HELMHOLTZ
Open Science

Helmholtz Open Science Briefing

3rd Helmholtz Open Science Forum „Helmholtz in the German National Research Data Infrastructure (NFDI)“

Report

Impressum

Die Onlineversion dieser Publikation finden Sie unter:
https://doi.org/10.48440/os.helmholtz.066

Verfasser:innen

Nina Leonie Weisweiler, Steffi Genderjahn, Roland Bertelmann, Jan Kröger, Mohammad Al-Turany, Alexandra Axtmann, Anton Barty, Christian Busse, Stephan Frickenhaus, Cordula Hege, Tamara Husch Lee, Uwe Jandt, Uwe Konrad, Sören Lorenz, Hela Mehrtens, Valentina Protopopova-Kakar, Gisela Schmidt, Carolina Schwedhelm, Antonia Schrader

Herausgeber

Helmholtz Open Science Office

Redaktion

Nina Leonie Weisweiler, Steffi Genderjahn, Roland Bertelmann, Christoph Bruch, Lea Maria Ferguson, Heinz Pampel, Antonia C. Schrader, Marcel Meistring, Lena Messerschmidt, 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

11.08.2023. Version. 1.0

Lizenz

HELMHOLTZ

Open Science

Content

Abstract

To promote dialogue on the National Research Data Infrastructure (NFDI) in the Helmholtz Association, the Helmholtz Open Science Office hosted two digital Forums in May and December 2021. The office has organized a third Forum on the topic on June 22, 2023. The objective of this event was to offer insights into the NFDI activities within the Helmholtz Association, presented from the internal perspectives of the Centers. Multiple Helmholtz Centers shared their experiences, fostering an interactive environment for questions and discussions. Furthermore, there were contributions highlighting the Base4NFDI basic service consortium.

Background

With the National Research Data Infrastructure (NFDI), the German federal and state governments are pursuing the goal of systematically making research data resources accessible in accordance with the FAIR principles.

The NFDI is established as a network of discipline-oriented consortia over a period of three years in three successive funding phases. In November 2022, the Joint Science Conference (GWK) decided to fund eight additional NFDI consortia proposed in the third and final round. Together with those selected in the previous two rounds, the National Research Data Infrastructure now has a total of 26 disciplinary consortia and one union of consortia (Base4NFDI).

To promote dialogue on the NFDI at Helmholtz, the Helmholtz Open Science Office already hosted two digital forum events in May 2021 (see report) and December 2021 (see report). Numerous NFDI consortia are realized with substantial Helmholtz participation.

On June 22, 2023 the Open Science Office arranged a third forum to offer insights into the NFDI activities within the Helmholtz Association, presented from the internal perspective of the Centers. Multiple Centers demonstrated their engagement in different NFDI consortia (figure 1). Furthermore, there were contributions highlighting the Base4NFDI consortium and the NFDI basic services project PID4NFDI. The speakers reported on their experiences with creating synergies and fostering exchange activities between different actors within Helmholtz and beyond. The presentations provided impetus for discussion on future and long-term development in the provision of infrastructure and service offerings, as well as cross-NFDI networking.

HELMHOLTZ
Open Science

Figure 1: Participation of the Helmholtz Centers in the NFDI, overview (as of March 2023)

HELMHOLTZ

Open Science

Program of the event of June 22, 2023

Time	Agenda	Speaker
10:00–10:10	Welcome	Roland Bertelmann Helmholtz Open Science
10:10–10:25	AWI's engagement in NFDI4Earth and synergies with DataHub Earth & Environment	Stephan Frickenhaus AWI
10:25–10:40	NFDI-related activities at DESY	Anton Barty DESY
10:40–10:55	NFDI at DKFZ	Christian Busse DKFZ
10:55–11:10	DZNE Involvement in GHGA	Gisela Schmidt DZNE
11:10–11:20	Coffee Break	
11:20–11:35	HMC and (Base4)NFDI	Sören Lorenz GEOMAR, HMC Speaker
11:35–11:50	HIFIS Perspectives on Base4NFDI	Uwe Jandt DESY, HIFIS Overall Management
11:50–12:05	Helmholtz Perspectives on Base4NFDI: PID4NFDI	Antonia Schrader Helmholtz Open Science Office
12:05–13:00	Lunch Break	
13:00–13:15	NFDI activities from the perspective of GEOMAR	Hela Mehrtens GEOMAR
13:15–13:30	GSI and (PUNCH4) NFDI	Mohammad Al-Turany GSI
13:30–13:45	Bridging the gap between user-friendliness and cutting-edge research: HZB in the NFDI	Tamara Husch Lee HZB
13:45–14:00	NFDI Activities @ HZDR	Uwe Konrad HZDR
14:00–14:10	Coffee Break	
14:10–14:25	NFDI activities at the HZI Dealing with Microbiota and Immunology	Cordula Hege HZI
14:25–14:40	NFDI activities at the GFZ	Valentina Protopopova-Kakar GFZ
14:40–14:55	NFDI networking at KIT	Alexandra Axtmann & Jan Kröger KIT
14:55–15:10	NFDI activities at MDC	Carolina Schwedhelm MDC
15:10–15:30	Wrap-up and final discussion	

Presentations

1. Stephan Frickenhaus: AWI's engagement in NFDI4Earth and synergies with DataHub Earth & Environment
2. Anton Barty: NFDI-related activities at DESY
3. Christian Busse: NFDI at DKFZ
4. Gisela Schmidt: DZNE Involvement in GHGA
5. Sören Lorenz: HMC and (Base4)NFDI
6. Uwe Jandt: HIFIS Perspectives on Base4NFDI
7. Antonia Schrader: Helmholtz Perspectives on Base4NFDI: PID4NFDI
8. Hela Mehrtens: NFDI activities from the perspective of GEOMAR
9. Mohammad Al-Turany: GSI and (PUNCH4) NFDI
10. Tamara Husch Lee: Bridging the gap between user-friendliness and cut-ting-edge research: HZB in the NFDI
11. Uwe Konrad: NFDI Activities @ HZDR
12. Cordula Hege: NFDI activities at the HZI. Dealing with Microbiota and Immunology
13. Valentina Protopopova-Kakar: NFDI activities at the GFZ
14. Alexandra Axtmann, Jan Kröger: NFDI networking at KIT
15. Carolina Schwedhelm: NFDI activities at MDC

Nationale Forschungsdaten Infrastruktur

NFDI4Earth

AWI's engagement in NFDI4Earth and synergies with the DataHub Earth & Environment

Stephan Frickenhaus, Co-Applicant, Task Area „2Interoperate“

3rd Helmholtz Open Science Forum on the National Research Data Infrastructure (NFDI):
Inside Perspectives from the Centers
22.6.2023

HELMHOLTZ

AWI Participation in NFDI

- **NFDI4Biodiversity**: Applicant Frank Oliver Glöckner
  - Contributions in Usecase Critterbase (Helmholtz Institute Oldenburg)
- **NFDI4Earth**: Co-Applicant Stephan Frickenhaus
  - TA3 Speaker, TA3.1 Measure lead
- **NFDI4DataScience**: Partner

NFDI4Earth

NFDI Consortium Earth System Sciences

ESS domain is characterized by

- high degree of standardization (e.g. OGC-standards for web services = Spatial Data Infrastructures SDIs, Geographic Information Systems, plenty of community defined fileformats and metadata standards)

- High degree of diversity of service provider technologies

Where AWI is engaged

NFDI4Earth2Participate
M1.1: Earth System Science Pilots
M1.2: Incubator Lab
M1.3: Education and Training Materials and Services
M1.4: NFDI4Earth Academy

Task Area 1

NFDI4Earth2Facilitate
M2.1: OneStop4All
M2.2: User Support
M2.3: Governmental Data
M2.4: Data in Long-Term Storage
M2.5: Advancing Tools

Task Area 2

NFDI4Earth2Interoperate
M3.1: Synthesis of a Sustainable NFDI4Earth Architecture
M3.2: Common Standards for FAIR ESS Data
M3.3: NFDI Commons
M3.4: International Networking & Embedding

Task Area 3

NFDI4Earth2Coordinate
M4.1: Coordination, Collaborative and Sustainable Governance of NFDI4Earth
M4.2: Towards a Cultural Change in ESS Research Data Management
M4.3: Central Support Services for the federated NFDI4Earth

Task Area 4

Similar activities in DataHUB Earth&Env.

Infrastructure perspective

TA-Speaker +
Measure „Architecture“

DataHub Erde und Umwelt

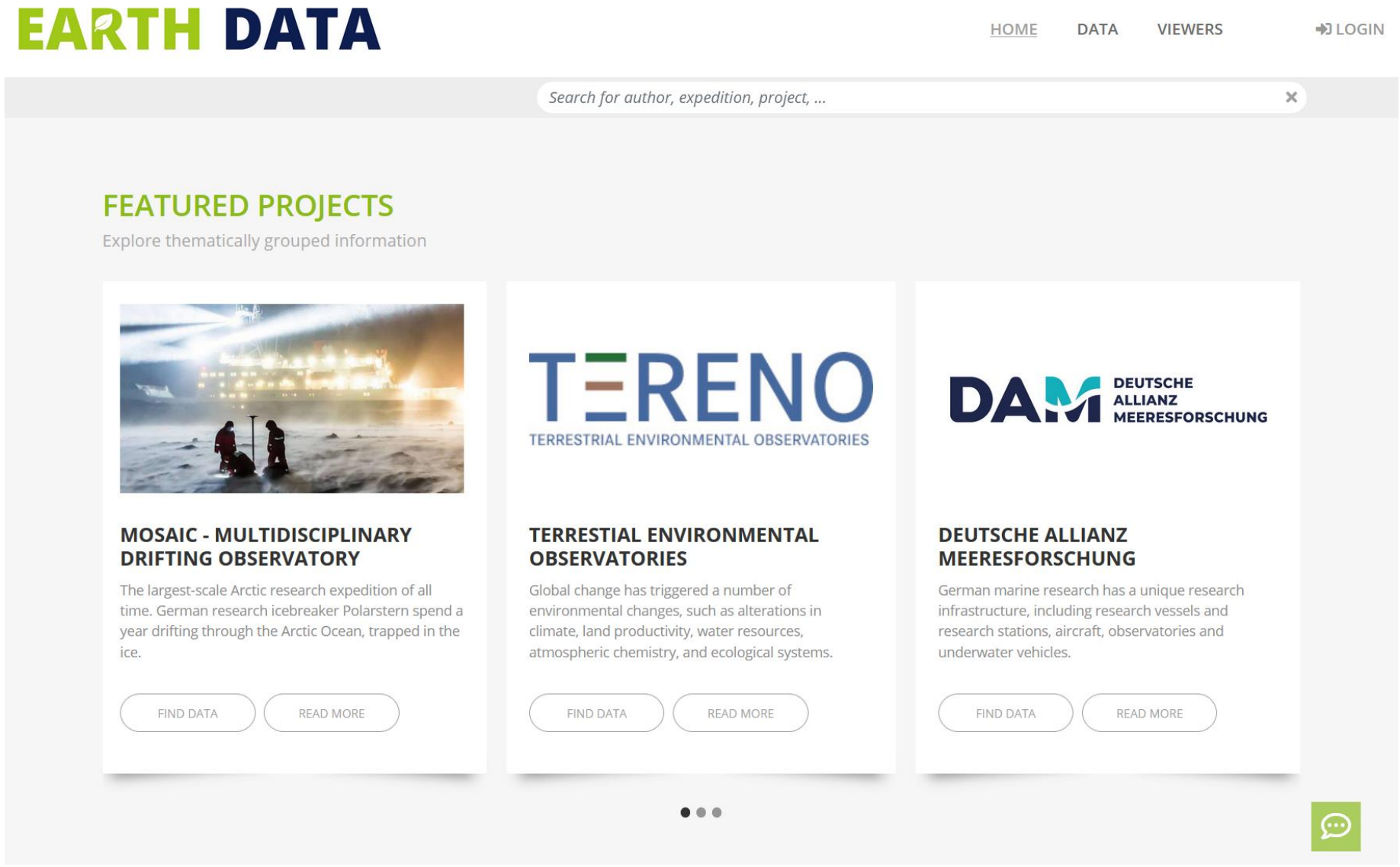
HELMHOLTZ SPITZENFORSCHUNG FÜR GROSSE HERAUSFORDERUNGEN

DataHub Research Field Earth and Environment

https://datahub.erde-und-umwelt.de/

HELMHOLTZ

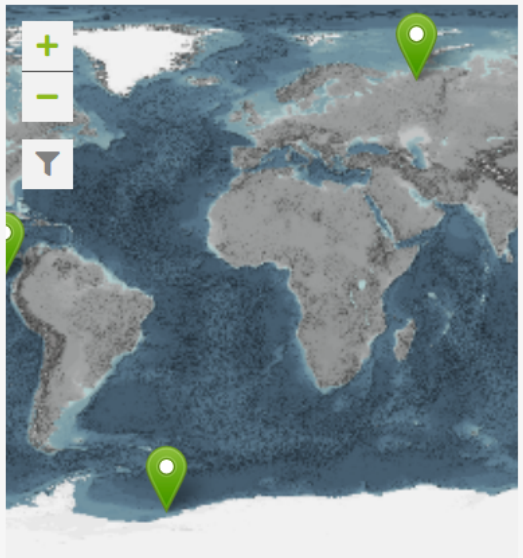
Portal/Viewers in earth-data.de, NFDI-contribution



Aggregator for repositories and SDIs

Search for Data and more...

MAP



Temporal Coverage

Resource Type

- Dataset (420,467)
- Text (50,319)
- Images (912)
- Videos (160)
- Maps (54)
- Software (14)

DATASETS | PUBLICATIONS | REPORTS | MAPS

SORT BY: RELEVANCE | DATE

Core/Section summary of ODP Hole 204-1250C (2005)

Rack, Frank R; Bohrmann, Gerhard; Tréhu, Anne M; Shipboard Scientific Party,

https://doi.org/10.1594/PANGAEA.255245

https://juser.fz-juelich.de/record/882435

Dicengxue-zazhi : jikan = Journal of stratigraphy

https://juser.fz-juelich.de/record/74936

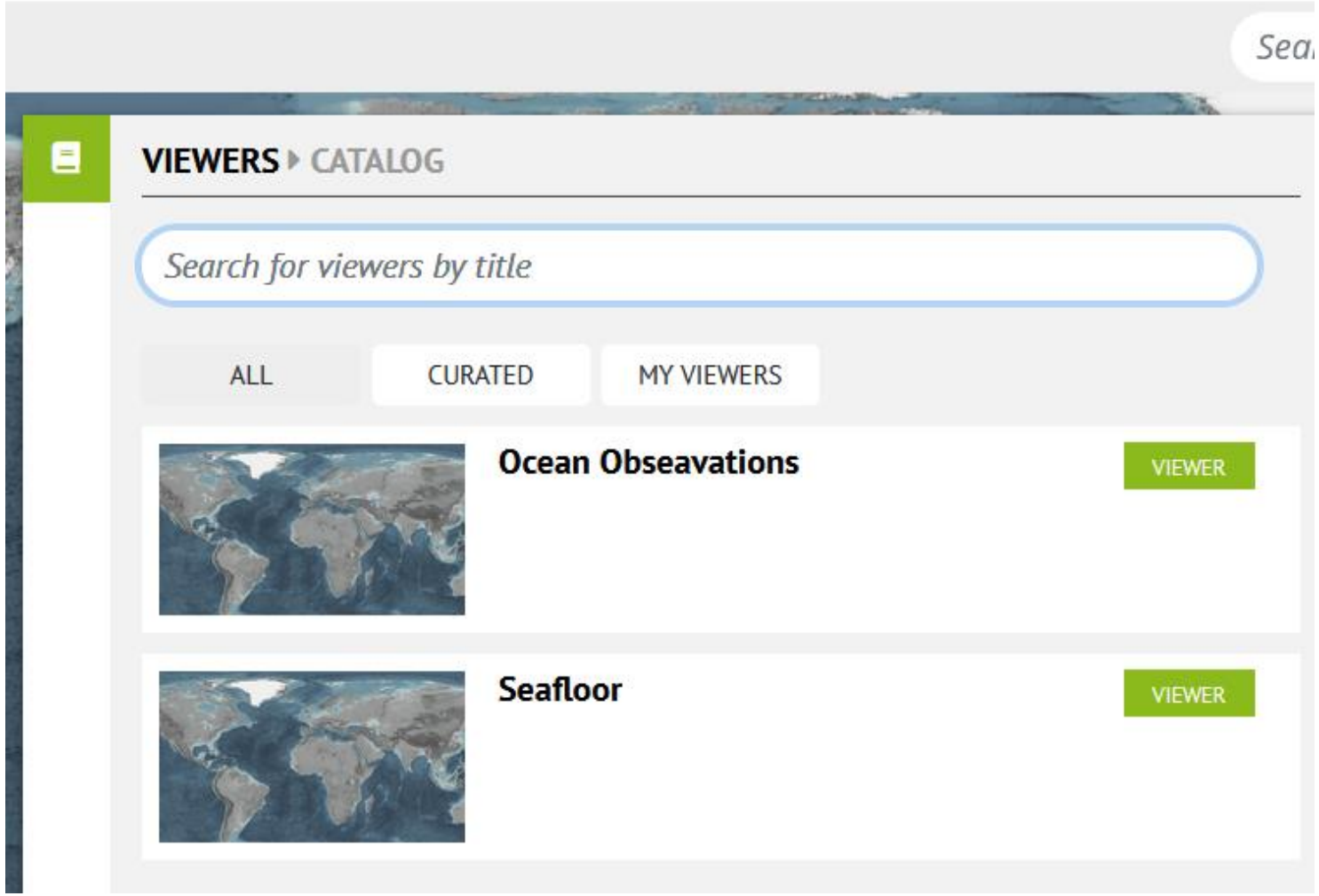
EPOCA Svalbard 2010 mesocosm experiment: depth-integrated (0-12m) phytoplankton community composition using marker pigments (CHEMTAX) (2013)

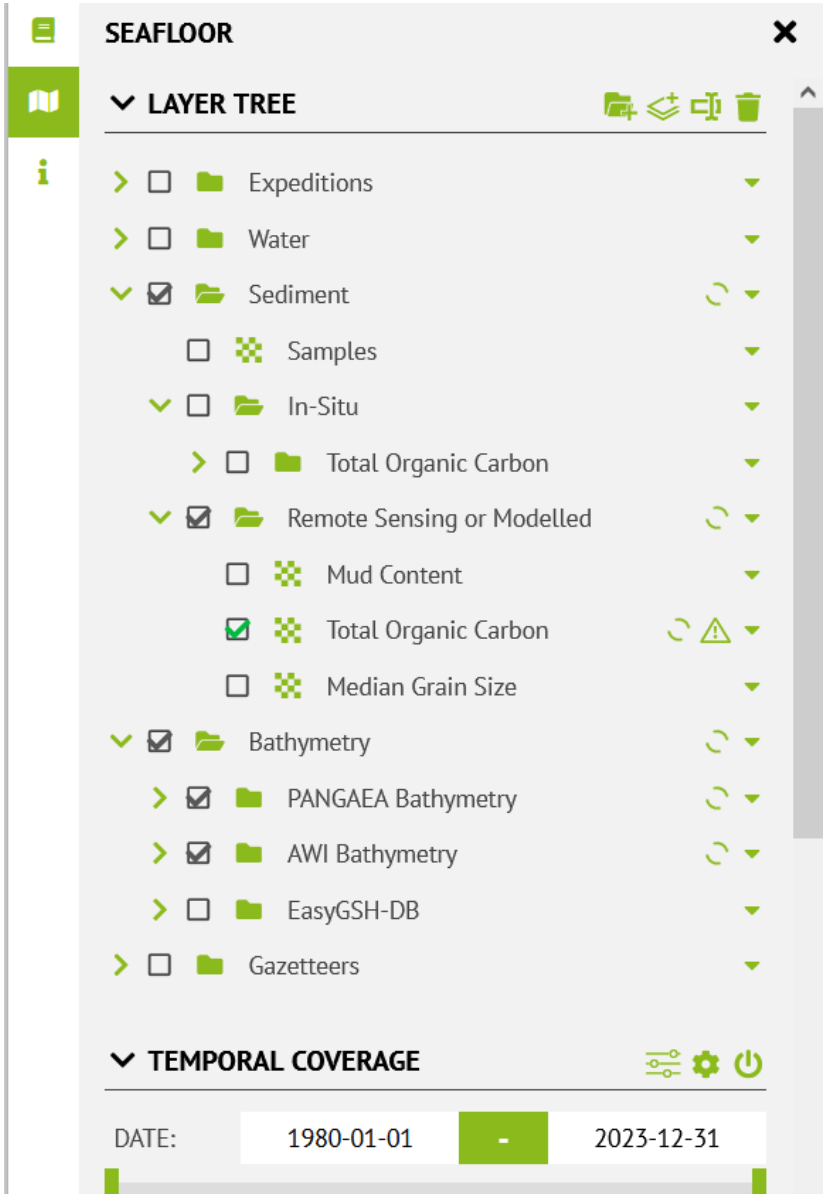
Schulz, Kai Georg

https://doi.org/10.1594/PANGAEA.815225

Viewer-Examples, embed SDIs by OGC-WCS

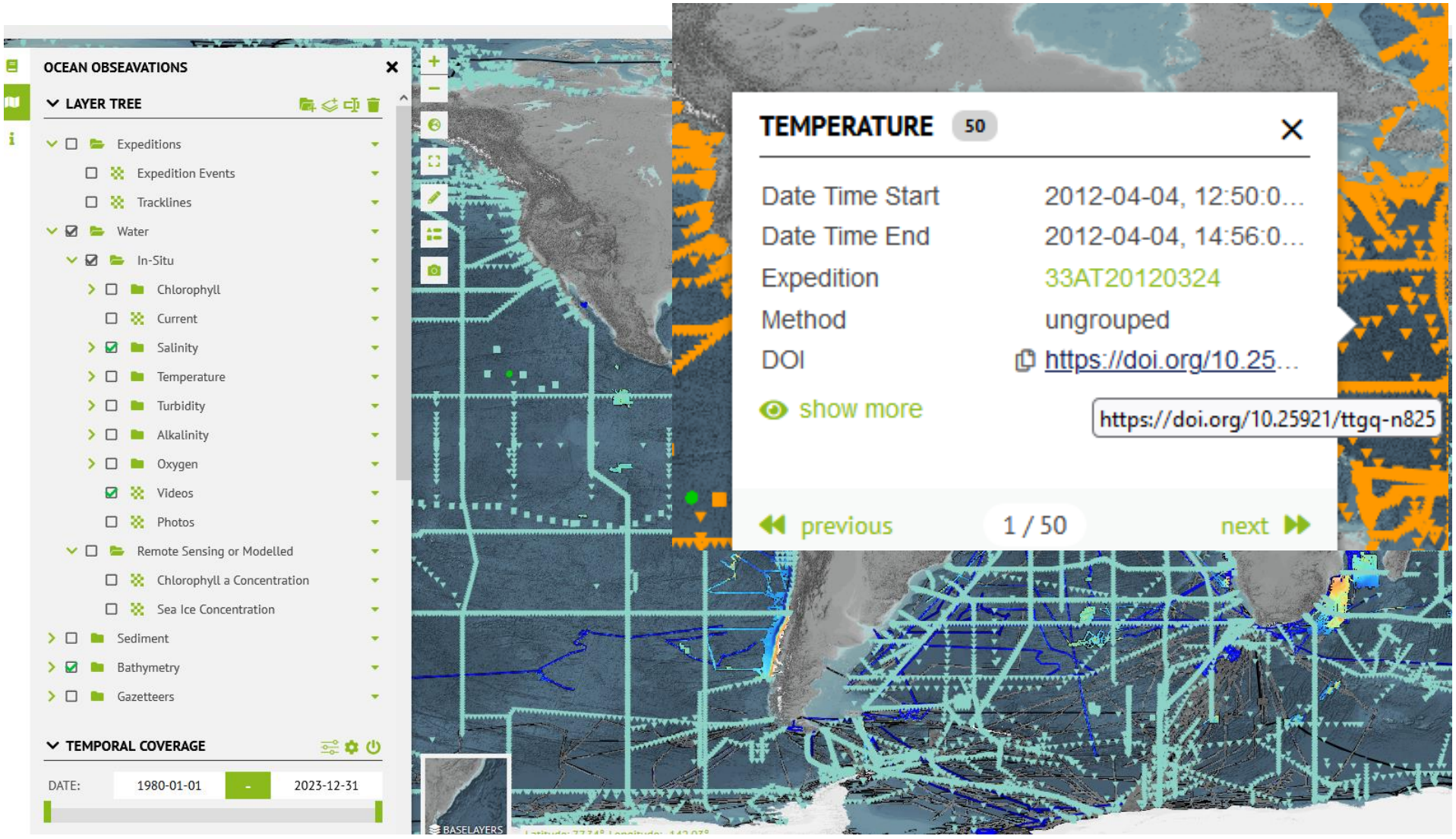
EARTH DATA





- Concept for Data Sharing/Exploration also in Projects, allows exploring governmental data (GDI-DE)

Data exploration across different sources



Interactive for meta-data exploration

Earth-data originates from DAM-portal marine-data.de
(whole german marine research community)

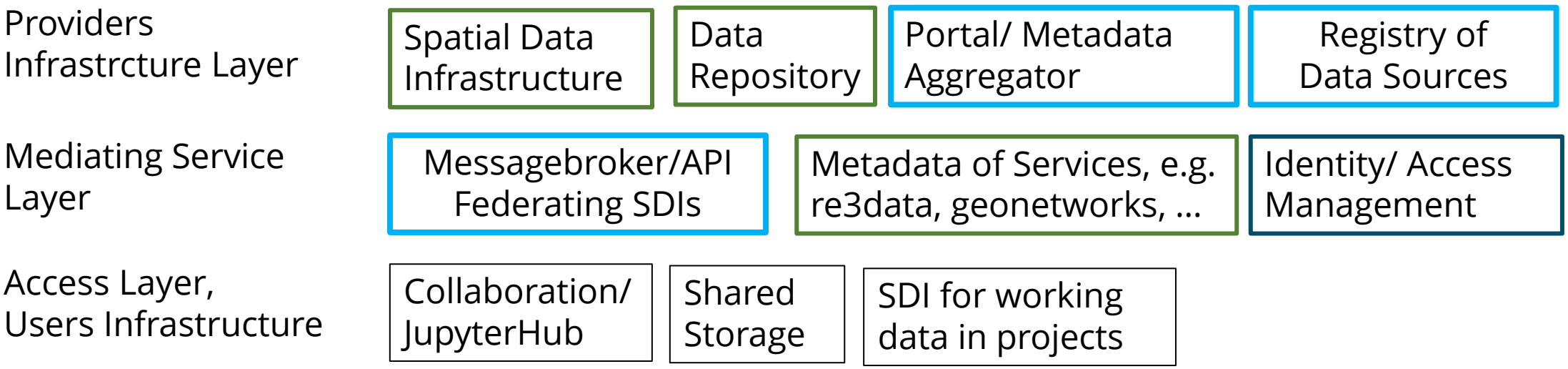
Interoperable Infrastructures in DataHub

beyond common file/ metadata formats and provision for download

DataHUB Earth&Environment implements:

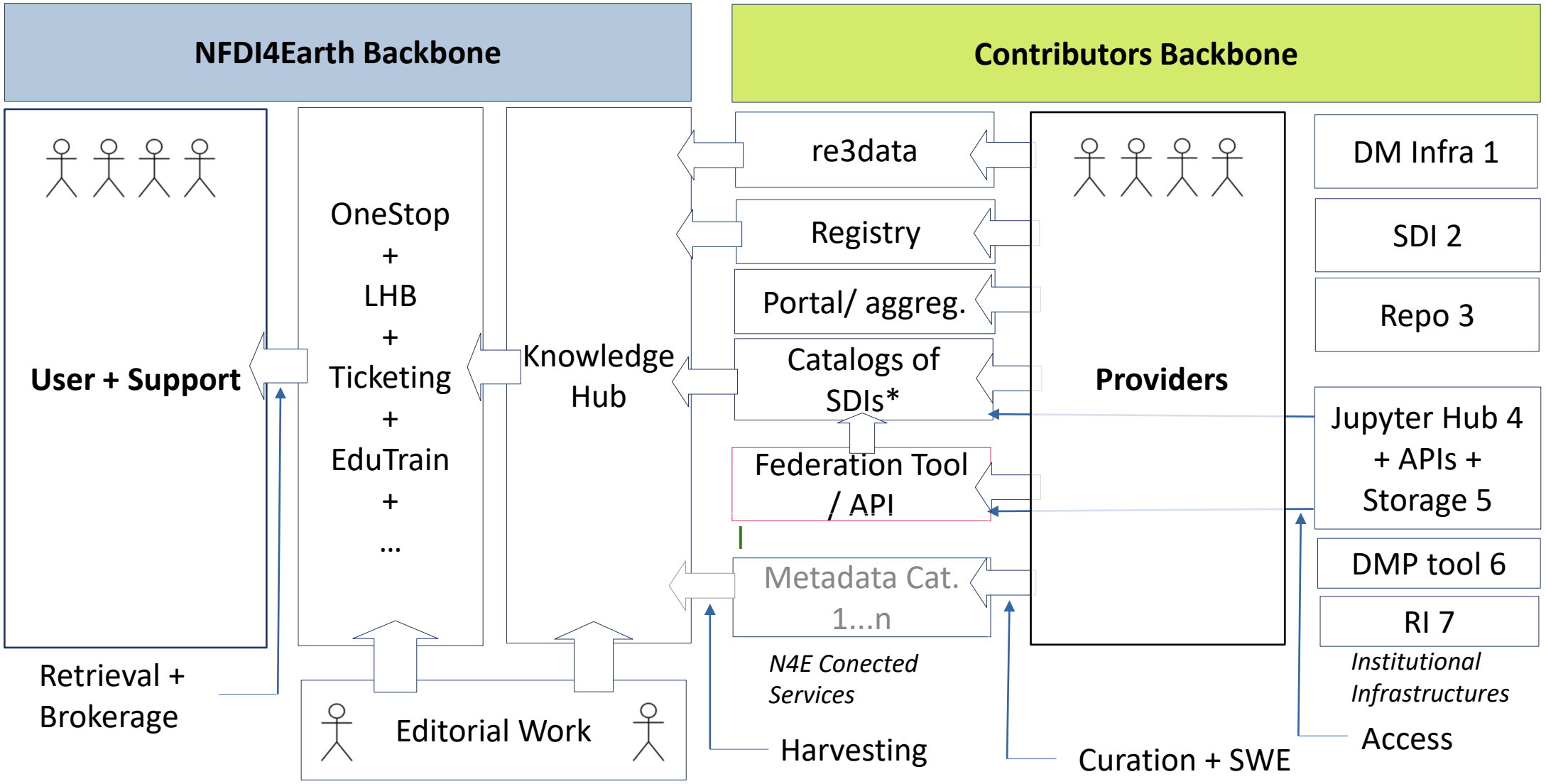
- Metadata aggregation ► earth-data.de
- Aggregation of data from different Spatial Data Infrastructures by **OGC-CSW**-standard (like those in GDI-DE)
  ► „Thematic Viewer“ demonstrators
- Sensor-metadata management within **HMC-project STAMPLATE** (**OGC-STA**) ► Time series management and access
- **Message broker-API** for federating across data sources
  ► part of **CAPRICE** NFDI4Earth Pilot in HPC

Layers of Architecture



- DataHub Earth & Environment
- Other NFDI-Participants' Services, GDI-DE, ...
- DFN, Helmholtz-AI, BASE4NFDI

Architecture Sketch Infrastructure Contributions



*e.g. GDI-DE, DataHUB ServCat

Some Actions in the Measure 3.1

- **Gap-Analyses** of Pilots and Incubators to improve services

- Develop and promote **NFDI4Earth-Label** on FAIR data + services,
  federated usability, international visibility, sustainablilty
  - Research on data quality in re3data
  - Use HMC FAIR Dashboard (FUJI-Tools)
  - Consultations with providers

- -> Push **standards for catalogs** for harvesting metadata into KnowledgeHub
  - STAC, WCS, STA

- embed DataHub developments

Thank You !

https://www.nfdi4earth.de/

Please contact Stephan.Frickenhaus.AT.awi.de

HELMHOLTZ

NFDI-related activities at DESY

Helmholtz Open Science Forum on the National Research Data Infrastructure (NFDI)

Anton Barty, DESY-FS Scientific Computing
22 June 2023 - online

https://gfz-potsdam-de.zoom.us/meeting/register/u5Eld--qrzwqE9KHQ8teHh4qquGGSd0SHxG2
https://os.helmholtz.de/veranstaltungen/foren/3-nfdi-forum/#c115105

HELMHOLTZ

DESY.

