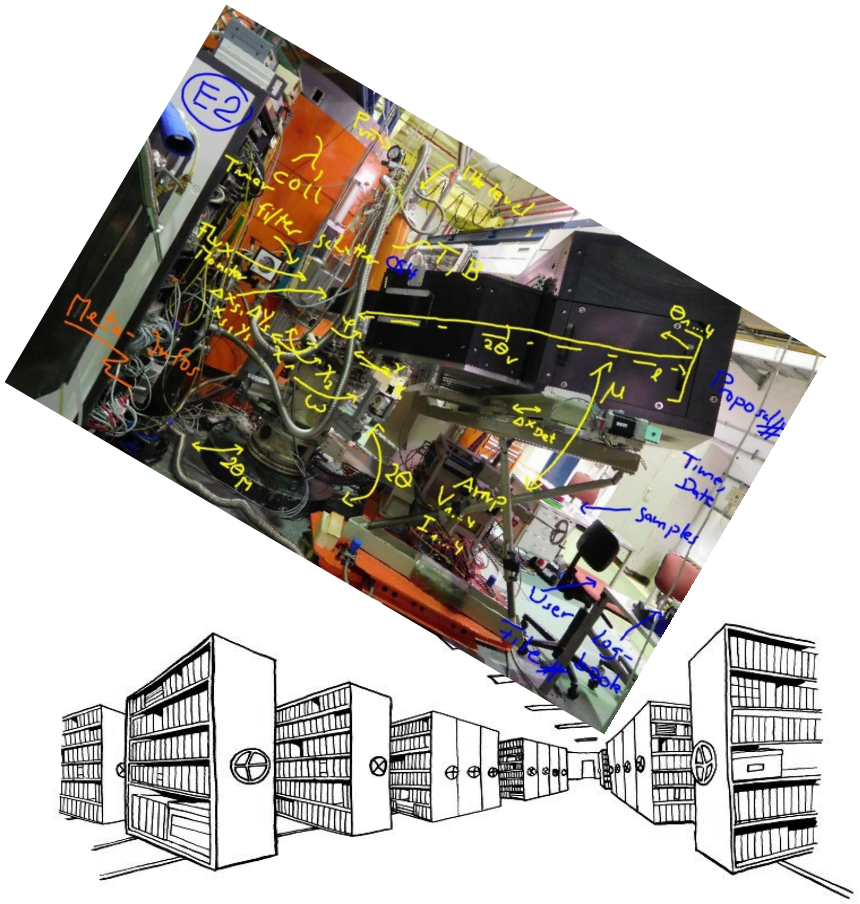
- **Standard format in Photon and Neutron Sciences is NeXus.**
- **Internal and external discussions about requirements for data analysis in NeXus.**
- **www.nexusformat.org**

- **Mapping of instrument variables/metadata to NeXus.**
- **Reading from different file formats writing in NeXus/HDF5.**
- **Transferring data to repository.**

- **Gather experiences (ELN users and IT)**
- **Formulate basic functionalities, configurations, and additional requirements.**

CONCLUSION

- **Requirement analysis**
  - Inside HZB -> Data stewards/(RDM group (IT))
  - Outside world -> RDM group (IT)
  - Strong coordination between the groups
- **Software**
  - Fast prototypes -> Data stewards (for now)
  - Stable and maintained -> RDM group (IT)
- **Metadata / Formats**
  - Usability for scientists -> Data stewards
  - Integration in infrastructure -> RDM group (IT)
- **Knowledge transfer**
  - all



Source: https://www.badv.bund.de/DE/LeichteSprache/OffeneVermoegensfragen/Archive/start.html

ENDE

CONTACT

- **Heike Görzig – heike.goerzig@helmholtz-berlin.de**

ROLES IN RDM AT THE FORSCHUNGSZENTRUM JÜLICH

FROM RDM TO OPEN SCIENCE

OCTOBER 20, 2022 I INES SCHMAHL

ROLES IN RDM AT THE FZ JÜLICH

Organisation

Main research areas:

Energy	• Renewable energy • Climate research
Information	• Neurosciences • Supercomputing
Bioeconomy	• Biotechnology • Plant sciences

Organisation:

11 Institutes

80 Institute departments

ROLES IN RDM AT THE FZ JÜLICH

Starting point

Situation:

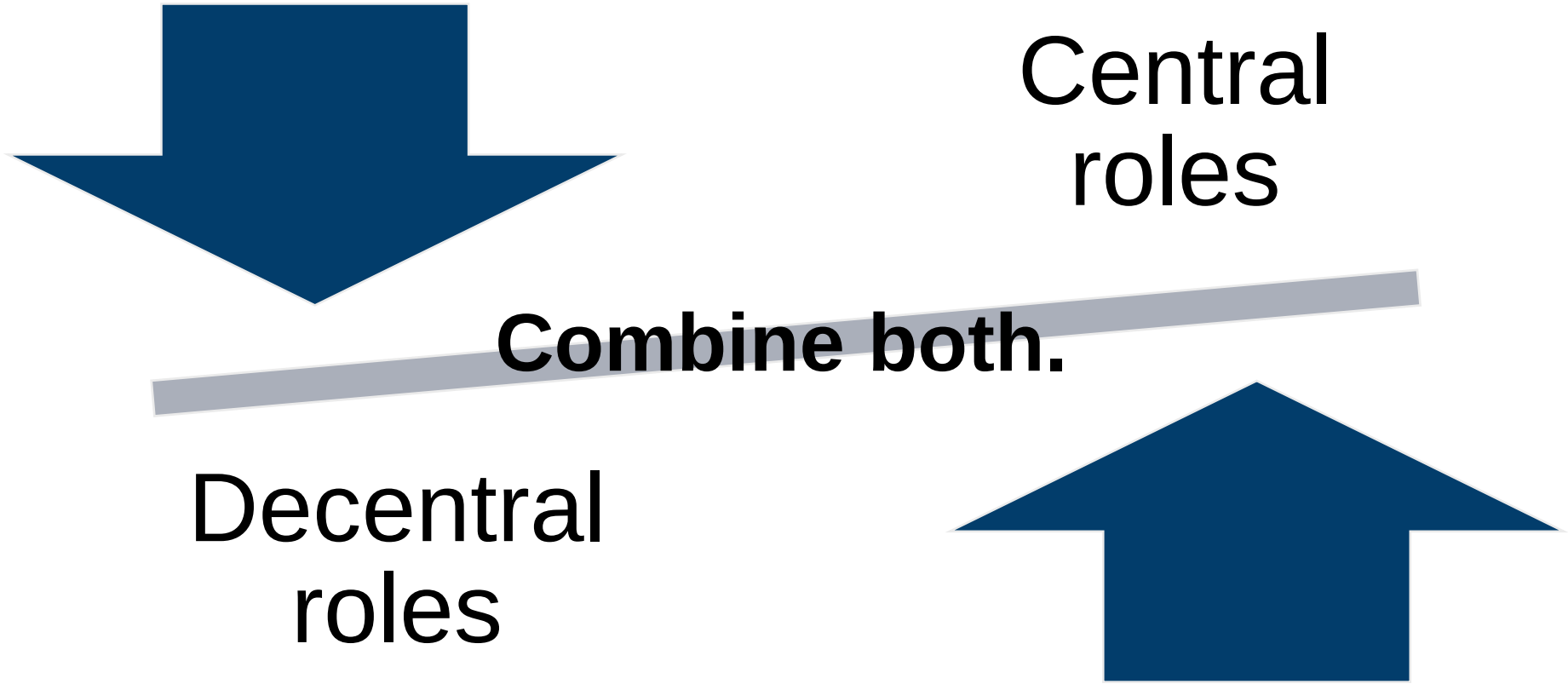
- Very heterogenous research topics
- Interdisciplinary research groups
- Cooperations with external partners

- Diverse workflows
- Particular standards and approaches in communities
- Different research funders with distinctive requirements

How can roles and responsibilities in RDM at the FZ Jülich be organised to fit all needs?

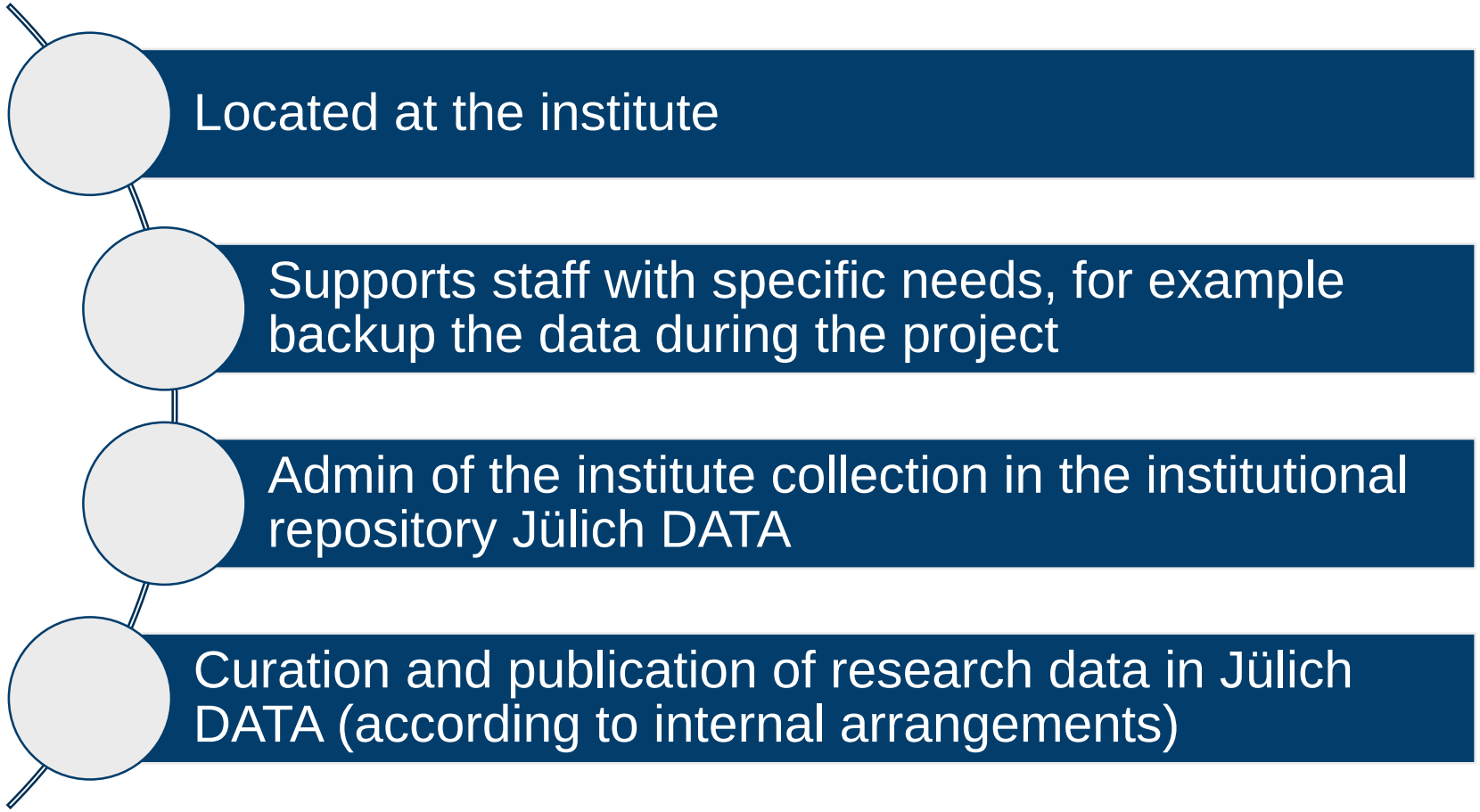
ROLES IN RDM AT THE FZ JÜLICH

Approach



ROLES IN RDM AT THE FZ JÜLICH

Decentral Role – DDM



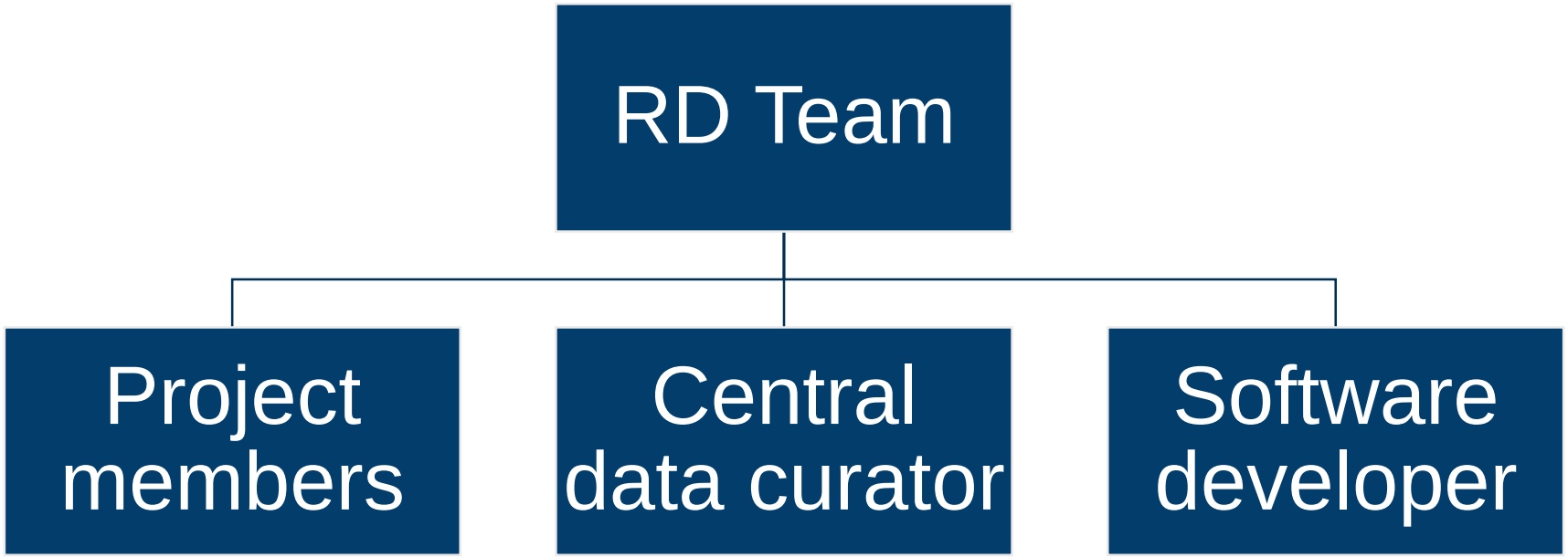
Tasks of the DDM (decentral data manager)

