National Research Data Infrastructure NFDI and Consortia

Thomas Schörner (DESY)
Workshop on Synergies in the NFDI and beyond
5 April 2022

Overview

Gemeinsame Wissenschaftskonferenz GWK

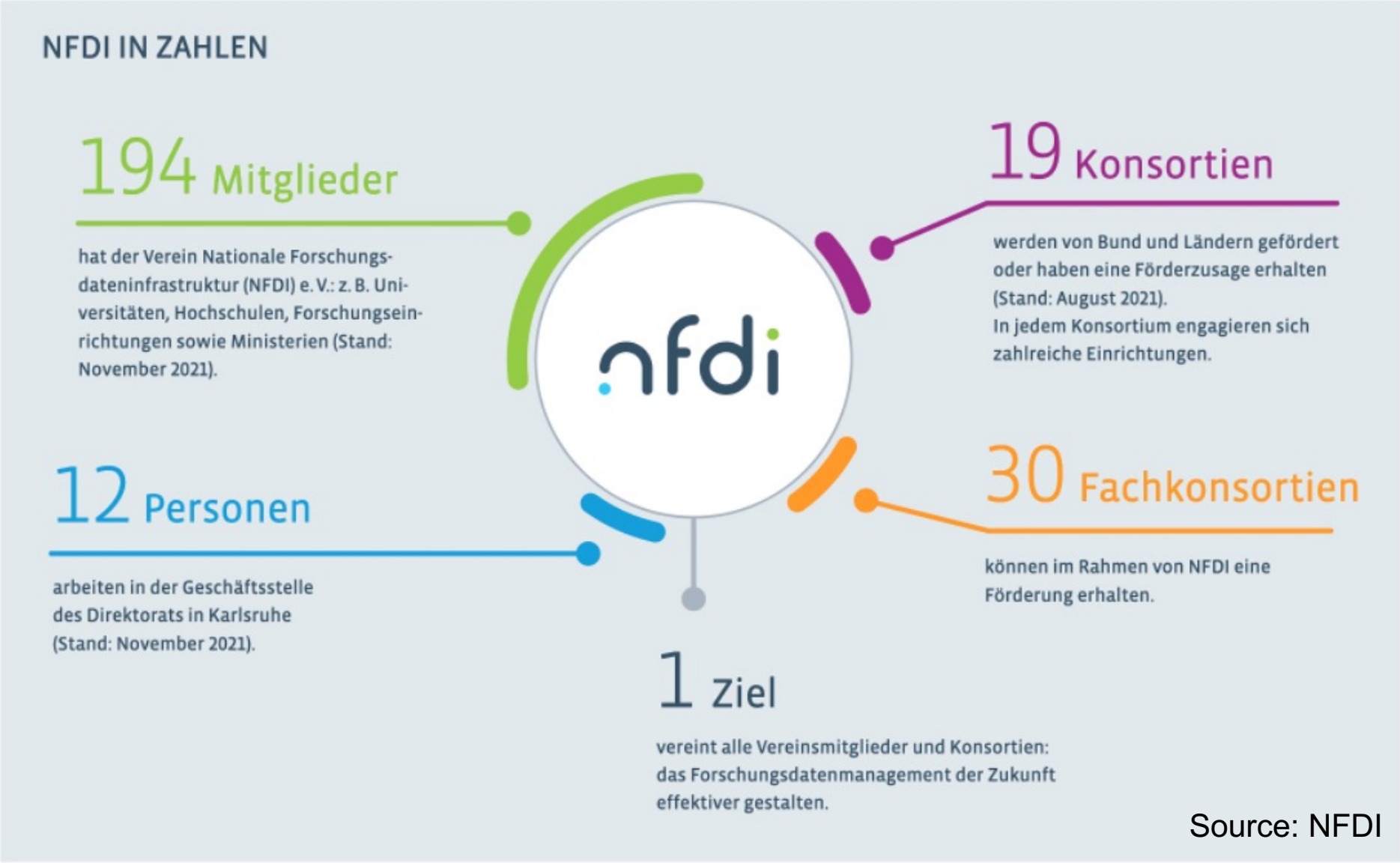
DFG Deutsche Forschungsgemeinschaft

Nationale Forschungsdaten-infrastruktur (NFDI)

- Sustainable utilisation of research data
- Establishment of FAIR data management
- Connection to European and international efforts (like EOSC)
- Bottom-up approach of independent consortia

See also DFG.de/nfdi and nfdi.de

NFDI in Numbers



Source: NFDI

Consortia so far …

… from rounds 1 and 2

1 Update Konsortien

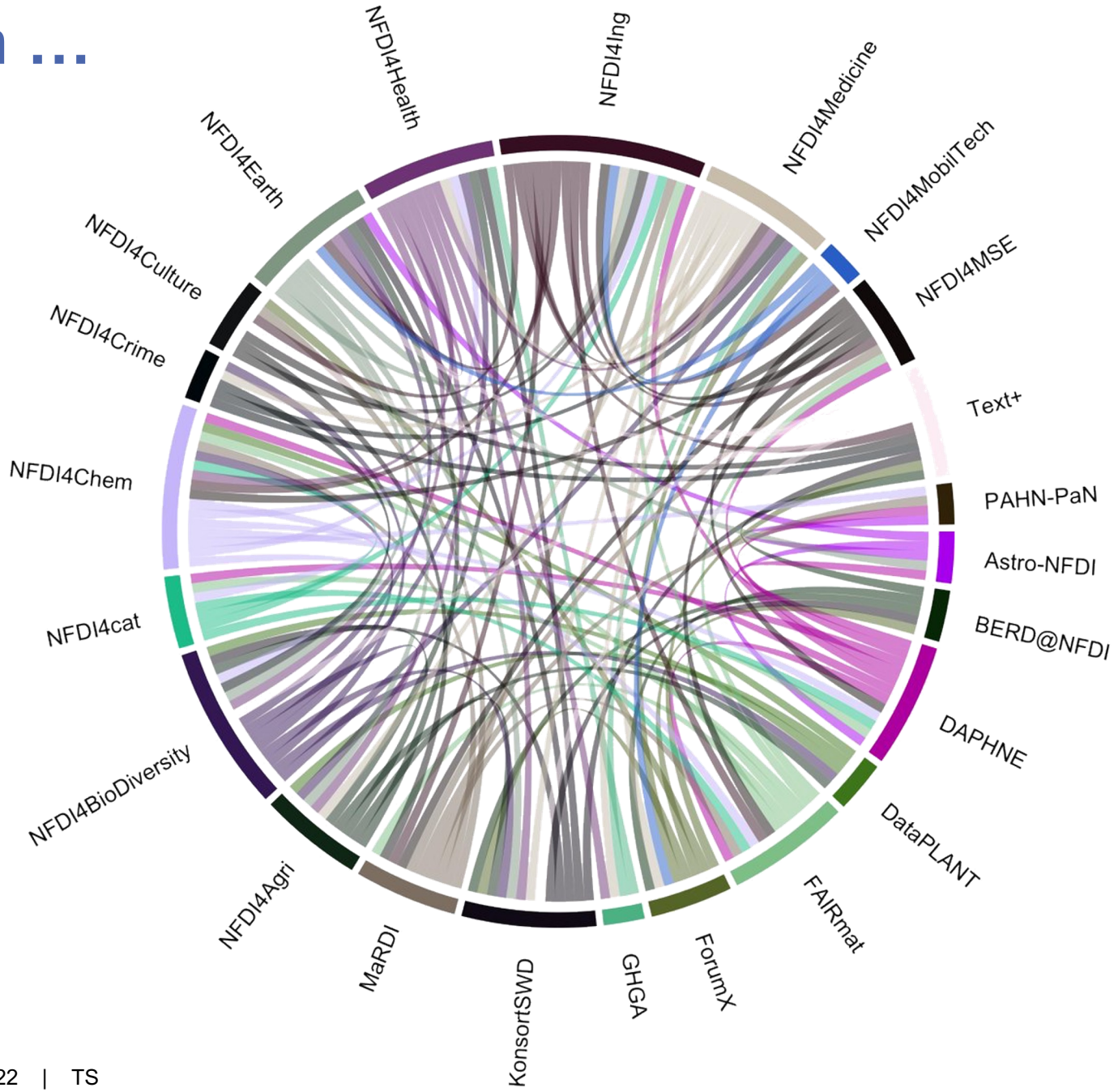
Nationale Forschungsdaten Infrastruktur

DAPHNE4NFDI	NFDI4Cat	NFDI4Microbiota
DataPlant	NFDI4Chem	NFDI-Matwerk
FAIRmat	NFDI4Culture	PUNCH4NFDI
GHGA	NFDI4DataScience	Text+
KonsortSWD	NFDI4Earth	BERD@NFDI
MaRDI	NFDI4Health	
NFDI4BioDiversity	NFDI4Ing	

