

**NFDI@MPG**

# **Report of Contributions**

Contribution ID: 2

Type: **not specified**

## Base4NFDI - Basic Services for the NFDI

*Friday, 21 April 2023 09:00 (30 minutes)*

### *Creating NFDI-wide basic services in a world of specific domains*

NFDI is a German initiative to set up research data infrastructures within all disciplines, covering Humanities and Social Sciences, Life Sciences, Natural Sciences and Engineering Sciences. To ensure sustainability, it will integrate national with international activities.

In addition to domain-specific NFDI consortia, Base4NFDI (<https://base4nfdi.de/>) has been formed. Base4NFDI is a unique joint effort of all NFDI consortia to develop and deploy NFDI-wide basic services. These services will be integrated into the emerging infrastructures at the European level, especially the EOSC. The target group for basic services is the wider NFDI-community and, in particular, operators of community-specific services. The resulting NFDI-wide basic service portfolio will be beneficial for all disciplines.

Decisions on basic services will be made by all consortia in the bodies of the NFDI Association. To generate proposals for basic services, Base4NFDI will draw on the expertise in the NFDI Sections. There exchange between consortia on cross-cutting topics happens. The sections provide infrastructural and technological expertise in combination with domain knowledge and act as incubators for identifying potential basic services. Currently, the following sections exist: 'Metadata, Terminologies and Provenance', 'Common Infrastructures', 'Education and Training', and 'Ethical, Legal and Social Aspects'(a further section dedicated to industry engagement is currently being discussed).

Development will commence with a service for Identity and Access Management (IAM). Establishing approved identities and organizationally defined access rights across service providers will be crucial for seamless data management workflows.

For development, Base4NFDI will rely on a three-stage process of 1) initialisation of potential basic services 2) integration of basic services candidates and 3) ramping-up for operation and becoming part of the NFDI basis service portfolio.

The work programme of Base4NFDI is clustered in four Task Areas. Task Areas 'Service requirements, design and development' and 'Service integration and ramping-up for operation' will accompany and support the basic services within the three phases of the process. Task Area 'Service coherence processes and monitoring' will overlook the whole process, and Task Area 'Project governance' will manage the project.

MPCDF has a co-leading role in Task Area 2: service integration and ramping-up for operation.

**Primary author:** RITZ, Raphael (MPCDF)

**Presenter:** RITZ, Raphael (MPCDF)

**Session Classification:** Session IV

Contribution ID: 3

Type: **not specified**

## NFDI4plants (DataPlant) - An NFDI consortium of plant research

*Friday, 21 April 2023 16:30 (30 minutes)*

DataPLANT's main goal is to provide its community with the tools and infrastructure to store, process, and share data in a FAIR (Findable, Accessible, Interoperable, Reusable) manner. The core element of DataPLANT's Research Data Management (RDM) system is the Annotated Research Context (ARC), which has a single-entry point logic starting with the input of data and metadata, allowing the integration of computational workflows, has storage and versioning functionality and allows an automated upload to repositories.

**Primary author:** WIJESINGHA AHCHIGE, Micha (Max Planck Institute of Molecular Plant Physiology)

**Presenter:** WIJESINGHA AHCHIGE, Micha (Max Planck Institute of Molecular Plant Physiology)

**Session Classification:** Session VI

Contribution ID: 4

Type: **not specified**

## NFDI4Cat - The NFDI consortium for digital catalysis

*Friday, 21 April 2023 10:00 (30 minutes)*

Data management in catalysis is currently organised mainly at institutional or working group level and based on local conventions. However, catalysis is complex and interdisciplinary research, so it would be important to create an overarching interface in the area of data management so that FAIR data can be easily exchanged between disciplines. Through this interface and the integration of previous data silos, it should also be possible to accelerate developments and gain new insights with the help of partially AI-based analysis functions. The key challenge for the NFDI4Cat consortium is to bring together the different disciplines of catalysis research in terms of data management.

**Primary author:** ADAMITZKI, Philipp (Max Planck Institute for Chemical Energy Conversion)

**Presenters:** LEITNER, Walter (Max Planck Institute for Chemical Energy Conversion); ADAMITZKI, Philipp (Max Planck Institute for Chemical Energy Conversion)