JÜLICH Forschungszentrum

ROLES IN RDM AT THE FZ JÜLICH

Central Role – RD Team

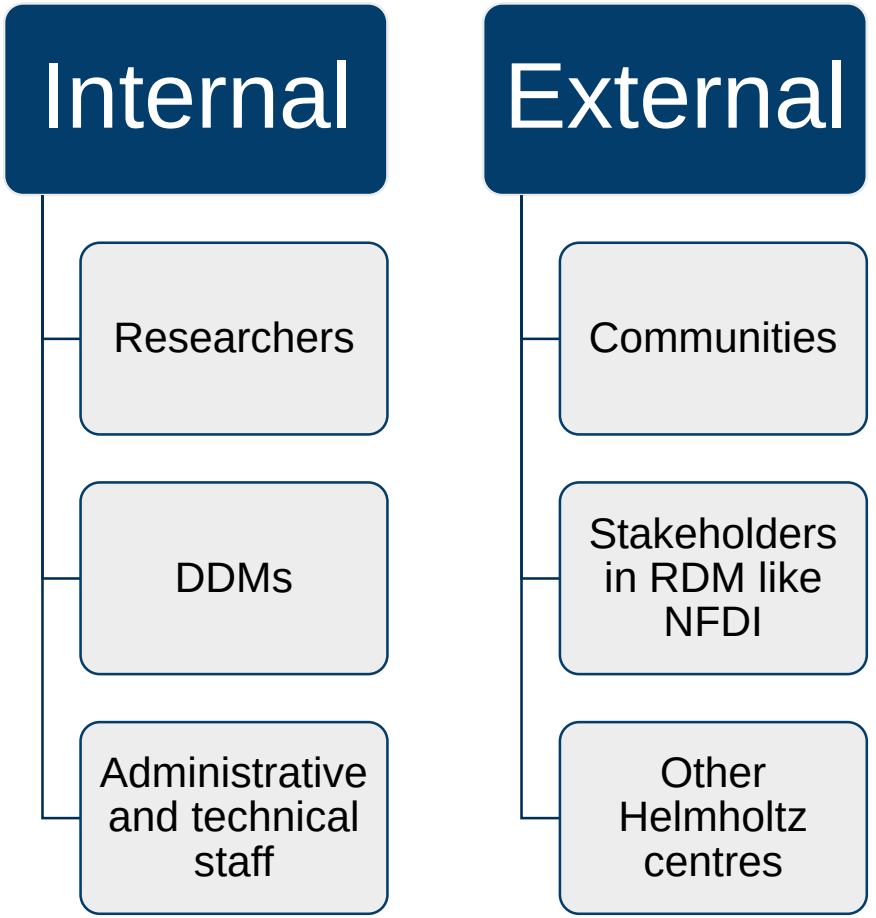
Positions in the team



ROLES IN RDM AT THE FZ JÜLICH

Central Role – RD Team

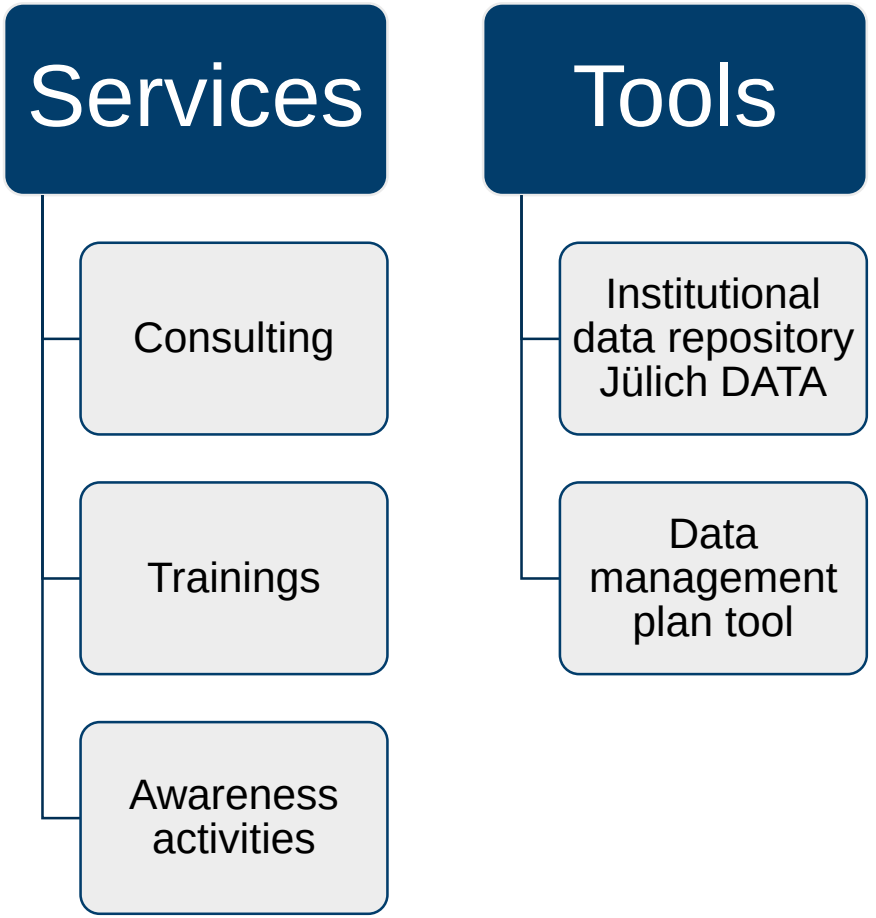
Central contact point



ROLES IN RDM AT THE FZ JÜLICH

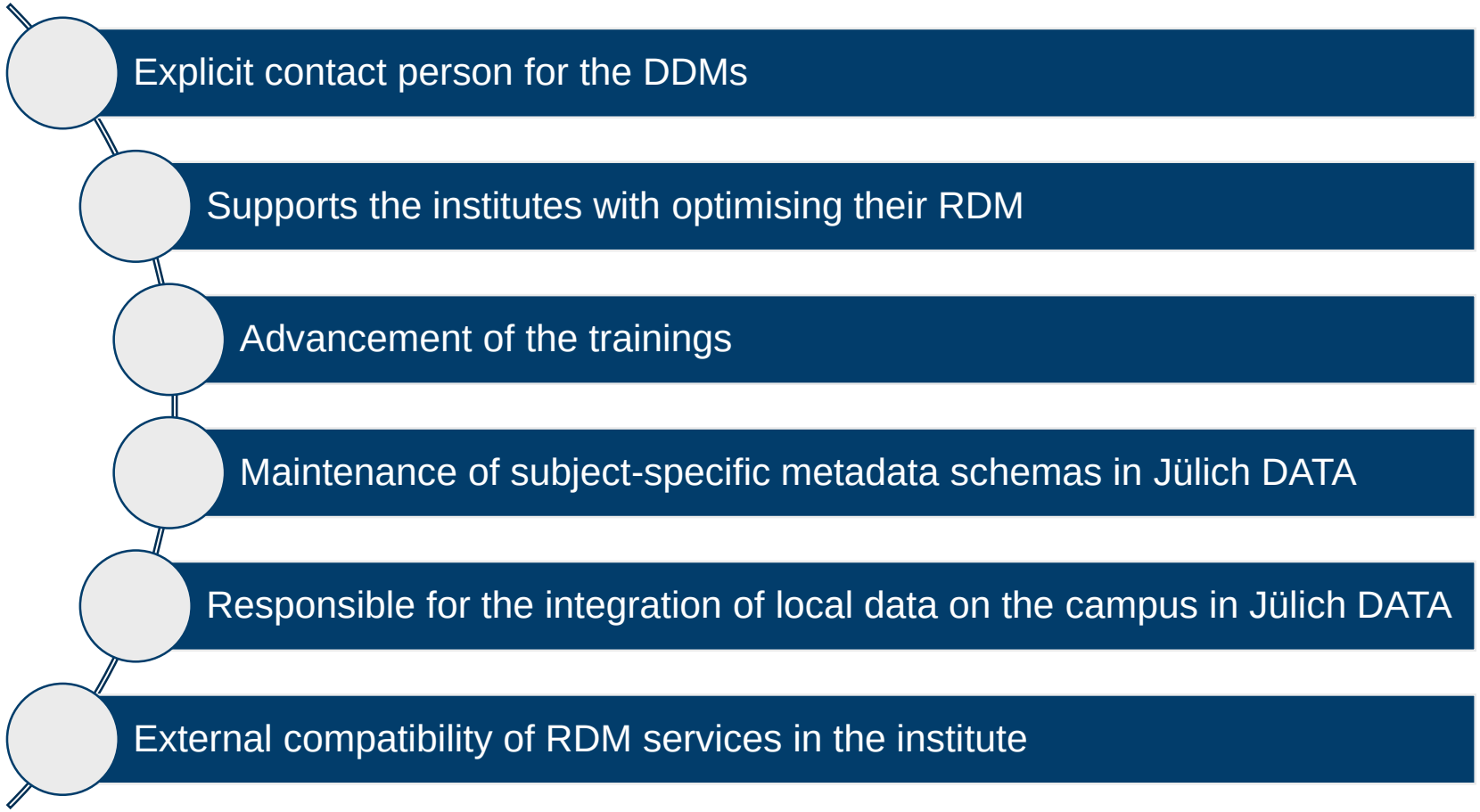
Central Role – RD Team

Operation of central tools and services



ROLES IN RDM AT THE FZ JÜLICH

Central Role – Central Data Curator



Tasks of the central data curator

JÜLICH Forschungszentrum

ROLES IN RDM AT THE FZ JÜLICH

Guidelines for Handling Research Data

Roles of the

- Central data curator
- DDMs

are laid down in the guidelines.

> 4. Responsibilites and roles
>
> "Each OU establishes the role of a data manager as the contact person for research data management in the OU. …
> Forschungszentrum Jülich will further establish the position of central data curator."
> (FZ Jülich (2019). Guidelines for Handling Research Data. https://go.fzj.de/data-policy.)

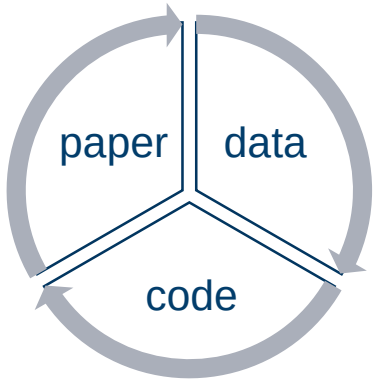
ROLES IN RDM AT THE FZ JÜLICH

From RDM to Open Science

Roles in RDM established

But managing and publishing data is not enough to ensure reproducibility of research.

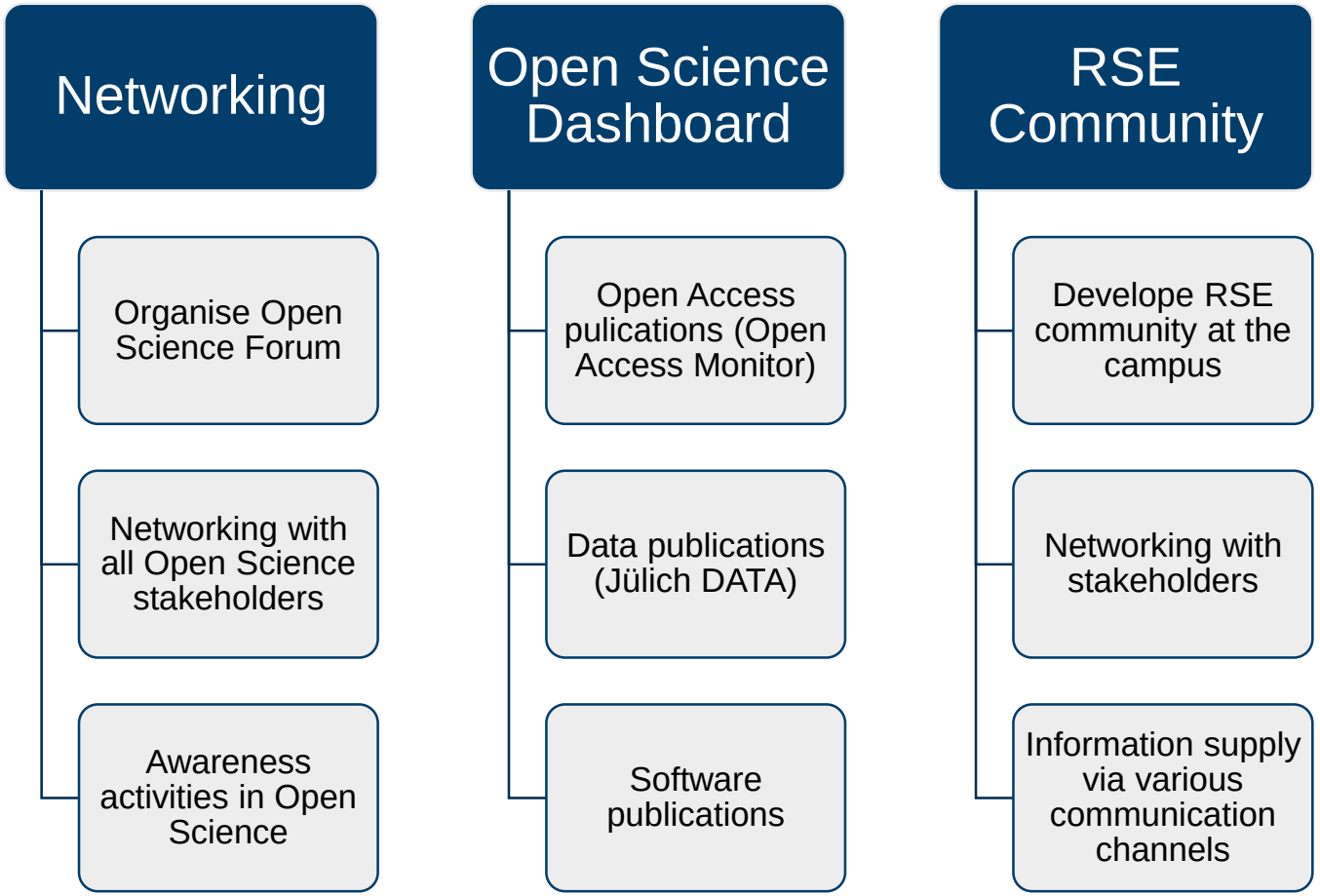
Roles must be enlarged to Open Science.