Source: NFDI

Collaboration …

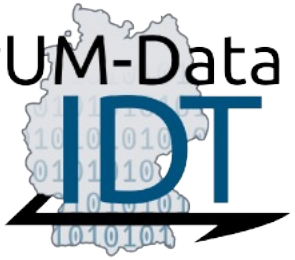
…in the NFDI …



Source: D. Strecker / Twitter

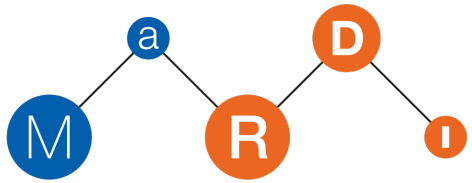
Physics-Related Consortia

…in the NFDI …



e.g. base services → later

Community committees



EUROPEAN OPEN SCIENCE CLOUD

Consortia

DAPHNE - DAta from PHoton and Neutron Experiments
16 institutions, 22 + 83 indiv. colleagues

Komitee Forschung mit Neutronen

Komitee Forschung mit Synchrotronstrahlung

Driven by the user committees KFN, KFS
Commitments of the German facilities +
ESS, ILL, ESRF, EMBL

Bring together

- KFS and KFN
- Large-scale photon and neutron research facilities
- Universities
- Research institutions
- Wider community

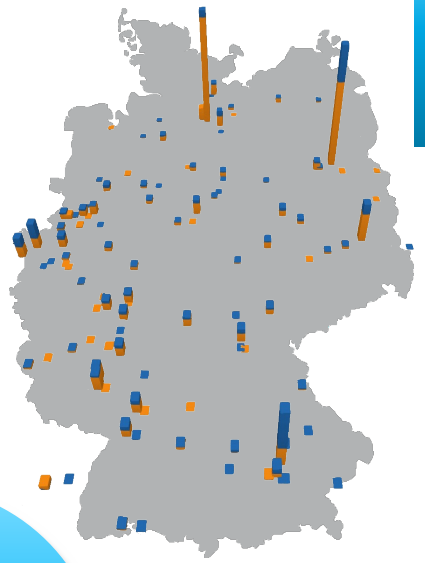
To make the growing volume of valuable measured data FAIR for the DAPHNE4NFDI community, for the whole NFDI and the scientific community.

1. Improve the **collection** of **metadata** about the measurement so that the **measured data** is **reusable** by the wider research community;
2. Implement **searchable curated databases** of raw, intermediate and processed data, **traceable** from **published and unpublished results**;
3. Develop a **curated repository of managed software** developed by leading research groups, accessible to any researcher so that others can repeat the data analysis pipelines and **re-use** the code in their own research;
4. Develop a **multidisciplinary data platform** for NFDI cross-consortia actions;
5. Provide **education** and **training** in research data management.

Research with photons and neutrons in numbers per year in Germany

per year in Germany

8 sources in Germany
33 sources in Europe
94 sources worldwide



3000
participants at facility
user meetings

Over 3000
experiments

28 PB data

3000 publications

5500 users

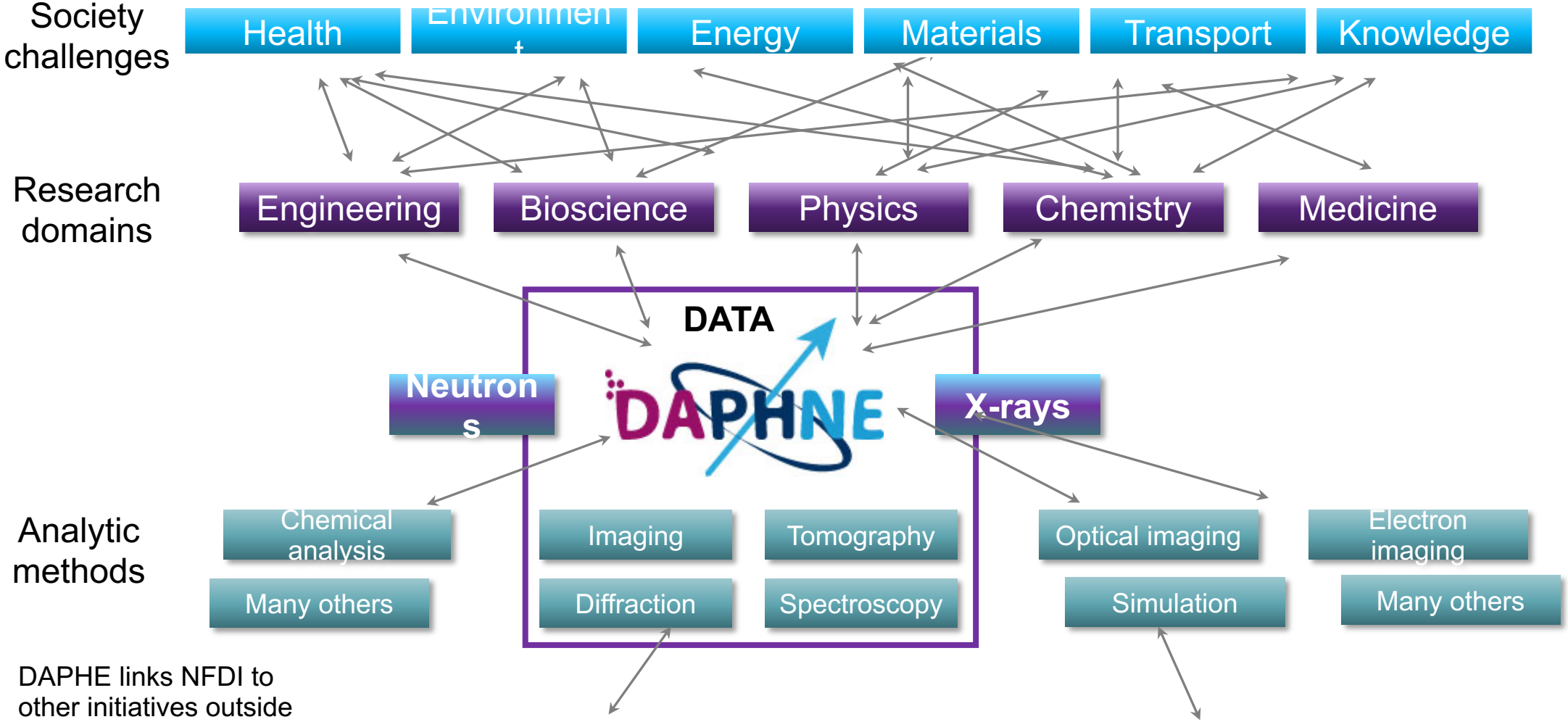
from 50 companies,
100 universities,
115 research institutions