Helmholtz Open Science Forum on the NFDI

June 22, 2023, 10:00 - 15:30

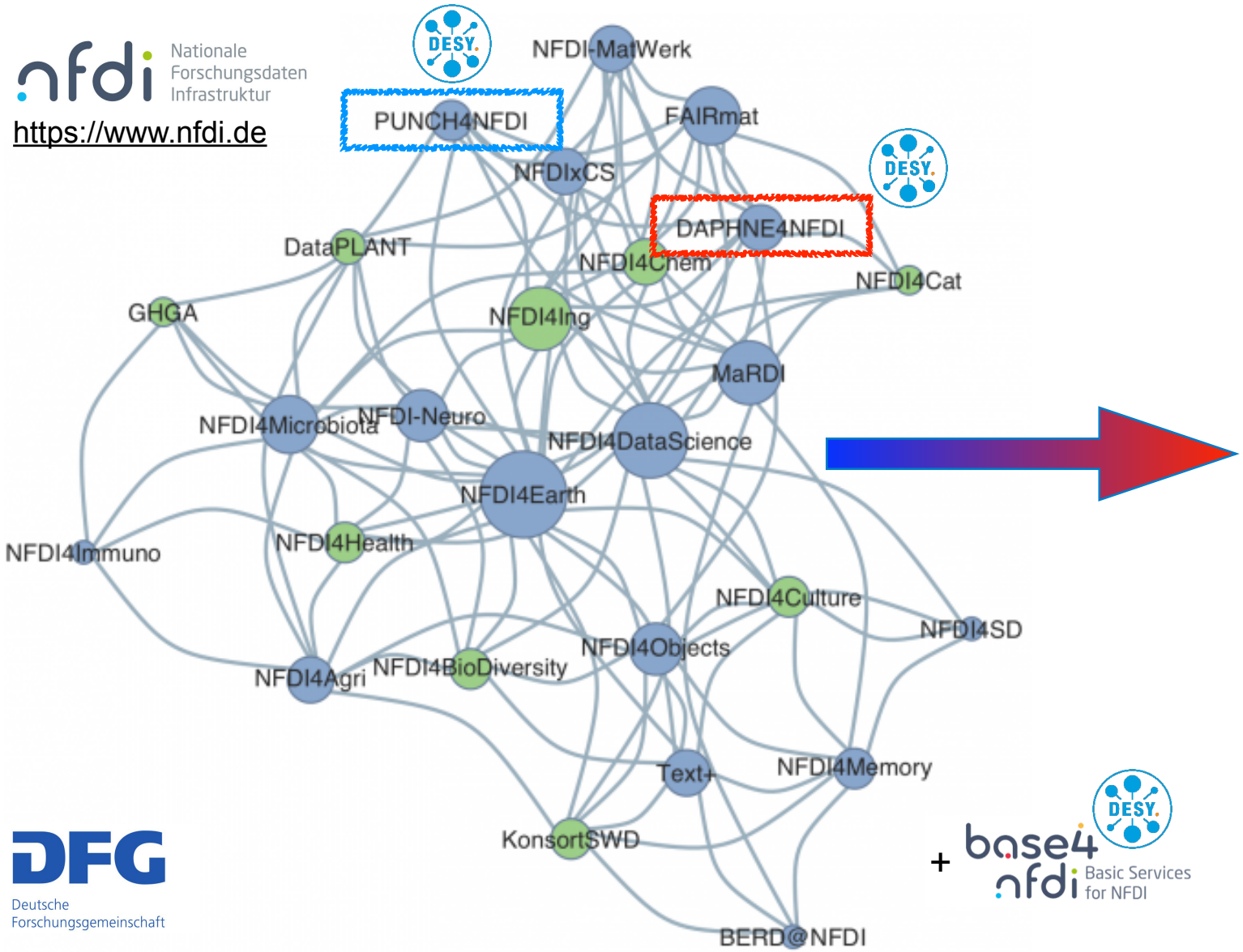
Time	Topic	Speaker
10:00–10:10	Welcome	Roland Bertelmann, Helmholtz Open Science
10:10–10:25	AWI's engagement in NFDI4Earth and synergies with DataHub Earth & Environment	Stephan Frickenhaus, AWI
10:25–10:40	Coordination of NFDI-related activities at DESY	Anton Barty, DESY
10:40–10:55	NFDI activities from the perspective of DKFZ	Christian Busse, DKFZ
10:55–11:10	NFDI activities from the perspective of DZNE	Gisela Schmidt, DZNE
11:10–11:20	Coffee Break	
11:20–11:35	Helmholtz Perspectives on Base4NFDI: HMC	Sören Lorenz, GEOMAR, HMC Speaker
11:35–11:50	Helmholtz Perspectives on Base4NFDI: HIFIS	Uwe Jandt, DESY, HIFIS Overall Management
11:50–12:05	Helmholtz Perspectives on Base4NFDI: PID4NFDI	Antonia Schrader, Helmholtz Open Science Office
12:05–13:00	Lunch Break	
13:00–13:15	NFDI activities from the perspective of GEOMAR	Hela Mehrtens, GEOMAR
13:15–13:30	NFDI activities from the perspective of GSI	Mohammad Al-Turany, GSI
13:30–13:45	Bridging the gap between user-friendliness and cutting-edge research: HZB in the NFDI	Tamara Husch Lee, HZB
13:45–14:00	NFDI activities from the perspective of HZDR	Uwe Konrad, HZDR
14:00–14:10	Coffee Break	
14:10–14:25	NFDI activities at the HZI	Alice McHardy, HZI
14:25–14:40	NFDI activities from the perspective of GFZ	Valentina Protopopova-Kakar, GFZ
14:40–14:55	NFDI networking at KIT	Alexandra Axtmann & Jan Kröger, KIT
14:55–15:10	NFDI Activities at MDC	Katharina Nimptsch & Carolina Schwedhel, MDC
15:10–16:30	Wrap-up and final discussion	

DESY runs two consortia in the wider NFDI network

Based around physical sciences topics, approved in second proposal round

nfdi Nationale Forschungsdaten Infrastruktur

https://www.nfdi.de



Horizontal cross-cutting sections within NFDI.eV

Common Basic Infrastructure Section:

- Long term archiving
- Identity management
- Federated cloud
- Research Software Engineering
- Data integration
- AI / ML

Metadata Section:

- Persistent identifiers
- Semantic interoperability and terminology services
- Ontology harmonisation and mapping
- Provenance verification

Ethical and legal issues Section

Training and education Section

Industry Engagement Section

DFG Deutsche Forschungsgemeinschaft

+ DESY. base4nfdi Basic Services for NFDI

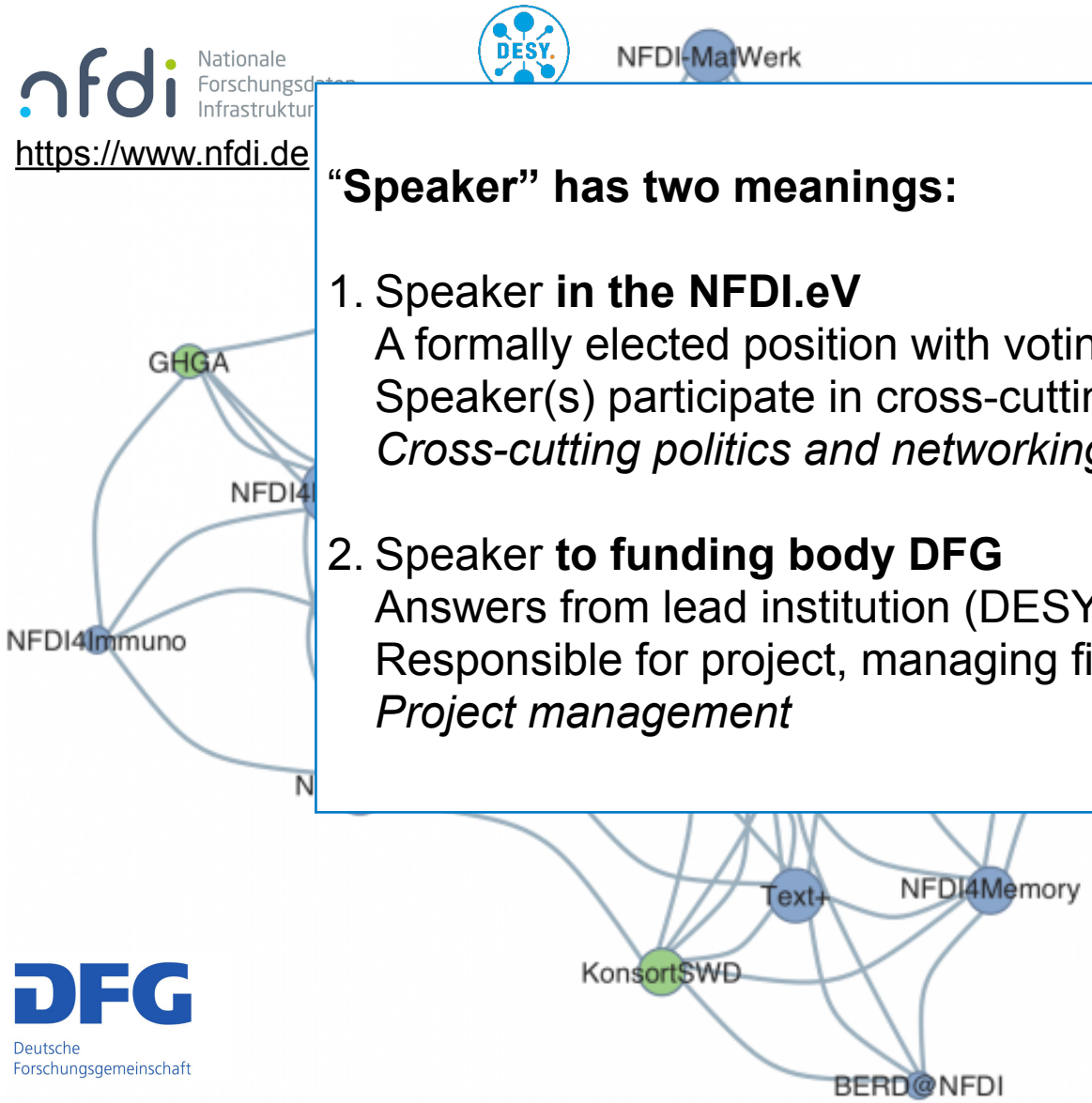
…meetings, meetings, and more meetings

DESY runs two consortia in the wider NFDI network

Based around physical sciences topics

nfdi Nationale Forschungsdaten Infrastruktur

https://www.nfdi.de



“Speaker” has two meanings:

1. Speaker **in the NFDI.eV**
   A formally elected position with voting rights in NFDI.eV meetings
   Speaker(s) participate in cross-cutting topical meetings, NFDI sections
   Cross-cutting politics and networking

2. Speaker **to funding body DFG**
   Answers from lead institution (DESY) to funding body (DFG)
   Responsible for project, managing finances, budgets, answerable to DFG
   Project management

Horizontal cross-cutting sections

Section:

gineering

ty and

n and mapping

on

Training and education Section

Industry Engagement Section

…meetings, meetings, and more meetings

DFG Deutsche Forschungsgemeinschaft

PUNCH4NFDI

PUNCH4NFDI - Particles, Universe, NuClei and Hadrons for the NFDI

Speaker (BLV): Thomas Schörner (**DESY**)
Speaker (NFDI): Andreas Haungs (KIT), deputy Christiane Schneide (**DESY**)

- **20 Co-applicant institutions:** DESY, FZJ, GSI, KIT, 11 universities, 2 Max-Planck institutes, 1 Leibniz institute, 2 other
- **22 Participant institutions:** HZDR, DLR, CERN, DPG, 12 universities, 3 Leibniz institutes, 1 Max-Planck institute, 2 other

- **7 Task Areas**
  - TA1: Management and governance (**lead DESY**)
  - TA2: Data management (**co-lead DESY**)
  - TA3: Data transformations
  - TA4: Data portal (**co-lead DESY**)
  - TA5: Data irreversibility
  - TA6: Synergies & services (**co-lead DESY**)
  - TA7: Education, training, outreach, citizen science

- **Funded personnel in total:** 179 person years
  - At DESY: 20.18 in total

+ strong links to ErUM data

- Active members at DESY: currently 14

Example: The A4 Experiment and the PATOF HMC Project

DESY+HIM

A4: Nuclear physics experiment on strange quark contribution to nucleon form factor in Mainz; dismantled since 2004

- Rich physics output, but potential for more.
- Remainders:
  - Data as ROOT files with exp.-specific class files in nexted folder structure with unclear context;
  - ASCII files provide minimal metadata, not machine-readable
  - Electronic logbook with context (XML)

"From the Past To the Future: Legacy Data in Small and Medium-Scale PUNCH Experiments – a Blueprint for PUNCH and Other Disciplines": PATOF HMC project (2nd HMC call)

- Make A4 data reusable and create template for future experiments

Status

- A4: Storage and metadata schema defined, ROOT-problems addressed, ASCII converted etc. ➔ FAIR
- Assessment according to FAIR data maturity model (right)
- PATOF: Finalise FAIRification of A4, put together cookbook of FAIR recipees and build "FAIR metadata factory"
- Note: Careful with discussion on data sets as KPIs (e.g. POF)

FDMM ID	Indicator	Priority	Comment
RDA-I1-01M	Metadata uses representation in standardized format	Important	☺ (DataCite)
RDA-I1-01D	Data uses representation in standardised format	Important	☺ (root/class)
RDA-I1-02M	Metadata uses machine-understandable knowledge representation	Important	☺ (xml/DataCite)
RDA-I1-02D	Data uses machine-understandable knowledge representation	Important	☹ (root/class)
RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	Important	☺ (DataCite)
RDA-I2-01D	Data uses FAIR-compliant vocabularies	Useful	☹ (root/class)
RDA-I3-01M	Metadata includes references to other metadata	Important	☺ (e.g. ORCID)
RDA-I3-01D	Data includes references to other data	Useful	☹ (root)
RDA-I3-02M	Metadata includes references to other data	Useful	☺ (via ELN)
RDA-I3-02D	Data includes qualified r		
RDA-I3-03M	Metadata includes qualifi		
RDA-I3-04M	Metadata include qualified references to other data		

FAIR data maturity model
https://zenodo.org/record/3909563

DAPHNE4NFDI

Data from Photon and Neutron Experiments

Speaker (BLV): Anton Barty (**DESY**)
Speaker (NFDI): Bridget Murphy (CAU),
Astrid Schneidewind (FZ Jülich)

- **18 Co-applicant institutions**

- **6 Task Areas**
  - TA1: Managing data production and collection
  - TA2: Data repositories and catalogues
  - TA3: Infrastructure for data evaluation and reuse
  - TA4: Outreach and Education
  - TA5: Politics and networking
  - TA6: Management (lead DESY)

- **Funded personnel in total:** about 40 FTE
  - Funded at DESY: currently 5
  - Active at DESY: about 15 across FS and IT (in FH)

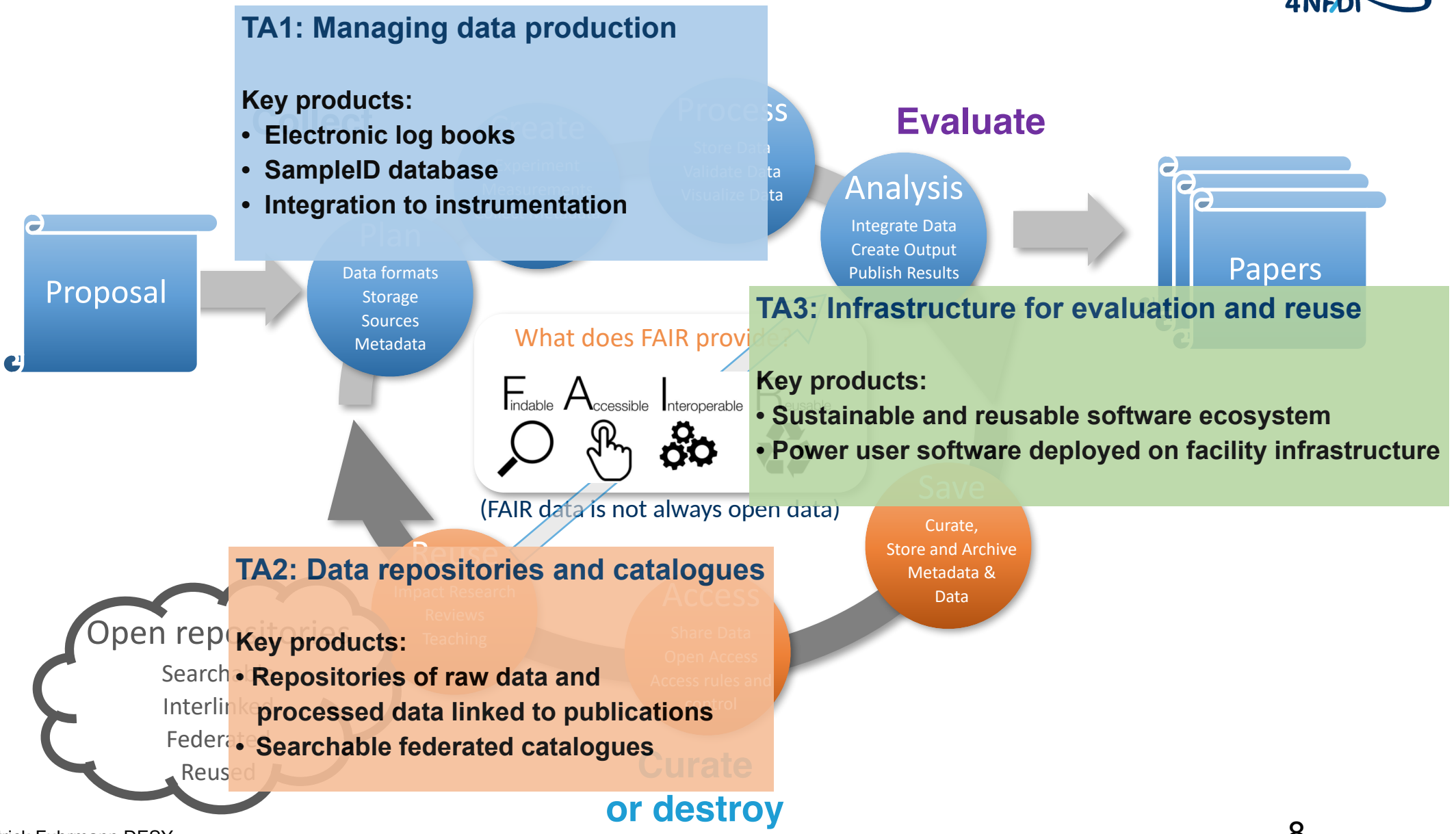
Funded partners

11 Universities

7 Large scale facilities

Voluntary partners

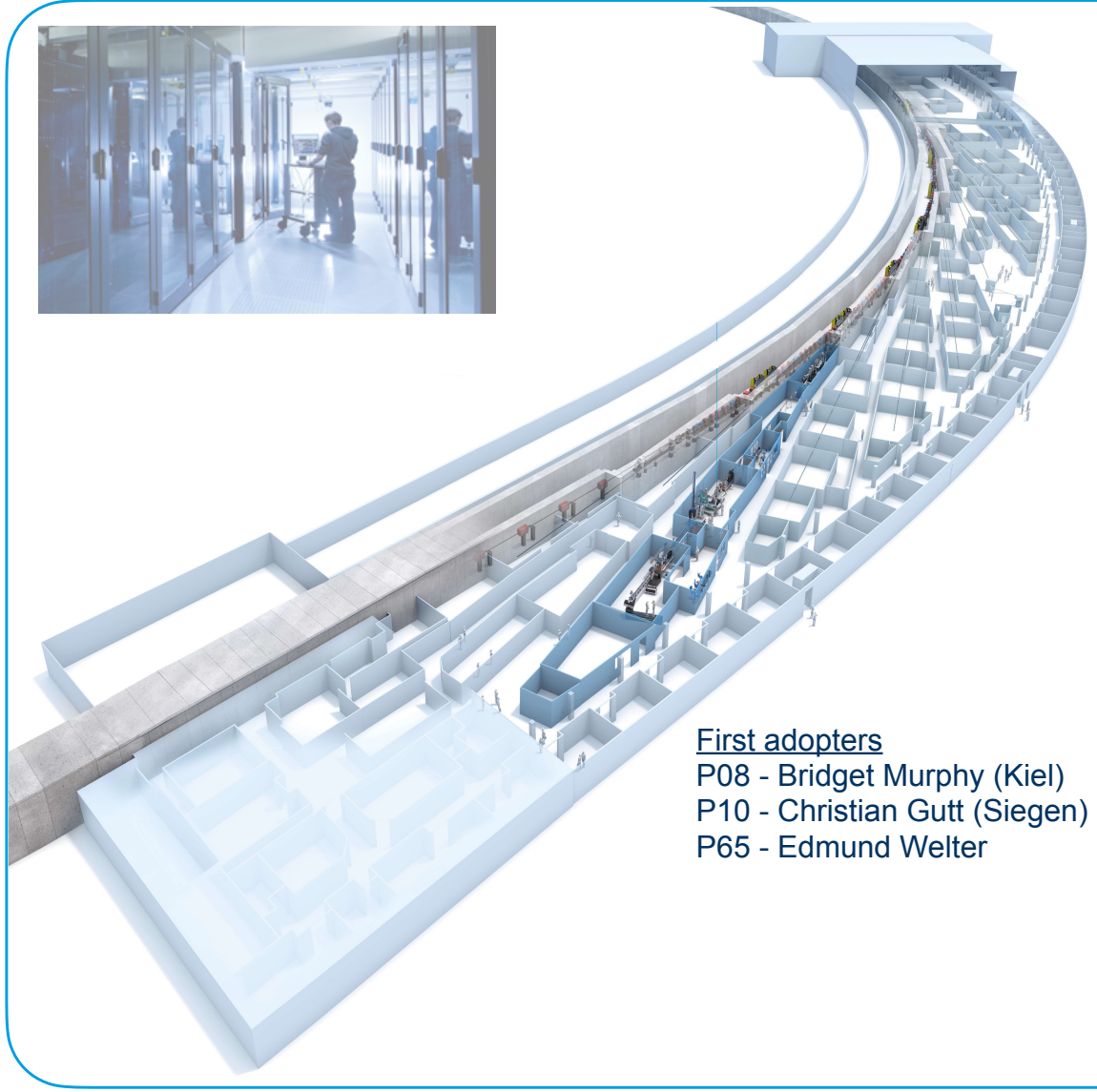
DAPHNE infrastructure for the research data life cycle



Graphic: Patrick Fuhrmann DESY

Specific DAPHNE4NFDI activities at DESY

My home institution



First adopters
P08 - Bridget Murphy (Kiel)
P10 - Christian Gutt (Siegen)
P65 - Edmund Welter

SciCat deployment:
- Improved data portal
- Prototype under development
- Data access management using federated login

Standardising file formats:
- Move beamlines to NeXUS standards
- Ontology harmonisation and semantic interoperability

Electronic logbooks:
- Integrate with control system and SciCat
- Sample PIDs (IGSN)
- Metadata harvesting, and more…
- Sample identifiers (unique sample IDs),
- Variety of community solutions a challenge

Open catalogue(s) of published data:
- A place to put any processed data related to publications
- Index by paper / DOI

Power user software:
- Upgraded software, sustainable, available to all users
- Uses latest file formats, proper RSE
- Validated standard analysis pipelines
- Space for hosting open software and data is needed

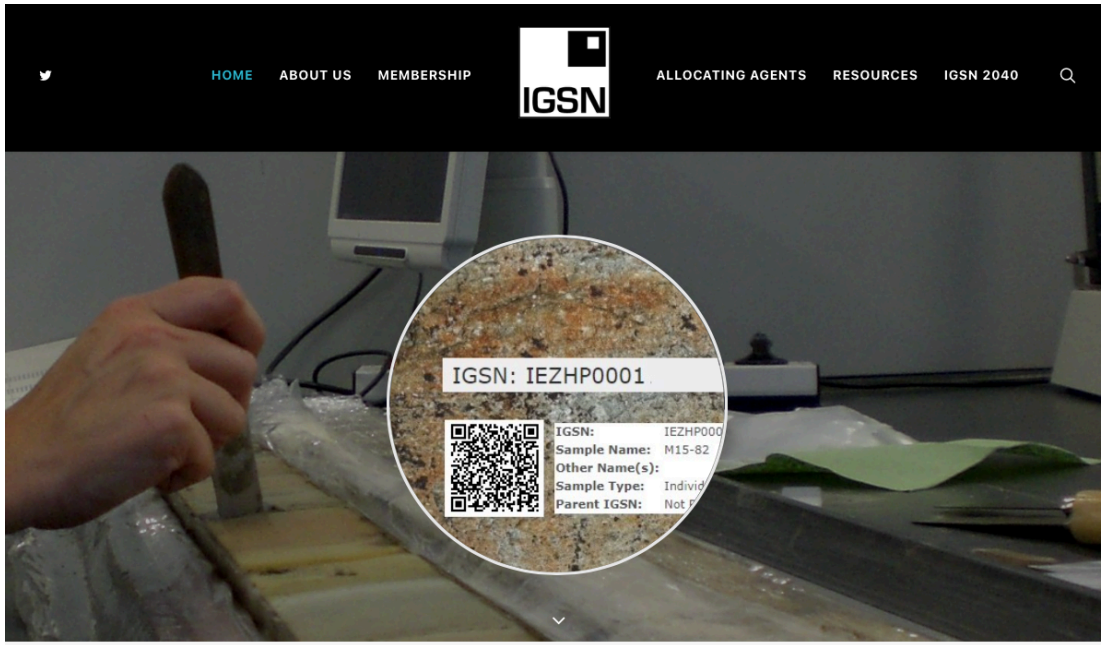
Daphne is introducing unique sample identifiers

Tracking samples from creation through to data and publication

- Uniquely identify samples so that they can be tracked through logbooks and datasets
- Identifier should be unique and persistent - even though samples themselves may not always persistent
- Must be simple, easy to use, minimal paperwork overhead

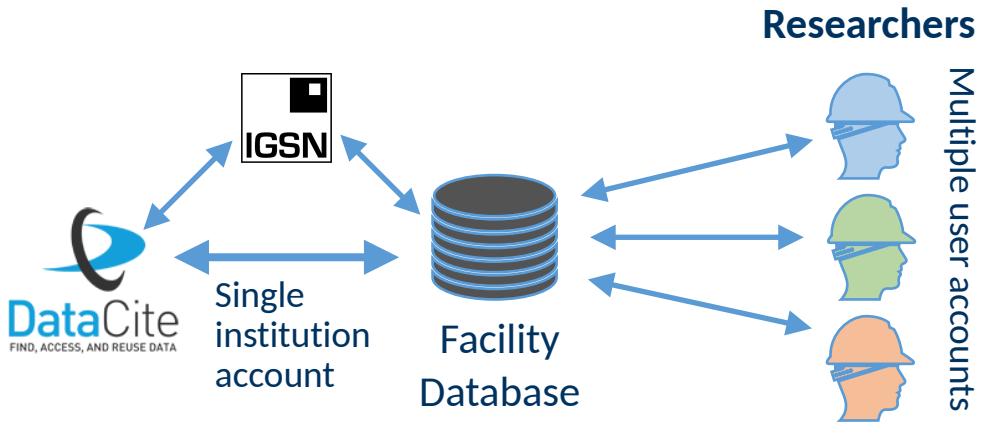
The IGSN* system has been developed for other disciplines

IGSN is a globally unique and persistent identifier for material samples.



https://www.igsn.org/ https://ardc.edu.au/services

* International ~~Geo~~ Generic Sample Number



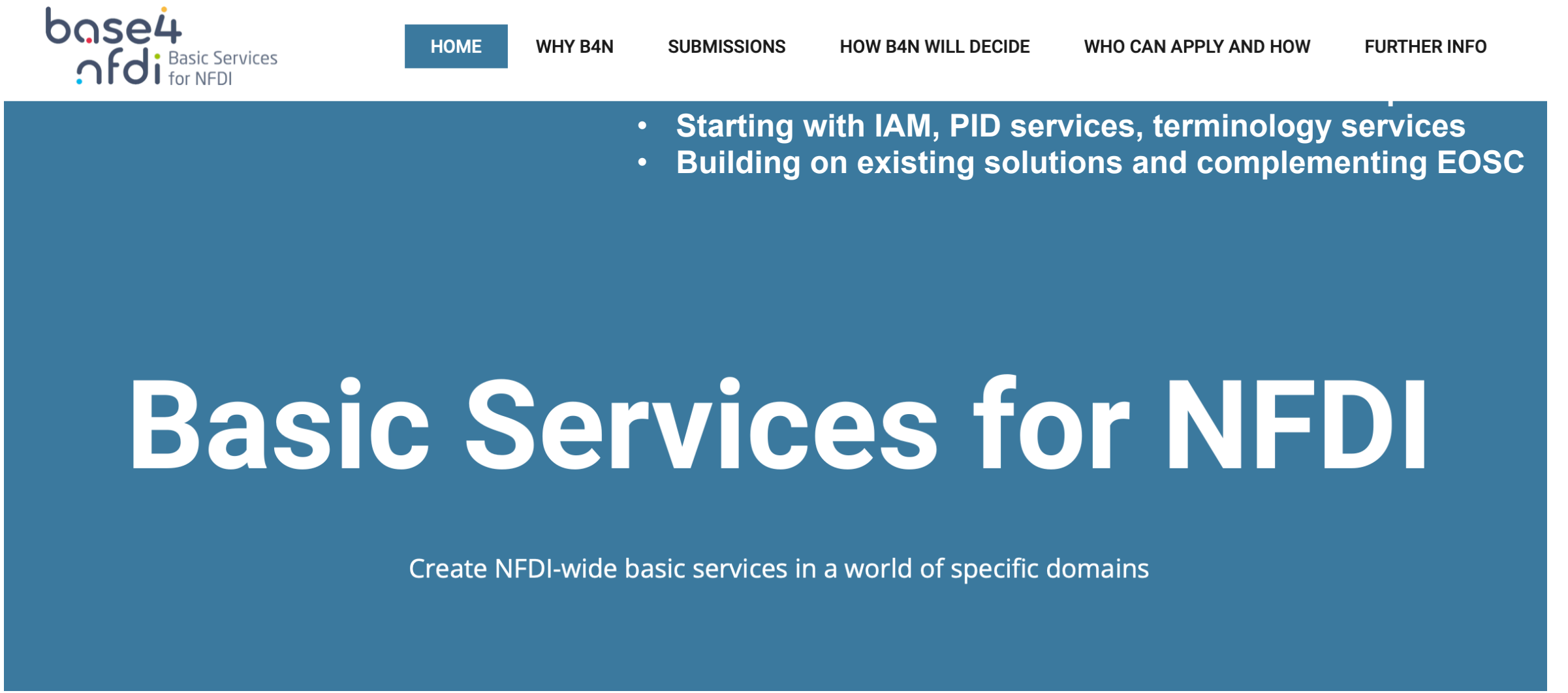
DataCite FIND, ACCESS, AND REUSE DATA https://datacite.org

IGSN now works with DataCite

In September 2021, IGSN e.V. and DataCite entered a partnership under which DataCite will provide the IGSN ID registration services and supporting technology to enable the ongoing sustainability of the IGSN PID infrastructure.