ROLES IN RDM AT THE FZ JÜLICH

From RDM to Open Science

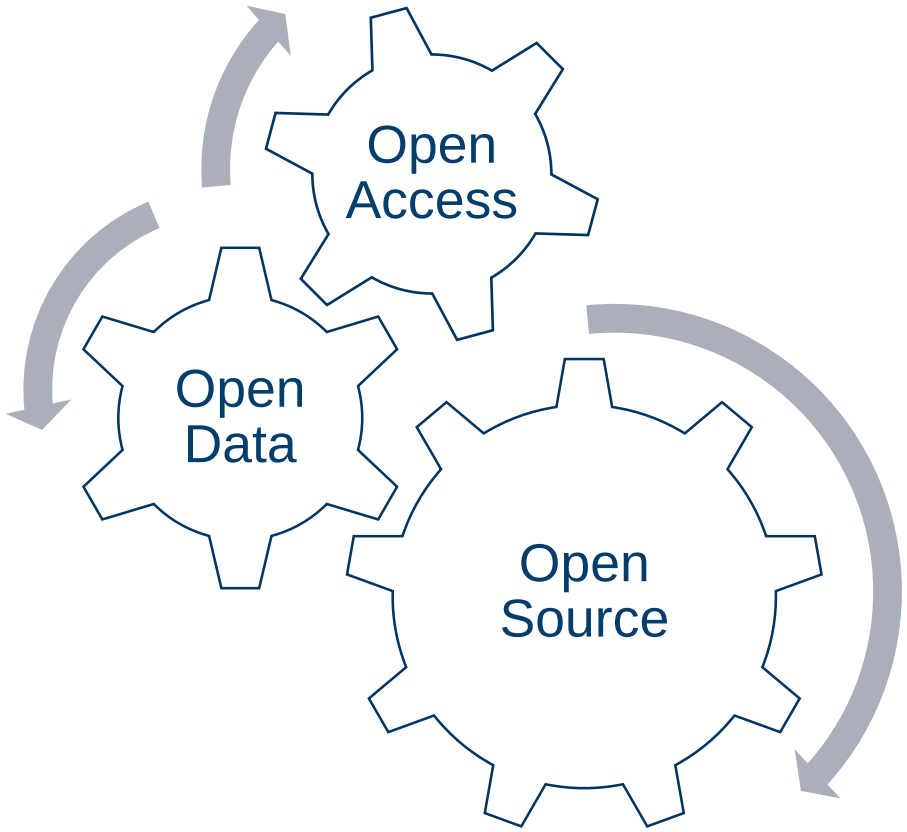
New roles



JÜLICH Forschungszentrum

ROLES IN RDM AT THE FZ JÜLICH

From RDM to Open Science



The goal is to bring the 3 topics together.

Researchers should get an overall support.

JÜLICH Forschungszentrum

THANK YOU!