Reaching a
community of
over 50 000

250 patents
In Europe

Our research community impacts on global challenges

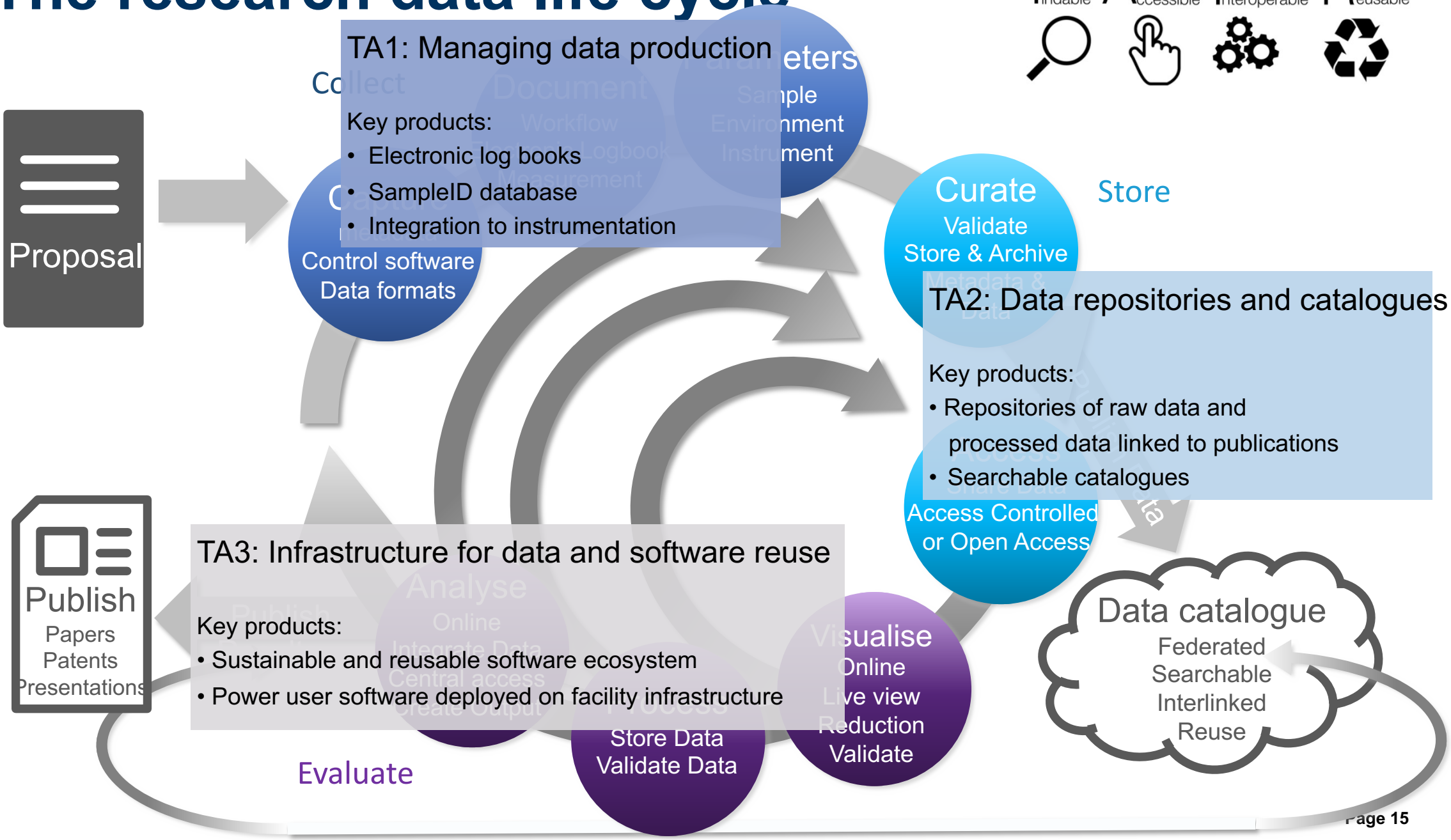
Impact extends far beyond the physics or materials science community



DAPHE links NFDI to other initiatives outside Germany

The research data life cycle

Findable Accessible Interoperable Reusable



Core products to focus on

Our deliverables to the community (and NFDI)

Which metadata to harvest, and how?
Community building

Proposal

Users

Instrumentation

Publish
Papers
Patents
Presentations

TA1: Managing data production

- Electronic log book
- Metadata
- User data curation

NeXuS/HDF5 DataDOI

ICAT SciCat

Metadata catalogue

harvesting

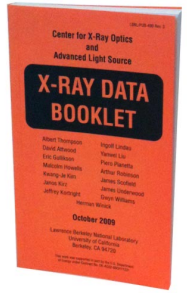
DiamondLightSource/ SynchWeb
ISPyB web application

TA2: Data repositories and catalogues

- SampleID database
- Archives of raw and processed
- Reference databases

data

API Open data portal



CXIDB Coherent X-ray Imaging Data Bank

WORLDWIDE PDB PROTEIN DATA BANK

TA3: Data and software reuse

- Machine learning interface
- Sustainable software repository
- Open software from power users