Base4NFDI – a Base Service Initiative Across Consortia

DESY and Helmholtz are engaged in NFDI basic services



Foundation of Base4NFDI

NFDI Sections identify cross-cutting needs of consortia

base4nfdi Basic Services for NFDI

- Act as **incubators for continuously identifying potential basic services**
- Combine **infrastructural / technological expertise** and **domain knowledge**

My personal summary:

BASE4NFDI is HIFIS, HMC, etc. for the NFDI community

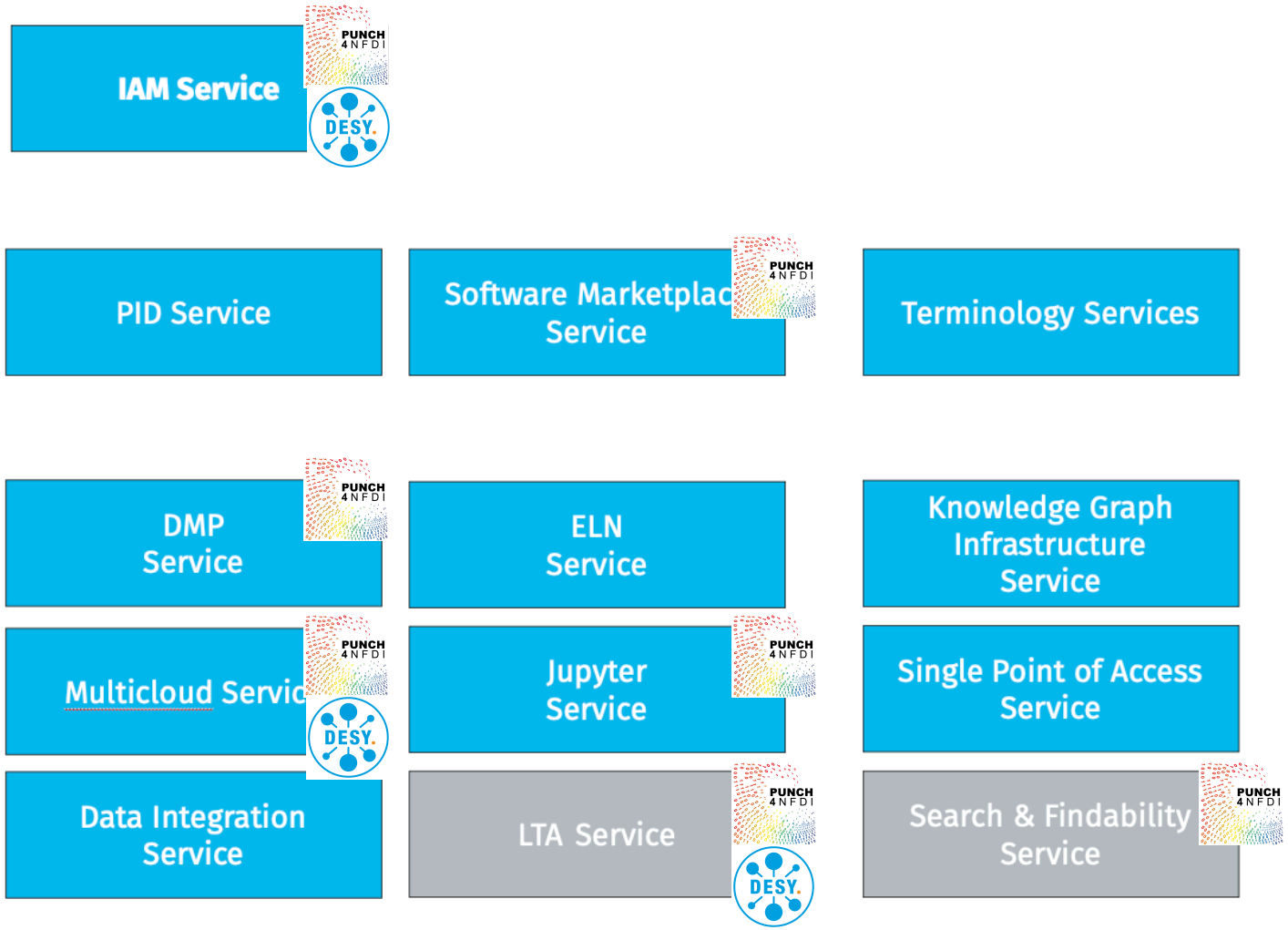
DESY:

- Co-Spokesperson Thomas Schörner & RFP Section Liason Officer
- Patrick Fuhrmann as a TEC member in Base4NFDI

Base4NFDI
Landscape of Basic Service Proposals

base4nfdi Basic Services for NFDI

nfdi Nationale Forschungsdaten Infrastruktur



Other activities at DESY related to NFDI

The NFDI projects are triggering a range of activities at the institutional level

- Data Policy, AAI implementation at DESY
- Role of the DESY library in data management
  DataCite monopoly - practical guidelines required
- Role of HIFIS in NFDI consortia
  PUNCH4NFDI, DAPHNE4NFDI, NFDI as a whole
- Networking with EOSC
  Use case in Call: *Next generation services for operational and sustainable EOSC Core Infrastructure* (HORIZON-INFRA-2023-EOSC-01-04) and ErUM-Data
- Organization and execution of the NFDI Verwaltungskreis @DESY in May 2023
- Active participation in NFDI Task Force Evaluation and Reporting
- Active contributions to the Physical Sciences for NFDI consortium
- EOSC: building on PaNOSC and ExPaNDS

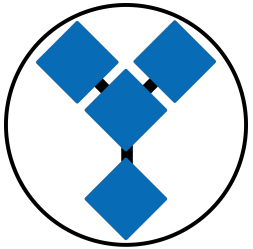
NFDI at DKFZ

Christian Busse [0000-0001-7553-905X]
Div. B Cell Immunology
German Cancer Research Center (DKFZ)

NFDI consortia at DKFZ

GHGA THE GERMAN HUMAN GENOME-PHENOME ARCHIVE

started October 2020
Applicant institution / 15 FTE



NFDI4Immuno

started March 2023
Applicant institution / 6 FTE

NFDI4 BIOIMAGE

started March 2023
Co-applicant institution / 2 FTE

dkfz.

Data types

- GHGA: Human genome, sequencing & phenome data
- NFDI4Immuno: Immunological data in general (cytometry, AB/TCR reactivity, cytokine levels, Ig/TCR repertoire sequencing)
- NFDI4BIOIMAGE: Microscopy and spatially-resolved data (not medical imaging)

Cross-consortia areas of expertise

- Spatial Omics
  - Numerous technical platforms (multiplex microscopy, multiplexed ion beam imaging, 10X Xenium, ...)
  - Linking biological information based on spatial position
- Sensitive data (GDPR)
  - Legal and ethical frameworks
    - Informed consent
    - Contractual relationships
    - Patient engagement
  - Controlled access: authentication/authorization infrastructure and standards
  - IT security: Necessary technical and organisational measures
  - Anonymization
- Enabling AI & ML

Facilitating exchange between NFDI and other RDM stakeholders

- Health & clinical research
  - genomDE / Medical Informatics Initiative (MII)
  - German National Cohort (NAKO)
- Infrastructure initiatives
  - I3D:bio - Information Infrastructure for BioImage Data
  - Genomic Data Infrastructure (GDI)
  - AIRR Data Commons (ADC)
  - German Network for Bioinformatics Infrastructure (de.NBI)

Questions

- How do NFDI activities integrate with the long-term objectives your Helmholtz center?
- How to build interoperable long-term infrastructures for AI?

Thank you for your attention!

More information:

- DKFZ: https://www.dkfz.de/en/datascience/
- GHGA: https://www.ghga.de
- NFDI4Immuno: https://www.nfdi4immuno.de
- NFDI4BIOIMAGE: https://nfdi4bioimage.de

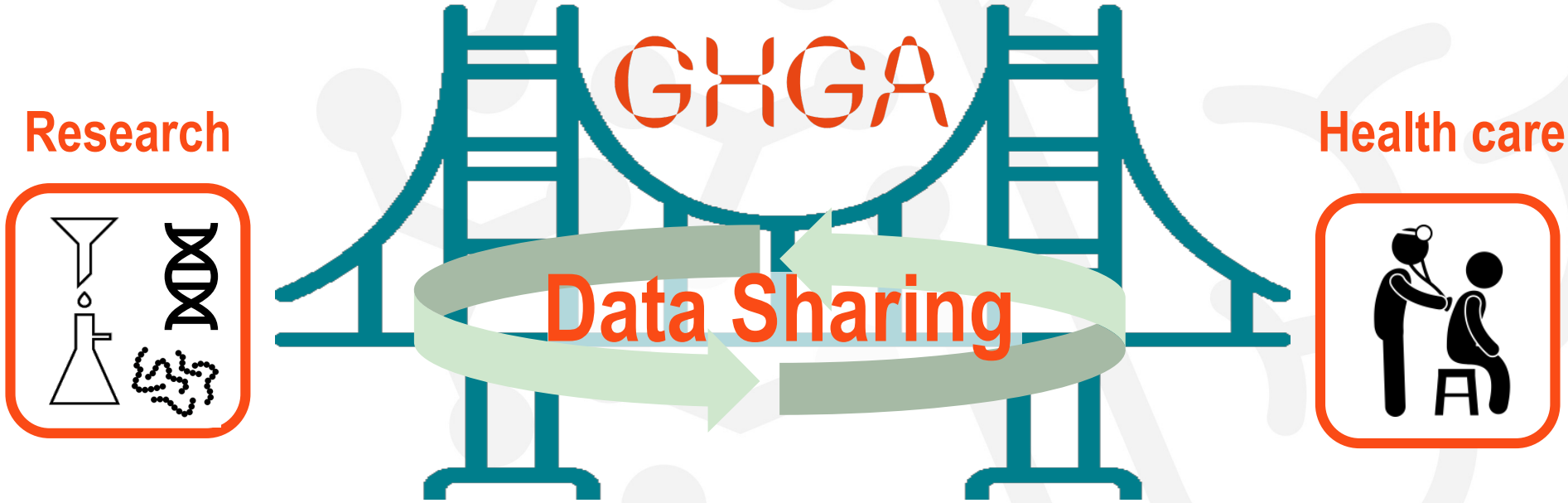
Das Deutsche Zentrum für
Neurodegenerative Erkrankungen

DZNE Involvement in GHGA

22. June 2023

Gisela Schmidt

The GHGA Vision: Enabling Genomic Medicine



Main Aims:
Establishment of a nationally coordinated, interdisciplinary infrastructure integrating genome research and healthcare

-- From Researchers for Researchers --

DZNE

DZNE Involvement in GHGA

Joachim L. Schultze
Director Systems Medicine
GHGA Co-Spokesperson

Thomas Ulas
Head of Bioinformatics at PRECISE
GHGA Deputy Workstream Spokesperson

Karoline Mauer
PhD Student in Systems Medicine
GHGA Deputy Workstream Coordinator

Involvement in the **GHGA Metadata Workstream**

- Active contribution to national harmonization and standardization of omics-data
- Development of the GHGA Metadata Model
- Members of the Operations Committee and the Steering Committee
- Via GHGA: interaction with NFDI4Health

DZNE

Projects that are planning to use GHGA

- In general: GHGA will be open for data submissions for all German research institutions
- Development with “Pioneer Projects” to develop infrastructure according to needs of the communities

GHGA Workstreams

Architecture

O. Kohlbacher
O. Stegle
L. Kuchenbecker
K. Breuer

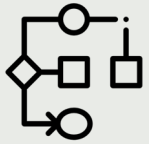
Data Hub: Operations & Data Stewardship

S. Motameny
I. Buchhalter
R. Schregle
P. Menges

Metadata

S. Nahnsen
T. Ulas
A. Iyappan
K. Mauer

Workflows



J. Gagneur
S. Ossowski / D. Hübschmann
C. Mertes
F. Heyl

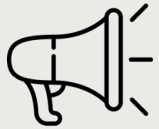
ELSI

E. Winkler
F. Molnar-Gabor
A. Bruns
S. Parker

Training

W. Huber
O. Kohlbacher
J . Philipp
N. Schatlowski

Public and Community Outreach



J. Walter
J. Winkelmann
U. Träger
N. Gasparoni

Project Management

O. Stegle
O. Kohlbacher
J. Eufinger
N. Schatlowski

Spokespersons / *Deputies*

Coordinators / *Deputies*

HELMHOLTZ
Metadata Collaboration

HMC and (Base4)NFDI

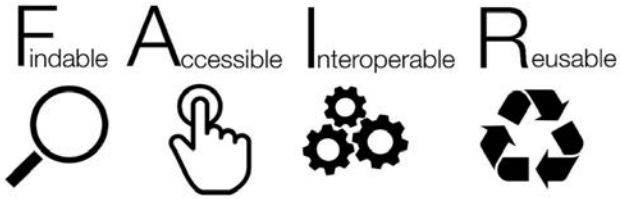
Sören Lorenz

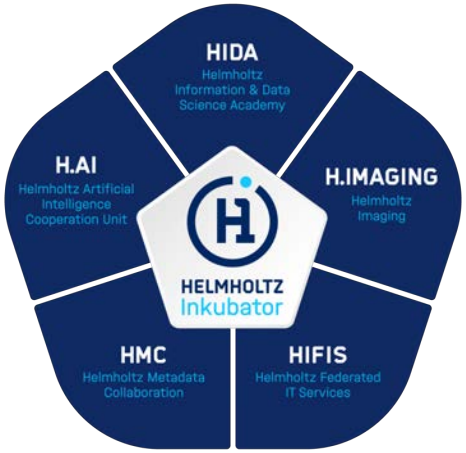
Open Science Forum NFDI,
22 June 2023

www.helmholtz-metadaten.de

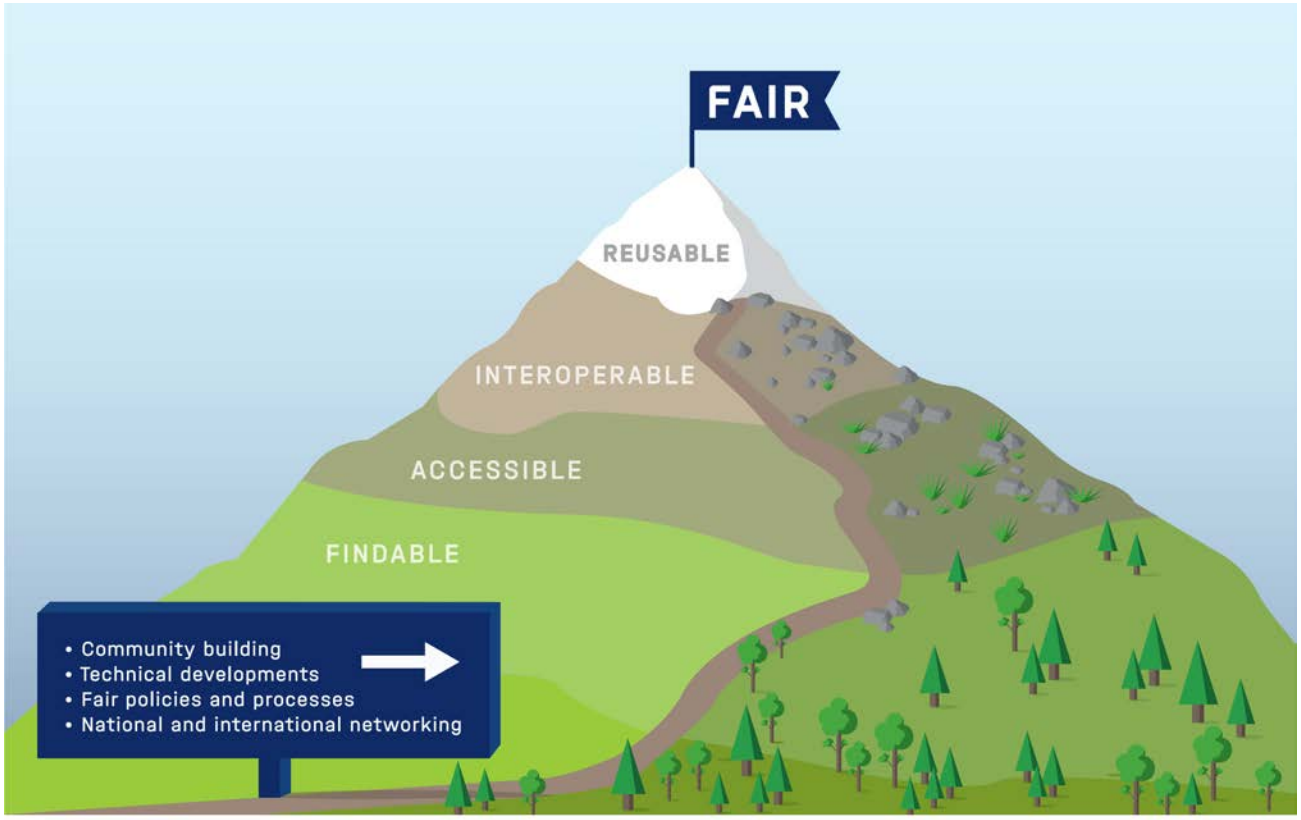
HMC Mission and Vision

- Turn FAIR into reality in Helmholtz and beyond
- Connecting the Helmholtz research data ecosystem on the level of people, organisations, and infrastructure
- Create a sustainable, distributed, semantically enriched Helmholtz data space.



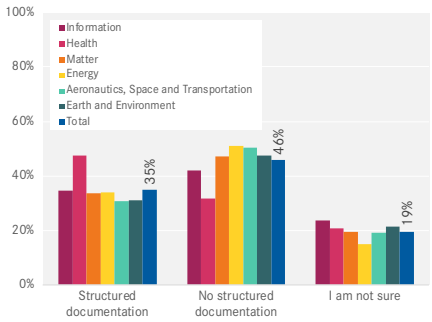


HMC Fields of Action



Some HMC measures

Community Survey



FAIR Assessment

<HMC> Informationportal

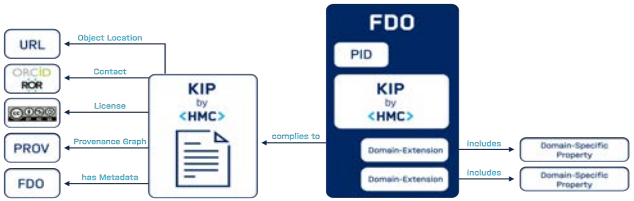
Metadata Tools

DIRSCHEMA

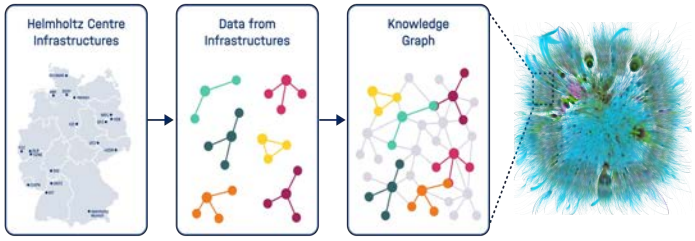
METADOR

...

FDO Implementation

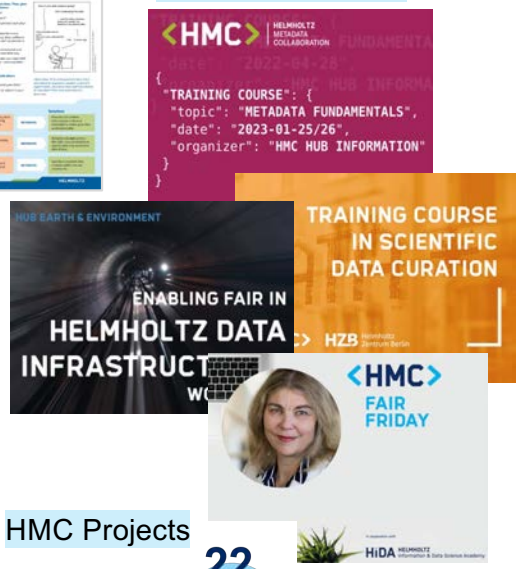


Connecting Data

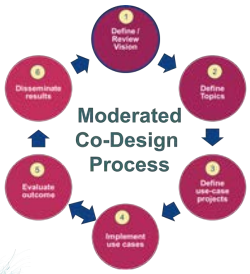


Documents

Trainings & Events



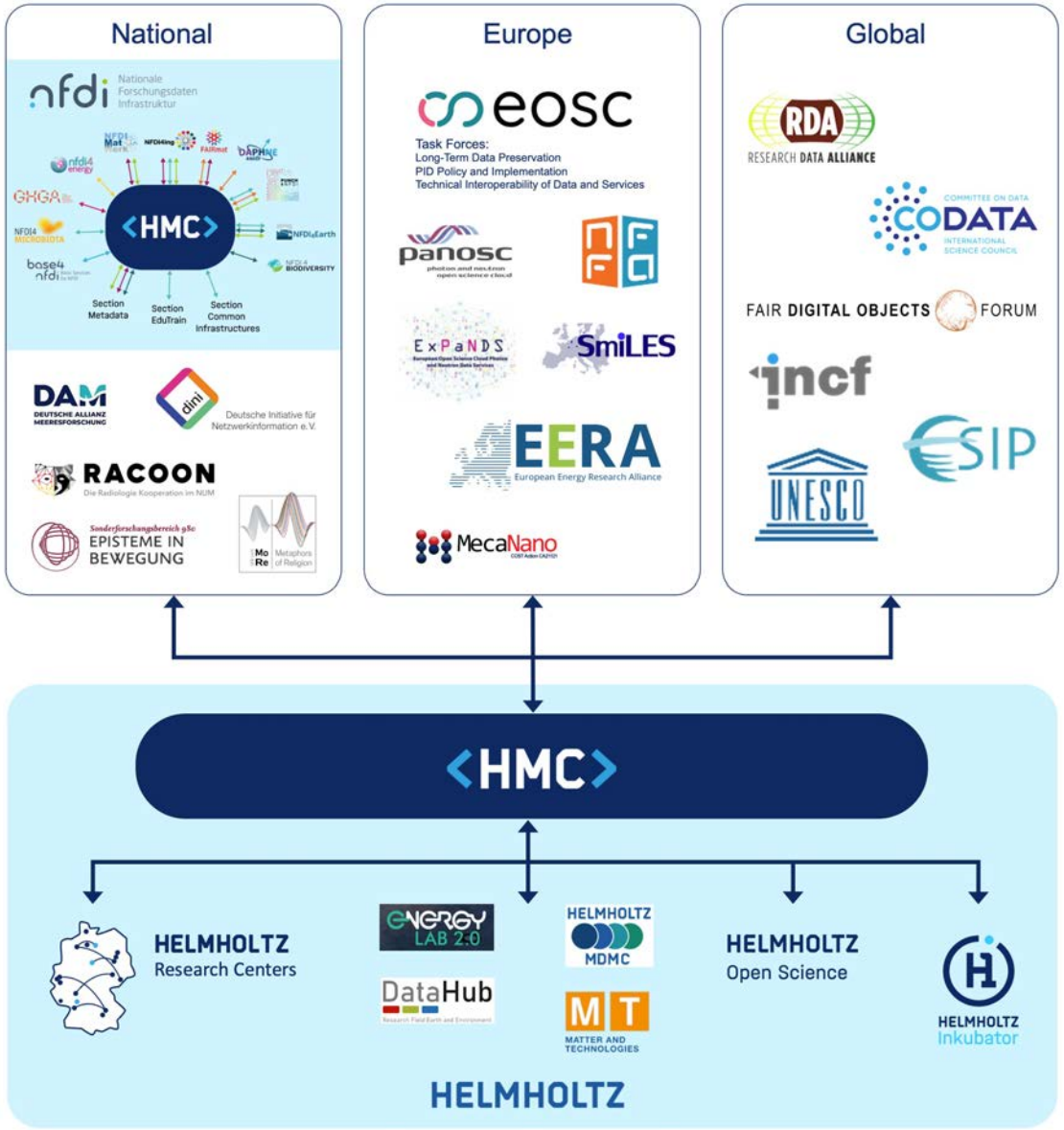
Community Processes



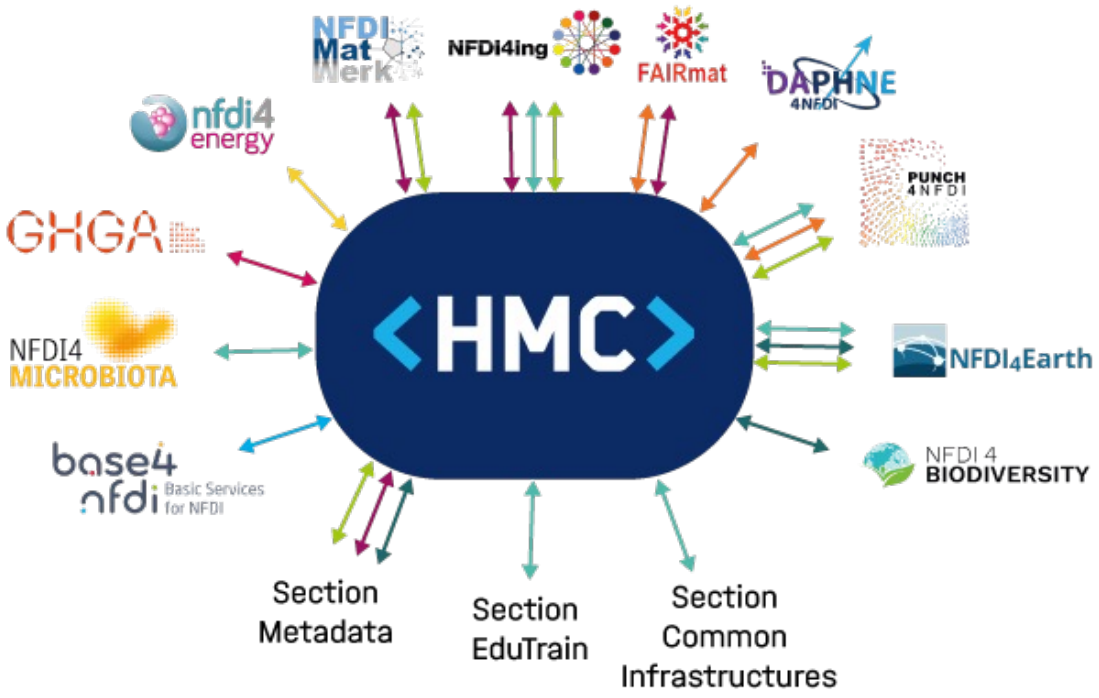
HMC Projects

HMC Network

- Helmholtz Incubator and Open Science Office
- NFDI-consortia & sections
- EOSC task groups
- Global RDM and domain communities



HMC and NFDI Consortia



Examples for HMC cooperation with NFDI cconsortia

NFDI MatWerk und FAIRmat	EM Glossary
NFDI MatWerk	Tools in action (FDO implementation), DISO development
Daphne4NFDI	Knowledge Transfer – EOSC Pan FAIR data recommendation
PUNCH4NFDI	Use case instrument to data publication), PATOF project
NFDI4Earth	Common standards for FAIR ESS data

EOSC PaN FAIR data recommendations

NFDI practical refinement of recommendations

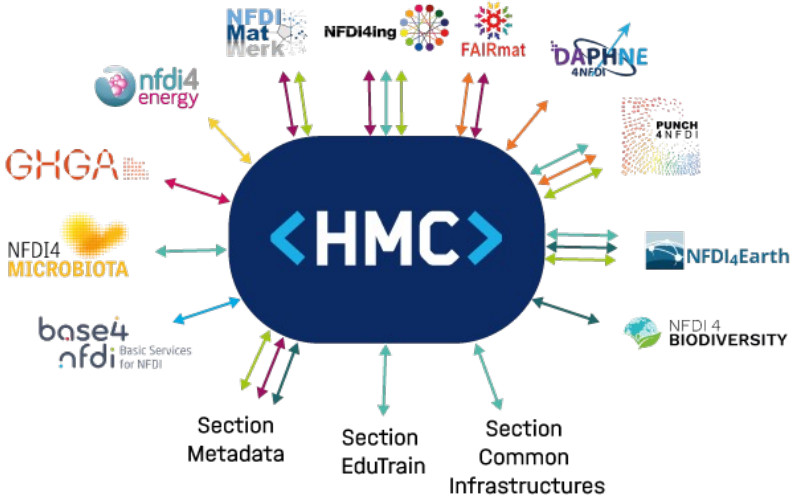
Implementation in Helmholtz data repository

<HMC> Knowledge Transfer

NFDI Sections

Five cross cutting sections established:

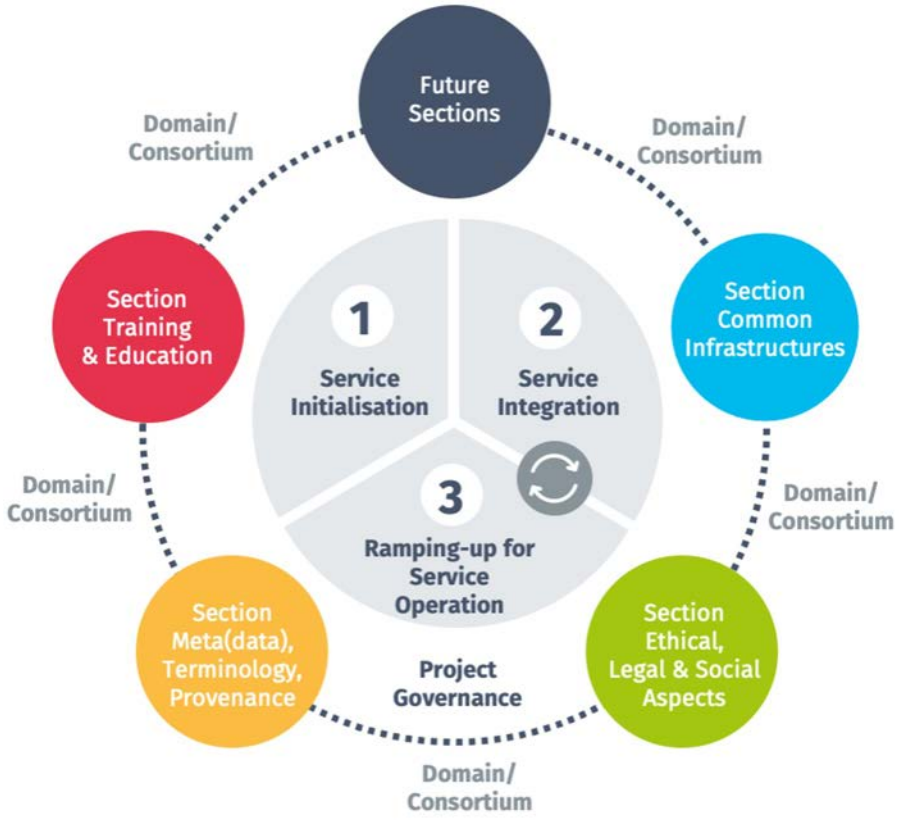
1. Common Infrastructures (section-infrastructure)
2. (Meta)data, Terminologies, Provenance (section-metadata)
3. Training and Education (section-edutrain)
4. Ethical, Legal and Social Aspects (section-ELSA)
5. Industry Engagement (section-industry)



- For participate in crosscutting discussions, the sections are the (only) organisational body within the NFDI e.V.

- Currently, HMC members are actively participating in section 1-3

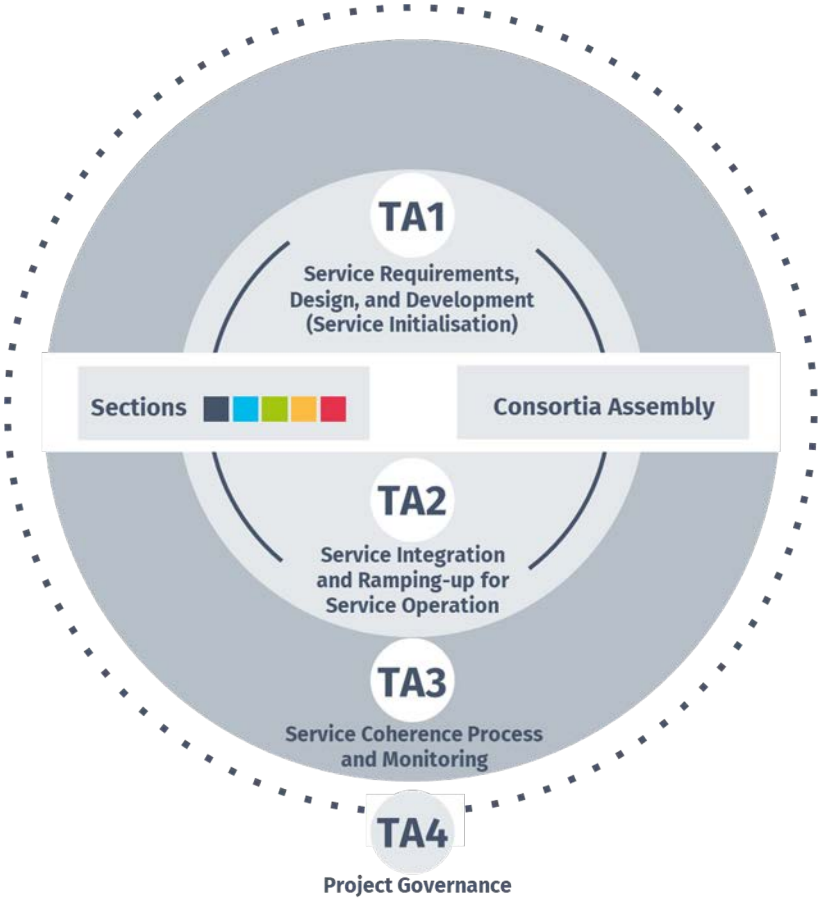
What is Base4NFDI?