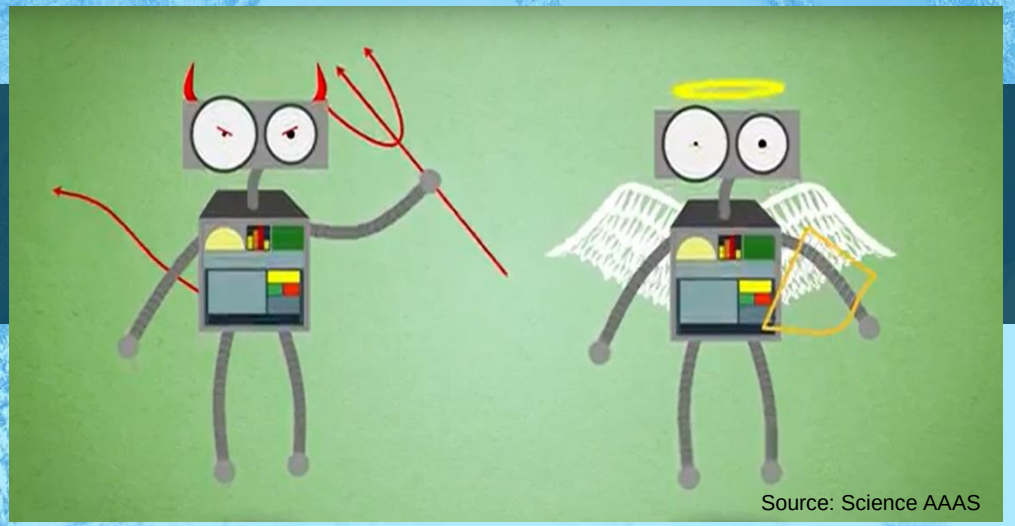
Ines Schmahl

Central Library, Forschungszentrum Jülich

forschungsdaten@fz-juelich.de

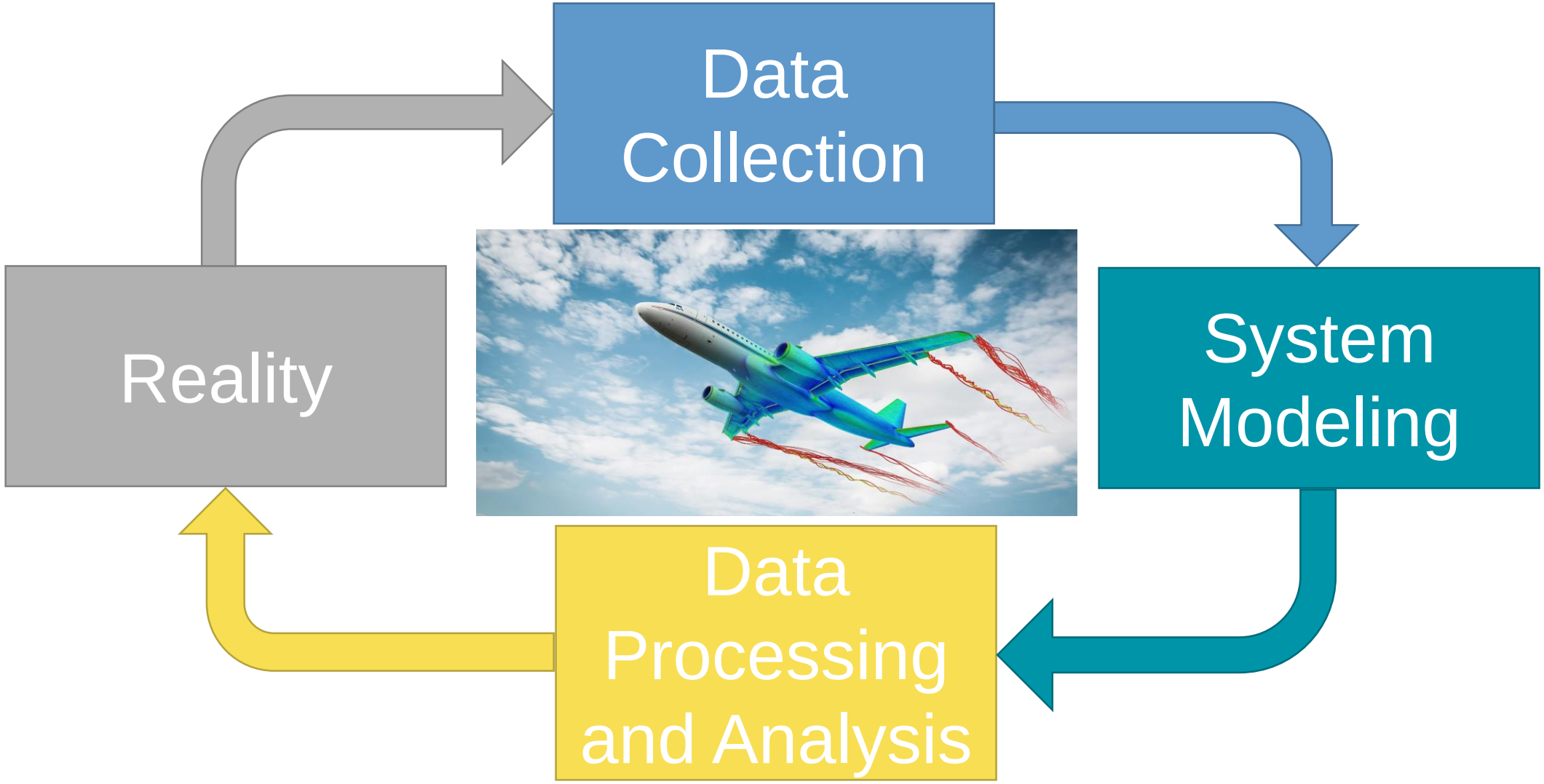
DATA MANAGEMENT PLANS

Angel or Devil for the Researcher Work

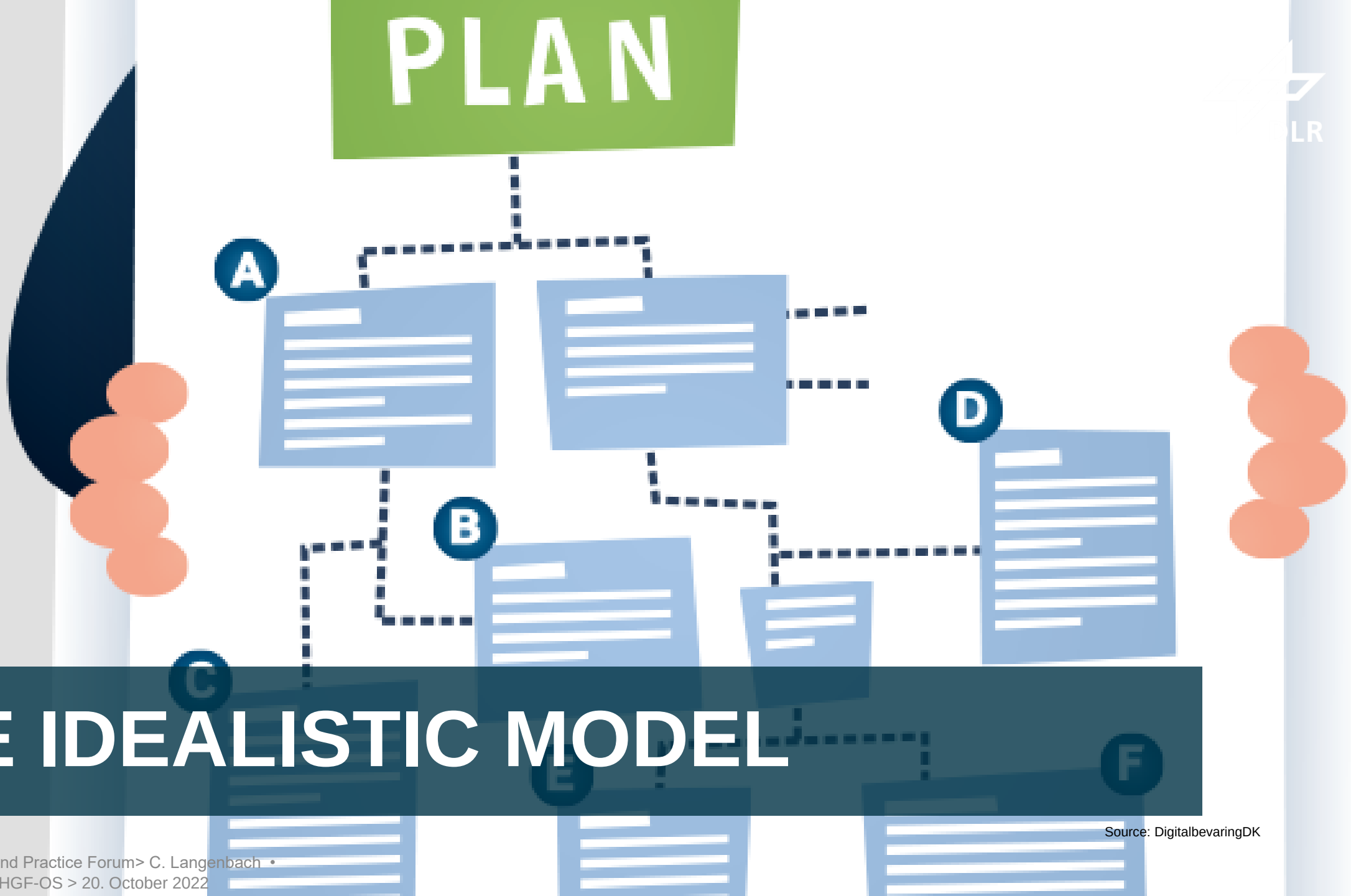


Source: Science AAAS

The digital twin example



Source: DLR

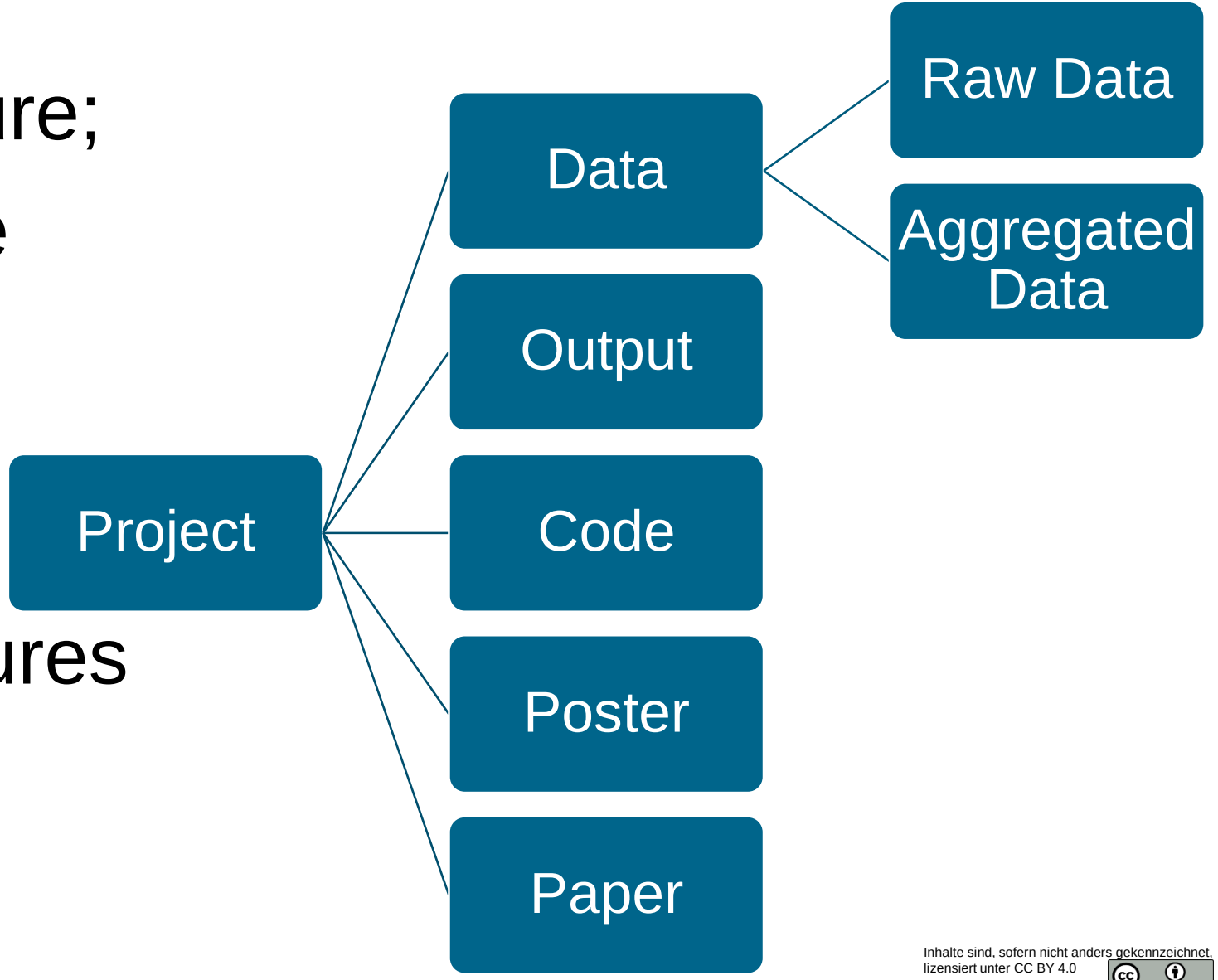


THE IDEALISTIC MODEL

Source: DigitalbevaringDK

The idealistic Model – Data collection and processing

- Clear folder structure;
- Meaning should be clearly evident;
- For me, also for other researchers;
- Hierarchical structures facilitate finding.



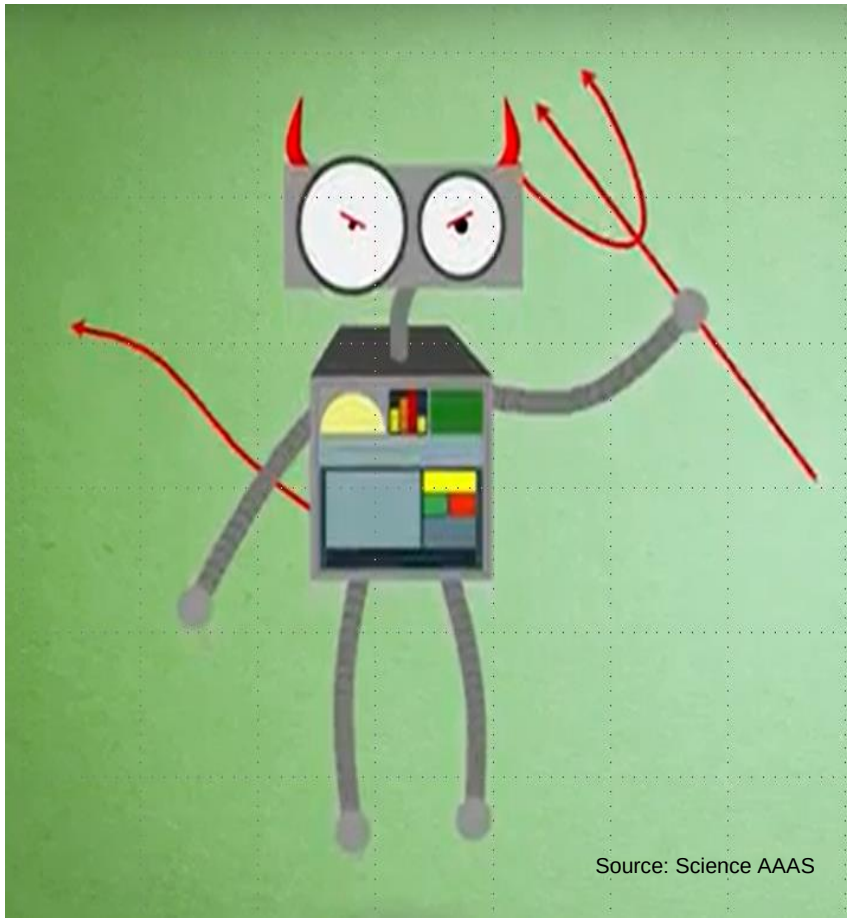
The idealistic Model – Data Management Plan?

I do not need to establish an DMP because…

- …I have a clear and easy folder structure;
- …it is a not data intensive project;
- …I keep track of where I am and where I am going;
- …I have a fantastic document management tool;
- …

The data cannot be publicly shared because…

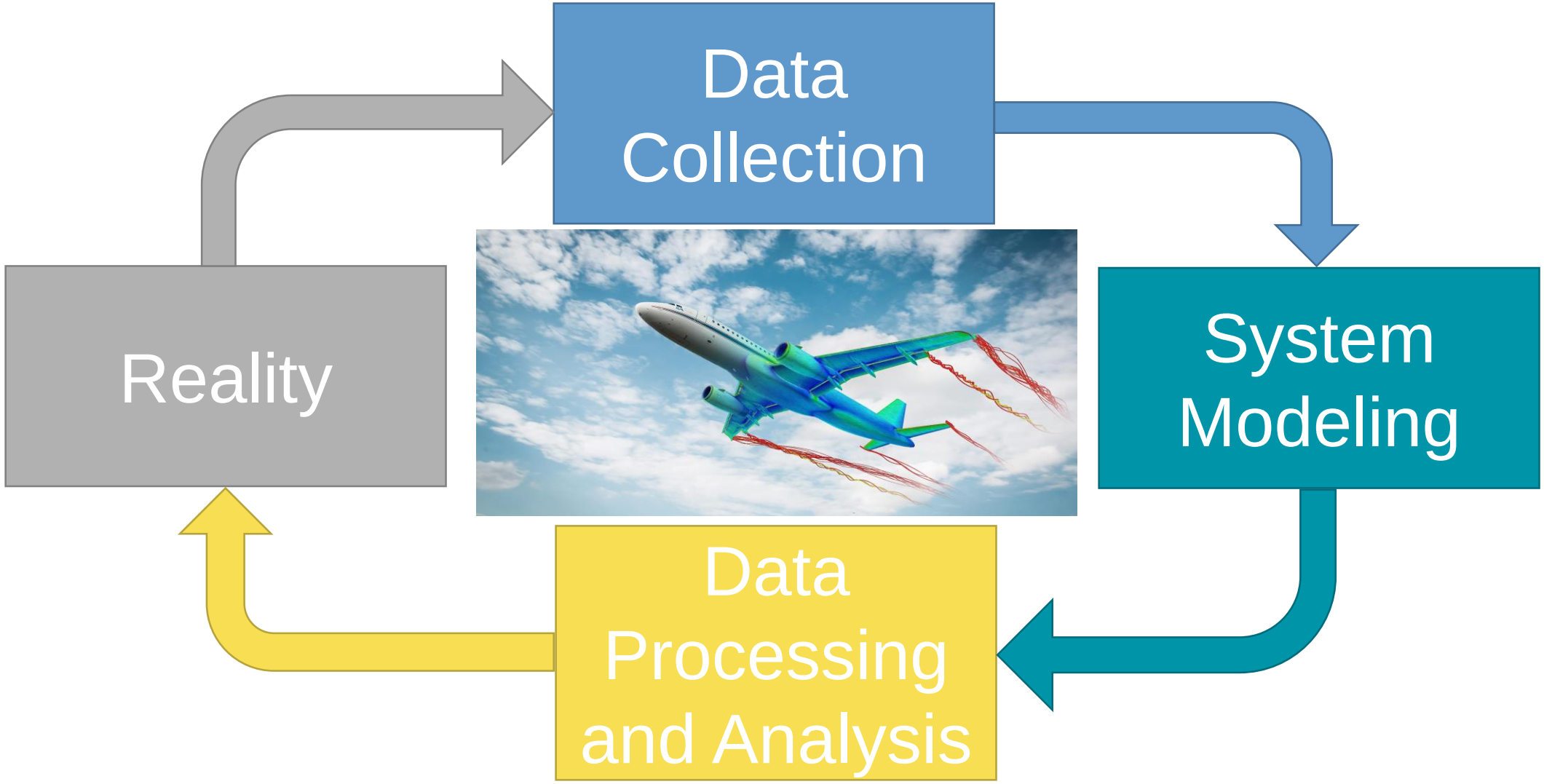
- ...it may be contains potentially identifying information of xxx;
- …it may be planned patents related to this research in the near future;
- …


Source: Science AAAS

THE REAL WORLD

Source: DigitalbevaringDK

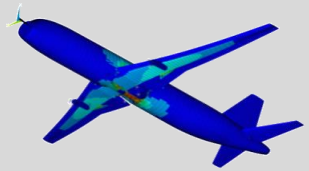
The real world – The digital twin example

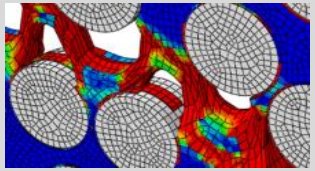


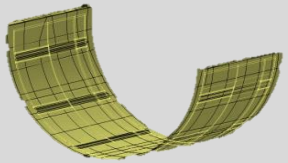
Source: DLR

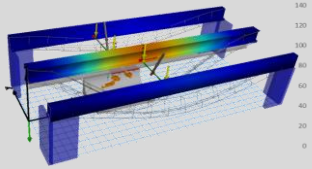
The real world – The system principle
Distributed disciplinary responsibilities and repositories

DISCIPLINARY-Sources


Design Data

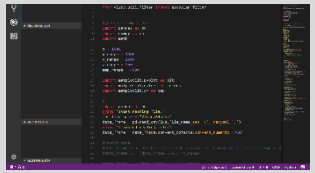

Material Data


CAD Models


Numeric Models


Mesurement Data


Reports / Publications


Source Code


Media

?
u.v.m.


External Storage


Network Drive


Local Storage


Intranet


Paper

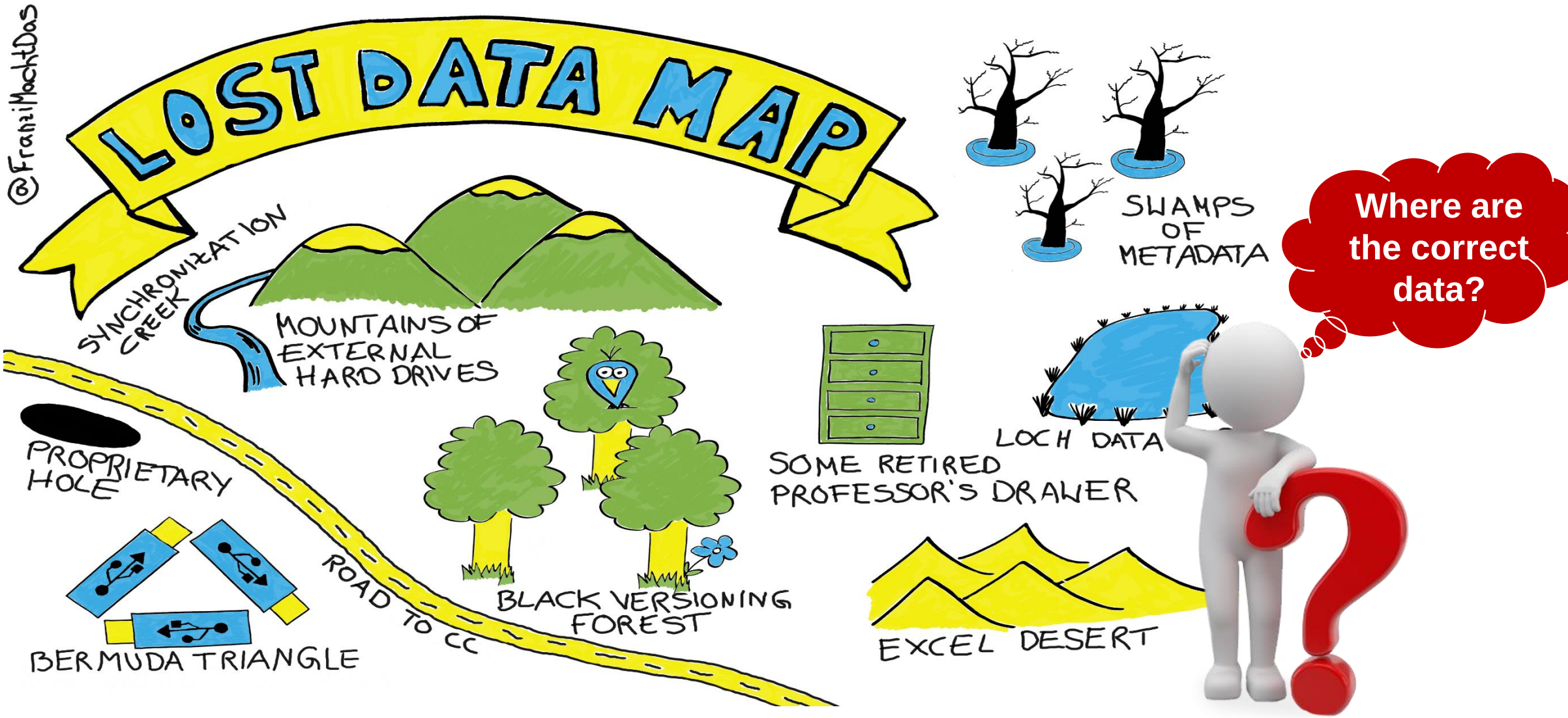

SCM Storage Class Memory


Employees

DATA STORAGE - Locations

Source: DLR

The real world –
Welcome at the lost data map! Where are your data?