CI/CD Containers

pandata

GitHub

GitLab

Data catalogue(s)
Federated
Searchable
Interlinked
Reuse

zenodo

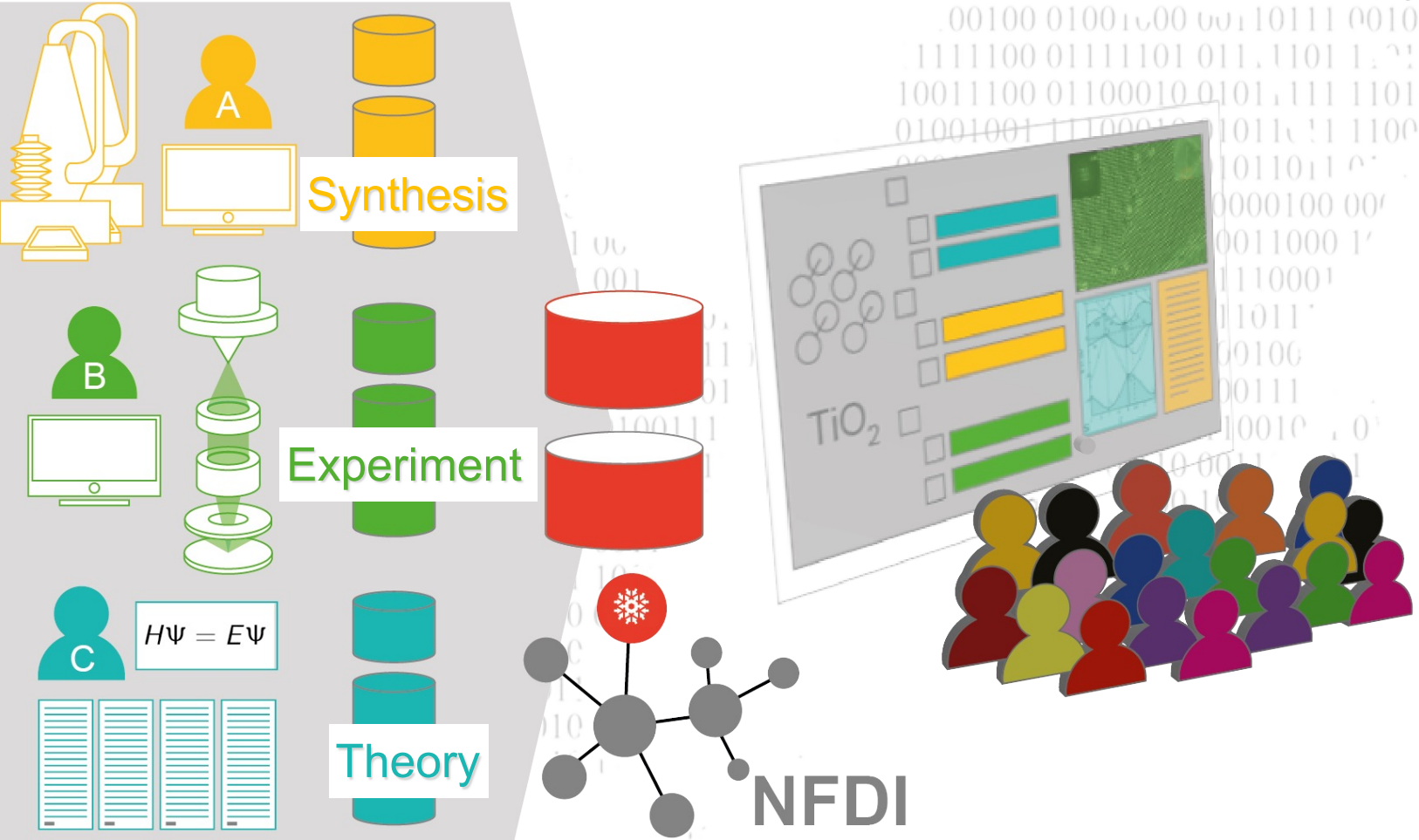
figshare

FAIRmat's goal

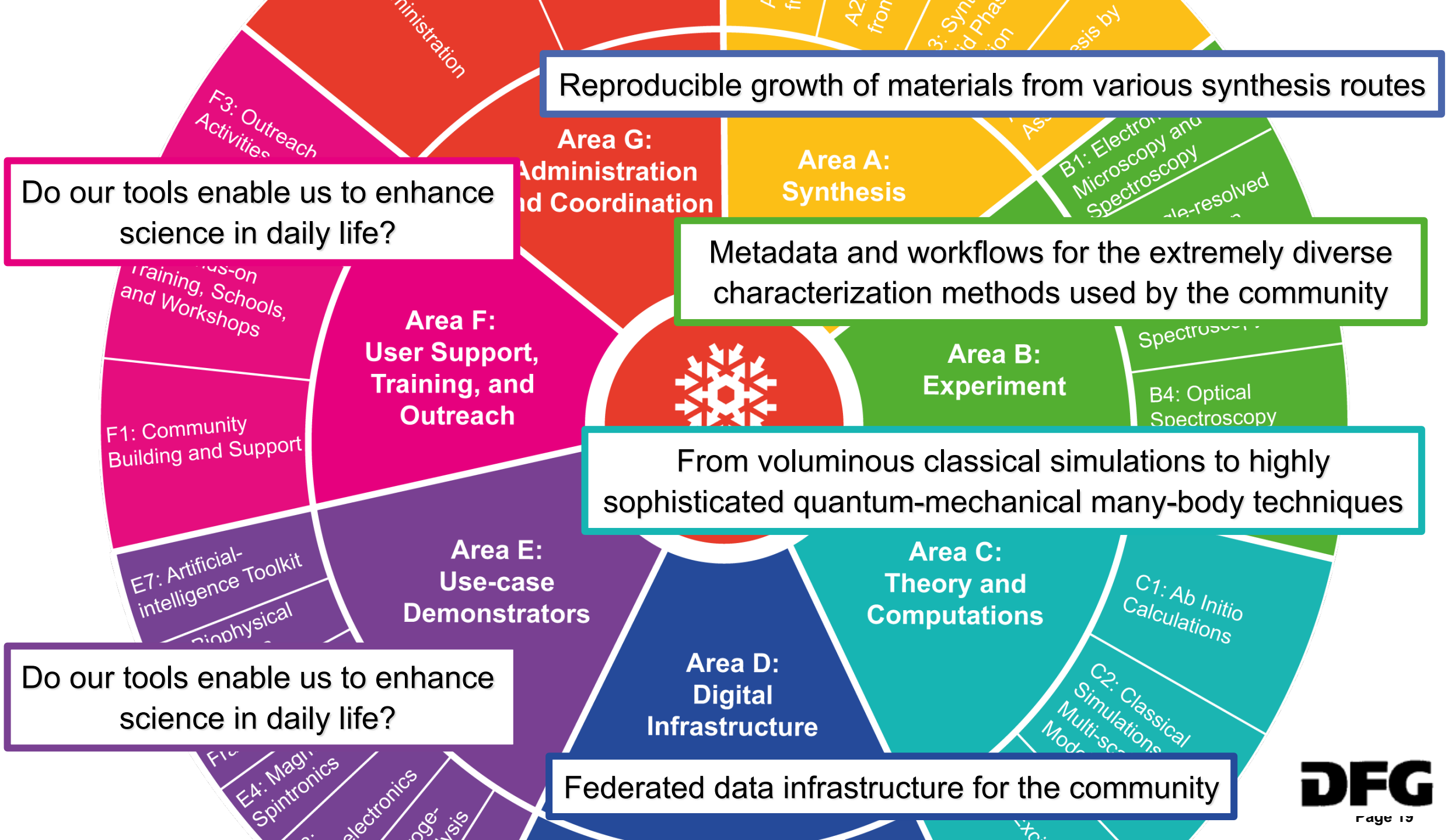
FAIR Data Infrastructure
for Condensed-Matter Physics
and the Chemical Physics of Solids

Federated data infrastructure

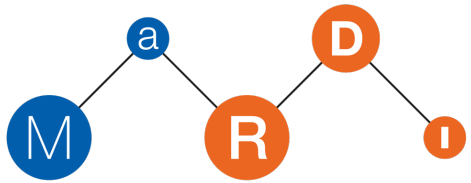
Science ▶ Federated data ▶ Centralized metadata ▶ Aggregated information ▶ Global community



An ***inclusive***, ***user-driven*** approach to develop easy-to-use tools and an infrastructure towards FAIR data processing, storage, curation, sharing, and AI readiness for future use of materials data


Reproducible growth of materials from various synthesis routes
Do our tools enable us to enhance science in daily life?
Metadata and workflows for the extremely diverse characterization methods used by the community
From voluminous classical simulations to highly sophisticated quantum-mechanical many-body techniques
Do our tools enable us to enhance science in daily life?
Federated data infrastructure for the community
Area G: Administration and Coordination
Area A: Synthesis
Area B: Experiment
Area C: Theory and Computations
Area D: Digital Infrastructure
Area E: Use-case Demonstrators
Area F: User Support, Training, and Outreach
F3: Outreach Activities
F1: Community Building and Support
E7: Artificial-intelligence Toolkit
E4: Magn, Spintronics
B1: Electron Microscopy and Spectroscopy
B4: Optical Spectroscopy
C1: Ab Initio Calculations
C2: Classical Simulations
Training, Schools, and Workshops

MaRDI: Mathematical Research Data Initiative (Overview)



Mission

Mathematical research data is vast, complex and multifaceted. It emerges within mathematical sciences but also in other scientific areas such as physics, chemistry, life sciences and the Arts. Standardised formats, data interoperability and application programming interfaces need to be developed to ensure the ease of use of data across broad disciplines.

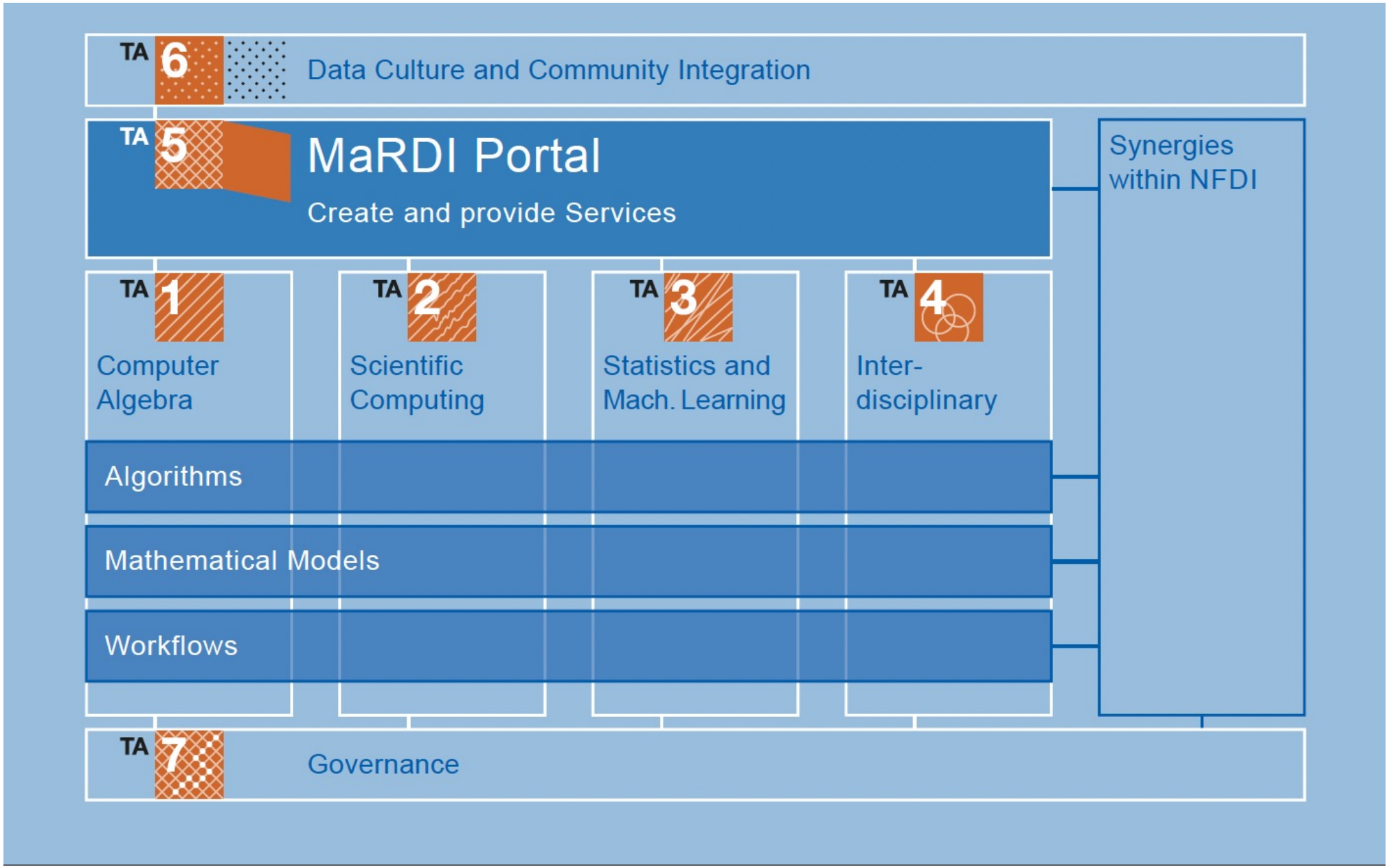
With this in mind, the Mathematical Research Data Initiative (MaRDI) is being established as **the** consortia initiative of mathematical science. Its mission is to:

1. develop a robust Mathematical Research Data Infrastructure that would be useful within mathematics and other disciplines as well as non-scientific fields.
2. set standards and confirmable workflows for certified Mathematical Research Data and
3. provide services to both the mathematical and wider scientific community.