- Base4NFDI is a joint initiative from all NFDI consortia
- the goal is to introduce and establish NFDI-wide basic services
- potential basic services are defined in the NFDI sections
- B4N provides a framework for the stepwise development of basic service candidates
- Three phases: initializing, integrating and ramping-up basic services
- setting up NFDI-wide basic service portfolio
- monitoring and evaluation
- transparent allocation of flexible funding

Source: www.base4nfdi.de

What will Base4NFDI do?

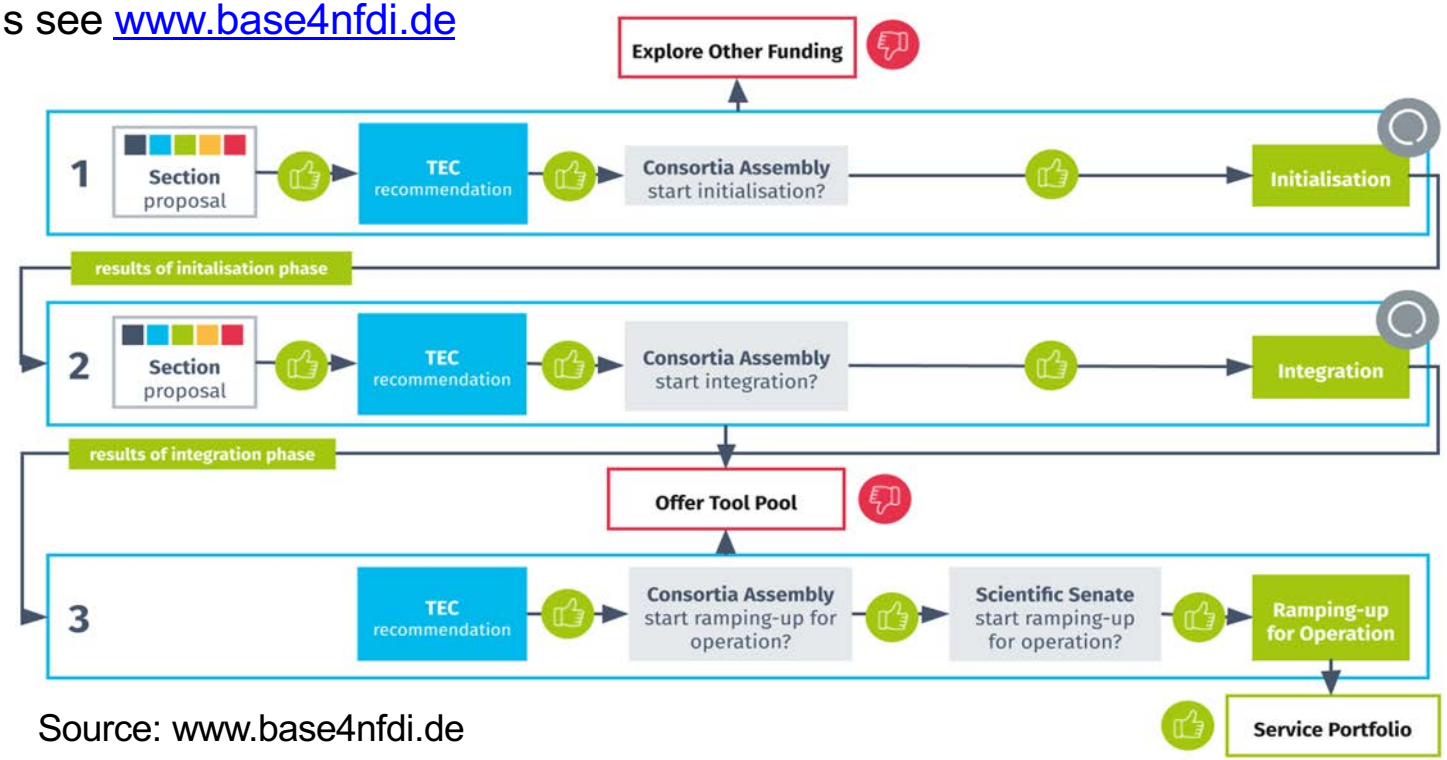


Source: www.base4nfdi.de

- TA1: Service Requirements, Design and Development (Service Initialisation)
- TA2: Service Integration and Ramping-up for Service Operation
  - maintaining basic service portfolio
- TA3: Service Coherence Process and Monitoring
  - providing a framework
  - monitoring and evaluation
- TA4: Project Governance
  - transparent allocation of flexible funding

Development process for basic services

- proposal-based procedure, proposals are developed in the sections of NFDI
- proposals are evaluated by the consortia and the TEC Board, final decision: Consortia Assembly
- for details see www.base4nfdi.de



Source: www.base4nfdi.de

HMC and Base4NFDI

- Already proposed topics*

  1 – DMP4NFDI – Data Management Plans

  2 – HaDES - Harvesting and Findability Enhancing Services

  3 – IAM4NFDI – Identity and Access Management

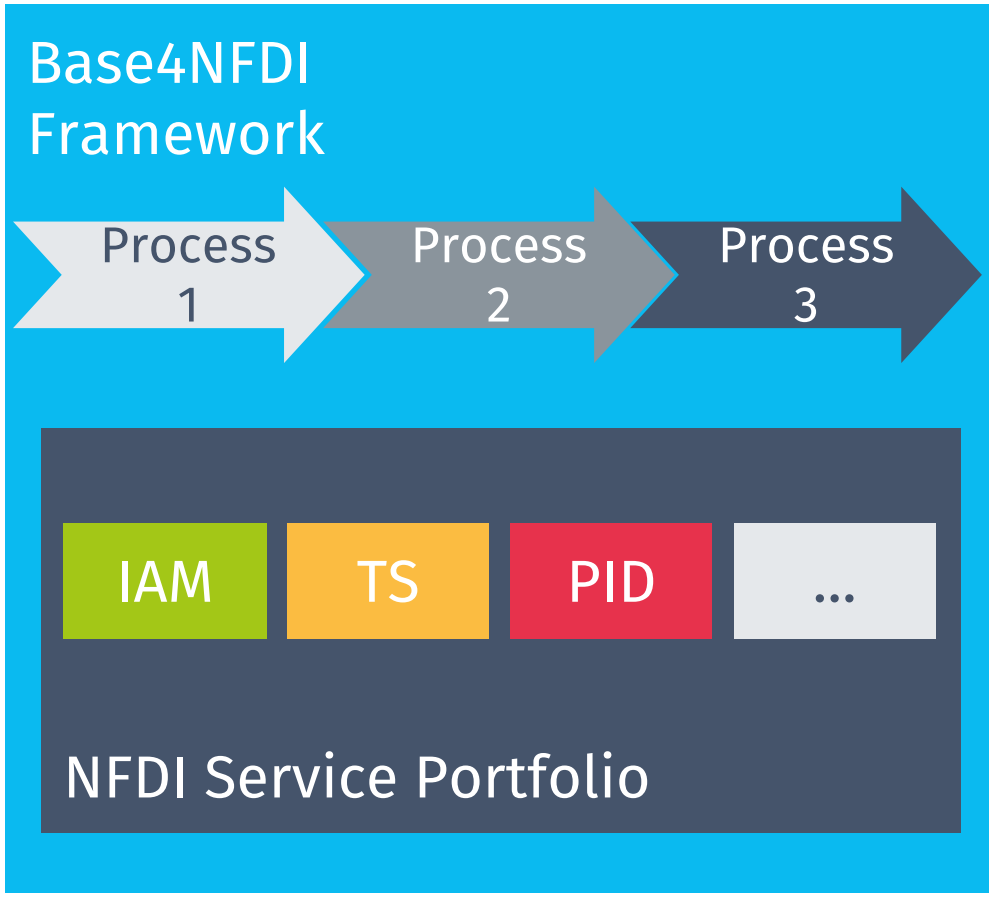
  4 – KGI4NFDI – Knowledge Graph Infrastructure for NFDI

  5 – PID4NFDI Persistent Identifier Services for NFDI**

  6 – TS4NFDI Terminology Services 4 NFD**

* all topics have a connection to HMC activities (except 3, which is purely HIFIS-related)

** currently under review



Source: www.base4nfdi.de

HMC and Base4NFDI - perspective

- HMC and HIFIS members where heavily involved in the Base4NFDI proposal (incl. co-spokesperson)
- HMC has a natural connectors to a broad range of discussed topics for basic services
- HMC is not eligible to apply for funding in base4NFDI -> not an institution and no affiliation to a consortium is possible
- HMC developments can be used for joint proposals for basic services by applicants from the Helmholtz community
- HMC can act as provider for blueprints or ready to use tools, services, or trainings with access to a broad cross-domain user group.
- Examples are PID implementation, terminology and ontology services, knowledge graphs (already in discussion with with possible external partners)
- engagement in the NFDI sections is key to ensure influence and alignment Helmholtz and NFDI developments

HELMHOLTZ

RESEARCH FOR GRAND CHALLENGES

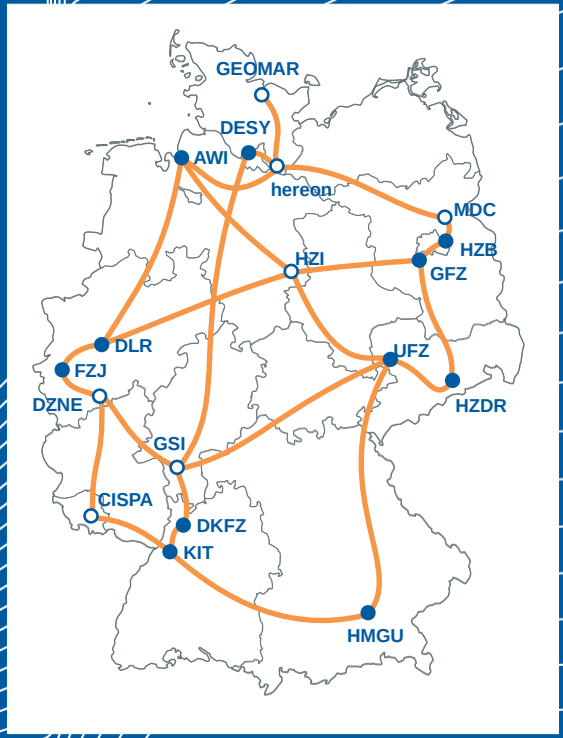
HIFIS | HELMHOLTZ FEDERATED IT SERVICES

HIFIS Perspectives on Base4NFDI

3rd Helmholtz Open Science Forum on NFDI

2023-06-22

Uwe Jandt (DESY), on behalf of the HIFIS Team

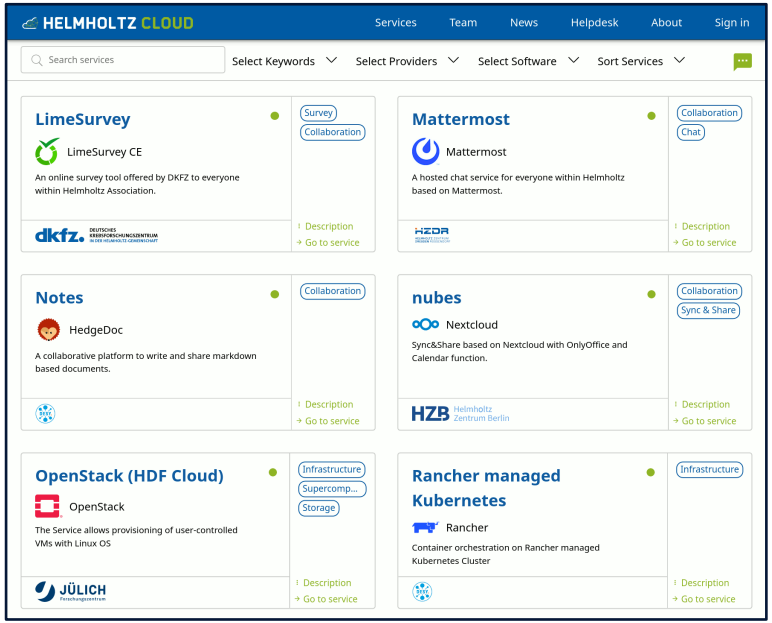


HIFIS for Scientists

➢ [Link to this poster]

HIFIS for Scientists

- **Helmholtz Cloud** and **Backbone**:
  - **Helmholtz ID / AAI** connects Helmholtz centres and their world-wide collaboration partners
  - **Collaboration, infrastructure, scientific** cloud and VPN services

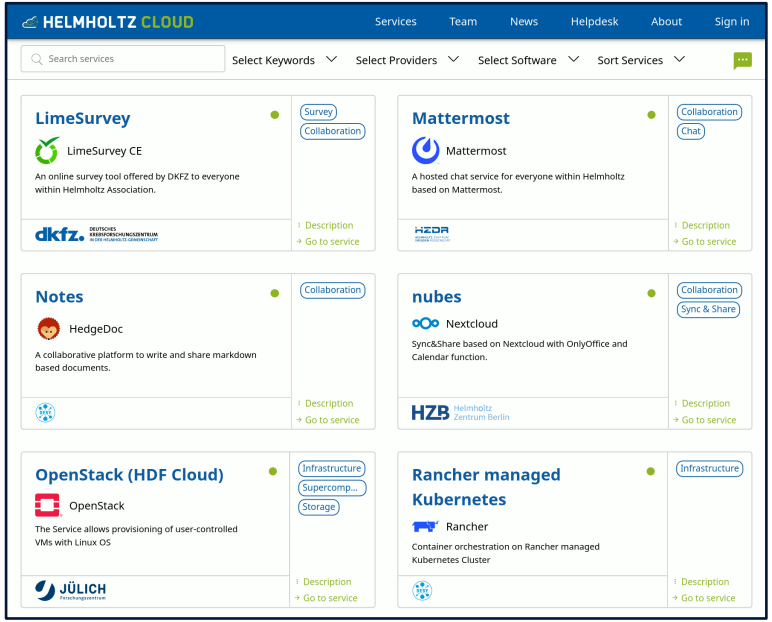


Cloud Portal: https://helmholtz.cloud

HIFIS for Scientists

- **Helmholtz Cloud** and **Backbone**:
  - **Helmholtz ID / AAI** connects Helmholtz centres and their world-wide collaboration partners
  - **Collaboration, infrastructure, scientific** cloud and VPN services

- **Research Software Engineering**:
  - Platform, Education & Training, Consulting, Community Services
  - High level of knowledge, quality, visibility and sustainability

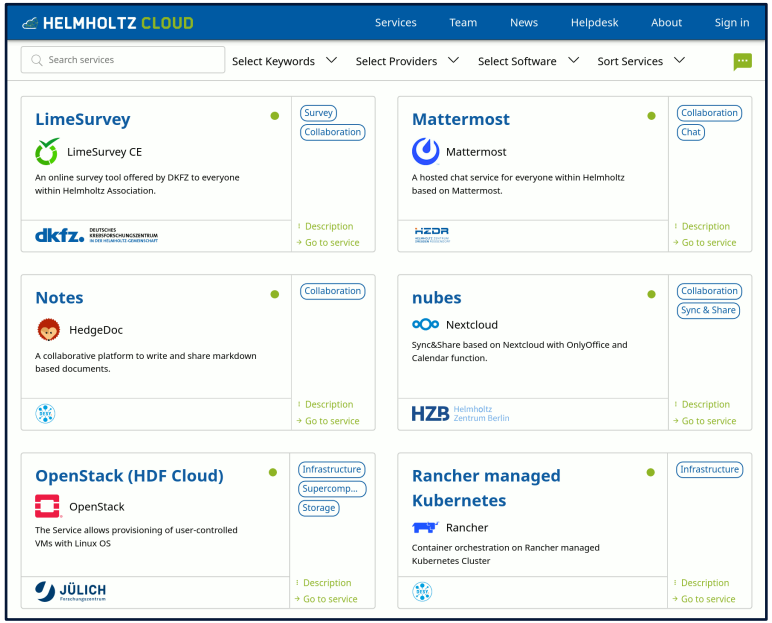


Cloud Portal: https://helmholtz.cloud

HIFIS for Scientists

Free of charge for Helmholtz + Partners

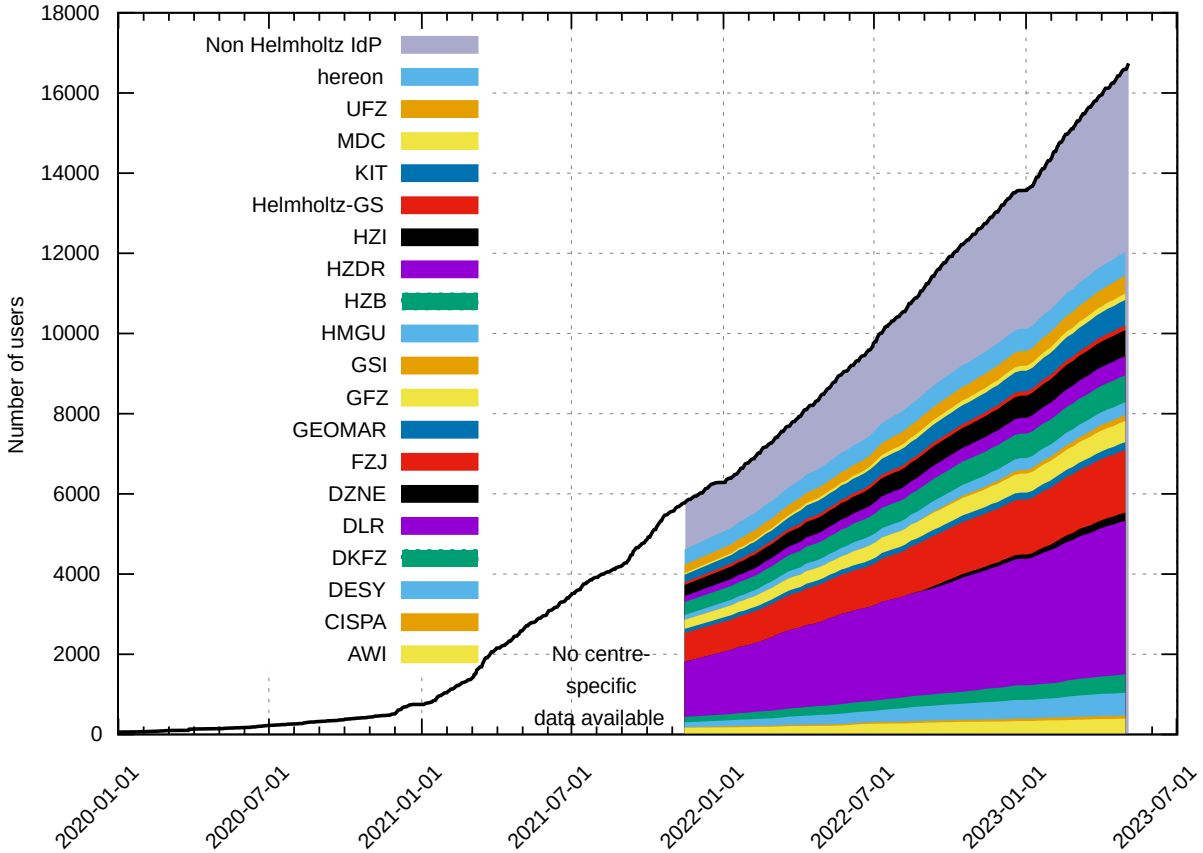
- **Helmholtz Cloud** and **Backbone**:
  - **Helmholtz ID / AAI** connects Helmholtz centres and their world-wide collaboration partners
  - **Collaboration, infrastructure, scientific** cloud and VPN services
- **Research Software Engineering**:
  - Platform, Education & Training, Consulting, Community Services
  - High level of knowledge, quality, visibility and sustainability



Cloud Portal: https://helmholtz.cloud

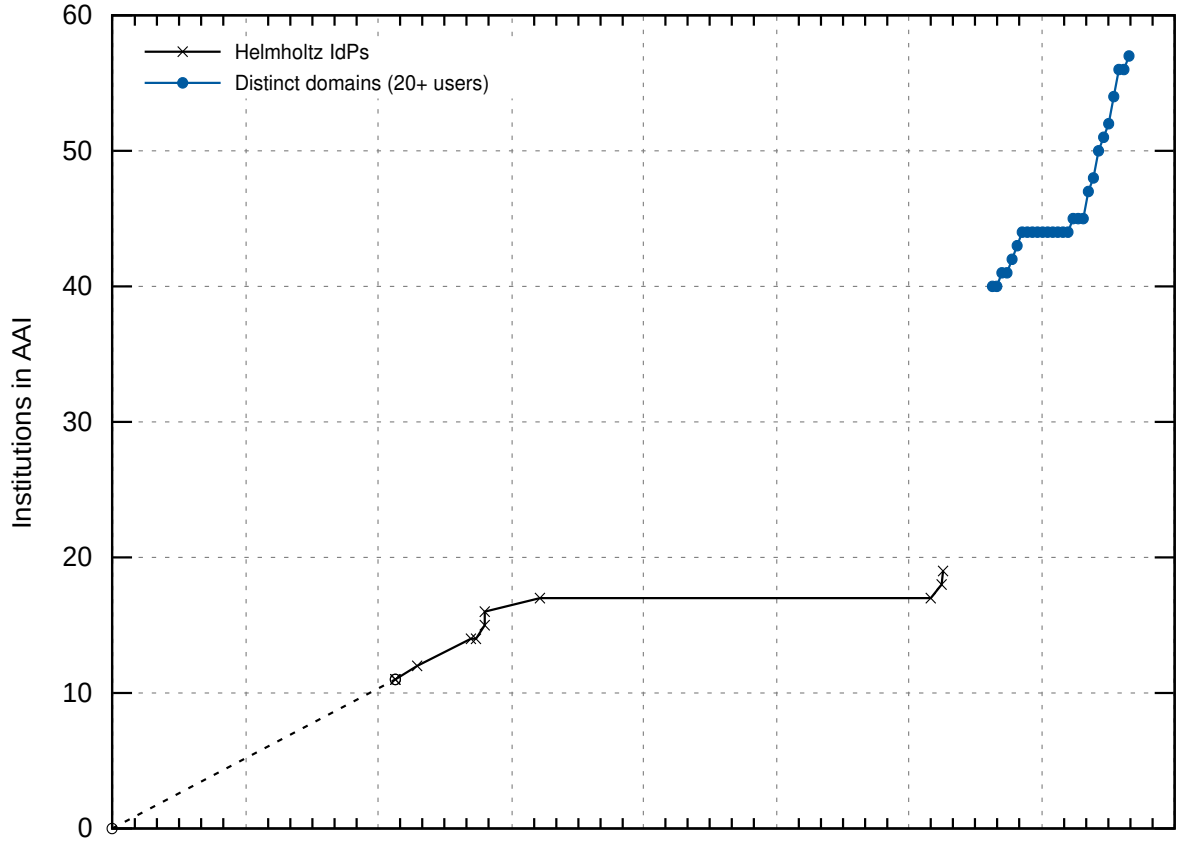
Helmholtz Digital Services for Science — Collaboration made easy.

Cloud + Backbone Usage: Individual users



- Link to Plots

Cloud + Backbone Usage: Institutions (>=20 users)



- Link to Plots

Fact Check - Overall

>17.000
individual users in AAI

~4.000
users from non-Helmholtz IdPs

31
Helmholtz Cloud Services

Helmholtz Digital Services for Science — Collaboration made easy.

>170%
average increased usage of each cloud service since onboarding

>70
collaborating groups (VO)

All 18 Centres
One Login!

>940
processed support tickets in 2022
(>>190 in 2023 so far)

>1.700
Course participants

Some overall feedback

*I just want to thank you for your great service which my colleagues and I are **using now frequently**! It is really great that we are establishing these useful services within Helmholtz and for cooperations beyond! Please go ahead ...*

*Thanks for this kind attitude of helping other members of the association facing problems, it feels like **being part of the big HIFIS family** ;)*

*This was a great experience. Having worked at a Helmholtz institution for many years, it is a bit difficult for senior staff to understand that **this is really a free and open service meant to support research** groups in different fields. This is quite uncommon in "old Helmholtz" …*

HIFIS4Base: Interfaces and Contributions

1) Helmholtz ID / Helmholtz AAI

a) Established **trust** between heterogenous communities within Helmholtz and increasingly outside, too

b) Is **compatible** to international frameworks (AARC)

c) Provides **procedural** framework (responsibility, security, group management, etc)

d) **Technical** capabilities to negotiate authentication and authorisation information between multiple sites

e) Has >> 10^5 **users**, all with their specific **workflows** and **requirements**

↔ Base4NFDI proposal, IAM4NFDI

HIFIS4Base: Interfaces and Contributions

2) Cloud Service Portfolio Management

a) Processes and Procedures to recruit, select, onboard and maintain services

b) Regulational framework (GDPR, VAT, ...)

c) Processes to technically onboard and maintain services

↔ TA2 Base4NFDI: Procedural Framework of Service Integration, Service Portfolio Management, Tool Pool

2b) Service Desk

a) Single point of contact for any user requests (support@hifis.net), with distributed back-end

HIFIS4Base: Interfaces and Contributions

3) Existing Cloud Services

a) From 2021, HIFIS onboarded primarily collaborative services:
They were **most-demanded**!

b) This fitted very well to the demands of mulitple NFDI consortia:
For Helmholtz-led consortia, HIFIS already provides many services for their daily work, e.g. [1-2].

- https://hifis.net/use-case/2022/09/14/DAPHNE4NFDI
- https://hifis.net/use-case/2022/06/28/use-case-PUNCH

HIFIS4Base: Interfaces and Contributions

3b) Existing Cloud Services – Service Types (Selection)

a) Collaborative Services –
Sync&Share, Collabtex, Helmholtz Codebase, Events Management, ...

b) Infrastructure –
Rancher managed Kubernetes, Codebase/Gitlab + CI/CD, Singularity, ...

c) (Super)compute –
HAICORE, GPU compute service, OpenStack, Jupyter ...

d) Storage –
Data Projects, B2Share, dCache, ...

e) Scientific Community Services –
Helmholtz Research Software Directory, Sensor Management System, ...

- https://helmholtz.cloud/services
- https://hifis.net/media/HIFIS_overview_en.pdf

So this goes hand in hand...

Base4NFDI

Image derived from Freepik

HIFIS Take home

I. Helmholtz Digital Services for Science — Collaboration made easy.

- Try it! Most Software and Cloud Services readily available for Helmholtz + Partners
- User-oriented workflows will be integrated continuously
- https://hifis.net/newsletter

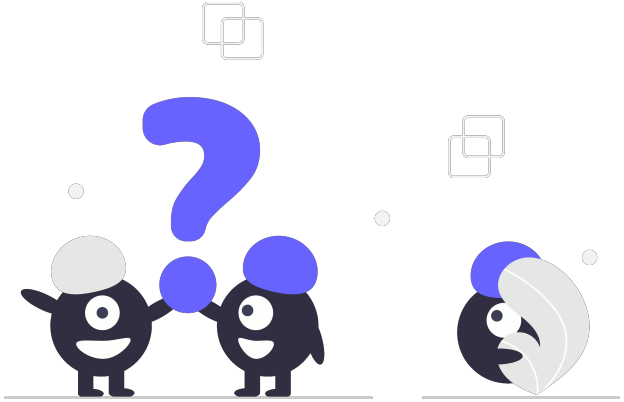
II. See what's there & disseminate

- https://hifis.net —> HIFIS for Scientists
- https://helmholtz.cloud
- https://hifis.net/media

III. Consult us:

- support@hifis.net

Questions? Discussion/Feedback



Katerina Limpitsouni
via undraw.co

HELMHOLTZ
Open Science

Helmholtz Perspectives on Base4NFDI: PID4NFDI

Antonia C. Schrader
Helmholtz Association,
Helmholtz Open Science Office

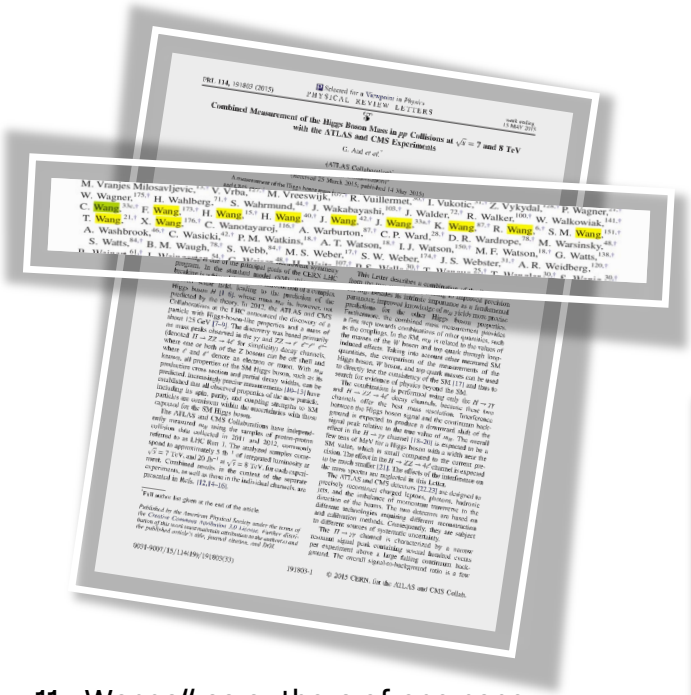
Online, June 22, 2023

Agenda

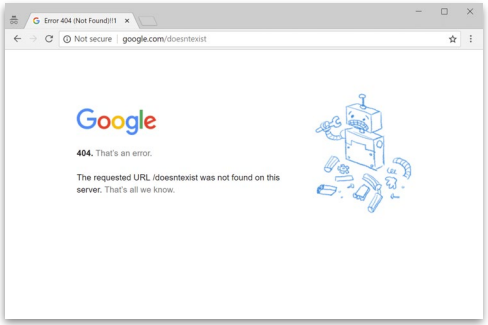
1. Introduction on Persistent Identifiers (PIDs)
2. PID4NFDI
3. PID Network Germany

Why PIDs are important:

https://doi.org/10.1103/PhysRevLett.114.191803



11 „Wangs“ as authors of one paper



Invalid URL

German National Library of Science and Technology

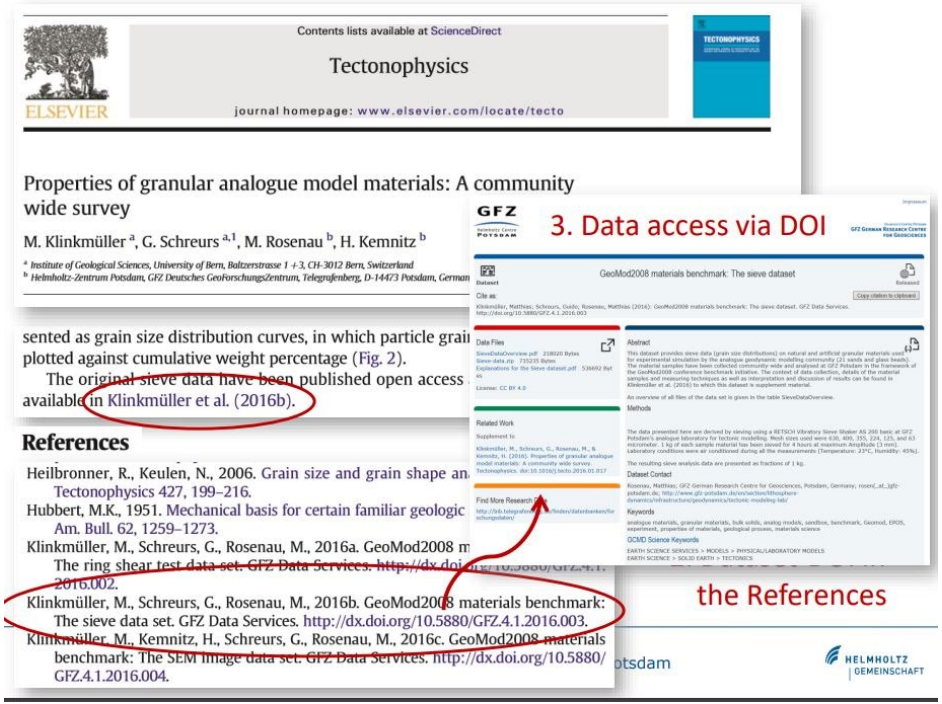
OTHER NAMES

Technische Informationsbibliothek (TIB) - Leibniz-Informationszentrum Technik und Naturwissenschaften - Universitätsbibliothek, German National Library of Science and Technology (TIB) - Leibniz Information Centre for Science and Technology - University Library, TIB, Technische Informationsbibliothek