**Session Classification:** Session IV

Contribution ID: 5

Type: **not specified**

## Local solutions for data acquisition and storage in heterogeneous catalysis

*Thursday, 20 April 2023 11:00 (30 minutes)*

A great potential in catalysis research is seen in an increasing integration of theory and experiment and the broader application of data science methods to experimental and calculated data.<sup>1</sup> The data exchange necessary for this requires progressive digitalization. Experimental data must be generated reproducibly and with sufficient diversity, and must be available in machine-readable form. At the Department of Inorganic Chemistry of the Fritz-Haber-Institut (FHI) der Max-Planck-Gesellschaft (MPG), we have developed and implemented a concept for a local data infrastructure during the past years. This work is integrated in our activities in FAIRmat, Use Case Demonstrators E2, "Heterogeneous Catalysis" in coordination with NFDI4Cat and in the BMBF project CatLab. The software solutions were developed in collaboration with the computer support group of the FHI and were initially based on a database that has been used intensively by the department for more than 20 years. The database was upgraded to a modern, flexible electronic laboratory notebook that meets the requirements of research in heterogeneous catalysis and enables data exchange via an Application Programming Interface (API).<sup>2</sup> For research projects, handbooks (Standard Operating Procedures (SOP's)) are developed, preferably in machine-readable form, detailing how experimental data are obtained, including the definition of benchmark catalysts. To facilitate the implementation of the handbook concept, automated systems for data acquisition and storage have been designed in the framework of a research project focused on the investigation of innovative catalysts for the efficient conversion of chemical energy into electrical energy and vice versa (CatLab).<sup>3</sup> Such systems consist of (i) EPICS for communication with devices and data acquisition, (ii) the database (archive), (iii) an archiving appliance for storing time series, (iv) Phoebe for creating graphical user interfaces, (v) Python/Bluesky/Jupyter notebooks for creating automation and data evaluation, and (vi) a S3 storage for long-term storage. The concept is explained using the examples of an automated reactor for catalyst testing and the automated storage of electron microscopy data.

(1) Marshall, C. P.; Trunschke, A., Achieving Digital Catalysis: Strategies for Data Acquisition, Storage, and Use. *Angewandte Chemie International Edition* 2023, submitted.

(2) Archive FHI MPG, <https://github.com/fhimpg/archive>.

(3) Automation solutions FHI MPG, <https://gitlab.fhi.mpg.de/fhi-ac/ertl>; <https://gitlab.fhi.mpg.de/fhi-ac/haber>; <https://gitlab.fhi.mpg.de/fhi-ac/velox>; <https://gitlab.fhi.mpg.de/fhi-ac/json-scripte>; <https://gitlab.fhi.mpg.de/fhi-ac/gaswarnanlage>; <https://gitlab.fhi.mpg.de/fhi-ac/berty>.

The collaboration with the Humboldt University Berlin, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), in the framework of the project FAIRmat –FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids, project number 460197019 and financial support by the Federal Ministry of Education and Research (BMBF) in the framework of the CatLab project, FKZ 03EW0015B is acknowledged.

**Primary author:** TRUNSCHKE, Annette

**Co-authors:** Mr MOSHANTAF, Abdulrhman; JUNKES, Heinz; Mr WESEMANN, Michael; Mr OPPERMANN, Patrick

**Presenter:** TRUNSCHKE, Annette

**Session Classification:** Session II

Contribution ID: 6

Type: **not specified**

## NFDI4Earth - the NFDI for Earth System Sciences

*Thursday, 20 April 2023 09:30 (30 minutes)*

NFDI4Earth is trying to bring together a Earth System Sciences through a diverse and large consortium of more than 60 institutions. The project is setting up a OneStop4All as a first contact point to find solutions for Earth System Science related data management tasks, connecting to important data portals, infrastructure providers and a User support network. The consortium has opened calls for pilot and incubator projects to activate the community and implemented an academy. The role of the MPI-BGC is in observing and developing “Advancing Tools” for large-scale geo-data processing.

**Primary author:** GANS, Fabian**Presenter:** GANS, Fabian**Session Classification:** Session I

Contribution ID: 7

Type: **not specified**

## The NFDI4BIOIMAGE consortium

*Thursday, 20 April 2023 14:30 (30 minutes)*

The NFDI4BIOIMAGE consortium was approved for funding in November 2022 and starts in March 2023. Led by Heinrich-Heine University Düsseldorf it brings together researchers, data stewards, and research software engineers from various German Universities and Research Institutes. The Max Planck Institute for Evolutionary Biology has Participant status. NFDI4BIOIMAGE is focused on the development, dissemination, and implementation of standards for data management, annotation, and processing of microscopy and bioimage analysis data. In this contribution I will outline the concept and work plan of NFDI4BIOIMAGE and delineate opportunities for collaborations and networking with the Max Planck Society in general and with the Max Planck BioImaging Core Unit Network in particular.