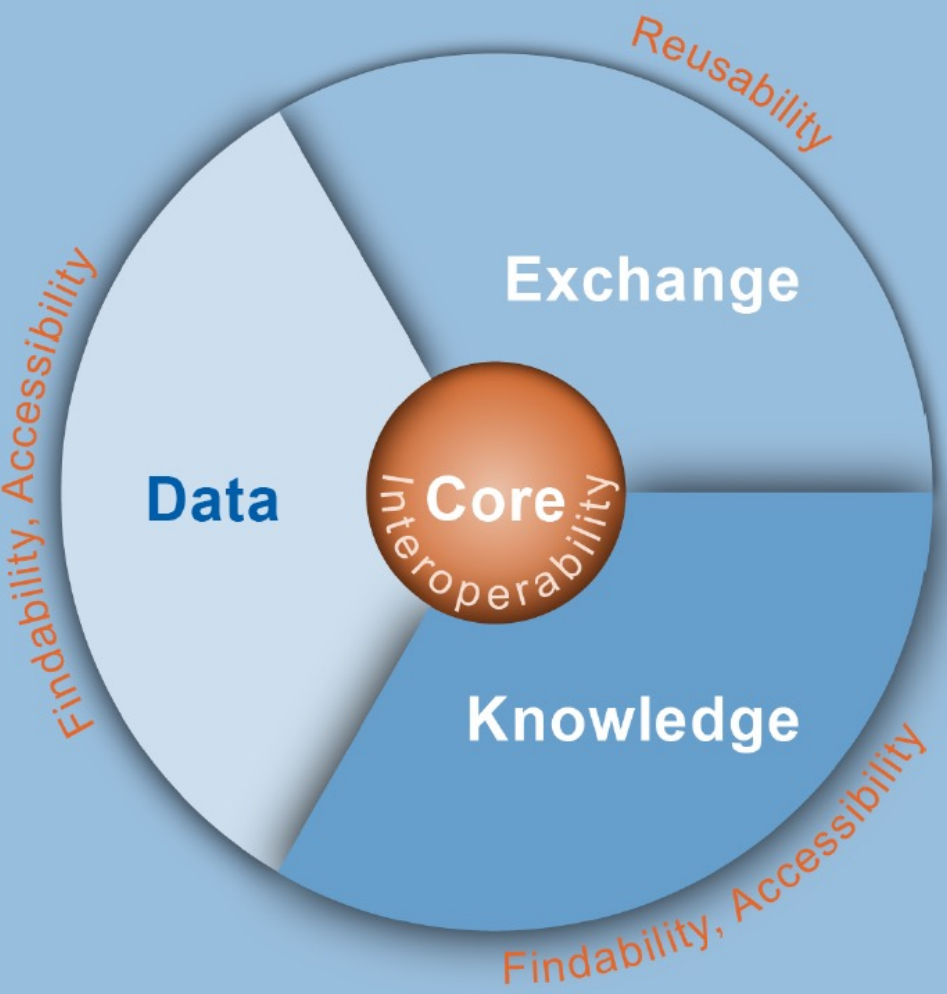
All of which is essential in creating and establishing collaborative platforms crucial for knowledge dissemination, quality control and scientific discourse.

MaRDI's Vision:
Building a community that embraces a FAIR data culture and research workflow through the sustainable realization of MaRDI findings.

MaRDI: Mathematical Research Data Initiative (Overview)



MaRDI Layer Architecture for FAIR Data



Example: Linear Solver $Ax=b$

X1: Core

- Data structures for matrices and vectors
- Representation formats
- Exchange formats
- Application programming interfaces (APIs)

X2: Data

- Test cases (matrices, solutions)
- Matrix properties (meta-data)
- Persistent identifiers (PIDs)