At least 4 different variants of the organization name

Why PIDs are important:

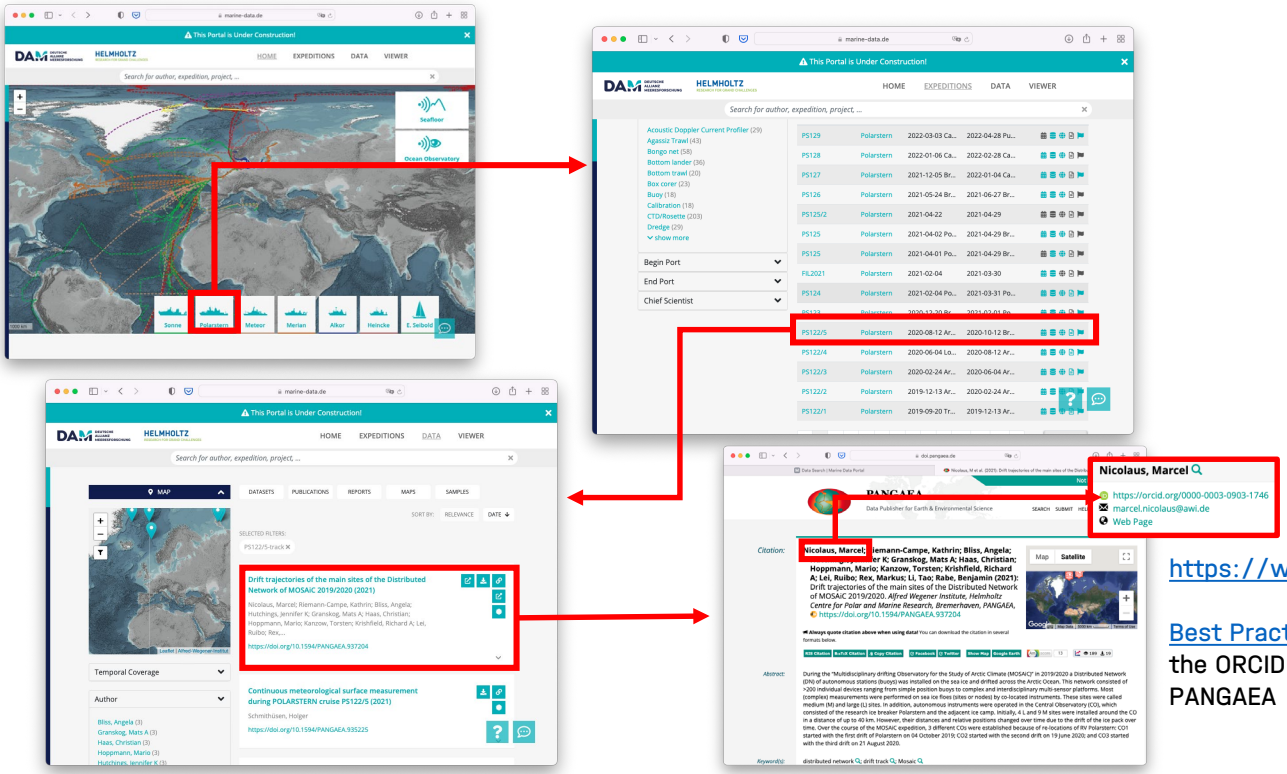
PIDs enable cross-references between the article and the dataset



Source:
Elger, K. (2016).
https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/rda_de_16_elger.pdf.
p. 25

Why PIDs are important:

https://marine-data.de/



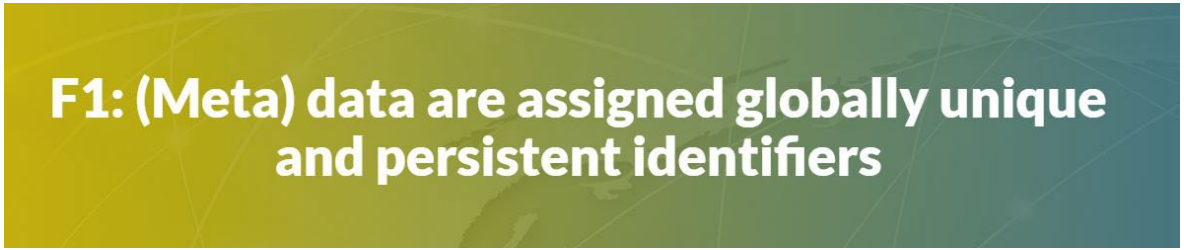
Interaction between DOI, ORCID & Data

https://www.pangaea.de

Best Practices Report of the ORCID integration in PANGAEA

This is why...

- References to potentially **any sort** of digital or non-digital **entity**
- Enable the **interlinking** between entities
- Contribute to the **integrity** of scientific communication and its **reproducibility**
- PIDs are central to make data **FAIR**.
- PID and their corresponding **metadata** provide a **comprehensive view** on all research process entities.



Source: https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers

A heterogenous PID landscape

International

DataCite

Crossref

doi Foundation

ePIC
Persistent Identifiers for eResearch

ORCID
Connecting Research and Researchers

National

urn:nbn:de:1111-20091210269 urn:nbn:de:
v:11-100287 urn:nbn:de:hbz:468-2007038
1417541491 urn:nbn:de:bsz:21-opus-2952
nbn:de:101:1-2010080220 urn:nbn:de:027
6 urn:nbn:de:0012:bverwg-17962 urn:nb

Handle.Net® Registry

A heterogenous PID landscape

Research Activity Identifier

Identifying and tracking of research projects and activities

https://www.raid.org.au

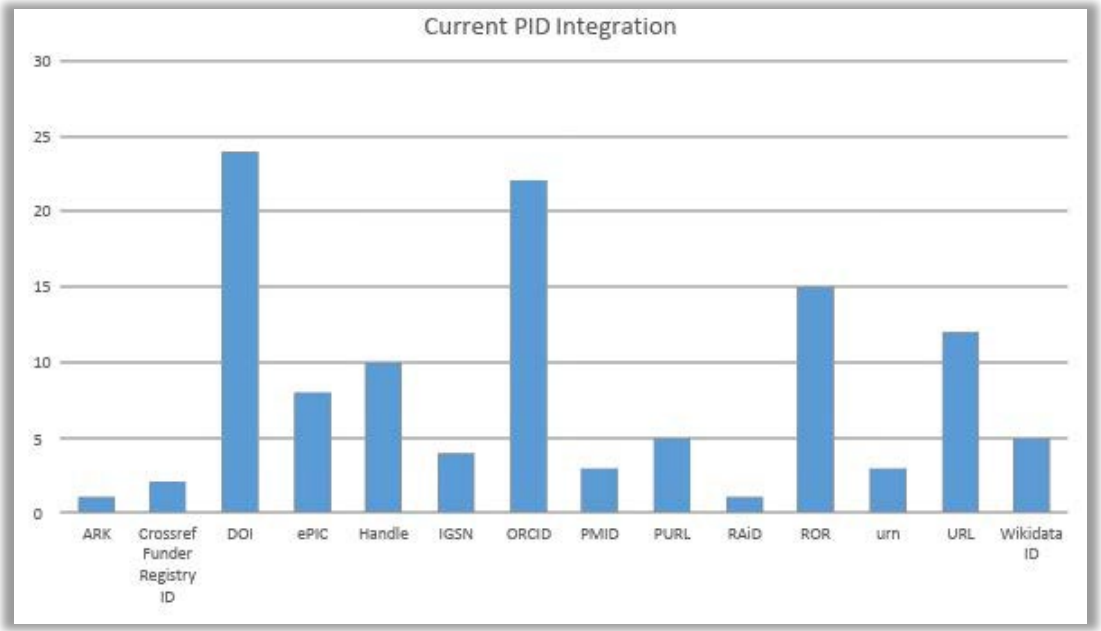
Research Resource Identifier

Identifying and cite key resources of biomedicine (e.g. antibodies, model organisms and software projects)

https://www.rrids.org

A heterogenous PID landscape in NFDI, too

Which kind of identifier or PID do you integrate?



From 15 response options; Calculation based on number of all answers given: no consolidation by consortium membership

Source: Hagemann-Wilholt, Stephanie. (2023). PID4NFDI: Survey on PID Practices. Main results. Zenodo. https://doi.org/10.5281/zenodo.7635791. p 3.

NFDI Working Group: Persistent Identifiers

- Founded in 2022
- Located in the Section Common Infrastructures
- Aim:
  - Develop a **common strategy** for the implementation and extension of community-based and broadly used PID services
  - Conducting technical & organisational measures as well as **training** and **education**, particularly around standardised and complete metadata
- Chairs: Philipp Wieder (GWGD) & Britta Dreyer (TIB/DataCite)
- Publications:
  - Charta of the working group
  - PID4NFDI: Survey on PID Practices. Main results.
  - Workshop on PIDs within NFDI

PID4NFDI

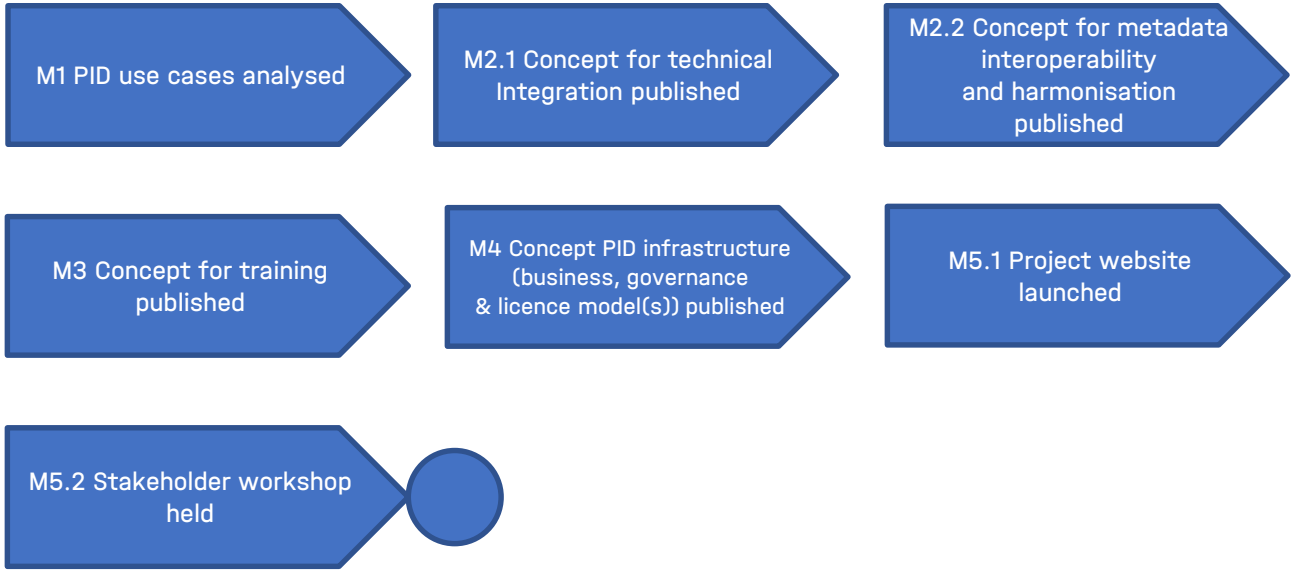
base4nfdi Basic Services for NFDI

- Proposal of founding **Persistent Identifier Services**
  for the German National Research Data Infrastructure (NFDI)
- Applicants:

DataCite GWDG Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen HELMHOLTZ Open Science TIB LEIBNIZ-INFORMATIONSZENTRUM TECHNIK UND NATURWISSENSCHAFTEN UNIVERSITÄTSBIBLIOTHEK

- The proposal was submitted for the Initialization Phase of Base4NFDI, and is currently under review for funding.

PID4NFDI: Initialisation phase



PID Network Germany

- Funded by the German Research Foundation (DFG)
- The project proposal is published (only in German): https://doi.org/10.48440/os.helmholtz.059
- Website (only in German): https://www.pid-network.de

- Projects aim:
  - Establishment of a **network** of already existing and currently forming actors around **persistent identifier in science and culture**
  - Promotion of the usage, implementation, standardization and compatibility to international PID systems on local, national and international levels
  - Development of a **PID roadmap for Germany**

- Project partner institutions:

DataCite CONNECTING RESEARCH, IDENTIFYING KNOWLEDGE
DEUTSCHE NATIONAL BIBLIOTHEK
HELMHOLTZ Open Science
TIB LEIBNIZ-INFORMATIONSZENTRUM TECHNIK UND NATURWISSENSCHAFTEN UNIVERSITÄTSBIBLIOTHEK
UNIVERSITÄT BIELEFELD Universitätsbibliothek
DFG

Close Collaboration

PID4NFDI

PID Network Germany

PID4NFDI will complement PID Network Germany –

Both projects agreed to closely collaborate.

Source: https://pixabay.com/de/illustrations/h%C3%A4ndedruck-hand-sch%C3%BCtteln-handeln-4784749

HELMHOLTZ
Open Science

PID NETWORK DEUTSCHLAND

Keep in touch

Website: www.pid-network.de

Email: info.pidnetwork@listserv.dfn.de

Email: section-infra-wg-pid@lists.nfdi.de

Please feel free to follow us on Social Media:

Mastodon (@PIDNetworkDE@openbiblio.social)

Twitter (@PIDNetworkDE)

HELMHOLTZ
Open Science

Thank you for your Attention!

Antonia C. Schrader

antonia.schrader@os.helmholtz.de

https://orcid.org/0000-0001-7080-634X

@AntoniaS@openbiblio.social

@AntoniaCamiS

HELMHOLTZ
SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

GEOMAR

NFDI activities from the perspective of GEOMAR

Hela Mehrtens, Klaus Getzlaff, Heiner Dietze, Sören Lorenz

22.06.023

3. Helmholtz Open Science Forum on the NFDI:
Inside Perspectives of the Centers

GEOMAR in DataHub – DAM – NFDI …

GEOMAR

nfdi Nationale Forschungsdaten Infrastruktur

NFDI4Earth

EUROPEAN OPEN SCIENCE CLOUD

HMC HELMHOLTZ METADATA COLLABORATION

Part of support infrastructure

Show Cases

Complementary initiatives

Visible data sets and products

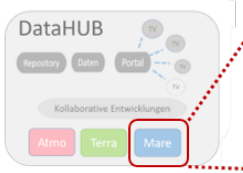
Workflows and processes

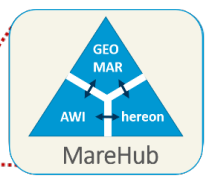
HELMHOLTZ

Data Hub

Mare Hub

DAM





PANGAEA
Data Publisher for Earth & Environmental Science

CAU, CEN, ICBM, IOW, MARUM, MPIC, BSH

GEOMAR

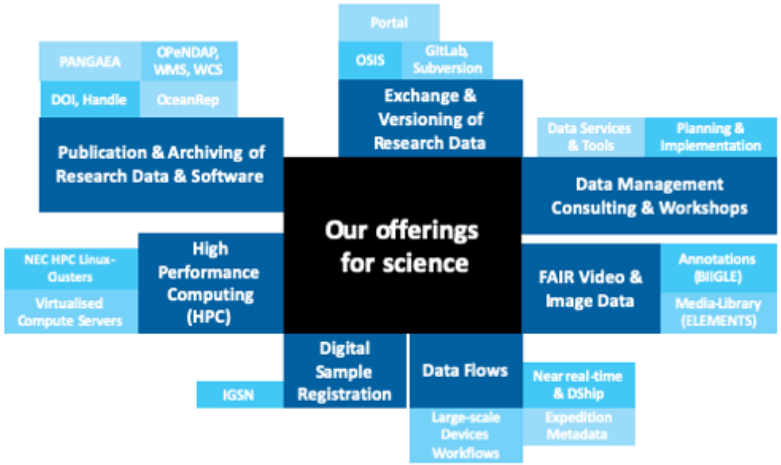
GEOMAR Strategy 2030 - Digitalisation

Short-term (1-2 years)	Digitalisation measures in research follow FAIR and open science principles on an institutional level (aims, financing, positions)	1. Standards and policies are implemented on an organisational and technical level 2. FAIR and open culture is institutionally promoted 3. Science support positions can be developed according to a (institutional) career path	1. Develop Research Data and Software Policy. 2. Design standards for data description and data flows 3. Describe job profiles in science support (IT-, data mgt. and data analytics)	

- Definition of data and support services including policies
- Harmonize concepts and set up standards in data management initiatives as DataHUB and NFDI4Earth
- Definition of roles and interaction of data stewards in departments and central data management

Data and support services

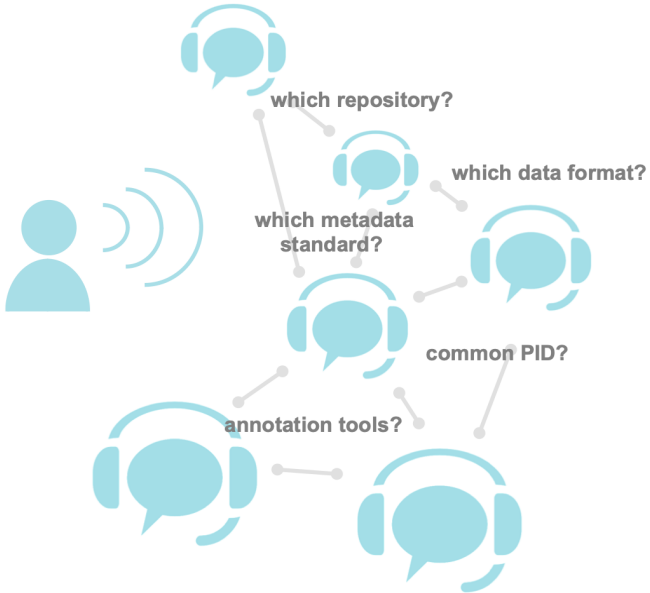
- Definition of RDM service catalogue as part of the IDCC services
- Support services play a major role



GEOMAR

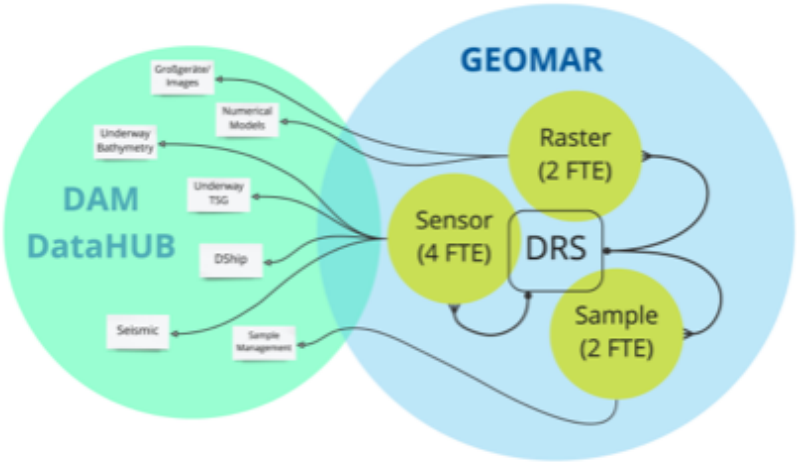
DAM, DataHUB and NFDI4Earth activities

- Harmonize concepts
  - Working groups e.g. on samples, observations, simulations, video/images
  - Common portal
  - Recommendations and SOPs
- Research data and software is available from different repositories
- Support is given by experienced data stewards



Data stewards

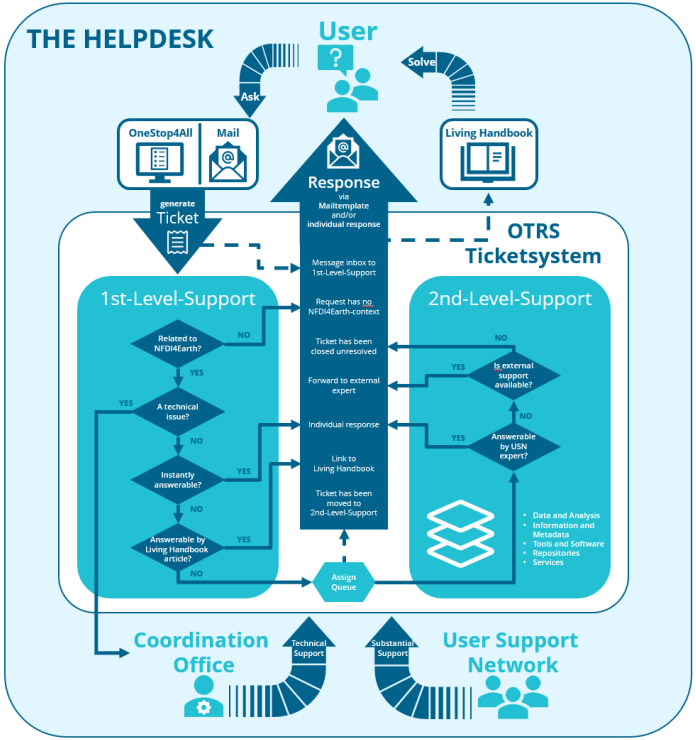
- Definition of roles and interaction of **data stewards** in departments and central data management



GEOMAR

NFDI4Earth
User Support Network

- a helpdesk is set up with agents from 8 partners
- GEOMAR data stewards are involved
- looking for new agents (e.g. Deutsche Allianz Meeresforschung DAM)



Base4NFDI4

Coordination of Service-Portfolio-Management

GEOMAR > Karriere & Campus > Karriere

Koordination (m/w/d) Service-Portfolio-Management Base4NFDI

Bewerbungsschluss: 28. Juni 2023

Das GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel ist eine von der Bundesrepublik Deutschland (90%) und dem Land Schleswig-Holstein (10%) gemeinsam finanzierte Stiftung des öffentlichen Rechts. Es gehört zu den international führenden Einrichtungen auf dem Gebiet der Meeresforschung.

Durch unsere Forschung und unser Engagement im Transfer von Wissen und Technologie tragen wir maßgeblich zum Erhalt der Funktion und zum Schutz des Ozeans für kommende Generationen bei.

Base4NFDI ist eine gemeinsame Initiative aller Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI) und hat die Bereitstellung von Basisdiensten für die NFDI zum Ziel. Base4NFDI gestaltet neue Formen der Dienstentwicklung und Dienstbereitstellung im deutschen Wissenschaftssystem.

Das Informations-, Daten- und Rechenzentrum des GEOMAR ist Teil dieses Projektes und entwickelt gemeinsam mit Partnern das Service-Portfolio-Management für Base4NFDI.

Für diese Aufgabe suchen wir zum nächstmöglichen Zeitpunkt eine

Koordination (m/w/d)

Karriere & Campus
- Willkommen
- Karriere
- Stellenangebote
- Ausbildung
- Praktikum
- GEOMAR als Arbeitgeber
- Campus

Kontakt

GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel

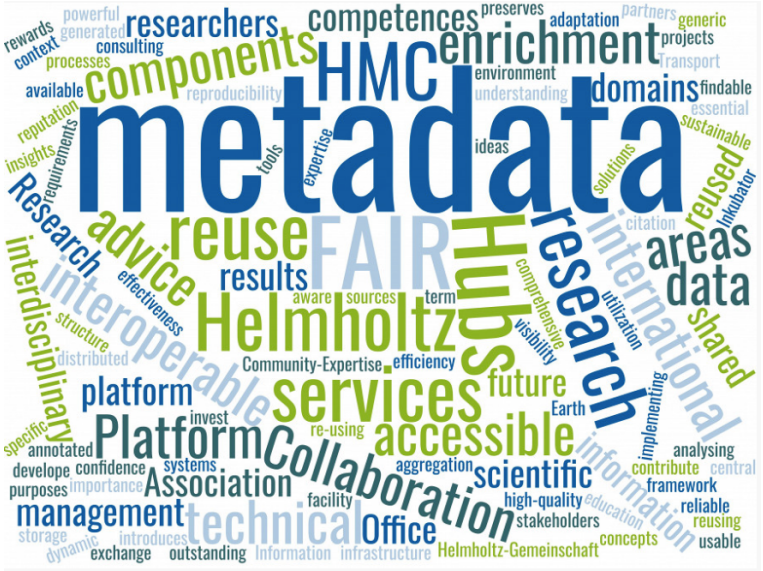
Personalabteilung
Wischhofstraße 1-3
24148 Kiel

bewerbung(at)geomar.de

GEOMAR

NFDI e.V.

- Involved in sections
  - Common infrastructures
  - (Meta)data, Terminologies, Provenance
  - Training and Education

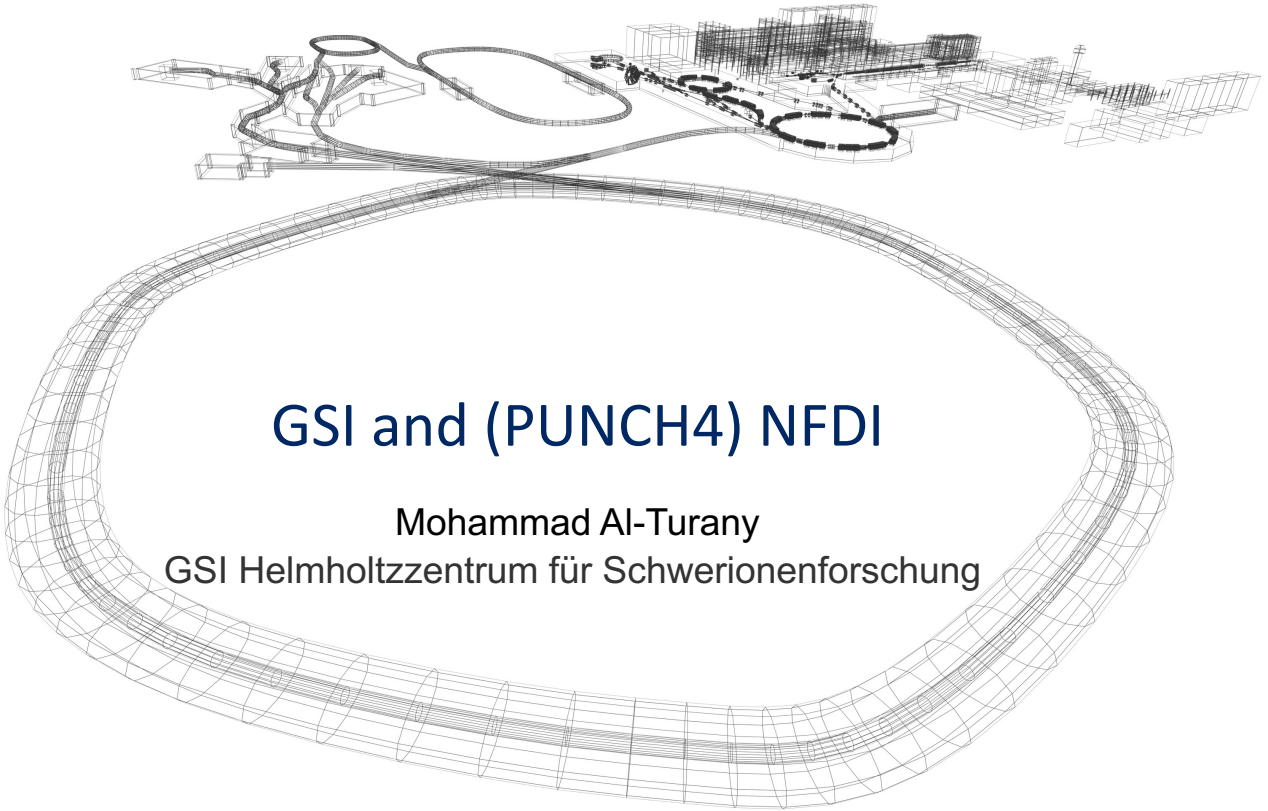


HELMHOLTZ
SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

GEOMAR

Thank you

https://www.geomar.de/en/research/research-data
https://marine-data.de
https://www.nfdi4earth.de/2facilitate/user-support-network



GSI and (PUNCH4) NFDI

Mohammad Al-Turany
GSI Helmholtzzentrum für Schwerionenforschung

NFDI Consortia at GSI

Particles, Universe, NuClei and Hadrons for the NFDI (PUNCH4NFDI)

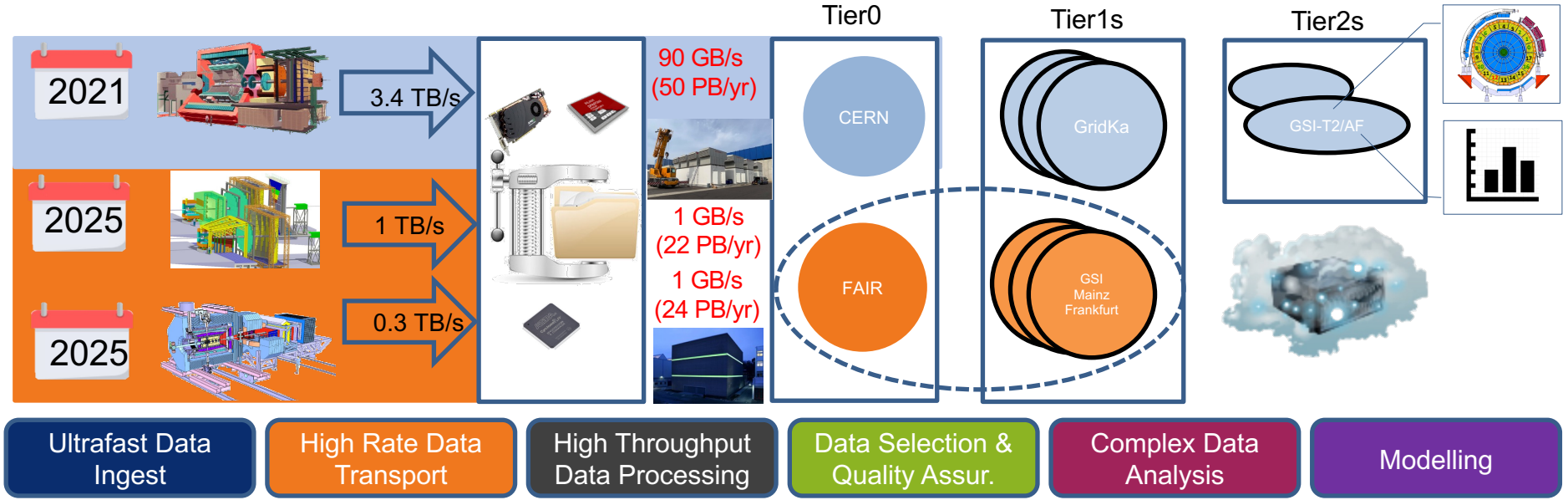
The prime goal of PUNCH4NFDI is the setup of a federated and "FAIR" science data platform, offering the infrastructures and interfaces necessary for the access to and (re)use of (meta)data and computing resources of the involved communities and beyond.