Source: FranziMachtDas; fotolia

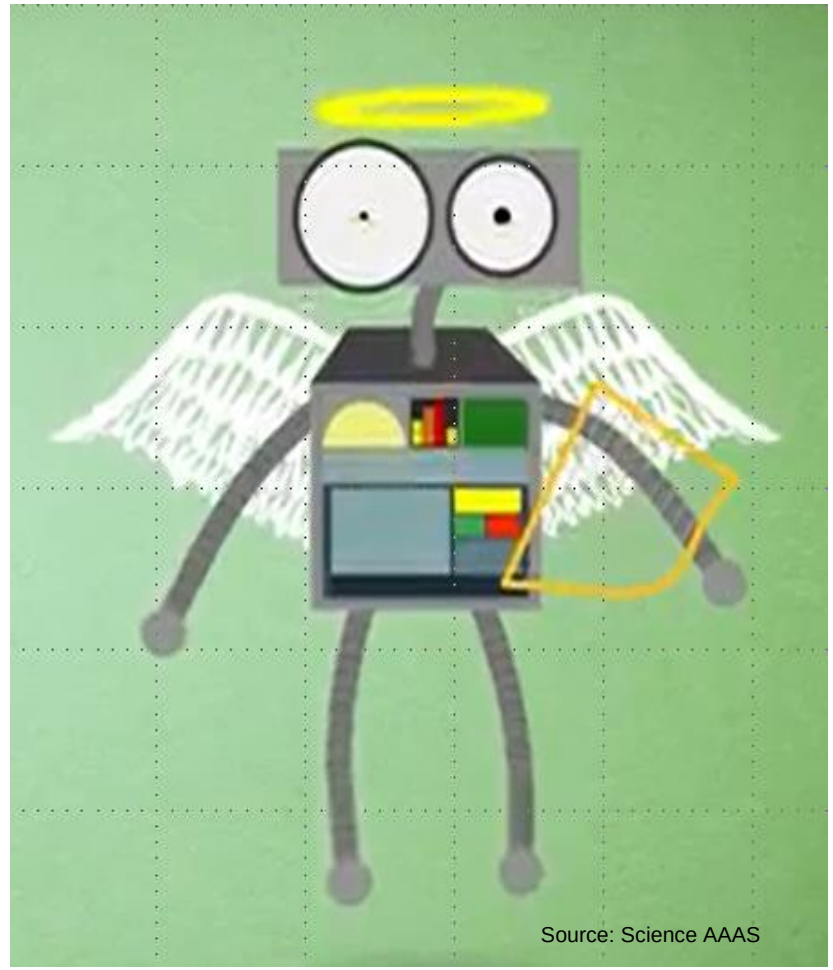
The real world – Data Management Plan!

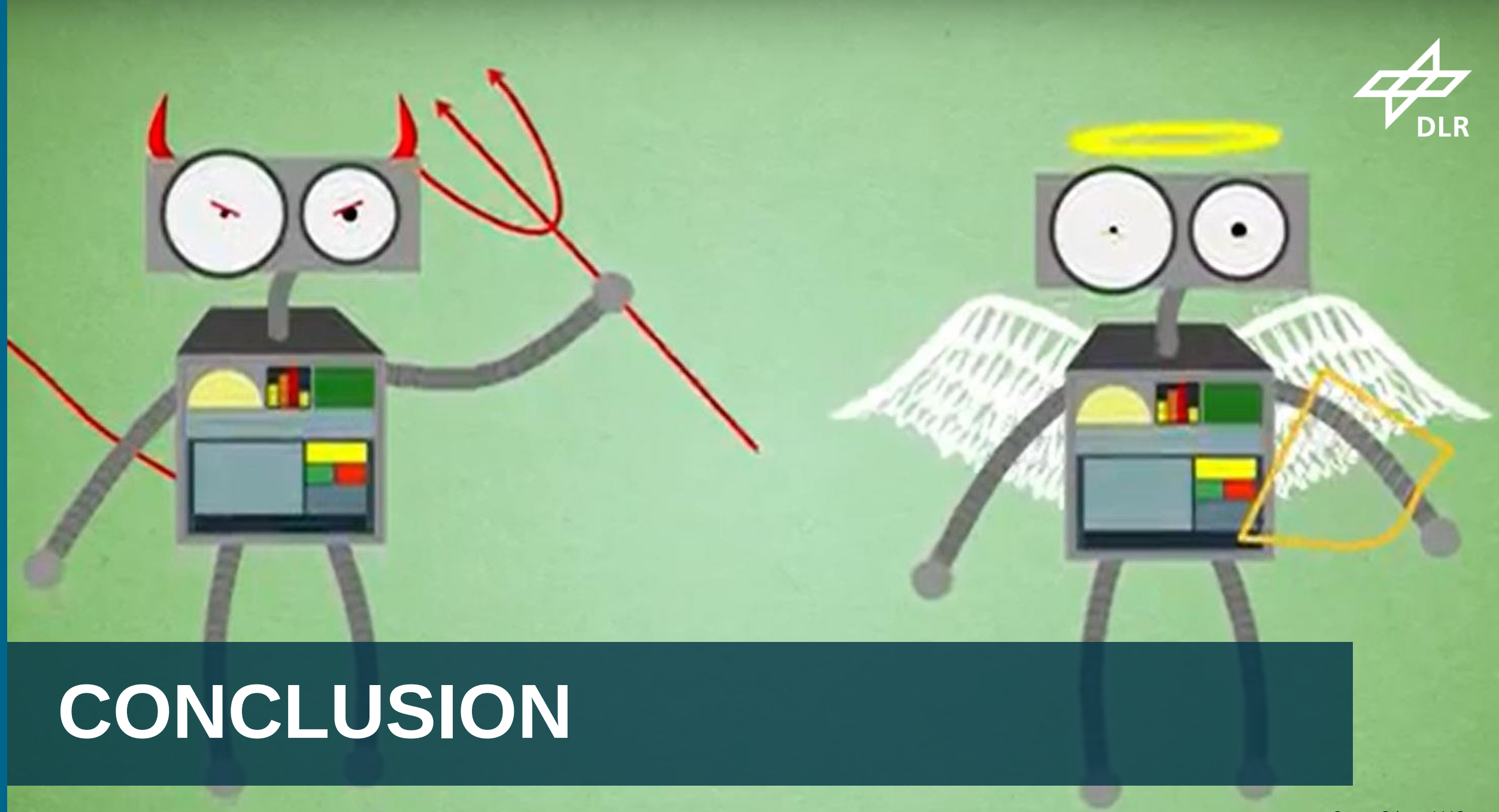
I do need to establish an DMP because…

- … all my R&D work is based on data;
- …I work with different types of data;
- …I apply different types of processing;
- …I am confronted with different legal situations (personal data, confidential data, open source data, biometric data, imaging data etc.);
- …I want to establish an effective data management and data sharing (Licensing, usage rights, valuation of data, quality control i.e. industry standards for collaborations);
- …

The data will be publicly shared because…

- ...they are necessary to validate my research findings;
- …they were generated of one-time events;
- …they must be kept for at least 10 years;
- …





CONCLUSION

Source: Science AAAS

DMP insights out of the DLR RDM
- Promotion by utilizing own RDMO instance -

Awareness

- DMP draws attention to the relevance of RDM;
- DMP offers a first introduction to the project and guided through it;
- DMP helps to be transparent.

Orientation

- DMP keep track of where you are and where you’re going;
- DMP helps to understand where the data comes from and where it’s going.

DMP insights out of the DLR RDM

Framework for actions

- DMP provides the reasons for defining specific measures and rules (e.g. institute-specific) in the project;
- DMP allow changes and keep track in the case of changes.

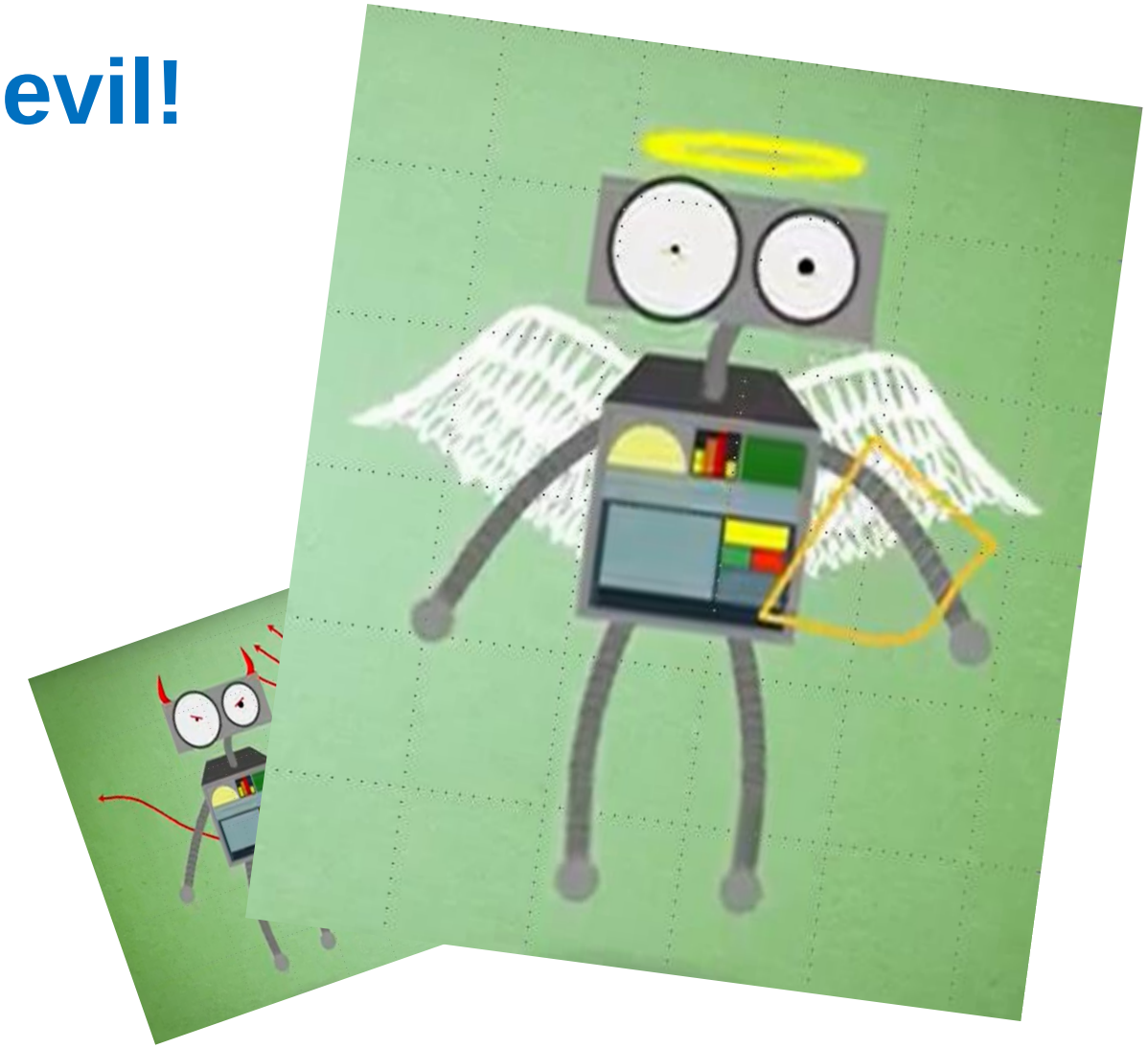
Support

- DMP supports collaboration;
- DMP supports to be honest about the data management costs.

Positioning

- DMP supports the strategic positioning of the project;
- DMP describe legal and institutional project boundaries.

Take your time for creating a really helpful DMP!
It is more Angel than Devil!



Dr.-Ing.