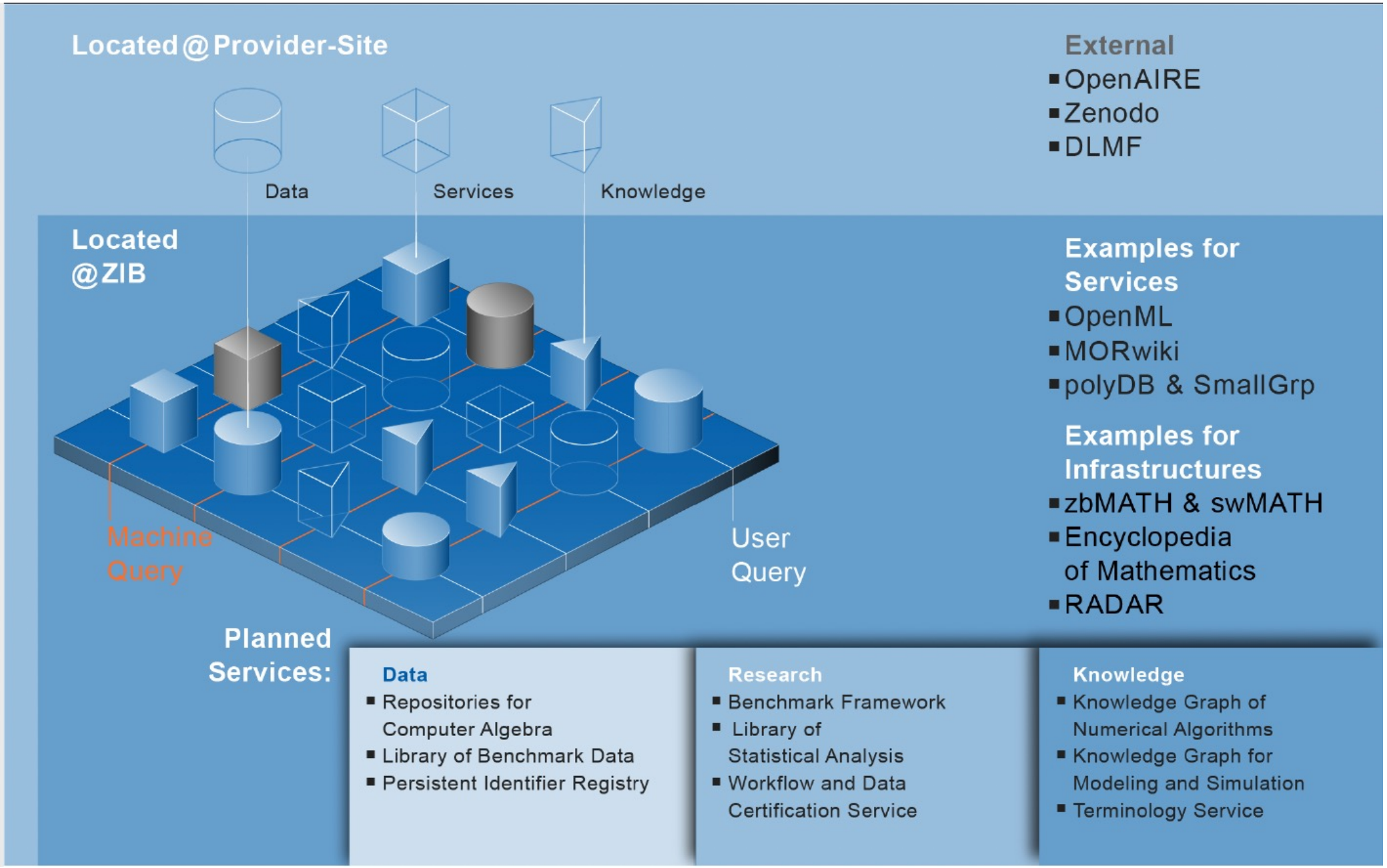
X3: Exchange

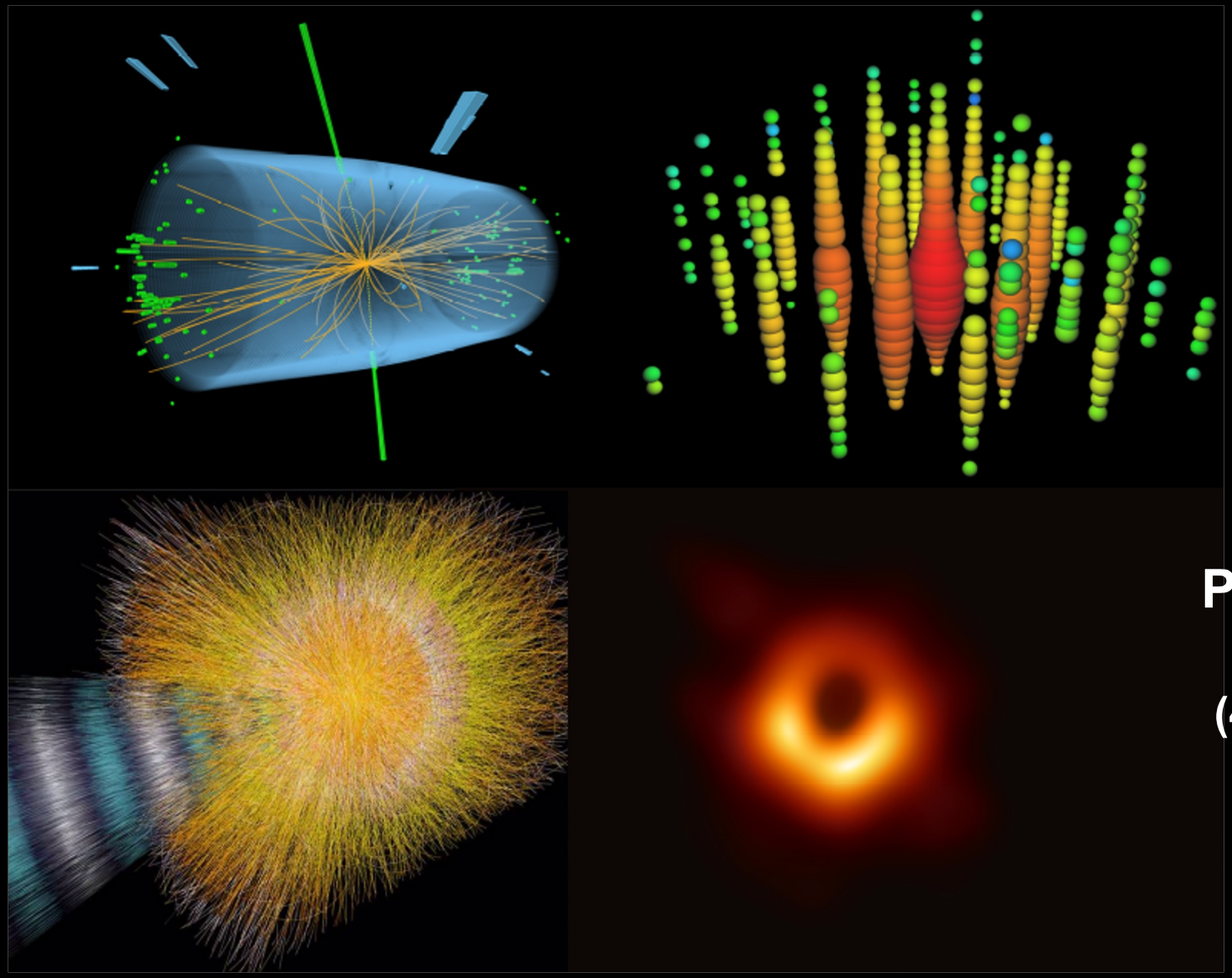
- Benchmark framework
- Pre-defined software environments
- Workflows, continuous-benchmarking

X4: Knowledge

- Ontology of linear problems and solvers
- Link to algorithms, publications software and test-data

An Federated Infrastructure and Value-Chain for New Services for Data, Research and Knowledge





PUNCH 4NFDI

Particles, Universe, NuClei and Hadrons for the NFDI

(42 partners, 10k scientists, support from KET, KAT, KHuK)

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 use of data and computing resources of the involved communities and beyond.

Global Internet Tr
420EB
FAIR@GSI
Online data
30 EB
PUNCH data are diverse
- in size and rate
- in complexity and purpose
- in abstraction level
FAIR@GSI
Physics
300 PB
Google
98PB
facebook
180PB
business emails sent worldwide
3,000PB
(3EB)
4EB
SKA1 mid
archive
PUNCH4NFDI expertise
- Big data and open data
- Data irreversibility and reduction
- Harnessing heterogeneous resources
- Highly collaborative globally distributed data management
Low
15.6ZB
HL-LHC
Challenges
- FAIRification: data & workflows
- “big data” and “open data”
- Irreversibility challenge and data loss
- heterogeneous data & infrastructures
- transfer of knowledge