* **FAIR:** **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable

NFDI activities from the perspective of GSI

- The NFDI activities at the GSI/FAIR are focused around our expertise in:
  - data management,
  - data processing (simulation, reconstruction and analysis),
  - (federated) infrastructure development

Online/Offline Data processing

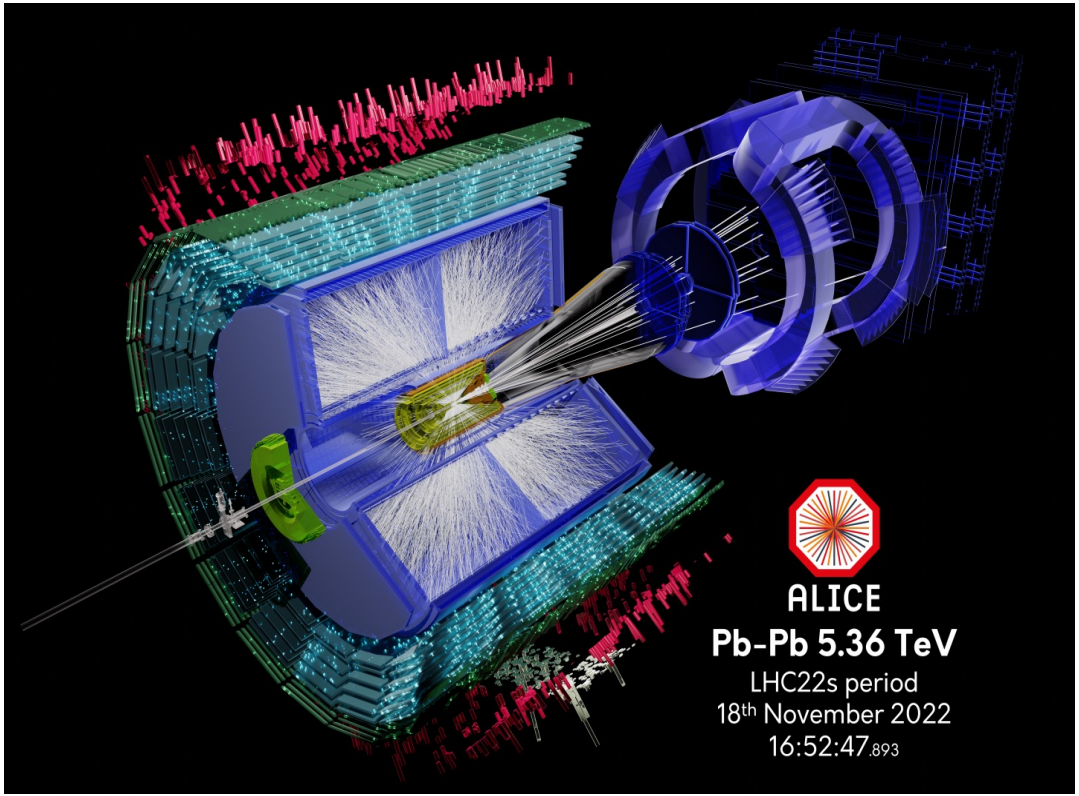


Ultrafast Data Ingest | High Rate Data Transport | High Throughput Data Processing | Data Selection & Quality Assur. | Complex Data Analysis | Modelling

Meanwhile HPC-Clusters are directly connected to the data acquisition systems and integrated into the online systems of the experiments

Heavy Ion beam in RUN3 at CERN

ALICE-FAIR common framework (ALFA) is used on (2200 GPUs and few thousends of CPUs) in the online farm directly connected to the ALICE detector.



Data management (TA2,4)

- Generic solutions with standardized protocols for archive/compute sites, suited for all physics communities involved (i.e.: Access to data, federated computing, automation)
- Globally distributed data lake with large storage and compute resources and portal access
- Opportunistic resource in federated science cloud

Infrastructure Development (TA2,4,6)

- Development of federated infrastructure :
  - Designing and implementing data storage and processing systems,
  - high-performance computing clusters,
  - high throughput data analysis pipelines.

- Ensure efficient and reliable data access, retrieval, and processing within the NFDI ecosystem.

Summary

PUNCH4NFDI facilitates data exchange and collaboration across disciplines, aligning with GSI's involvement in ESCAPE, EURO-LAPS and Helmholtz Matter and Technologies.

HZB Helmholtz Zentrum Berlin

BRIDGING THE GAP BETWEEN USER-FRIENDLINESS AND CUTTING-EDGE RESEARCH

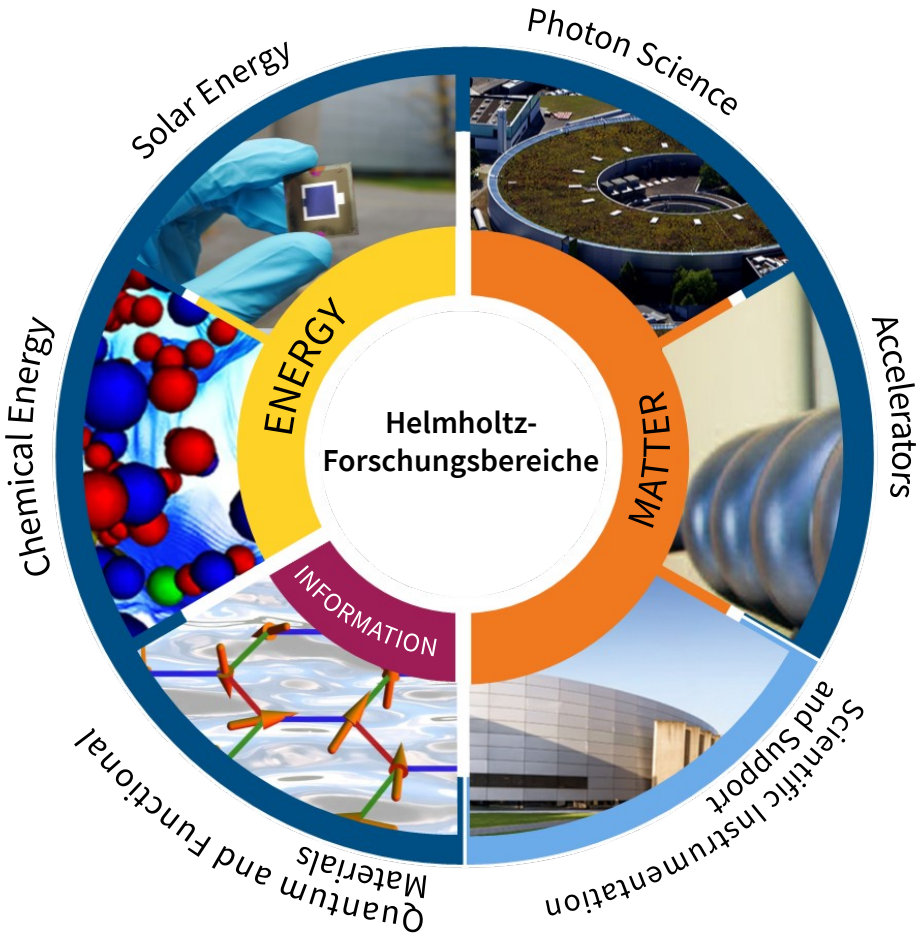
HZB in the NFDI

22.06.2023

Helmholtz Open Science Forum

HZB Helmholtz Zentrum Berlin

Research at HZB

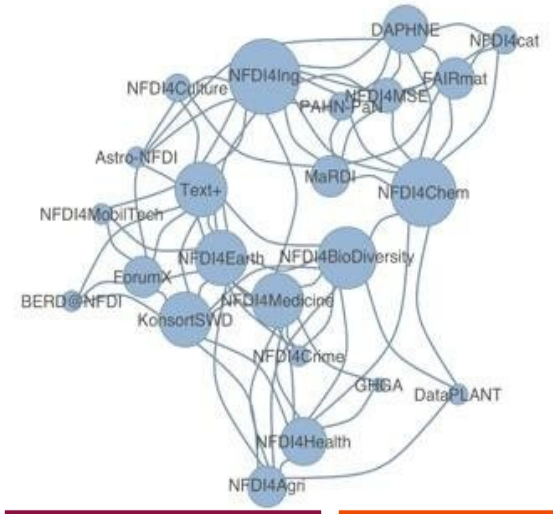


HZB Helmholtz Zentrum Berlin

HZB in the NFDI

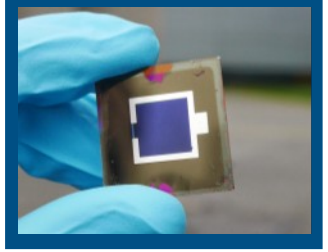
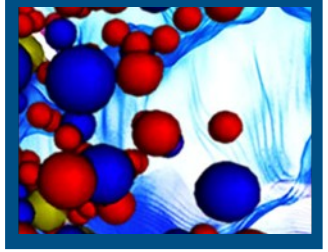
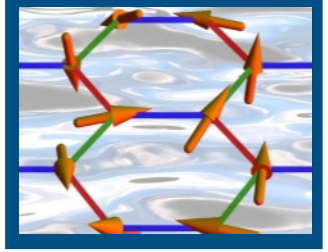
FAIRmat

Task Leader E3:
Thomas Unold



DAPHNE 4NFDI

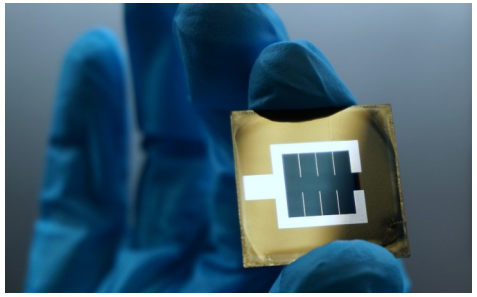
Co-Spokespersons:
Ingo Manke,
Heike Görzig

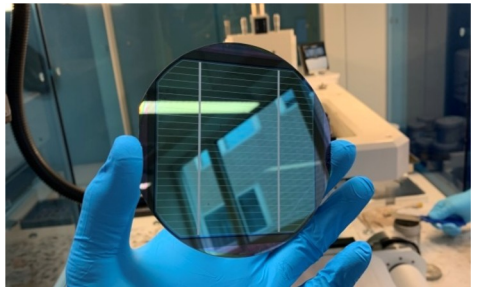
Solar Energy	Chemical Energy	Quantum and Functional Materials	Photon Science	Accelerator	Sci. Instrumentation and Support
					

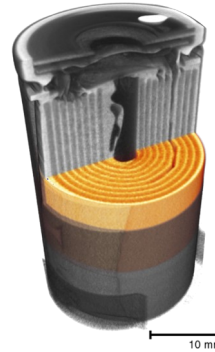
nfdi Nationale Forschungsdaten Infrastruktur

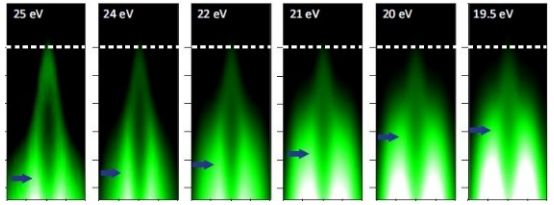
Representative HZB:
Tamara Husch Lee

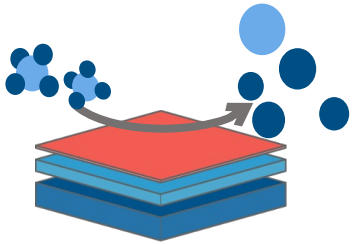
Cutting-edge Research, e.g,. on Energy Materials

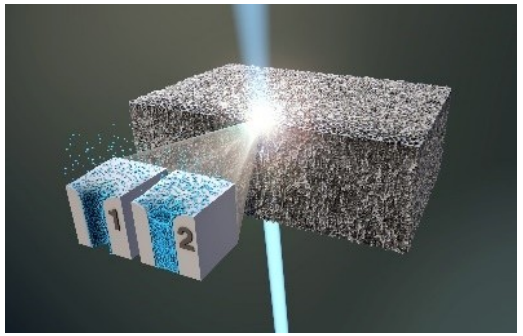












- Researchers embedded in **respective domain communities** (eg. PV community with own standards and practices)
- Communities in Germany organized in NFDI consortia

FAIRmat

Member

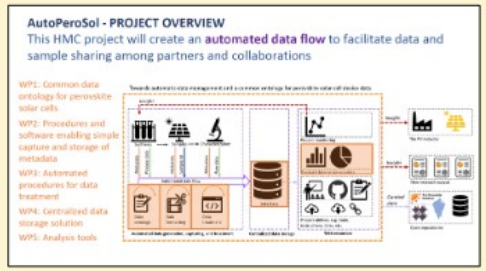
nfdi4 energy

NFDI4Cat NFDI for Catalysis-Related Sciences

NFDI MatWerk

Observing developments

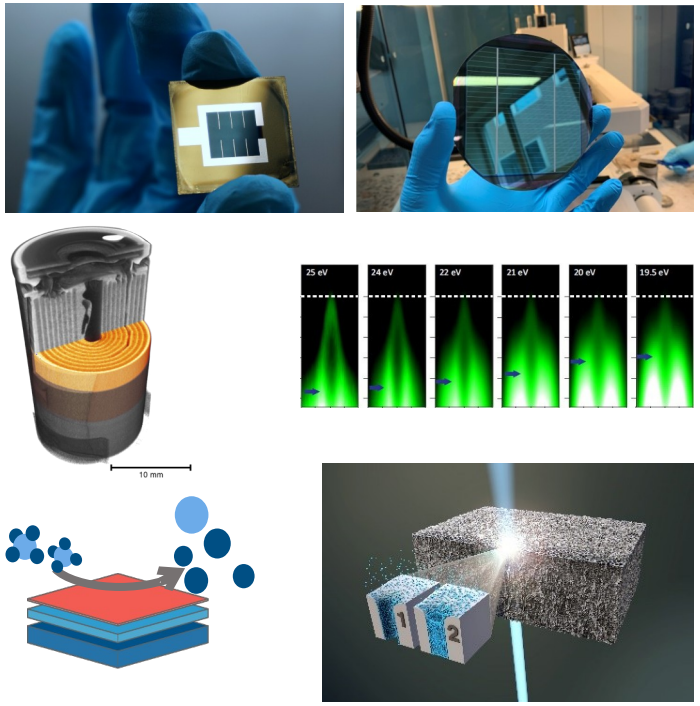
- HZB – Task E3: Use Cases - Optoelectronics
- Active in context of Helmholtz Metadata Collaboration

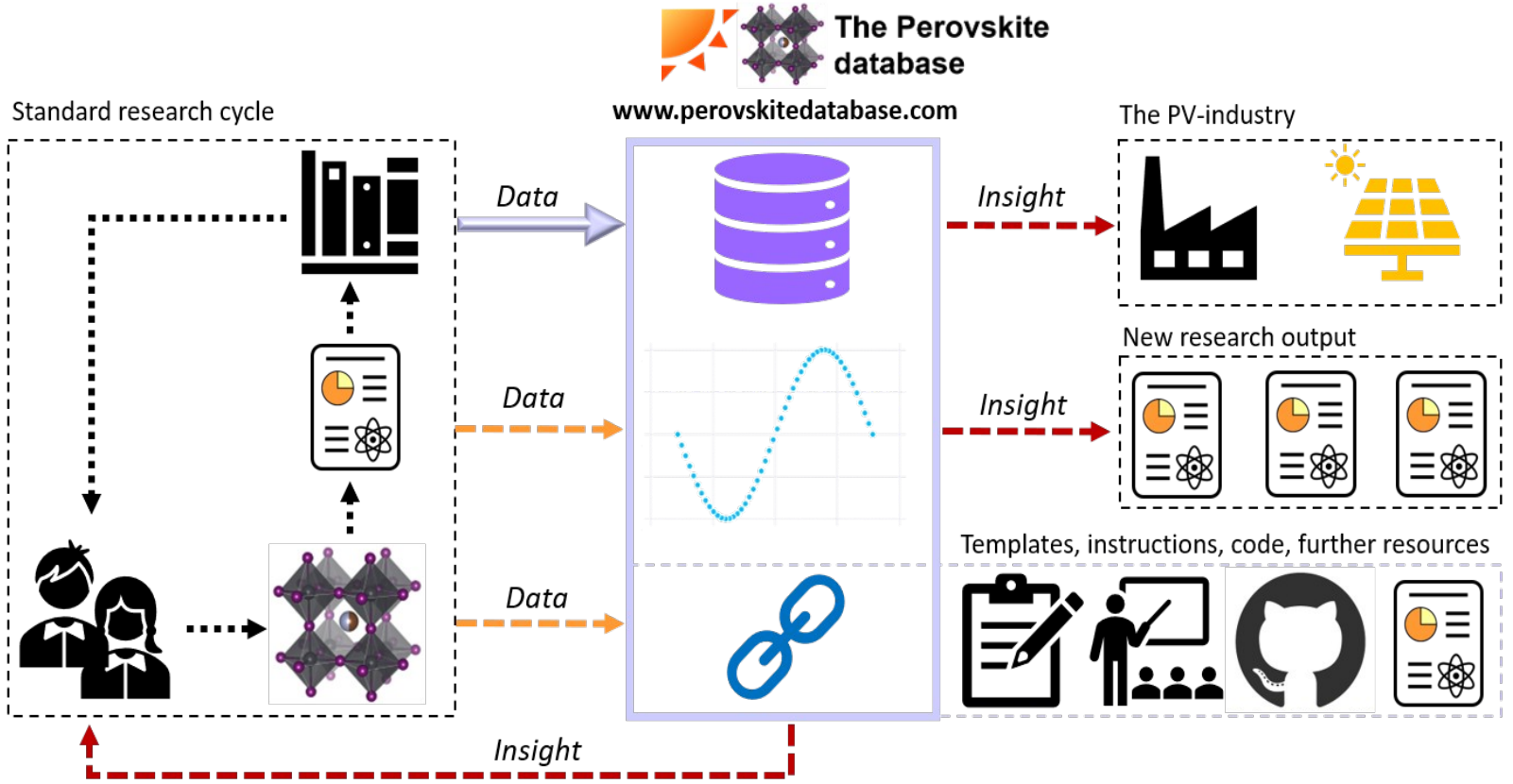


AutoPeroSol - Towards automatic data management and a common ontology for perovskite solar cell device data

HZB, KIT, FZJ – Energy, Matter

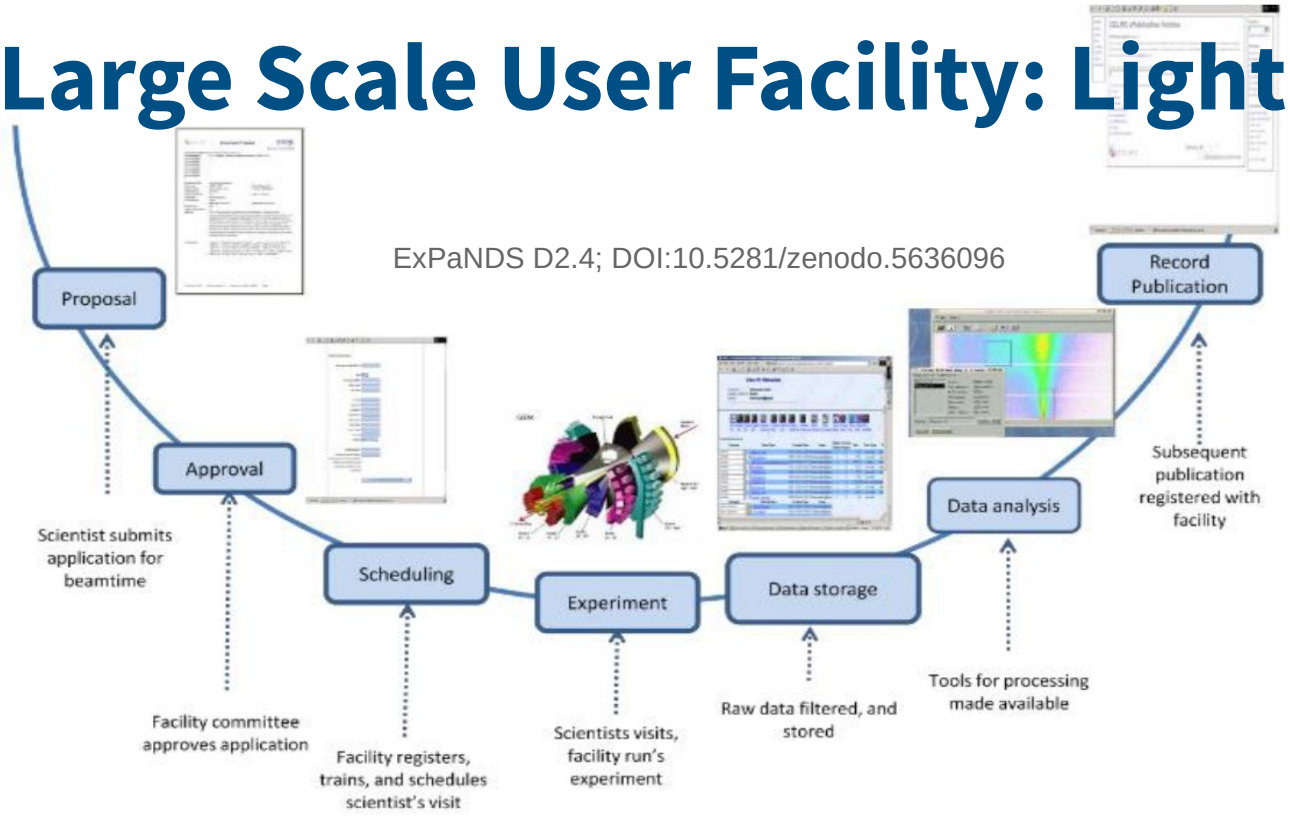
Research Cycle – In-house Researchers

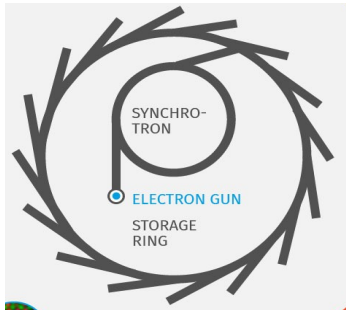




HZB Helmholtz Zentrum Berlin

Large Scale User Facility: Light Source BESSY II

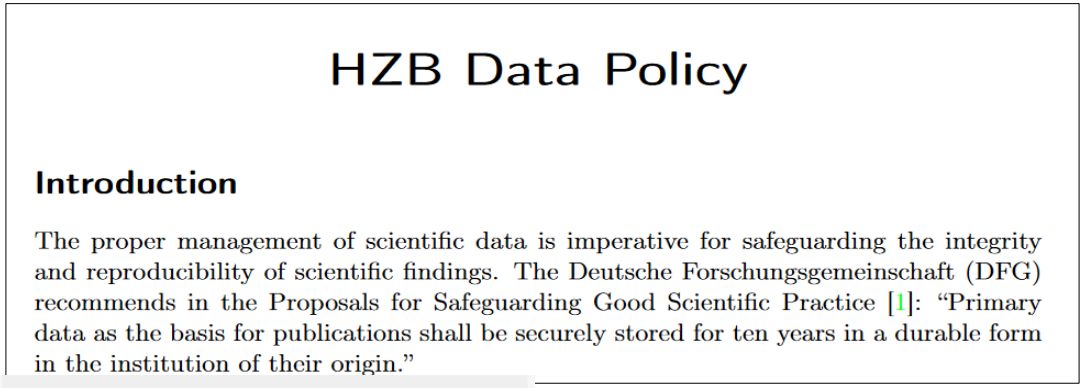




- Use cases even more diverse than in-house research
- **38 beamlines** in user operation with **~ 2700 user visits** p.a. from ~30 countries
- Embedded in **community of large-scale facilities** (Community in Germany organized in DAPHNE4NFDI)
- HMC Hub Matter at HZB

HZB BESSY II Light Source

Research Data Management at BESSY II



HZB Data Policy

Introduction

The proper management of scientific data is imperative for safeguarding the integrity and reproducibility of scientific findings. The Deutsche Forschungsgemeinschaft (DFG) recommends in the Proposals for Safeguarding Good Scientific Practice [1]: "Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin."

HZB Helmholtz Zentrum Berlin

HMC HELMHOLTZ Metadata Collaboration

Research Data Policy Review for HZB

The Research Data Policy of the Helmholtz Center Berlin (HZB) was analyzed for adherence to the FAIRsFAIR Data Policy checklist and key principles within a Proton and Neutron Research Infrastructures Data Policy of the ExPaNDS Project. This enhances the quality, reproducibility, and transparency of research results and facilitate collaboration and data sharing among researchers and cross-facility. If you require assistance with data policies, please reach out to hub-matter@helmholtz-berlin.de

ICAT

NeXus

- Building further on engagement in PaNOSC and ExPaNDS: E.g., Data policy forms framework built on community work in PaNOSC
- DAPHNE4NFDI: Guide work going forward
- Ongoing work on metadata modelling, data catalogues, PIDs, storage, authentication & authorization

Special Interest: Use Cases at Interface of NFDI Consortia

- HZB as one of few institutions that (1) operates a synchrotron and (2) conducts cutting-edge research (which in many cases makes use of synchrotron, ie user community and user facility at same place)
- We are very interested in bridging domain communities (eg catalysis) to synchrotron - Example: Helmholtz Demonstrator Project ROCK-IT (led by DESY, along with KIT, HZDR) which will need to engage with:

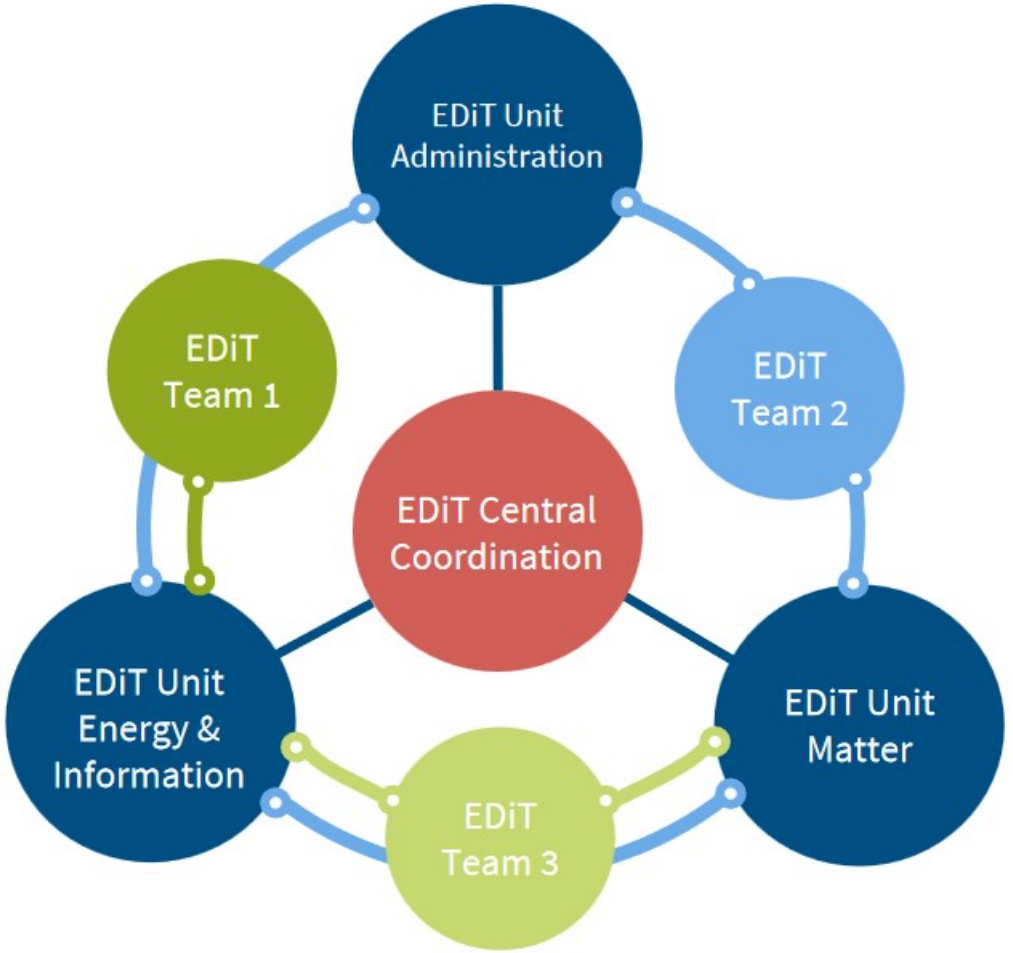
FAIRmat

DAPHNE 4NFDI

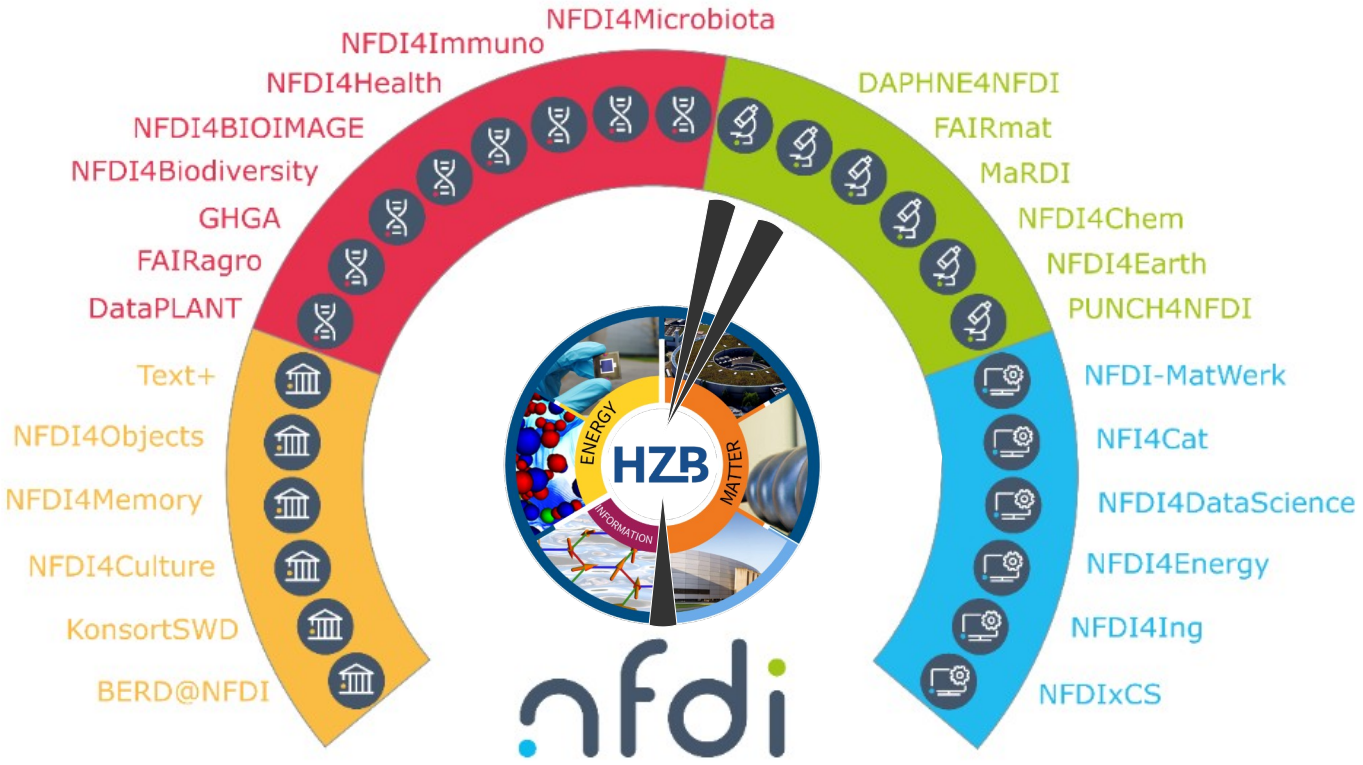
NFDI4Cat
NFDI for Catalysis-Related Sciences

Towards Organizing Work on NFDI

- Push to **structure digitalization topics** through initiative from board of directors: Enabling digital transformation (EdiT)
- (Self-)organization: Scientists with same requirements and interest to contribute actively at this stage (e.g., also through HMC proposals)
- Inform about ongoing developments, from participation in different venues (e.g., scientific retreats/department meetings) to 1:1 conversations



HZB in the NFDI



- HZB member in **two NFDI consortia** (FAIRmat, DAPHNE4NFDI)
- **Interested in the developments in many more**, specifically also in connection with large-scale user facilities
- Encountering very similar challenges as with other research data management efforts