Christian Langenbach

Research Data Manager DLR

DLR

Max-Reichpietsch-Str. 1-3, 51147 Cologne

Phone: +49-2203-601-2704

Mail: christian.langenbach@dlr.de

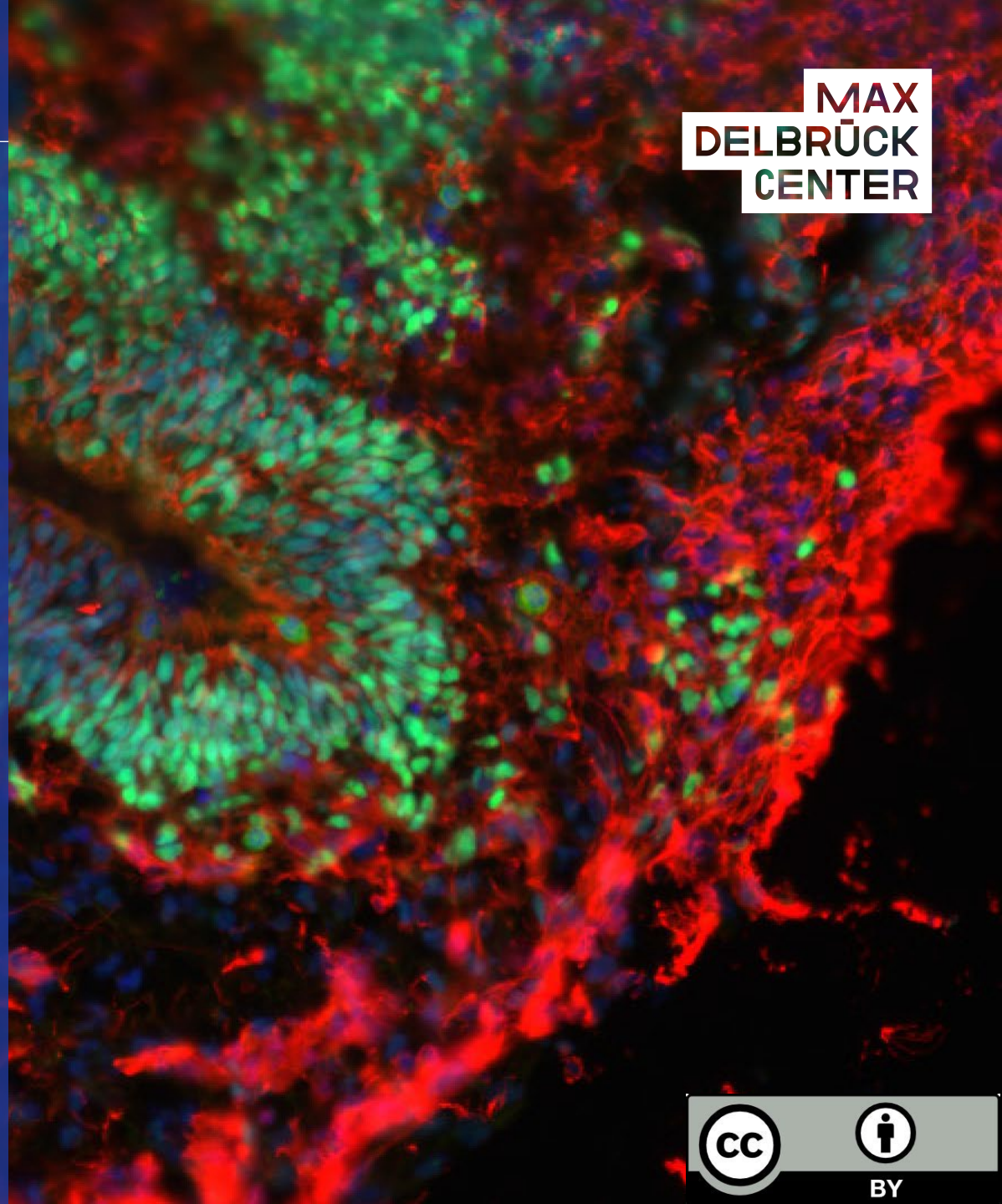
Imprint

Thema:	**Data Management Plans** Angel or Devil for the Researcher Work
Datum:	20.10.2022
Autor:	Dr.-Ing. Christian Langenbach
Institut:	Wissenschaftliche Information - Forschungsdatenmanagement
Bildcredits:	„DLR (CC BY-NC-ND 3.0)“; „Science AAAS“; „fotolia“; „DigitalbevaringDK“; „Mau, Franziska. (2019). Sketchnote: Lost Data Map. Zenodo. https://doi.org/10.5281/zenodo.4388672“

FAIR WIZARD AT THE MDC

Inga Patarčić

Helmholtz Open Science Practice Forum
20.10.2022



MDC RESEARCH DATA MANAGEMENT UNIT

The Max Delbrück Center for Molecular Medicine is one of the world's leading biomedical research centers (1992).

The MDC has **88 labs** and **833** researchers who analyze how the human body works in both health and disease.

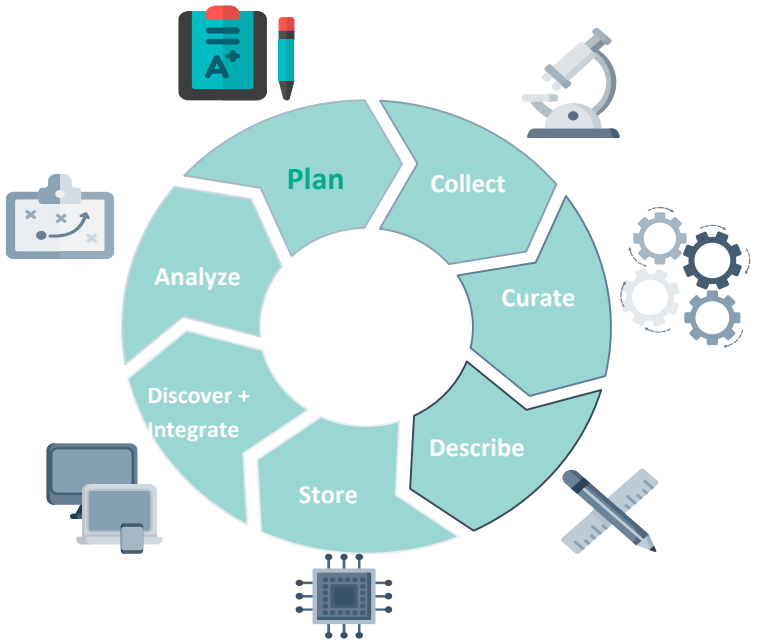
The Research Data Management Unit was established **in 2020** as a part of the MDC's Scientific Infrastructure Department.

DATA MANAGEMENT PLAN

is...

A formal declaration on how an activity's datasets are to be handled throughout activity's lifetime and on what terms those datasets will survive the activity.

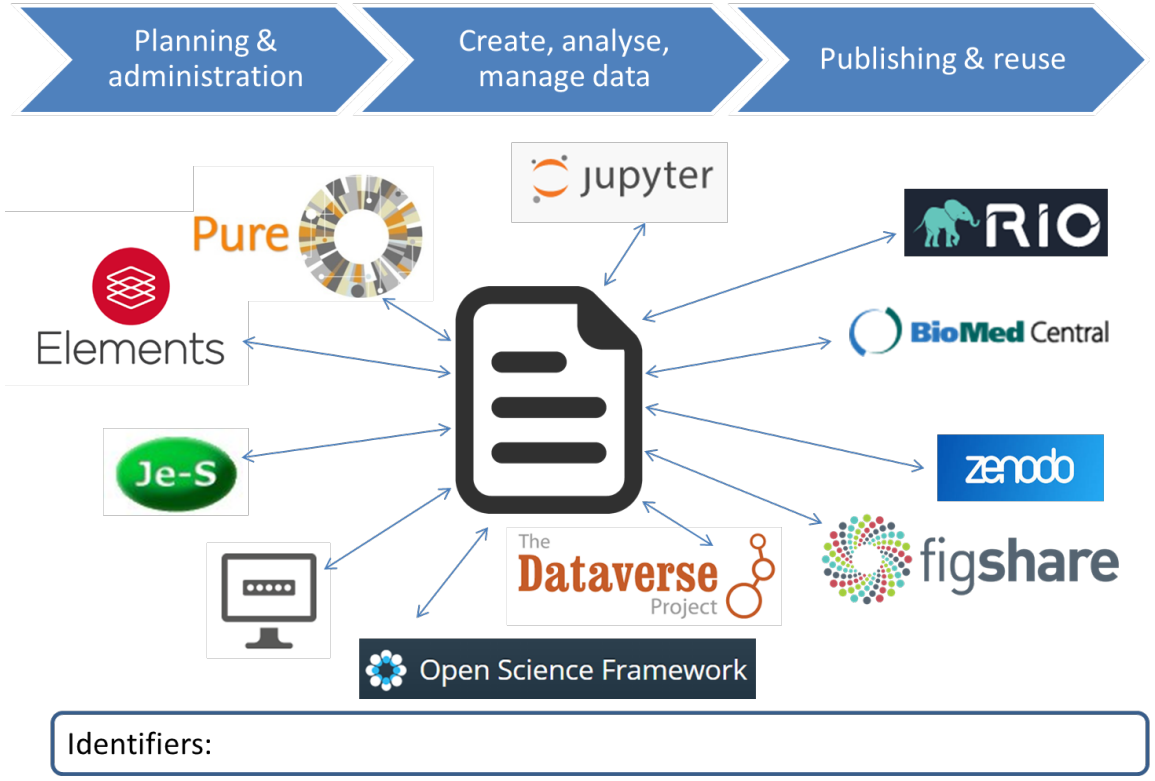
- May span several grants or a coincides a single grant
- Datasets might be interim results, outpost or inputs of the activity.



BUT!

DMP ≠ extensive static document with a long text created to satisfy grant agency

- Machine-actionable DMPs
- System connections
- Versioning of DMPs
- Notification of changes
- Archiving and publishing plans
- Assessing the FAIRness of data
- Usage statistics



WHICH TOOLS CAN BE USED TO WRITE A DMP?

Many.

DMPOnline

Based on the DMPRoadmap (DCC and the University of California Curation Center).

DMPTool

DMPTool is a service of the California Digital Library

RDMO

DMP tool developed by AIP, FHP und KIT-Bibliothek

Argos

online machine-actionable tool developed by OpenAIRE

DS Wizard

machine-actionable tool based on FAIR principles by ELIXIR CZ and ELIXIR NL

...

TOOL SELECTION

We created a comparison matrix.

Tool	DMPTool	DMPOnline Basic/Enh	RDMO	FAIR Wizard	ARGOS
Archiving and Publishing Options	yes	yes	yes	yes	yes
Template Selection	yes	yes	yes	yes	yes
Template/Documents Creation	limited	limited	limited	yes	limited
Templates for German funders	possible	possible	yes	possible	possible
Support for admins	not tested	fast	slow	fast	not tested
Project Phases Acknowledged	no	no	no	yes	no
Making ToDoS	no	no	yes	yes	limited
Machine Actionable tool	yes	yes	yes	yes	yes
DMP ID	yes	yes	no	no	no
Version control	no	no	no	yes	yes
Integration with other tools	yes	yes	no	yes	possible
Unser frendliness/interactivnes	poor	poor	poor	good	good
Assessing FAIRness of data plans	no	no	no	yes	no
Teaching potential	no	no	no	yes	no
Usage statistics	no	no	no	yes	no

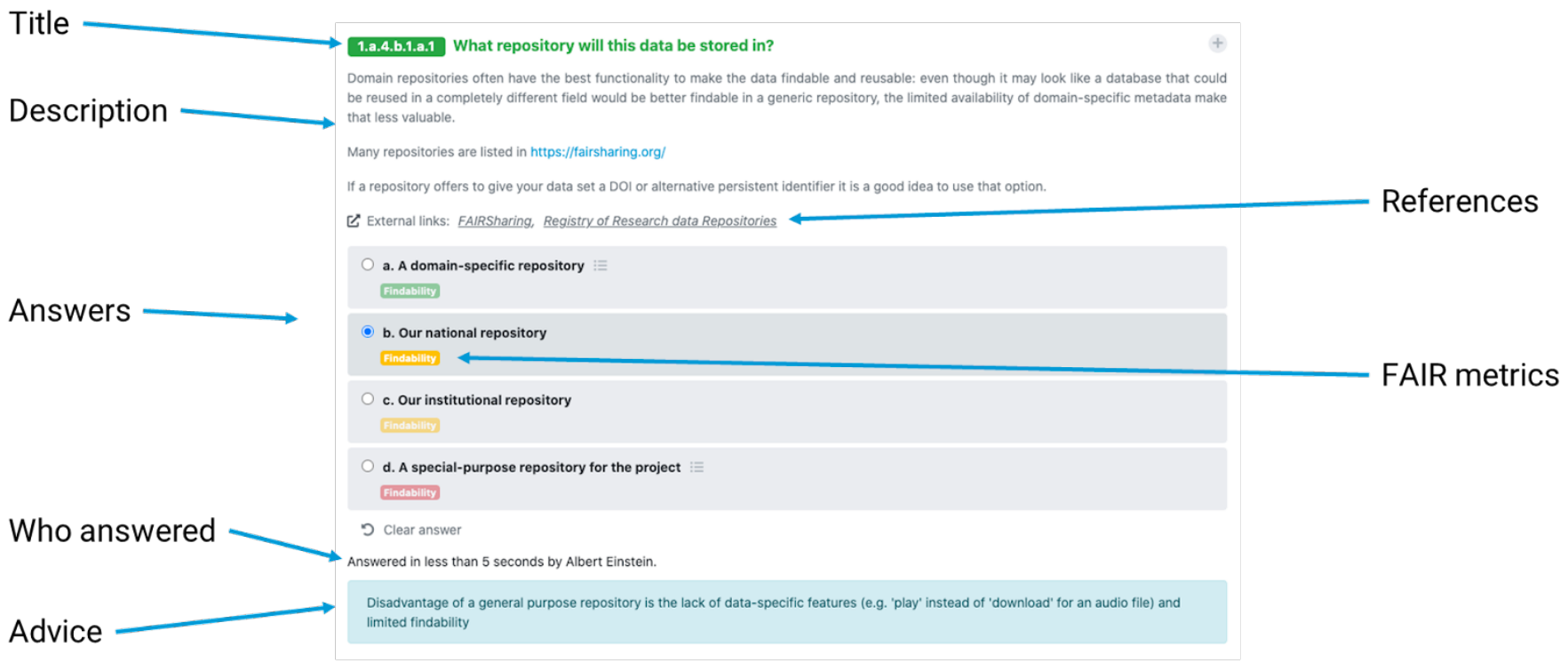
Figure: Comparison matrix examining DMPTool, DMPOnline, Argos, RDMO and FAIR Wizard

FAIR WIZARD PROVIDES SUMMARY REPORT

Assessing FAIRness of the data

The MDC "is committed to a goal of making data created as part of the research process compliant with the FAIR principle" (Policy Framework for Research Data Management, 2021).