…

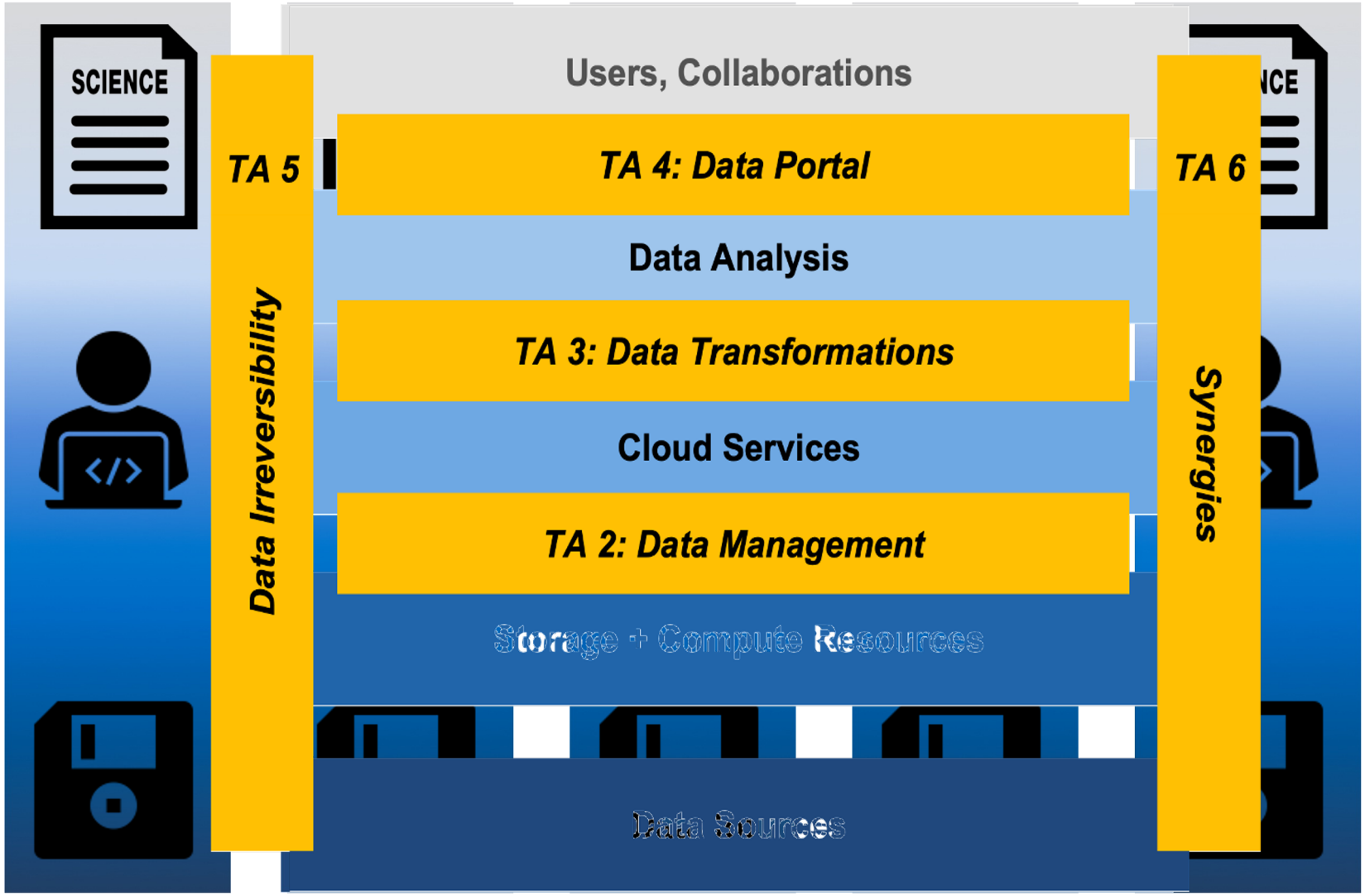
Analysis tools

Compute resources

Data access

Data storage

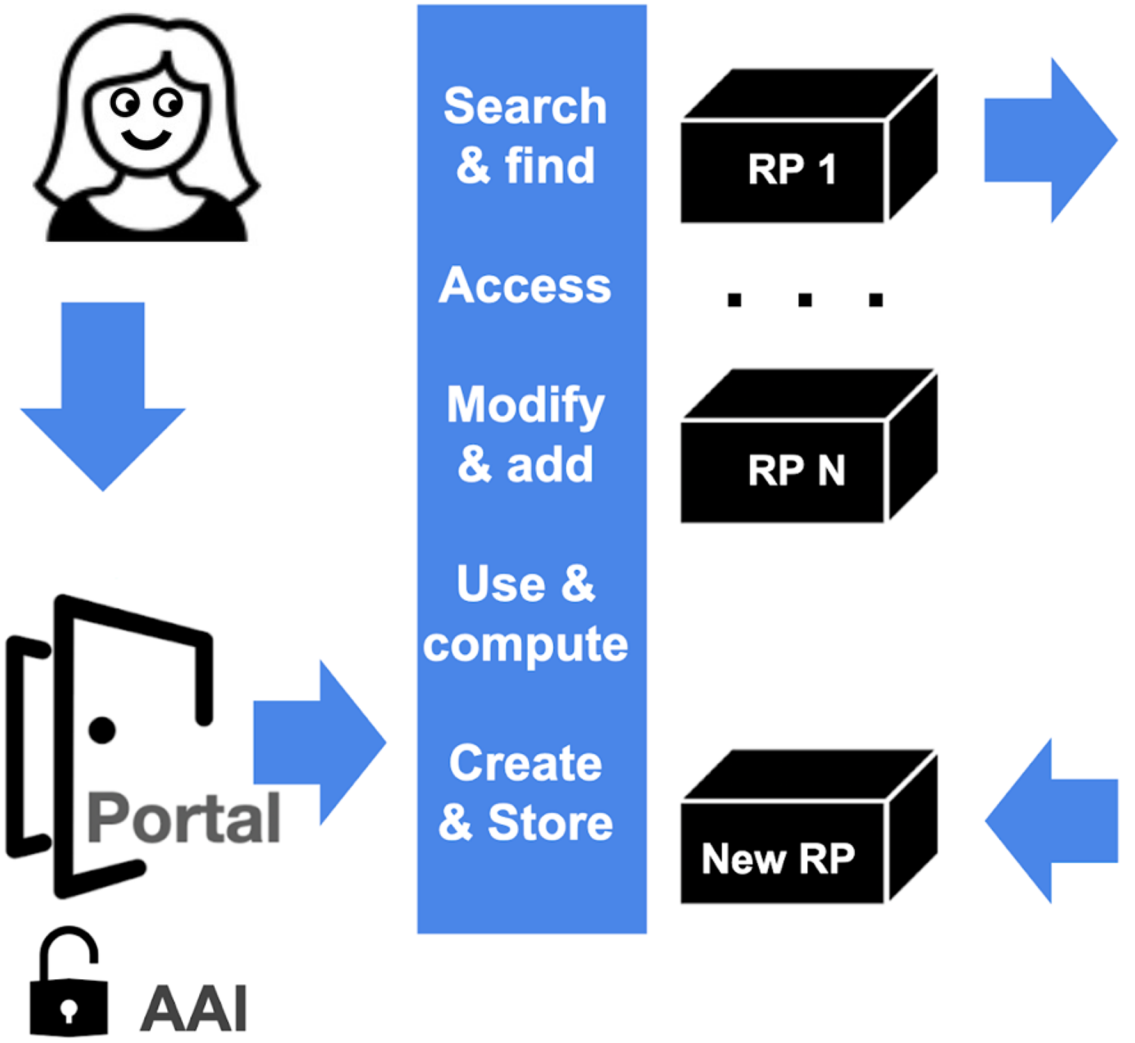
Data sources



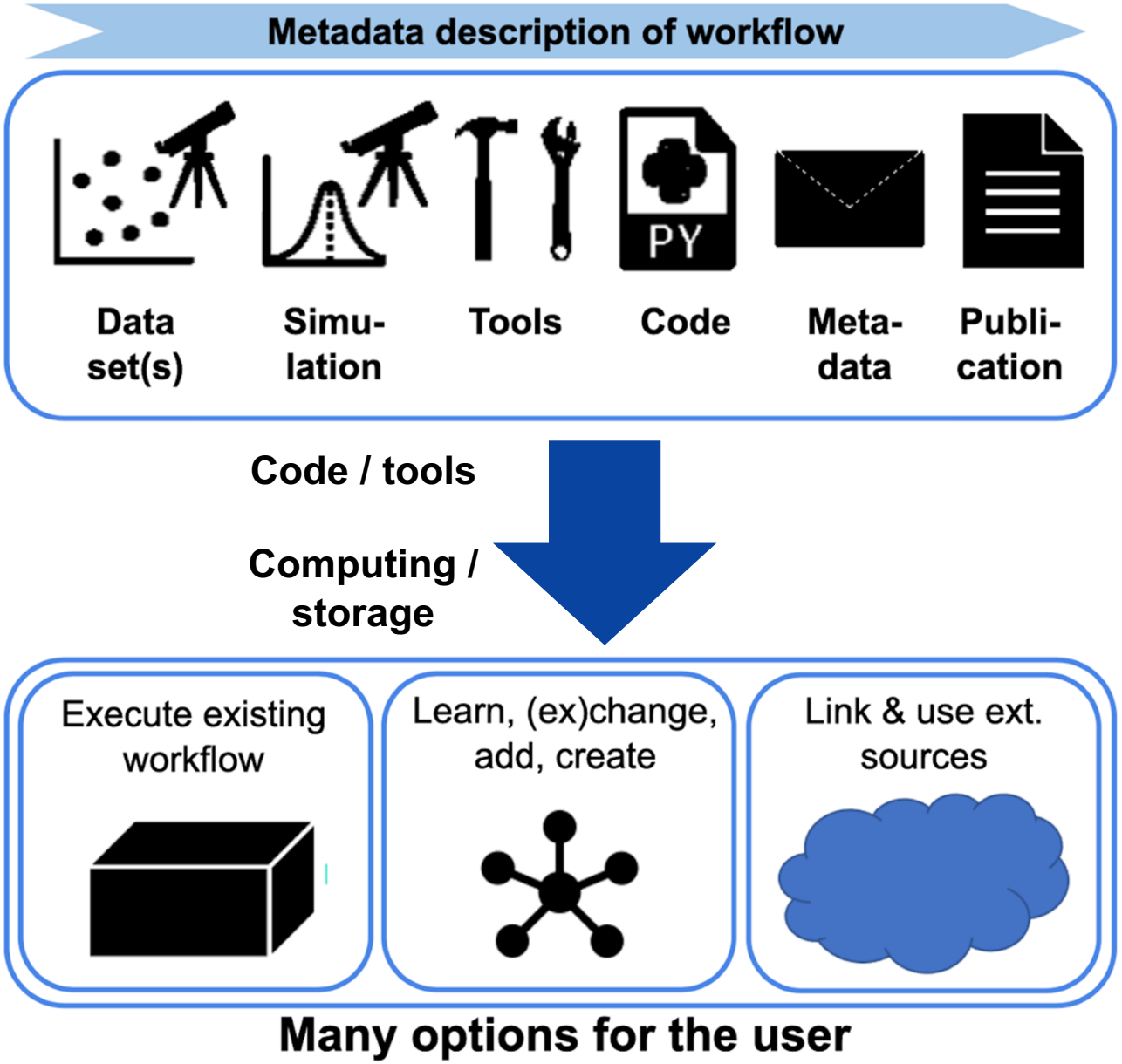
Flexibility, efficiency, scalability:
data volumes, number of users and analyses, heterogeneous resources; data combinations

PUNCH-SDP

The science data platform for RPs



Research product contains executable workflow



Cross-Cutting Topics, Sections, Base4NFDI

History

Cross-cutting topics, and Berlin and Leipzig-Berlin declarations

September 2019: Berlin declaration

In the spirit of our common interest to build what could become an international beacon for sharing the benefits of high-quality research data and research data management, the signing NFDI consortia strongly support the idea that NFDI must build on a coordinated and inclusive effort fostering the FAIR data principles. Recalling the challenges that interoperability and sharing of data faces, the 21 consortia signing this declaration agreed on a common list of cross-cutting topics.