NFDI Activities @ HZDR

3rd Helmholtz Open Science Forum on the NFDI, June 22nd 2023

U. Konrad, M. Bussmann, T. Kluge, K. Tippey, H.-P. Schlenvoigt

u.konrad@hzdr.de · www.hzdr.de

HZDR NFDI Engagement

- The HZDR is mainly involved and active in four NFDI consortia:
  - **DAPHNE**: Research with synchrotron radiation and research with neutrons
  - **PUNCH**: Particle, astro-, astroparticle, hadron and nuclear physics
  - **FAIRmat**: Condensed-matter physics and the chemical physics of solids
  - **base4NFDI**: NFDI-wide basic services in a world of specific domains

Goals

- FAIR Data Management for both, large **User Facilities** and **Research Activities**
- Provide simple and convenient access to both simulation and experiment data for **humans and machines**
- Reuse and further develop solutions achieved in **international and national projects** (EOSC, ExPaNDS, PaNOSC, ELIXIER)
- Promote **Helmholtz Solutions** (DMA, platforms HIFIS, HMC, HIP…)

DAPHNE Activities (example laser-particle acceleration)

- SciCat: collect data first in MongoDB, publish to SciCat as 2nd step
- Cataloguing of metadata from particle-in-cell simulations via openPMD and PICMI standard
  - Directly from simulation codes such as PIConGPU (SMILI, WARP-X)
  - Use for simulation resource planning, machine learning on simulation output, to experiments
- Cataloguing of metadata from laser-plasma experiments (SciCat, openPMD)
  - Web app development for experiment data logging and metadata enrichment
  - Built on use case of FWKT, but to be generalized and extensible for any kind of experiment/facility
- Metadata flow control at laser-plasma experiments (combining NEXUS & openPMD)
  - Metadata from machine to experiment log and raw data (delayed by machine OS)
  - Use for live visualization and live analysis workflows

Goals

- Simple and convenient access to both simulation and experiment data for humans and machines
- Solutions for User Facilities (e.g. ELBE, DRACO, PENELOPE, EU-XFEL HiBEF, ELI, LaserLab Europe)

PUNCH Activities

- The PUNCH4NFDI consortia is led by DESY
- Challenge to reduce data in particle and astroparticle physics
- Using simulation input to reduce data using machine learning and surrogate models
- Use modern hardware such as GPUs and FPGAs to have efficient representations of data, data compression and reduction
- Using openPMD for in-memory streaming and compressionof data from plasma simulations with PIConGPU to machine learning models (neural solvers)
- Use Alpaka for portable, performant computations on CPUs, GPUs and FPGAs in collaboration with CERN CMS and ALICE experiments as well as DESY, GSI
- Use LLAMA for efficient data representations (in collab. with CERN ROOT team)

Goals

- Fast, intelligent data reduction (Alpaka, LLAMA libraries also used in EuroHPC)
- Optimal data representations

NFDI activities at the HZI

Dealing with Microbiota and Immunology

Cordula Hege
Department Computational Biology for Infection Research

Helmholtz center for infection research (HZI)

Main campus: Braunschweig (Stöckheim)

Braunschweig Integrated Centre of System Biology (BRICS)

Technische Universität Braunschweig

BRICS
Braunschweig Integrated C
of Systems Biology

HZI is leading FAIR data activities

nfdi Nationale Forschungsdaten Infrastruktur

DZIF German Center for Infection Research

NFDI4Microbiota

Alice McHardy

NFDI4MICROBIOTA

Vision: Enabling microbiologists to easily translate research data into a deep understanding of microbial species and their interactions on a molecular level

Mission: Becoming a central hub to support researchers from the German microbiology research community with data access, data analysis, (meta)data standards and training

nfdi Nationale Forschungsdaten Infrastruktur

HZI HELMHOLTZ Centre for Infection Research

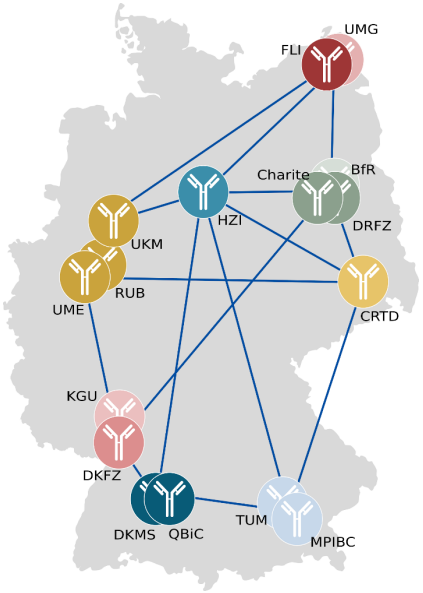
HELMHOLTZ Health

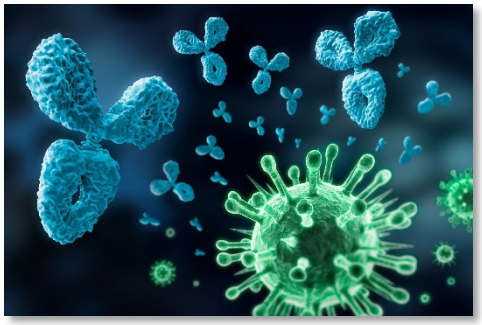
NFDI4Immuno

Michael Meyer-Hermann

NFDI4Immuno

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION





Goal of the consortium:

Attain an infrastructure facilitating integrated data analysis by the community for the advancement of immunological research

nfdi Nationale Forschungsdaten Infrastruktur

HZI HELMHOLTZ Centre for Infection Research

HELMHOLTZ Health

Interaction between the NFDI

Ongoing Interaction

- Conversation between project managers
- Participation on local Braunschweig NFDI Meeting

Cooperation

- Development of solutions for the parallel analysis of microbiota and the immune system
- Harmonisation of metadata and ontologies describing the host as well as identification and enhancement of formalised description of sampling procedure
- Collaborate on employing data analysis pipeline, exchange of computational methods

nfdi Nationale Forschungsdaten Infrastruktur

3th Helmholtz Open Science Forum: Helmholtz in the National Research Data Infrastructure Germany (NFDI)

Helmholtz Centre Potsdam

GFZ German Research Centre for Geosciences

Geo.X The Research Network for Geosciences in Berlin and Potsdam

RIFS Research Institute For Sustainability

NFDI4 MICROBIOTA

NFDI4Earth

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

GFZ Helmholtz Centre Potsdam

RIFS Potsdam Research Institute for Sustainability Helmholtz Centre Potsdam

Geo.X

bibliothek Wissenschaftspark Albert Einstein

HELMHOLTZ

GFZ German Research Centre for Geosciences

Participant in NFDI4Microbiota

Section 3.7 Geomicrobiology

The section Geomicrobiology investigates geomicrobiological processes on the Earth's surface and in the 'deep biosphere'.

- Microorganisms as engineers for the Earth's surface development
- Microbial carbon dynamics in the climate system
- The deep biosphere
- Astrobiology - Habitability and biosignatures
- Laboratory for Integrated Geological and Biological Research (GeoBioLab)

GFZ German Research Centre for Geosciences

Participant in NFDI4Microbiota

Section 3.7 Geomicrobiology

Expertise/activities in NFDI4Microbiota:

- Tool development for FAIR (metagenomic) data management and data analysis
- Teaching for data analysis and best practice analysis
- Representing geomicrobiology research topics
- Representing geomicrobiology needs for data infrastructure and data sharing with other research fields (earth surface, deep biosphere sciences)

Dr. Alexander Bartholomäus - alexander.bartholomaeus@gfz-potsdam.de

https://nfdi4microbiota.de/

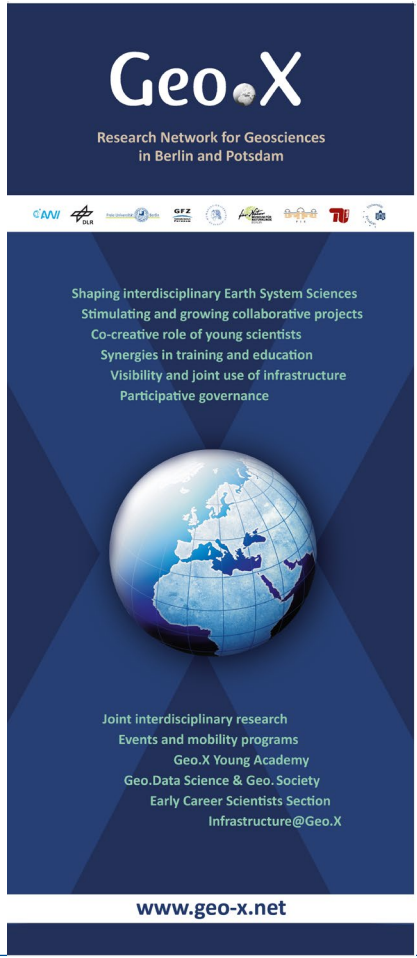
NFDI4Earth

Geo.X

Co-Applicant in NFDI4Earth via GFZ

- **Geo.X research network** of four universities and five non-university research institutions in the Berlin and Potsdam area
- Covering the **entire spectrum** of geoscientific disciplines
- **Close cooperation** in research, teaching and shared infrastructure
- Promotion of **early career scientists:** ECS Section, Geo.X Young Academy and **NFDI4Earth Academy**

More information: www.geo-x.net & @Geo_X_

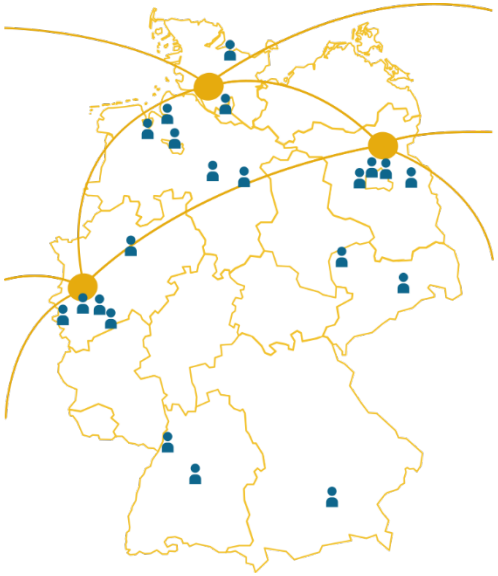


Geo.X

Co-Applicant in NFDI4Earth via GFZ

NFDI4Earth Academy – Research and learning network of Early Career Scientists integrating Earth System and Data Science

- **NFDI$_4$Earth Academy's coordinating backbone:** Research networks Geo.X, DAM, and Geoverbund ABC/J
- **39 Fellows** from 24 German universities and research institutions
- **Think tank** and **peer-mentored environment**
- **Open academy program** for the fellows and NFDI$_4$Earth community based on agile bottom-up processes
- **Connectivity** to all NFDI$_4$Earth initiatives and external training programs



https://www.nfdi4earth.de/2participate/academy & @nfdi4earth

Dr. Jonas Kuppler (Coordinator) - jonas.kuppler@gfz-potsdam.de
Dr. Hildegard Gödde (Co-spokesperson NFDI4Earth) - hildegard.goedde@gfz-potsdam.de

GFZ German Research Centre for Geosciences

Department 5: Geoinformation

Section 5.1 Library and Information Services

- NFDI4Earth coordination

Section 5.2 eScience Centre

- Digital transformation in the geoscience using state-of-the-art information technologies
- Developing services for the scientific sections and research units at GFZ and Helmholtz centers
- Involved in the strategic and organizational development in the areas of research data, scientific software development and IT research infrastructures

Section 5.3 IT Services and Operation

- IT Support and consulting
- Operation of the IT infrastructure and IT services
- Data storage, archiving and backup
- Data exchange and collaborations

DataHub
Research Field Earth and Environment

NFDI4Earth

GFZ German Research Centre for Geosciences

Section 5.1 Library and Information Services

Joint Library Wissenschaftspark Albert Einstein for the institutes GFZ, PIK, AWI Planetary Geodynamics, RIFS.

WG Library

- **Information services:** On-site library collection, branches, journals, databases, e-books, document delivery service.
- **Support of the publication process:** Open access publication, open access funds, DEAL agreements, publication costs, copyrights, reference management
- **Publication management at** GFZ, RIFS , PIK via OA-repository *public: □ publication lists on personal websites
- **Library as publisher:** DOI and URN registry, GFZpublic□ Publication platform for reports, proceedings and monographs
- **Provision of metrics:** bibliometric evaluations, performance figures for POF & QUIBS
- **Events:** library toures, coffee lectures, workshops, book a librarian
- **Training & InternshipsPraktika**

lis@gfz-potsdam.de

NFDI4Earth

GFZ Helmholtz Centre Potsdam

RIFS Potsdam Research Institute for Sustainability Helmholtz Centre Potsdam

GeoX

bibliothek Wissenschaftspark Albert Einstein

HELMHOLTZ

GFZ German Research Centre for Geosciences

Section 5.1 Library and Information Services

The team comprises expertise from librarians, research data managers, software engineers and geoscientists

WG Research Data

- **GFZ Data Services**
- **Service and support** for citable research data and software publications
- **DOI** and **IGSN** registration and catalogue services
- **Data and metadata curation**, standardized data description templates
- **Outreach** "Open Science in- and outside of GFZ": talks, workshops, discussions, international networks

Projects and activities

- DFG OA Costs 2022-2024
- FID GEO (WGEO (WG Lib+ WGData)
- HMC – RF EaE FAIR WISH
- DataHub
- NFDI4Earth
- EPOS MSL, RI@GFZ

GFZ German Research Centre for Geosciences

Co-applicant in NFDI4Earth

5.0 Geoinformation/5.1 Library and Information Services

- Measure 3.4 **International Networking & Embedding** – Co-Applicant

  Dr. Valentina Protopopova-Kakar (Coordinator) **-** valentina.kakar@gfz-potsdam.de
  Dr. Wolfgang zu Castell (Co-spokesperson) – wolfgang.castell @gfz-potsdam.de

- Measure 4.2 **Towards a Cultural Change in ESS Data Management** – co-lead

  Dr. Kirsten Elger (GFZ Data Services) – kirsten.elger@gfz-potsdam.de
  Melanie Lorenz (FID GEO) – melanie.lorenz@gfz-potsdam.de

- Measure 2.2 **User Support Network** (GEOMAR-Co-Applicant, KIT, others)

  Dr. Florian Ott - florian.ott@gfz-potsdam.de

- Measure 3.3. **NFDI Commons** – Living HandBook, Long Tail Data

nfdi@gfz-potsdam.de

https://www.nfdi4earth.de/

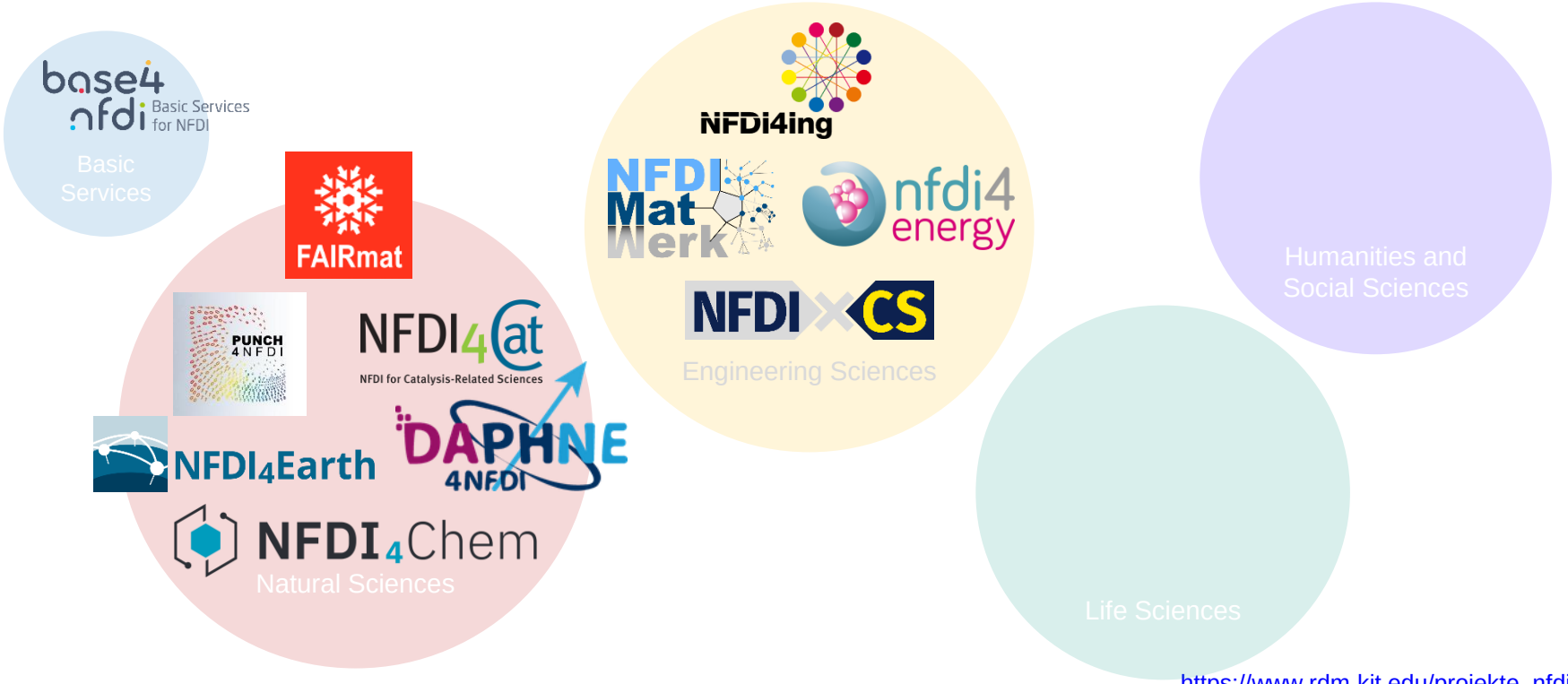
Karlsruher Institut für Technologie

NFDI networking at KIT

Dr. Alexandra Axtmann, Jan Kröger

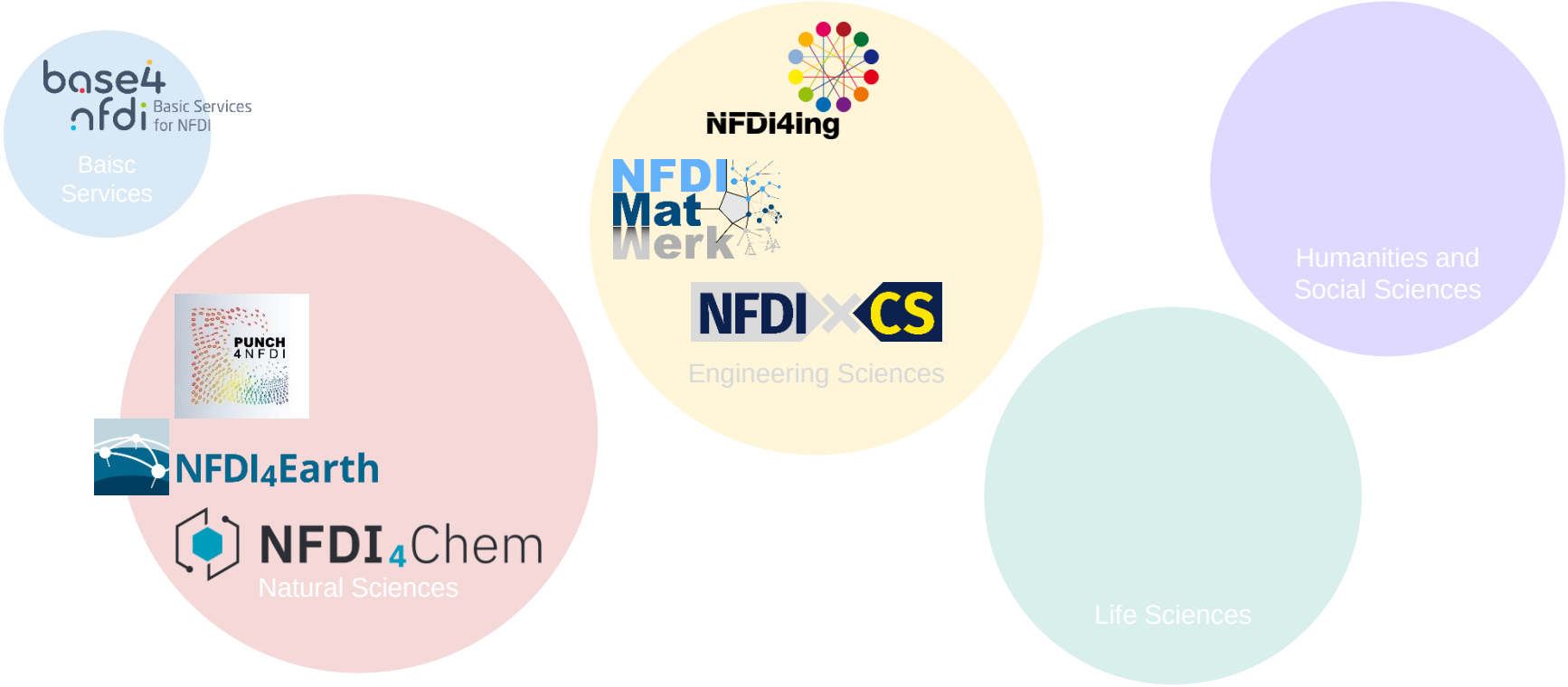
3. Helmholtz Open Science Forum on the NFDI: Inside Perspectives from the Centers
June 22, 2023

KIT's NFDI participations



https://www.rdm.kit.edu/projekte_nfdi.php

SCC’s NFDI participations



Internal Networking – Obstacles

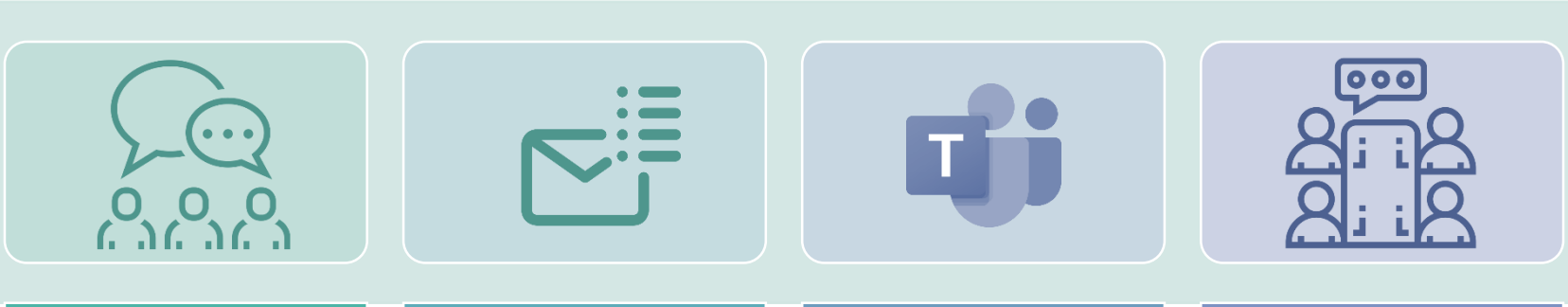
Obstacles

- Different beginnings, different knowledge levels
- Divergent work packages
- Missing well-established NFDI coordination
  > Partial overview of personel

But!

IT steering committee research and innovation +
Service team RDM@KIT
> Task to contribute to networking activities

How-to shape? Our Activities



RDM Forum

- well-established exchange forum on RDM topics
- with participation of NFDI consortia

Mailinglist for all NFDI involved

- Many to many
- actual “one to many”

Team in MS Teams

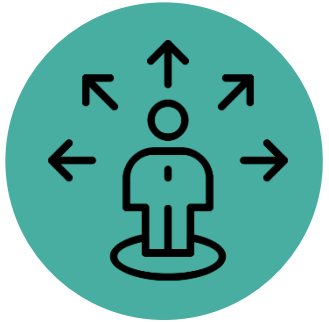
- Low-lewel contact opportunity for everyone
- Rarely used

First in-person-workshop

- as kickoff
- Problem finding a common date

NFDI networking within the SCC

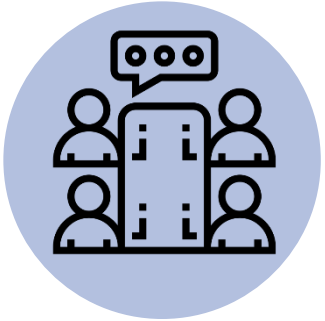
- 7 out of 11 consortia
- One responsible coordinator:
  Dr. Isabella Bierenbaum
  - central NFDI contact person
  - Regular (bi-weekly) meetings for internal information exchange (general and specific topics (AAI/IAM, storage, …)



Kickoff workshop NFDI@KIT, further plans

- Mid April 2023, 2 hours
- Barcamp-like
- Findings:
  - Many unknown people
  - Discussions, communication of topics to others
  - Activation of marginally involved scientists

- **Further plans**:
  - **In-person-meetings (centrally organized)**, quarterly (2 h) with previous topic inquiry, goal: general exchange and information
  - **thematic meetings** (self-organized by NFDI participants), as needed, announced via internal channels and NFDI RocketChat



Thank you for your attention!

Contact

Dr. Alexandra Axtmann
alexandra.axtmann@kit.edu
Serviceteam RDM@KIT

Jan Kröger
jan.kroeger@kit.edu
Digital Office

Website: rdm.kit.edu

NFDI activities at MDC

Carolina Schwedhelm, Katharina Nimptsch,
Sofia Siampani, Tobias Pischon
Molecular Epidemiology Research Group

Helmholtz
Open Science Forum
22.06.2023

MAX DELBRÜCK CENTER

nfdi 4 health

Introduction to NFDI4Health
Context & aim

BMBF → DFG call in 2019.

Aim: to improve management, accessibliity, storage, and sustainability of scientific and research data in all areas of research

Aim: to increase the value of research in epidemiology, public health, and clinical trial-based medicine, by making high quality personal health research data from Germany internationally accessible according to the FAIR data principles.

First funding period: 5 years (10/2020 – 09/2025)

Introduction to NFDI4Health
Structure

- Research institutes involved in data collection and analysis of personal health data

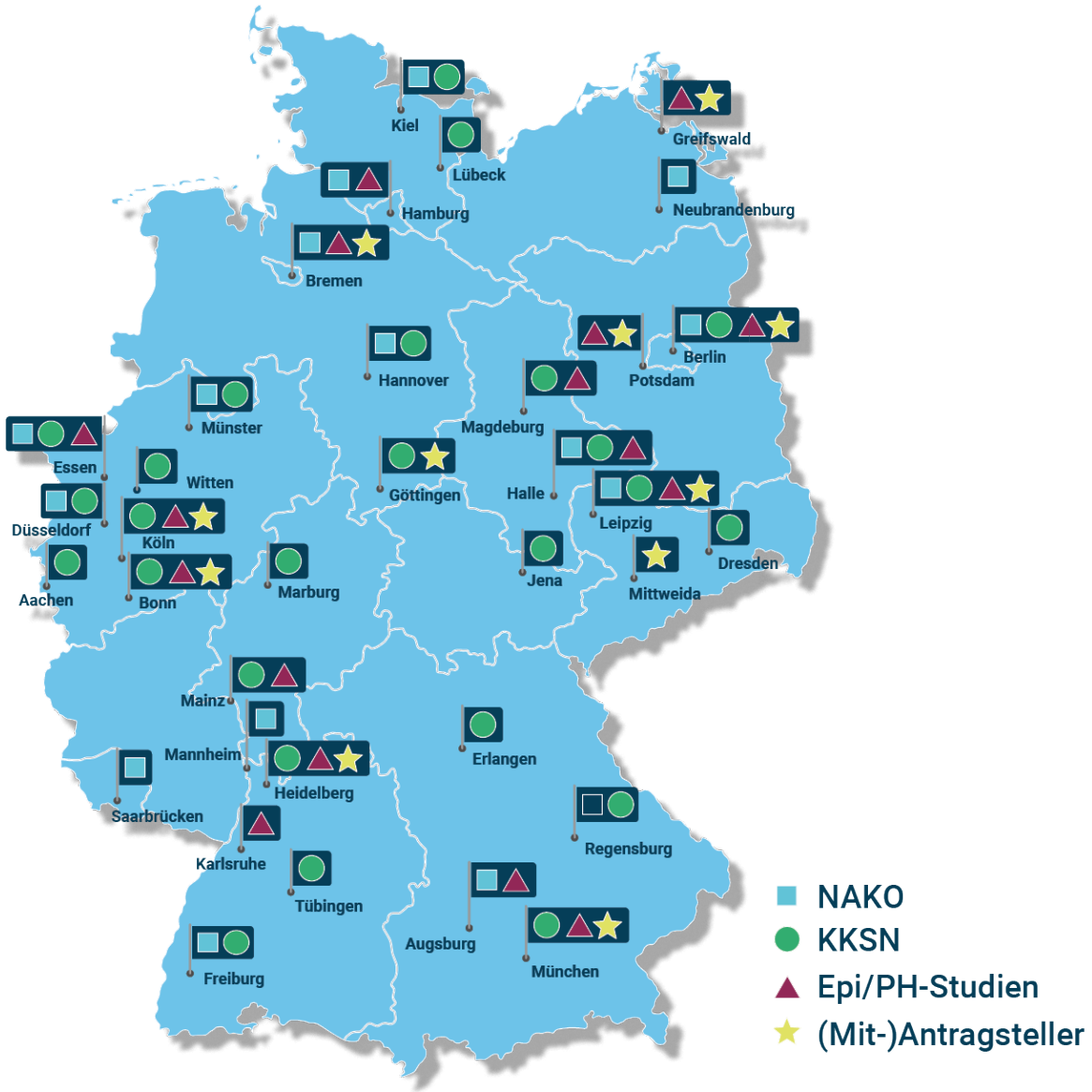
- Infrastructure facilities / Experts in development of standards and methods

- Institutions with focus on involving the user community

Introduction to NFDI4Health

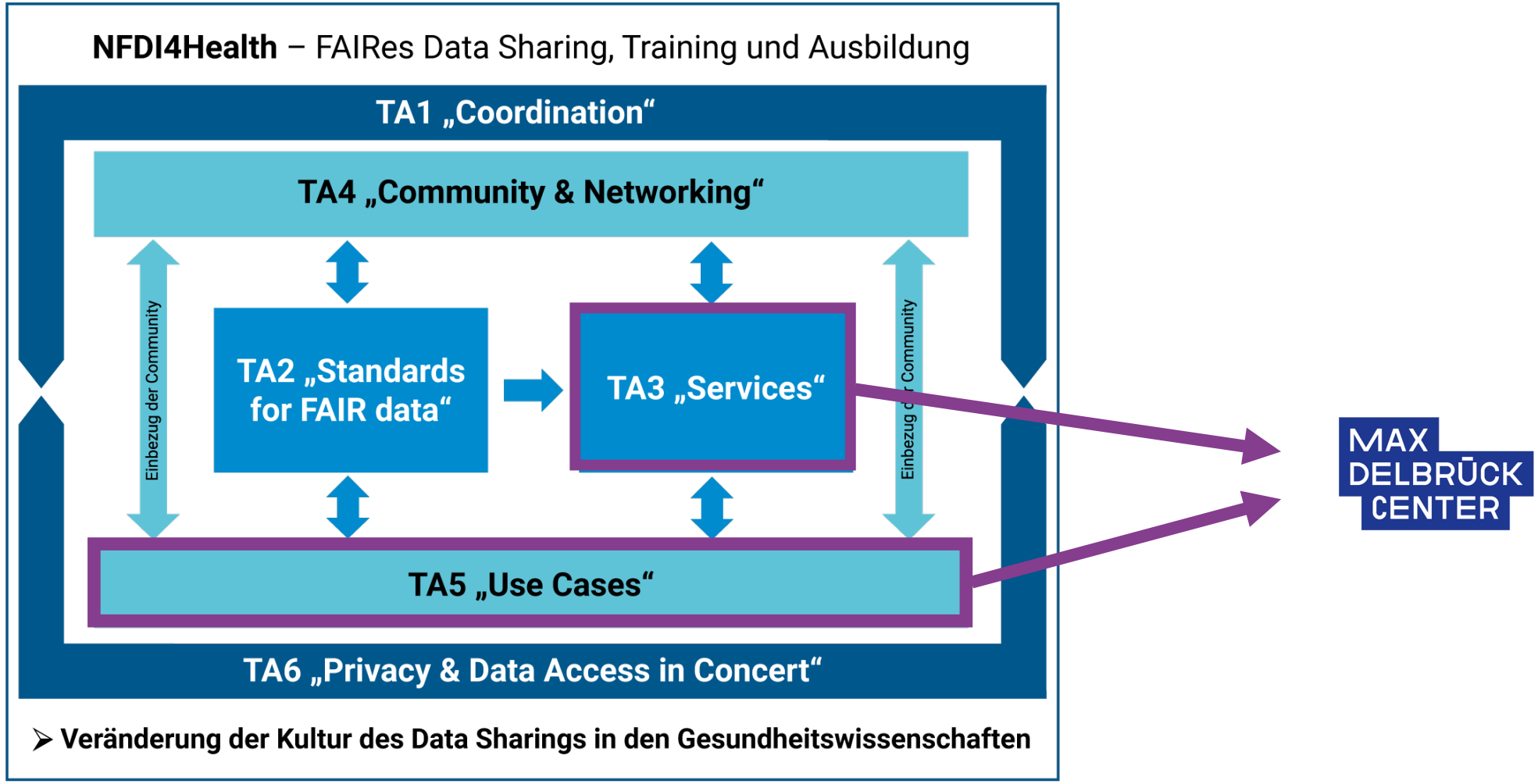
Participating dataholders

- **Epidemiological and public health studies**
  - 26 local studies with > 400,000 ▲ participants
  - NAKO Health Study ■
- **Clinical studies**
  - 24 university study centers ●
- **Registries**
- **Administrative Databases**