FAIR WIZARD GUIDES SCIENTISTS TOWARDS BETTER RESEARCH PRACTICES



IMPLEMENTATION

tbc

- In purchasing process.
- On-boarding PhD students during the introduction week in November.
- Introduced to all scientists – Dec 2022.
- Experiences?

OPEN FOR QUESTIONS

Contact: inga.patarcic@mdc-berlin.de

MAX DELBRÜCK CENTER

Data Management Plans

DMPs as Living Documents

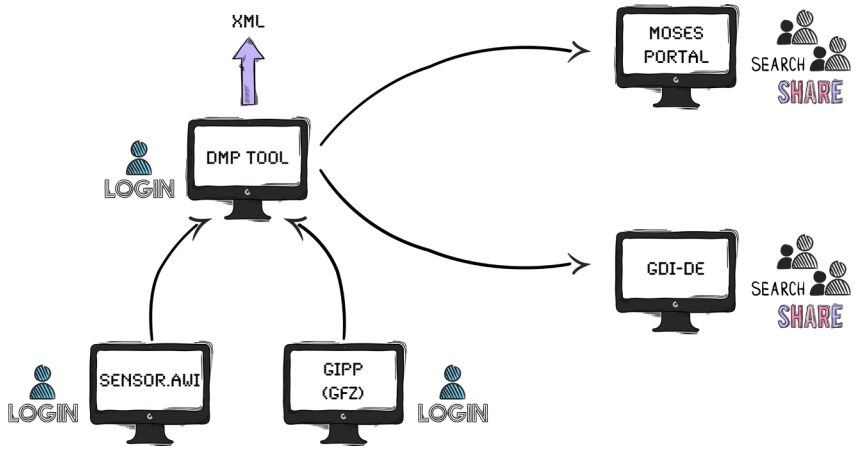
Hannes Fuchs / GFZ Potsdam / eScience Centre

MOSES

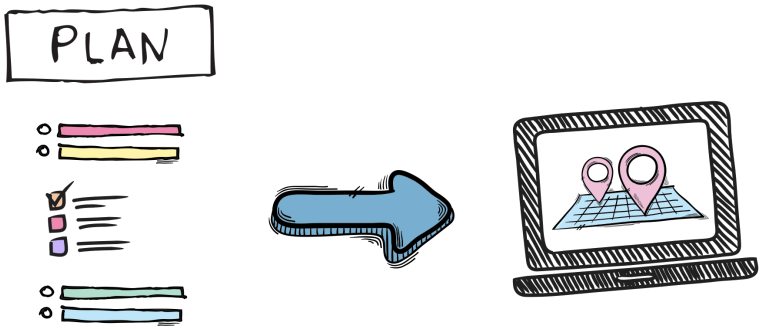
- Mobile and modular observing system in Helmholtz
- Workshops with Users (ex. Scientists) in early project phase
- Evaluation of DMP Tools → First Campaigns with RDMO
- RDMO could not meet requirements
- Connect / Integrate with existing and upcoming Tools
- Implements Data Policy as DMP

MOSES DMP Tool

- Roles with different permissions
- Snapshots
- Integration of established tools
- Templating support
- Export (ISO 19115)

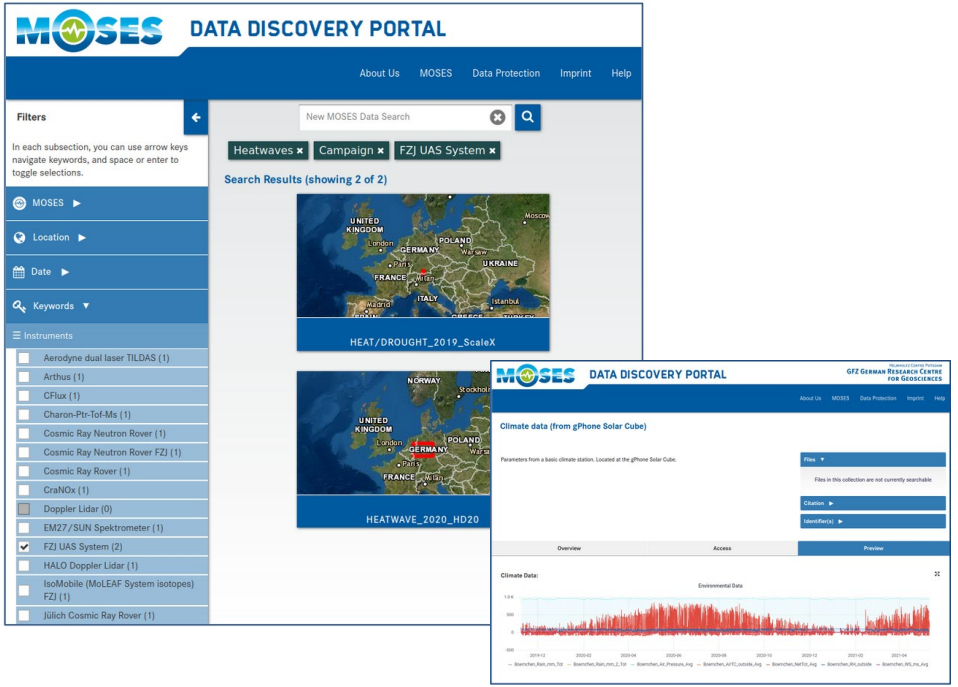


MOSES DMP Tool



- DMP for every campaign
- Accompany campaign (before, during and afterwards)
- Provides metadata for campaign and data
- Documentation of data exchange
- Automatic transfer to Portal

DMP Tool and Data Discovery Portal



Lessons learned

- Add information of existing systems automatically
- Make entered information reusable (ISO export)
- Snapshots / History / Rollback
- Limit information to enter to the „most important“ ones
- Copy whole DMP and/or parts of existing one(s) to new one

Outlook

- Export of data into EarthData Portal (replacing DDP)
- Add more comfort features
- RDMO may replace the MOSES DMP Tool in future

Contact

Website:https://moses-dmp.gfz-potsdam.de/
Contact: moses-dmp@gfz-potsdam.de
MOSES project homepage: http://www.moses-helmholtz.de

HELMHOLTZ METADATA COLLABORATION

HZB Helmholtz Zentrum Berlin

INFORMATION | HEALTH | MATTER | ENERGY | AST | E&E | DATA COMMONS | HMC OFFICE | HMC PROJECTS

Monitoring Data Publications: A Dashboard Approach in HMC Hub Matter

Markus Kubin[1,2]

// Astrid Gilein[2] // Mojeeb R. Sedeqi[1,2] // Alexander Schmidt[2] // Tempest Glodowski[2] // Gabriel Preuß[1,2] // Oonagh Mannix[1,2]

[1] Helmholtz Metadata Collaboration (HMC) Hub Matter

[2] Helmholtz-Zentrum Berlin für Materialien und Energie

www.helmholtz-metadata.de

Helmholtz Metatdata Collaboration (HMC)

- Make Helmholtz data FAIR: findable, accessible, interoperable and reusable [1]
- Provide services for sustainable and efficient metadata handling.
- Develop, share and consolidate community-expertise in metadata across Helmholtz.
- **Hub Matter**: Physics, Chemistry, Materials, Large scale facilities, User programs, …

- Turning FAIR into reality [2] on all levels to enable reuse of data
- Develop HMC as a research infrastructure platform!

[1] M Wilkinson et al. *Sci Data* **3**, 160018 (2016). doi: 10.1038/sdata.2016.18
[2] Turning FAIR into Reality, Final Report and Action Plan from the European Commission Expert Group on FAIR Data , doi: 10.2777/1524

Monitoring Research Data Management Practices

> *Good decisions depend on good measurement*
>
> – Ben S. Bernanke, 2012 –

- Survey of data practices in Helmholtz (2021)
- FAIR data assessment of instruments (2021)
- **Pilot Dashboard: Open & FAIR Data** (2022)
- Qualitative interview campaign (2022/23)

Monitoring Data Publishing Practices

1. How much data is published?
2. In which repositories is data published?
3. How FAIR is this data?

How to find data publications?

Libraries

- **Registration** of data publications by research centers **started rather recently**

Repositories

- We may **not know** the **repositories**, yet
- Metadata relevant to us may be missing

Text mining in research articles

- Data availability statements, data citations
- e.g. **Charité Metrics Dashboard** [3,4]

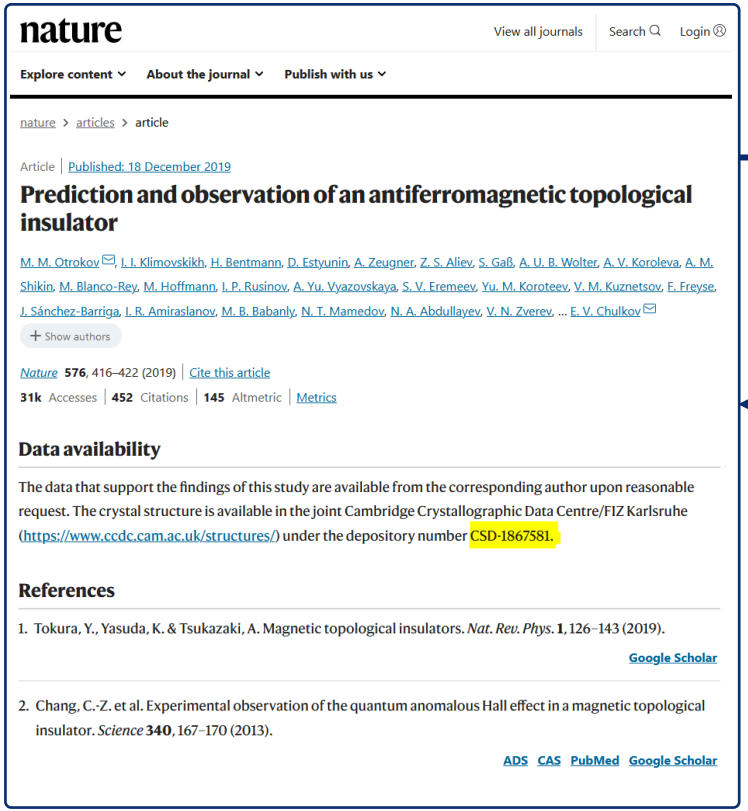
Registries and Data Bases

- **Crossref, DataCite**, (OpenAire)
- **SCHOLIX** [5]

[3] Charité Dashboard on Responsible Research, https://quest-dashboard.charite.de/#tabStart, accessed 2022-10-19
[4] A. Iarkaeva et al., „Semi-automated extraction of information on open datasets mentioned in articles“, protocols.io (2022). doi: 10.17504/protocols.io.q26g74p39gwz/v1
[5] SCHOLIX, “A Framework for Scholarly Link eXchange.” http://www.scholix.org, accessed 2022-10-19.

Linked data publications

Research article



nature

View all journals | Search | Login

Explore content | About the journal | Publish with us

nature › articles › article

Article | Published: 18 December 2019

Prediction and observation of an antiferromagnetic topological insulator

M. M. Otrokov, I. I. Klimovskikh, H. Bentmann, D. Estyunin, A. Zeugner, Z. S. Aliev, S. Gaß, A. U. B. Wolter, A. V. Koroleva, A. M. Shikin, M. Blanco-Rey, M. Hoffmann, I. P. Rusinov, A. Yu. Vyazovskaya, S. V. Eremeev, Yu. M. Koroteev, V. M. Kuznetsov, F. Freyse, J. Sánchez-Barriga, I. R. Amiraslanov, M. B. Babanly, N. T. Mamedov, N. A. Abdullayev, V. N. Zverev, … E. V. Chulkov

+ Show authors

Nature **576**, 416–422 (2019) | Cite this article

31k Accesses | **452** Citations | **145** Altmetric | Metrics

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The crystal structure is available in the joint Cambridge Crystallographic Data Centre/FIZ Karlsruhe (https://www.ccdc.cam.ac.uk/structures/) under the depository number CSD-1867581.

References

1. Tokura, Y., Yasuda, K. & Tsukazaki, A. Magnetic topological insulators. *Nat. Rev. Phys.* **1**, 126–143 (2019).
   Google Scholar
2. Chang, C.-Z. et al. Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167–170 (2013).
   ADS CAS PubMed Google Scholar

Data availability

Crossref

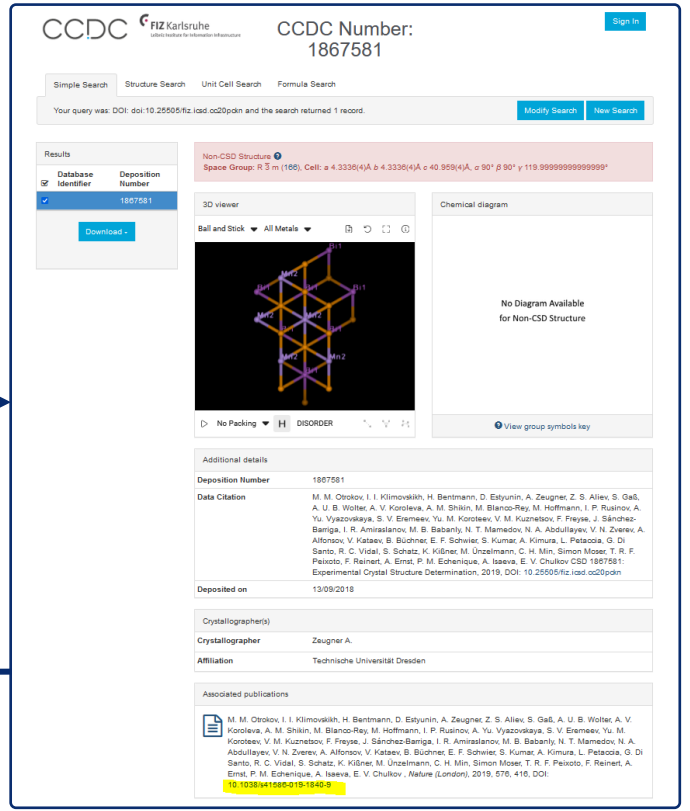
Scholix links

SCHOLIX

Associated paper

DataCite
FIND, ACCESS, AND REUSE DATA

Data publication



CCDC | FIZ Karlsruhe

CCDC Number: 1867581

Sign In

Simple Search | Structure Search | Unit Cell Search | Formula Search

Your query was: DOI: doi:10.25505/fiz.icsd.cc20pckn and the search returned 1 record.