July 2020: Leipzig-Berlin declaration

Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung

Maik Bierwirth; Frank Oliver Glöckner; Christian Grimm; Sonja Schimmler; Franziska Boehm; Christian Busse; Andreas Degkwitz; Oliver Koepler; Heike Neuroth

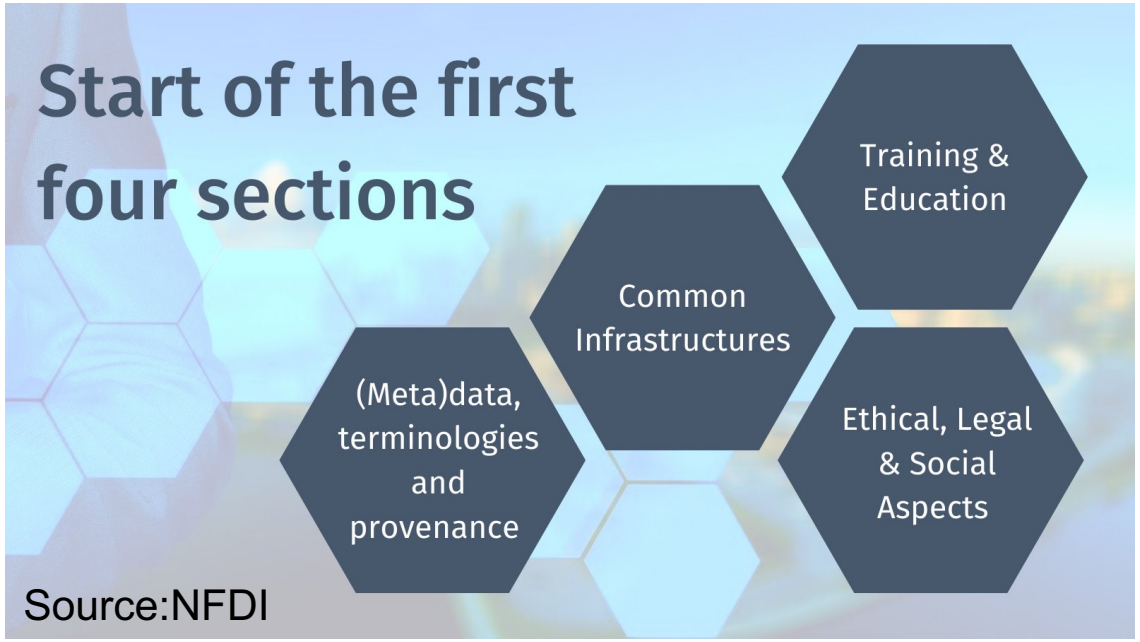
Für den wissenschaftsgeleiteten Aufbau der Nationalen Forschungsdaten-Infrastruktur (NFDI) muss sich die Infrastruktur gemeinsam mit der Forschung weiterentwickeln. Die dafür notwendigen, wechselseitigen Abstimmungen müssen auf Basis tragfähiger Prozesse und Strukturen sichergestellt werden. Themen, die für mehrere Fachkonsortien relevant sind, müssen im Sinne einer nachhaltigen Funktionalität kooperativ und über einzelne Konsortien hinweg bearbeitet werden. Dieses Dokument identifiziert solche Querschnittsthemen und Wege zu ihrer Bearbeitung in der NFDI.

Sections…

Collaboration towards the goals of the NFDI

Although the individual NFDI consortia are dedicated to research data management in a wide variety of disciplines, such as the natural, cultural and social sciences, they have many topics in common. Sections are legally dependent departments of the NFDI Association in which these cross-sectional topics are worked on across the boundaries of the consortia.

Cross-cutting issues are identified and prioritised with representatives of the consortia. The strategy-led process, initiated by the NFDI Directorate and the NFDI Consortium Assembly, can lead to the establishment of sections. In the sections, the association members work together to develop cross-consortium standards, metadata standards and formats.



DFG Call

For base service consortia

Information for Researchers No. 8

National Research Da
NFDI-wide Basic Servi

Supplement to the third round

In November 2018, the Federal Gov
National Research Data Infrastructu
systematically indexed, secured in t
and national borders in accordance
the academic community, the NFDI
consortia acting on their own initiat
proposals, the Deutsche Forschung
invites proposals for funding to sup

- Subject-specific and methodological consortia themselves cannot provide the requested basic service(s) on a long-term basis and, in particular, they are unable to do so for potentially all consortia.
- Basic service initiatives must be closely aligned with the shared needs of the subject-specific consortia, and this needs orientation must be appropriately demonstrated in the proposal.
- In principle, all consortia have an interest in using the basic service; this must be shown with the proposal. The actual use of a basic service in subject-specific and methodological consortia can vary over time, as can the contributions of individual consortia to a basic service. Unless services are intended for all NFDI consortia as a matter of principle, however, they cannot be funded as basic services.
- Basic service initiatives must develop their technical solutions by means of a process that is supported by the subject-specific and methodological consortia and they must ensure acceptance by the consortia. This process of negotiation is fundamental to a basic service initiative: it must be carried out in advance of proposal submission based on a basic service needs assessment and must be documented in the proposal. This process has to be continued on an ongoing basis over the course of the project in order to meet changing needs; plans to this effect are to be presented in the proposal.

21 March 2022: “Letter of Intent” Deadline

For proposals for base service consortia

Only one LoI was handed in to the DFG – carried by all 19 funded consortia and almost all applicants for the third round: Base4NFDI

(One of the logo options)

Basic idea: Base4NFDI is NOT (in the first place) a provider of a specific base service but

- gives a framework for the identification and integration of base services,
- provides flex funds for the “scouting” and (maybe) prototypical implementations of certain base services,
- collaborates closely with the section in identifying candidate services, and
- is fully integrated into the decision structures (senate, consortium assembly) of the NFDI Association that contribute to the decisions.

Currently: Writing of proposal with deadline 29 April. Decision together with 3rd-round consortia in November. Expectation (?): Funding at the order of 1 or 2 scientific consortia.

Base4NFDI Setup

TA1 Requirements, Design and Development

Sehr viele BD-Kandidaten aus Sektionen zu erwarten, Gewährleistung der Auswahlkriterien u Sicherstellung guter Bedingungen für Initialisierung, enge Rückbindung an Sektionen, Ressourcen für Unterstützung der Sektionen

TA2 Integration and Implementing Operational Services

Ausgewählte BD-Kandidaten werden mit bestehenden Infrastrukturen in den Communities integriert (insb. Service Scouts) und im Erfolgsfall für den Dauerbetrieb vorbereitet, enge Rückbindung an Sektionen

TA3 Service Negotiation, Development Process Coordination

Definition und Durchführung Auswahlprozesse, externe Begutachtung, Verhandlung mit Stakeholdern

TA4 Project Governance

Basic Services Board <> CTO, Netzwerke auch außerhalb der NFDI, Evaluation Prozesse und Begleitforschung, Outreach und PR, Administration

12

"Service onion"

Personnel Setup of Base4NFID

TA1 Requirements, Design and Development (3 SP)

Sonja Schimmler (Fraunhofer), **Axel Klinger** (Leibniz), **Alexander Sczyrba** (Uni Bielefeld)

Ingenieurwissenschaften - Lebenswissenschaften

TA2 Integration and Implementing Operational Services (3 SP)

Brigitte Mathiak (Leibniz), **Sören Lorenz** (Helmholtz), **Raphael Ritz** (MPG)

Geistes-/Sozialwissenschaften - Naturwissenschaften

TA3 Service Negotiation, Development Process Coordination (3 SP)

Juliane Fluck (Leibniz), **Reinhard Altenhöner** (SPK), **Thomas Schörner** (Helmholtz)

Lebenswissenschaften - Geistes-/Sozialwissenschaften - Naturwissenschaften

TA4 Project Governance (3 SP)

Regine Stein (Uni Göttingen), **Lars Bernard** (TU Dresden), **Bernhard Miller** (Leibniz)

Geistes-/Sozialwissenschaften - Naturwissenschaften

Conclusions

Conclusions

Very similar goals in the consortia; partly overlapping methods, and increasingly use of the same infrastructures

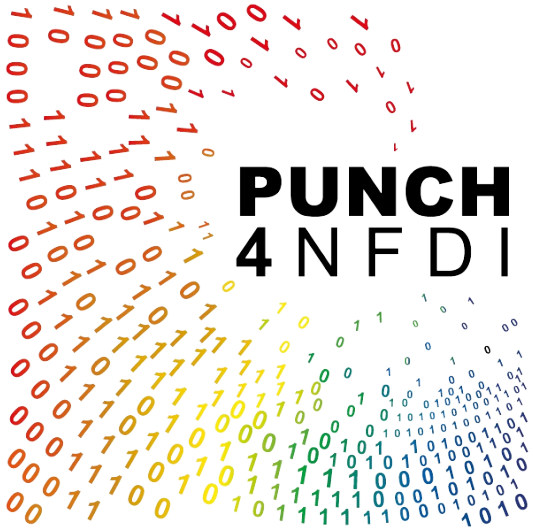
Work programmes of scientific consortia is a starting point, addressing numerous discipline-specific aspects

The true (potential) power of the NFDI emerges from
- the true embedding of the consortia into their communities
- the effective use of existing (infra)structures
- the cross-cutting collaboration between all NFDI consortia

Today try to explore connections
- between physics-related and other consortia
- between NFDI and ErUM
- between (parts of) NFDI and (parts of) Helmholtz
- …?

Thank you!

The PUNCH4NFDI Consortium

Spokesperson:
PD Dr. Thomas Schörner (DESY, thomas.schoerner@desy.de)
DESY, Notkestr. 85, D-22607 Hamburg

Contact:
Mail: punch4nfdi@desy.de
Web: www.punch4nfdi.de
Twitter: #punch4nfdi