**Primary author:** FORTMANN-GROTE, Carsten (MPI for Evolutionary Biology)

**Presenter:** FORTMANN-GROTE, Carsten (MPI for Evolutionary Biology)

**Session Classification:** Session III



Contribution ID: 8

Type: **not specified**

## NFDI-MatWerk –National Research Data Infrastructure for Materials Science & Engineering

*Friday, 21 April 2023 16:00 (30 minutes)*

The performance of any engineering material depends critically on its strongly heterogeneous and process-dependent microstructure, ranging from crystal defects at the atomic level, through microscale secondary phases up to macroscale pores. Furthermore, processes on timescales ranging from picoseconds up to centuries need to be addressed. This inherent multiscale character of materials needs to be represented in corresponding data models and combined with measurements and simulations that capture these processes.

Due to the vast number of different experimental, computational and analytical methods used to reveal these dependencies, the MatWerk community has developed a large variety of data tools and workflows. NFDI-MatWerk aims to provide a federated digital materials environment that gives scientists full control over their data while enabling and incentivizing data sharing and offering highly performant, complex search queries and analysis runs based on a materials knowledge graph connected to a database infrastructure. Integrated development environments ensure that data processing in workflows follow the same FAIR standards as the data. These concepts will be outlined within the presentation to foster synergies between the consortia.

**Primary authors:** BITZEK, Erik (Max-Planck-Institut für Eisenforschung); HICKEL, Tilmann (Max-Planck-Institut für Eisenforschung)

**Presenter:** HICKEL, Tilmann (Max-Planck-Institut für Eisenforschung)

**Session Classification:** Session VI

Contribution ID: 9

Type: **not specified**

## MaRDI - Mathematical Research Data Initiative

*Thursday, 20 April 2023 09:00 (30 minutes)*

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 ease the use of data across 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:

- Develop a robust Mathematical Research Data Infrastructure that will be useful within mathematics and other disciplines, as well as non-scientific fields.
- Set standards and confirmable workflows for certified Mathematical Research Data and
- provide services to both the mathematical and wider scientific community.

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

**Primary authors:** BACHER, Tabea (Max Planck Institute for Mathematics in the Sciences); BENER, Peter; HEILAND, Jan

**Presenters:** BACHER, Tabea (Max Planck Institute for Mathematics in the Sciences); SAAK, Jens (Max Planck Institute for Dynamics of Complex Technical Systems)

**Session Classification:** Session I

Contribution ID: 10

Type: **not specified**

## Knowledge organisation and integration of heterogeneous data in Art History

*Thursday, 20 April 2023 10:00 (30 minutes)*

The Bibliotheca Hertziana (BHMPI) has been involved with the process of conceiving a Consortium for the cultural-historic sciences in the Humanities, which led to the foundation of NFDI4Culture in June 2020. The NFDI4Culture mission is to systematically develop, make accessible, and sustainably secure research data from art, music, architecture, theatre, dance, film, and media studies into a demand-oriented infrastructure for research data on tangible and intangible cultural assets. BHMPI is a long-term partner and aims at official membership in the Society.

Though the most prominent BHMPI digital research assets in art history are its library and its photographic archive, the work of several related research departments throughout the past decades has resulted in a myriad of databases characterised by unique valuable content, implicit data inter-linking through authority control (e.g. GND), but otherwise compartmentalised implementations and management. These include the digitised rare book collection, the Zuccaro information system on Italian art history, digitised and geo-referenced historical maps of Rome and Naples, and several catalogues based upon Linked Open Data, such as Mapping Sacred Spaces (Medieval church interiors) and Magnetic Margins (history of science census).

Through the foundation of a Digital Humanities Lab, and under the auspices of NFDI4Culture, BHMPI is developing a data infrastructure, compounded with curatorial and re-engineering workflows, able to manage and publish data from these art-historical projects and from future ones. The goal is for these data to be consumable by third parties - through standardised interfaces like RDF, IIIF and SPARQL - as if being part of one connected knowledge graph, allowing related iconographic content from the photo archive and bibliographical/citational contexts from the library to be delivered and discovered. The challenges involved include: (1) dealing with the highly variable scales of the datasets; (2) prioritising the efficiency of textual search, geographical querying, and bibliographical lookup; (3) ensuring data integration whilst minimising the overhead of cross-source transforms; (4) not disrupting the established curatorial practices of existing projects. An experimental implementation, interoperating with the photo archive reengineering work underway and based on these standards and widespread data schemas for GLAM and Humanities research, such as CIDOC-CRM and FRBR, is selectively being made available.