Modify Search | New Search

Results

Database Identifier | Deposition Number

1867581

Download -

Non-CSD Structure

Space Group: R $\bar{3}$ m (166), Cell: a 4.3336(4)Å b 4.3336(4)Å c 40.959(4)Å, α 90° β 90° γ 119.99999999999999°

3D viewer

Ball and Stick | All Metals

No Packing | H | DISORDER

Chemical diagram

No Diagram Available for Non-CSD Structure

View group symbols key

Additional details

Deposition Number	1867581
Data Citation	M. M. Otrokov, I. I. Klimovskikh, H. Bentmann, D. Estyunin, A. Zeugner, Z. S. Aliev, S. Gaß, A. U. B. Wolter, A. V. Koroleva, A. M. Shikin, M. Blanco-Rey, M. Hoffmann, I. P. Rusinov, A. Yu. Vyazovskaya, S. V. Eremeev, Yu. M. Koroteev, V. M. Kuznetsov, F. Freyse, J. Sánchez-Barriga, I. R. Amiraslanov, M. B. Babanly, N. T. Mamedov, N. A. Abdullayev, V. N. Zverev, A. Alfonsov, V. Kataev, B. Büchner, E. F. Schwier, S. Kumar, A. Kimura, L. Petaccia, G. Di Santo, R. C. Vidal, S. Schatz, K. Kißner, M. Unzelmann, C. H. Min, Simon Moser, T. R. F. Peixoto, F. Reinert, A. Ernst, P. M. Echenique, A. Isaeva, E. V. Chulkov CSD 1867581: Experimental Crystal Structure Determination, 2019, DOI: 10.25505/fiz.icsd.cc20pckn
Deposited on	13/09/2018

Crystallographer(s)

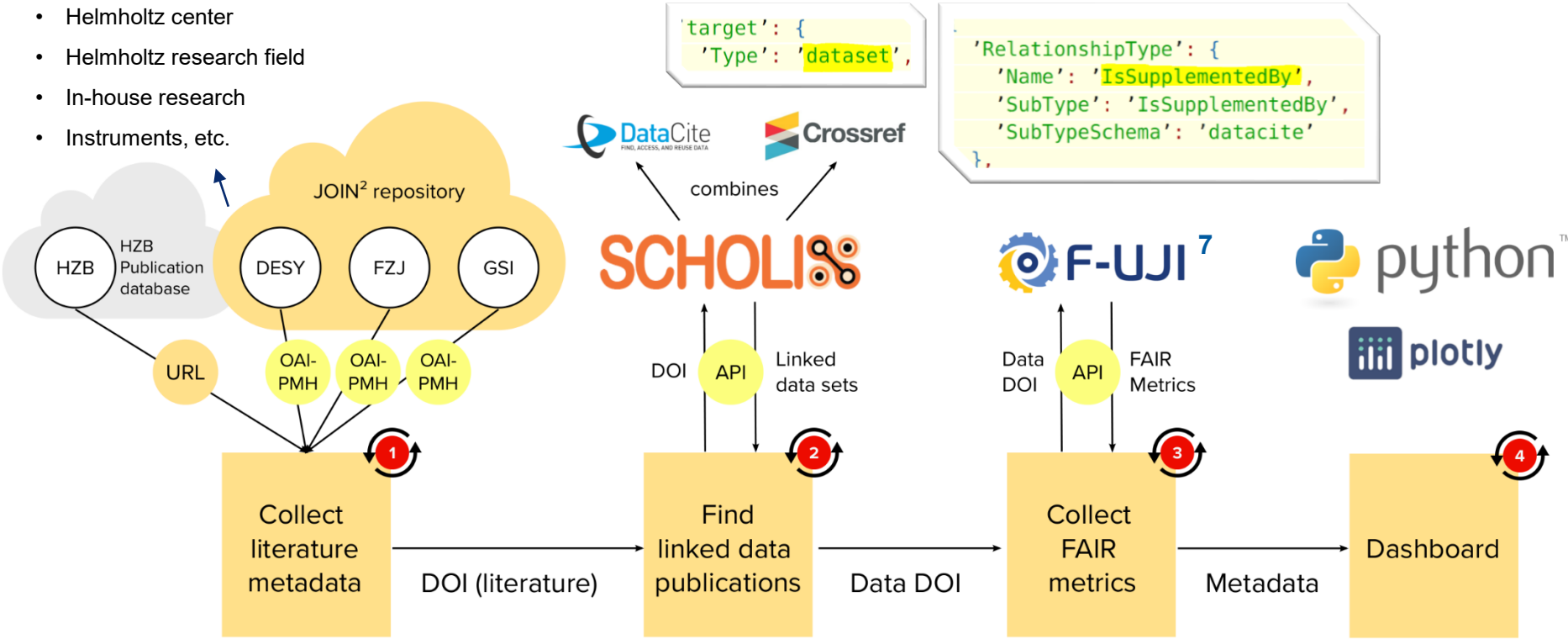
Crystallographer	Zeugner A.
Affiliation	Technische Universität Dresden

Associated publications

M. M. Otrokov, I. I. Klimovskikh, H. Bentmann, D. Estyunin, A. Zeugner, Z. S. Aliev, S. Gaß, A. U. B. Wolter, A. V. Koroleva, A. M. Shikin, M. Blanco-Rey, M. Hoffmann, I. P. Rusinov, A. Yu. Vyazovskaya, S. V. Eremeev, Yu. M. Koroteev, V. M. Kuznetsov, F. Freyse, J. Sánchez-Barriga, I. R. Amiraslanov, M. B. Babanly, N. T. Mamedov, N. A. Abdullayev, V. N. Zverev, A. Alfonsov, V. Kataev, B. Büchner, E. F. Schwier, S. Kumar, A. Kimura, L. Petaccia, G. Di Santo, R. C. Vidal, S. Schatz, K. Kißner, M. Unzelmann, C. H. Min, Simon Moser, T. R. F. Peixoto, F. Reinert, A. Ernst, P. M. Echenique, A. Isaeva, E. V. Chulkov . *Nature (London)*, 2019, 576, 416, DOI: 10.1038/s41586-019-1840-9

Data Harvesting & Dashboard [6]

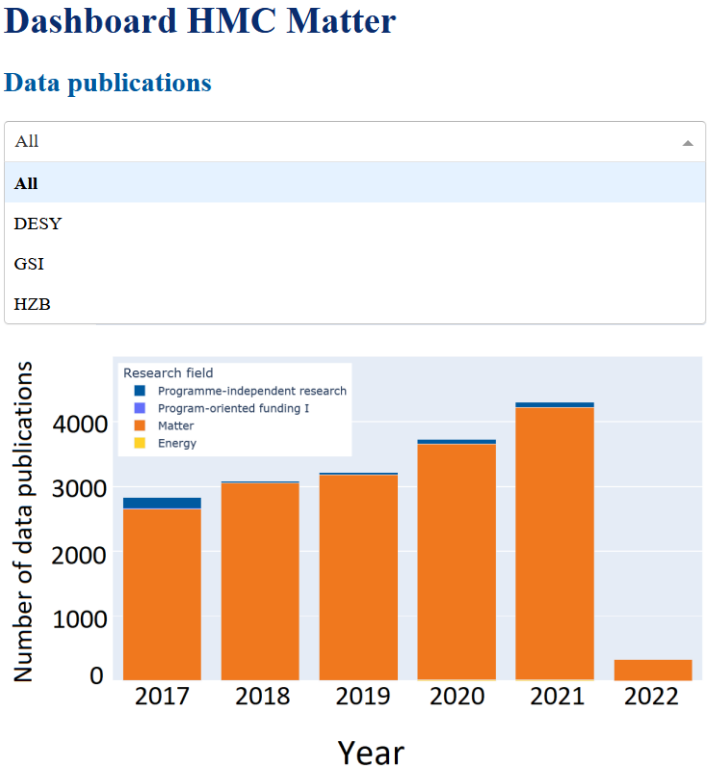
- Helmholtz center
- Helmholtz research field
- In-house research
- Instruments, etc.



[6] A Gilein, Virtueller Werkzeugkasten für Data Mining und FAIR-Data-Metriken von Datenpublikationen (2022). doi: 10.5281/zenodo.7219635

[7] A Devaraju and R Huber: F-UJI - An Automated FAIR Data Assessment Tool (2020). doi: 10.5281/zenodo.4063720

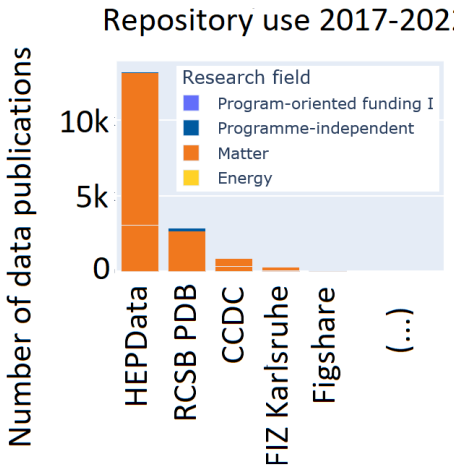
Pilot Dashboard on Open and FAIR data

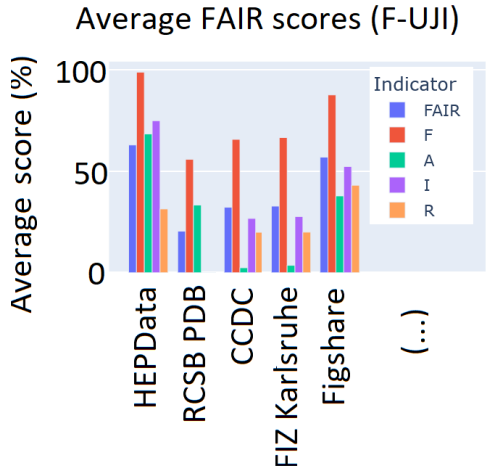


Open Data: Repository usage by research field, over time etc.

FAIR Data: Automized FAIR scores using **F-UJI** [7]

- **Detailed insights** from currently 16 FAIRsFAIR metrics [8]
- Biases and limitations





[7] A Devaraju and R Huber: F-UJI - An Automated FAIR Data Assessment Tool (2020). doi: 10.5281/zenodo.4063720
[8] Devaraju et al., FAIRsFAIR Data Object Assessment Metrics (2020). doi: 10.5281/zenodo.6461229

Manual Validation of the Data harvested

Statistical validation of the Dashboard data

- **100/100 SCHOLIX links** category "**A**" in the dashboard are **correct**
- **99/100 "Type: dataset**" in the dashboard are **correct**
  - 1% Generic repositories → inconsistencies in "Type: dataset"
  - 99% Disciplinary repositories → consistent use of "Type: dataset"
- **~ 99% of data in the dashboard correctly assigned**

Room for improvement:

- Manual validation of "Type: dataset" for generic repositories
- Inconsistencies in SCHOLIX link-categories "**B**" and "**C**"
  - 1/3 - 2/3 could be assigned SCHOLIX link-category "**A**"
  - Even more data publications out there!
- Add further filters (e.g. in-house)

RelationshipType

A) „IsSupplementedBy" / „IsSupplementTo" [9]

B) „IsRelatedTo"

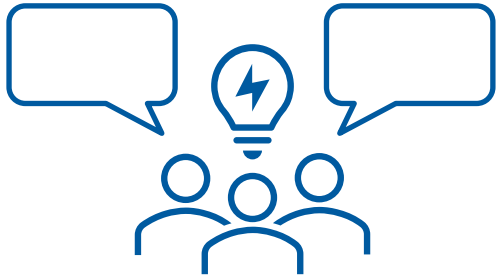
C) „References" / „IsReferencedBy" [9]

[9] DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. doi: 10.14454/3w3z-sa82

Considerations for the Open Science Community

Counting data publications is sensitive

- Need consensus: What counts as a „data publication“?
  - Raw data / processed data / data shown in an article
  - Standardized (meta)data quality
  - Persistent identifiers → splitting of data sets?



Data harvesting approaches

- Manual validation (curation) recommended
- Depends on data quality in the sources
  - Harmonize registration of “RelationshipType”, “dataset”

Standardized (meta)data quality

- Reusability and Interoperability
- Support by RD professionals

Summary and Outlook

Summary

- Pilot Dashboard on Open and FAIR data
  - 10k + data publications found so far
  - Data concentrated in few repositories
- Dashboard is one of several approaches
- Identified action items for improvement

Outlook

- Make dashboard & code available
- Improving the dashboard (DB, UX, …)
- More Helmholtz centers to be included
- Potentially include software publications

Coming soon!

- Report of HMC Community Survey 2021
- Qualitative interview campaign (2022 / 23)
- Advanced HMC Survey (2023)

Thank you

Get in touch …

Markus Kubin - markus.kubin@helmholtz-berlin.de

HMC Hub Matter - hmc-matter@helmholtz-berlin.de

Group page …

https://helmholtz-metadaten.de/en/matter/contact-us

Twitter …

@helmholtz_hmc / #HMCMatter

Acknowledgements:

Data harvesting & dashboard team

Astrid Gilein (HZB)
Mojeeb R. Sedeqi (HMC, HZB)
Alexander Schmidt (HZB)
Tempest Glodowski (HZB)
Gabriel Preuß (HMC, HZB)
Oonagh Mannix (HMC, HZB)

Helmholtz libraries

HZB, DESY, GSI, FZJ, JOIN², HZDR