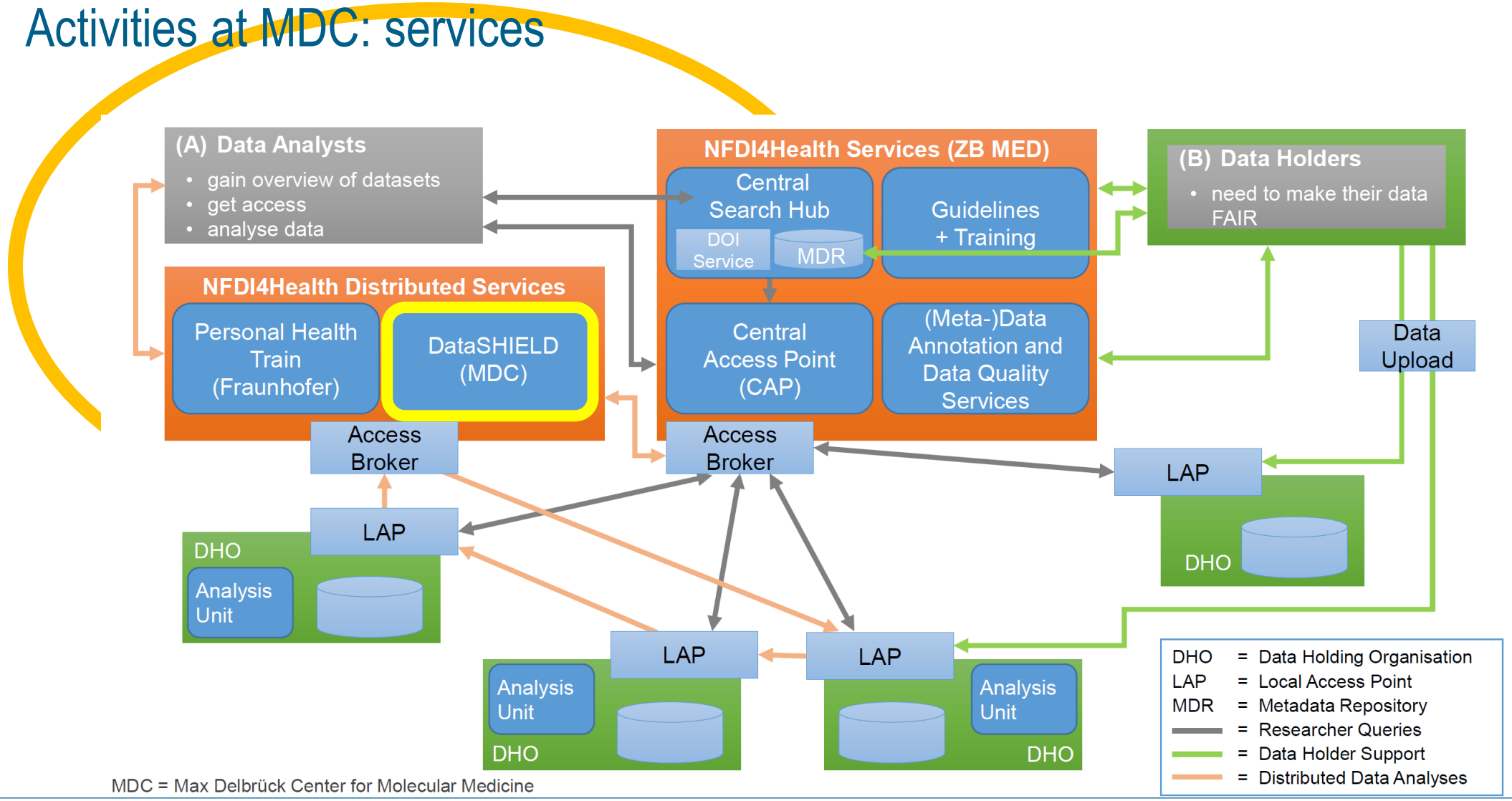
Introduction to NFDI4Health

Task Areas



Introduction to NFDI4Health

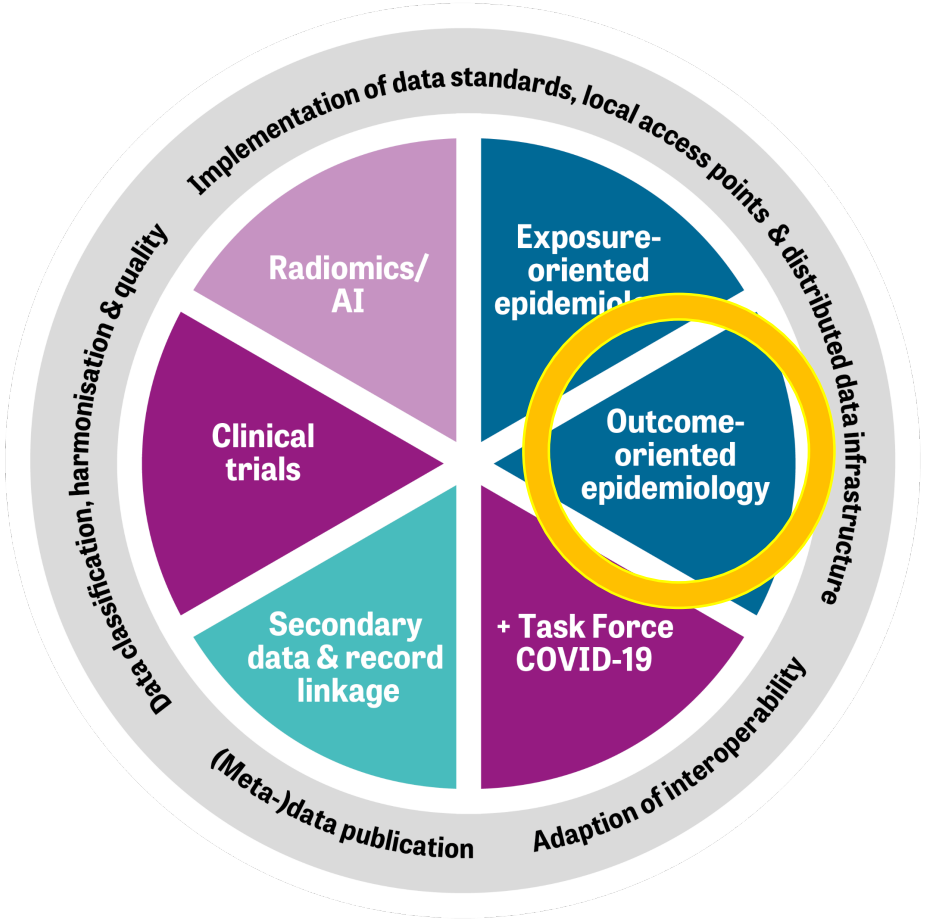
Activities at MDC: services



MDC = Max Delbrück Center for Molecular Medicine

Introduction to NFDI4Health

Activities at MDC: use cases



- Great wealth of personal health data from multiple studies → increased value of data if combined
- **Problem:**
  - Multiple data sources
  - Different definitions
  - Different assessment methods
- **Major objectives:**
  - Standardized metadata publication of studies
  - Standardized data harmonisation
  - Distributed data analysis

Data harmonisation & federated analysis

Our experience: successful projects

DataSHIELD
Secure Bioscience Collaboration

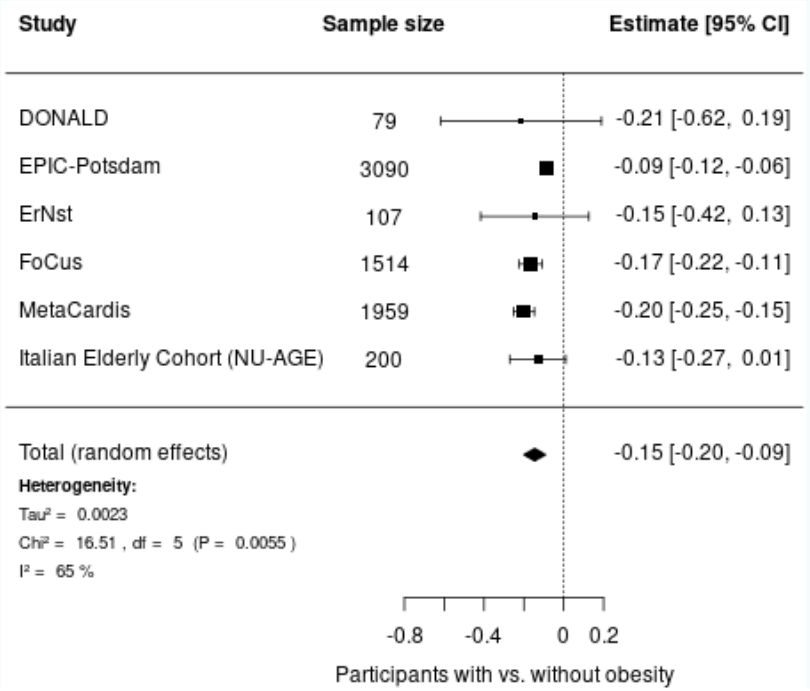
ENPADASI case study *(Pinart et al. Eur J Nutr 2021)*

- Macronutrient composition and circulating non-HDL cholesterol; a federated individual-level analysis in 8 European studies (n=5,919)

Multivariable adjusted associations between replacement of 5% of energy from carbohydrates with total fats or types of fats and non-HDL-C among adolescents and adults from eight European studies

INTIMIC case study *(Schwedhelm et al. Manuscript in preparation)*

- Gut microbiota composition and obesity in 8 European studies (n=7,656)



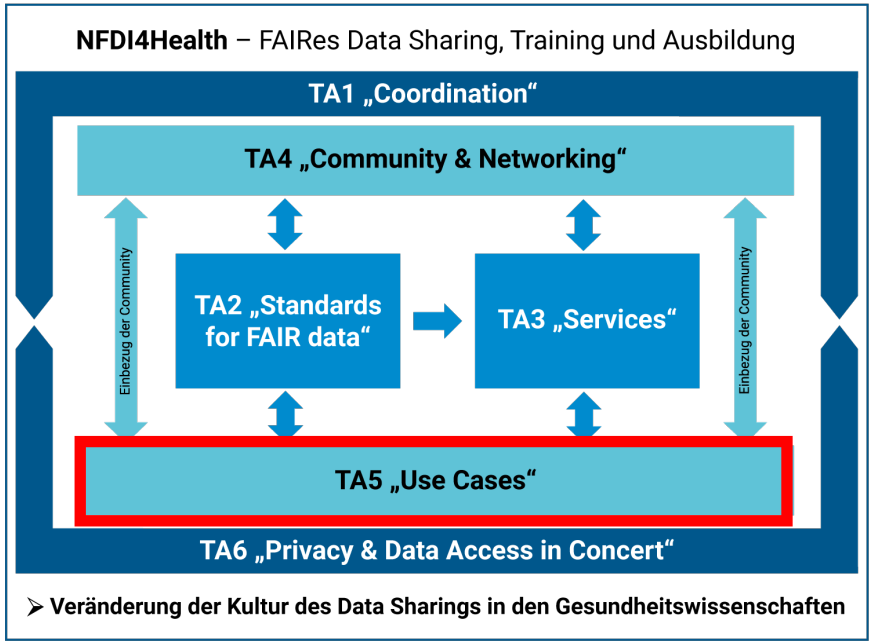
Pilot studies: data harmonisation & federated analysis

Aims

1. Pilot study 5.1.1: To investigate systematically the methodological limitations in the derivation of dietary patterns

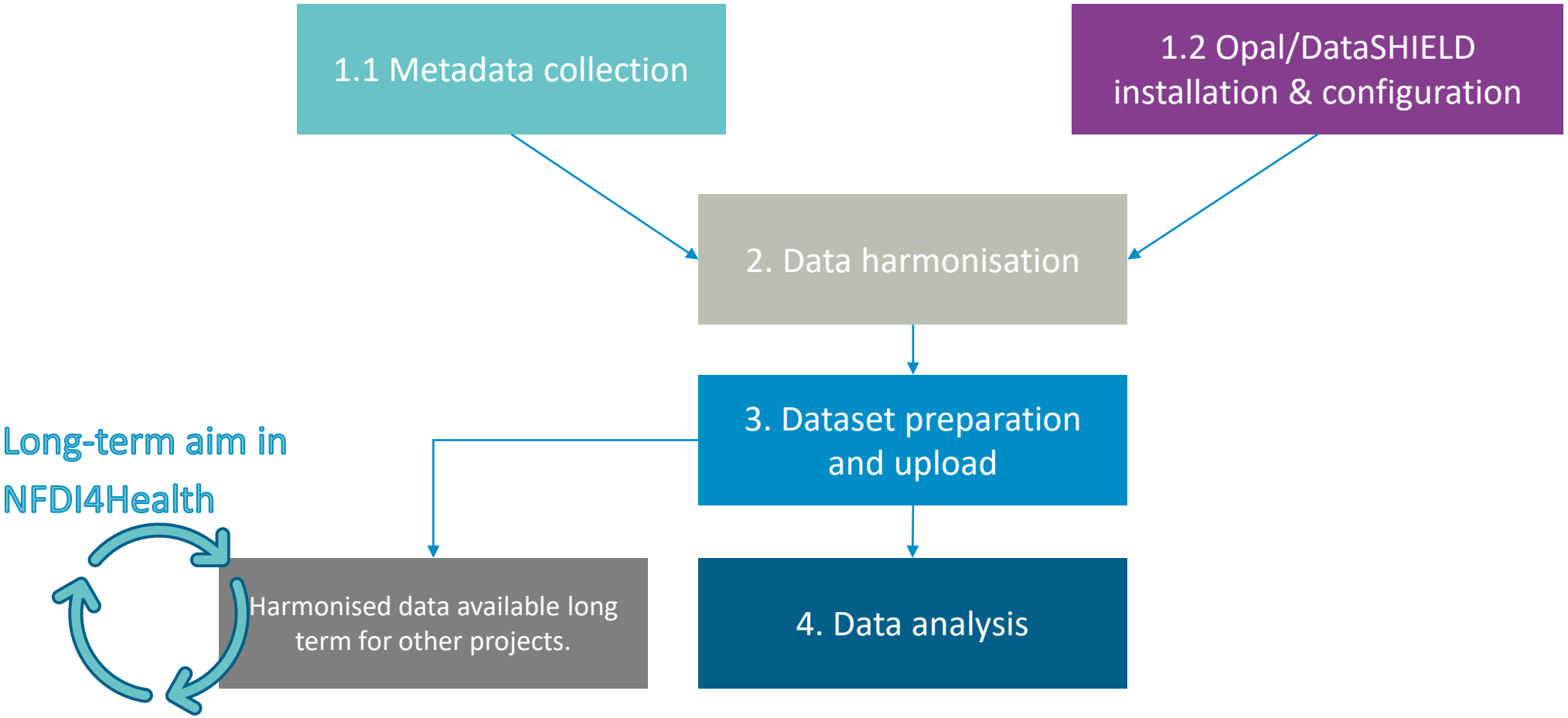
2. Pilot study 5.1.2: To investigate the association of dietary suger intake and body composition

3. Pilot study 5.2.1: To investigate the association of dietary sugar intake and glycemic load with chronic disease risk



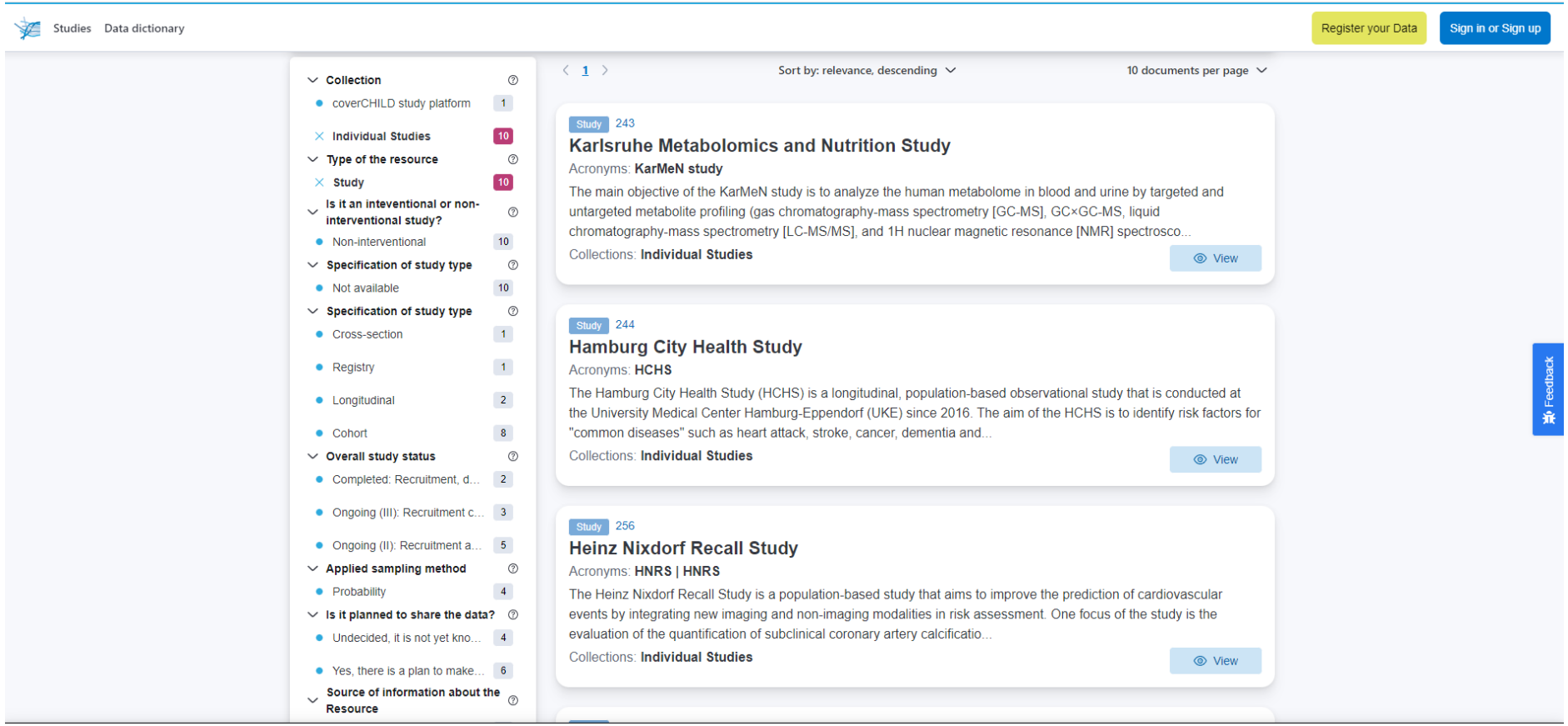
Pilot studies: data harmonisation & federated analysis

Steps



Pilot studies: data harmonisation & federated analysis

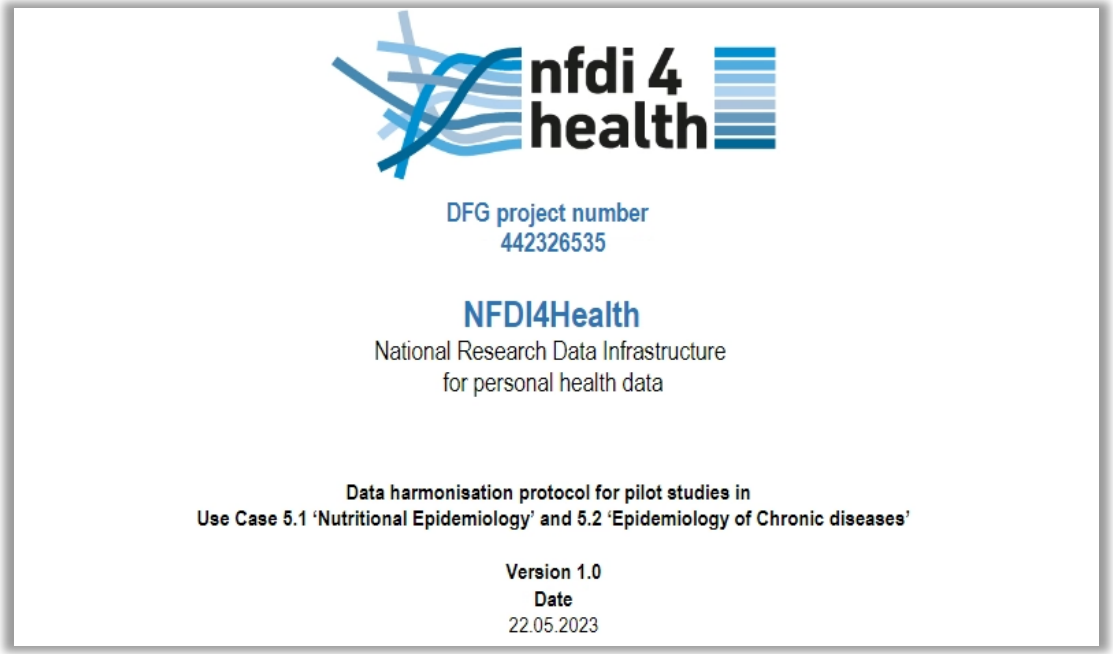
Metadata publication: German Central Health Study Hub



https://csh.nfdi4health.de/

Pilot studies: data harmonisation & federated analysis

Data harmonisation



nfdi 4 health

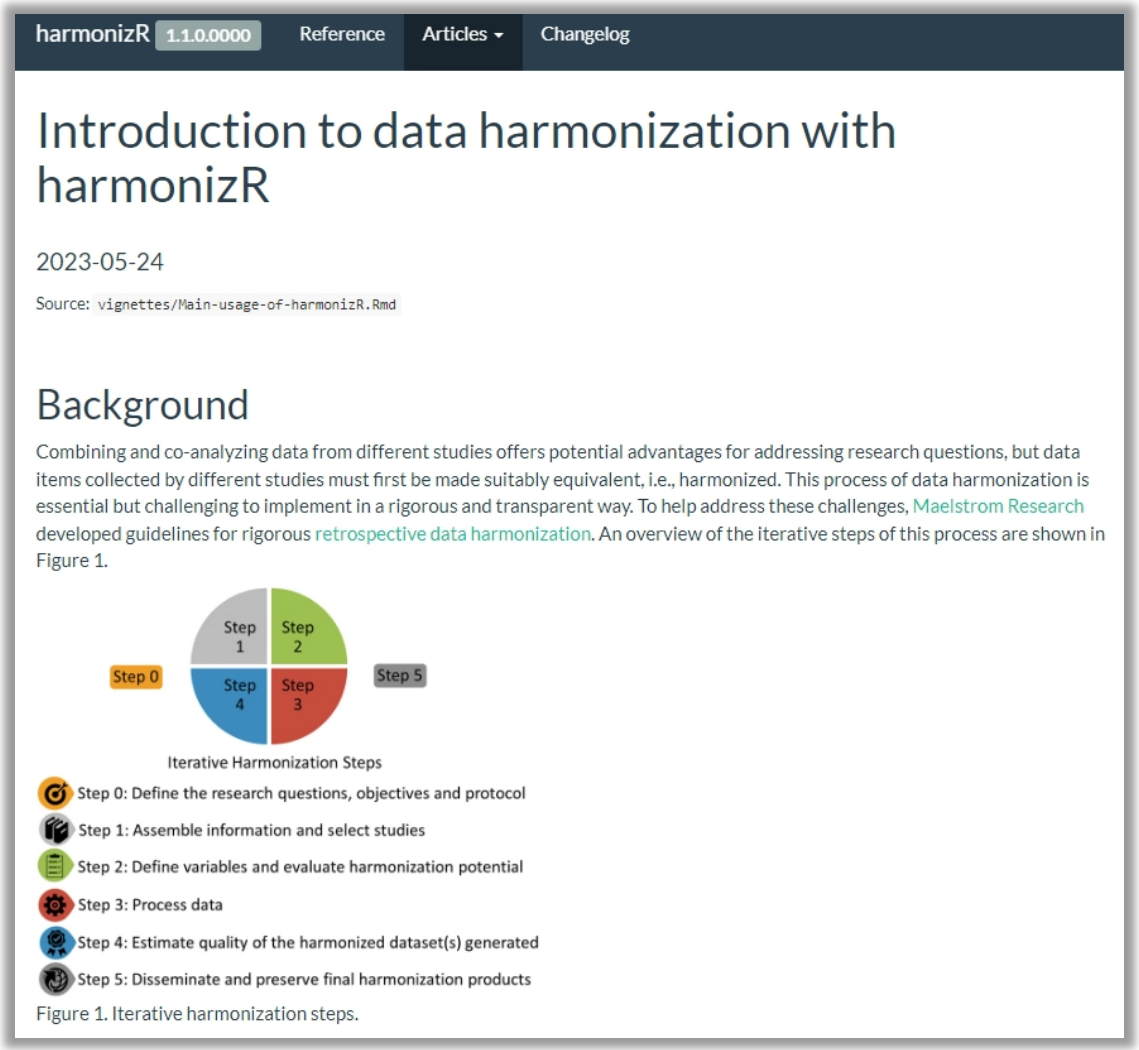
DFG project number
442326535

NFDI4Health
National Research Data Infrastructure
for personal health data

Data harmonisation protocol for pilot studies in
Use Case 5.1 'Nutritional Epidemiology' and 5.2 'Epidemiology of Chronic diseases'

Version 1.0
Date
22.05.2023

maelstrom



harmonizR 1.1.0.0000 | Reference | Articles | Changelog

Introduction to data harmonization with harmonizR

2023-05-24

Source: `vignettes/Main-usage-of-harmonizR.Rmd`

Background

Combining and co-analyzing data from different studies offers potential advantages for addressing research questions, but data items collected by different studies must first be made suitably equivalent, i.e., harmonized. This process of data harmonization is essential but challenging to implement in a rigorous and transparent way. To help address these challenges, Maelstrom Research developed guidelines for rigorous retrospective data harmonization. An overview of the iterative steps of this process are shown in Figure 1.

Iterative Harmonization Steps

- Step 0: Define the research questions, objectives and protocol
- Step 1: Assemble information and select studies
- Step 2: Define variables and evaluate harmonization potential
- Step 3: Process data
- Step 4: Estimate quality of the harmonized dataset(s) generated
- Step 5: Disseminate and preserve final harmonization products

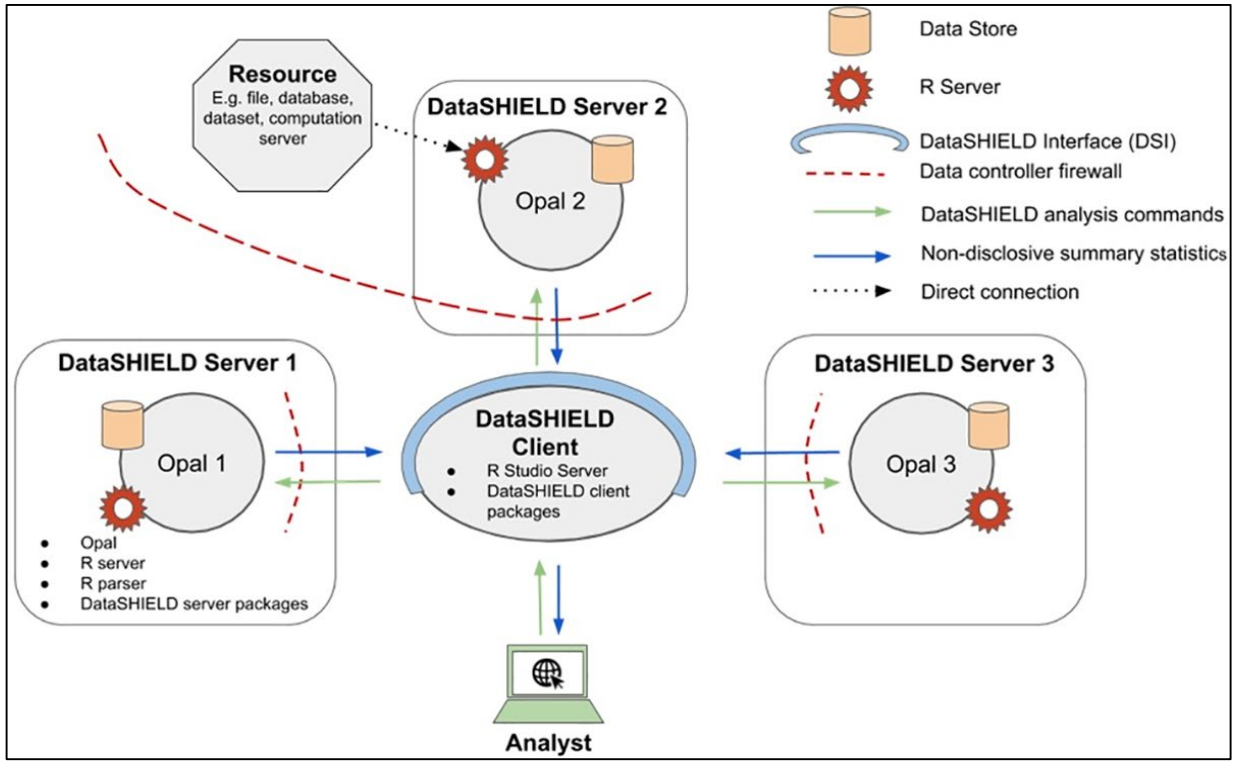
Figure 1. Iterative harmonization steps.

https://maelstrom-research.github.io/harmonizR-documentation/articles/Main-usage-of-harmonizR.html

Pilot studies: data harmonisation & federated analysis

Data analysis infrastructure

DataSHIELD
Secure Bioscience Collaboration



Marcon et al. PLOS Comput Biol 2021
Gaye et al. Int J Epi 2014

Pilot studies: data harmonisation & federated analysis

List of participants and participation status

Study name	Data Holding Organisation	Pilot study participation	Progress		
			Metadata collection	DataSHIELD	Data harmonisation
CARLA	UM Halle	1, 2, 3			
GINIPlus	HMGU	1, 2			
LISAPlus		1, 2			
KORA		1, 2, 3			
EPIC-Heidelberg	DKFZ	1, 2, 3			
Gutenberg Health Study	UM Mainz	1, 2, (3?)			
Hamburg City Health Study	UKE	NA			
Heinz-Nixdorf Recall Study	IMIBE	1, 2, 3			
KarMeN	MRI	1, 2			
NAKO	NAKO	1, 2, (3?)			
ActivE	MDC	1, 2			
DEGS1	RKI	1, 2, 3			
DONALD	U Bonn	1, 2, (3?)			
EPIC-Potsdam	DIfE	1, 2, 3			
IDEFICS/i.Family	BIPS	1, 2, (3?)			
LIFE-Adult	U Leipzig	1, 2, 3			
SHIP/SHIP-Trend	UM Greifswald	1, 2, 3			

= in progress = completed

Thank you for your attention!

Contact:

Dr. Carolina Schwedhelm

carolina.schwedhelm@mdc-berlin.de

Molecular Epidemiology Research Group
MDC

MAX DELBRÜCK CENTER

nfdi 4 health

HELMHOLTZ

Open Science