**Primary author:** ADAMOU, Alessandro (Bibliotheca Hertziana - Max Planck Institute for Art History)

**Presenter:** ADAMOU, Alessandro (Bibliotheca Hertziana - Max Planck Institute for Art History)

**Session Classification:** Session I

Contribution ID: 11

Type: **not specified**

## **Human brain observatories: the power of the collective to unravel the mysteries of the human mind**

In the past decade, much of the progress in human neuroscience has come under scrutiny due to issues related to reproducibility of findings, low powered studies and also the large degree of flexibility in the analysis pipelines. Several mechanisms have been put in place to address those challenges, e.g., preregistrations and registered reports. We have focused on another neglected aspect, the fact that theories and schools of thought develop in parallel leading to silos instead of cross-talk. To counteract those tendencies we are exploring an open science adversarial collaboration powered by multi team science. Implementing this initiative led us to face many obstacles concerning data sharing and integration of studies across laboratories. We discovered the importance of metadata, and the meandering road to define them. While team science may produce many advances, a big portion of the infrastructure still needs to be developed. The next frontier for human cognitive neuroscience lies in developing that infrastructure to materialise human brain observatories.

**Primary author:** MELLONI, Lucia (MPI for Empirical Aesthetics)

**Presenter:** MELLONI, Lucia (MPI for Empirical Aesthetics)

Contribution ID: 12

Type: **not specified**

## NFDIxCs - data management for and with Computer Science

*Friday, 21 April 2023 09:30 (30 minutes)*

NFDIxCs is a new consortium starting in 2023 which focuses on the needs of the Computer Science (CS) community. It aims to provide services and workflow to support the FAIR principles for the complex domain-specific data objects from the vast field of Computer Science. This includes producing reusable data objects which contain not only various types of CS data including the associated metadata, but also the corresponding software, context and execution information in a standardized form. A key concept is the use of so-called Research Data Management Containers which can be of any size, structure and quality. NFDIxCs will, in collaboration with other scientific disciplines, support the application of CS methods such as Big Data, Artificial Intelligence and Machine Learning. In addition, in the areas of high-performance computing and computer architecture, which in turn also contribute to the further development of genuine CS methods. The talk will provide an overview on the scope and structure of NFDIxCs consortium.

**Primary author:** YAHYAPOUR, Ramin**Presenter:** YAHYAPOUR, Ramin**Session Classification:** Session IV

Contribution ID: 13

Type: **not specified**

## Particles, Universe, NuClei and Hadrons for NFDI

*Thursday, 20 April 2023 11:30 (30 minutes)*

PUNCH4NFDI is the NFDI consortium of particle, astro-, astroparticle, hadron and nuclear physics, representing about 9.000 scientists with a Ph.D. in Germany, from universities, the Max Planck Society, the Leibniz Association, and the Helmholtz Association. PUNCH physics addresses the fundamental constituents of matter and their interactions, as well as their role for the development of the largest structures in the universe - stars and galaxies. This talk will provide a brief overview over the ongoing activities.

**Primary author:** KRAMER, Michael (Max-Planck-Institut für Radioastronomie, Bonn, Germany)

**Presenter:** KRAMER, Michael (Max-Planck-Institut für Radioastronomie, Bonn, Germany)

**Session Classification:** Session II

Contribution ID: 14

Type: **not specified**

## NFDI4Memory (online)

*Thursday, 20 April 2023 15:00 (30 minutes)*

The 4Memory consortium<sup>1</sup> focuses on the field of history and those disciplines that make use of historical data as part of their methodology.

Historical data includes “texts ranging from antiquity to the modern era, images, photos, audio and video recordings, statistics, structured data, metadata, ontologies, and hypertexts” as well as “personal data, spatial structures, and changes in classification systems and categories over time.” The main goals of 4Memory are to integrate this broad spectrum of historical data and to sharpen and apply historical source criticism to those data, as a basis for the production and communication of historical knowledge in the digital space.

To achieve these goals, 4Memory works to establish standards and norms for historical research data, to facilitate access to and preservation of these data in an interoperable “Data Space”, and to ensure their quality and reusability by establishing data literacy in historically oriented humanities.

Producers and users of historical data from historical research, memory institutions, and information infrastructures are participating in 4Memory. The consortium is scheduled to begin its work in March 2023.

<sup>1</sup> <https://4memory.de/>

**Primary author:** HENNICKE, Steffen (MPI for the History of Science)

**Presenter:** HENNICKE, Steffen (MPI for the History of Science)

**Session Classification:** Session III

Contribution ID: 15

Type: **not specified**

## FAIR data infrastructure for core-level spectroscopy

*Thursday, 20 April 2023 14:00 (30 minutes)*

To exploit data generated in the fields of condensed-matter physics and chemical physics of solids as well as catalysis research, a FAIR data infrastructure is necessary. FAIRmat's goal is to provide this infrastructure. FAIRmat integrates synthesis, experiment, theory, digital infrastructure and applications to pursue this goal.

From an experimental point of view, the tasks include the generation of application definitions for the corresponding experimental techniques as well as generating parsers for different file formats. In addition, standardized workflows should be designed and the connection to electronic lab notebooks and data storage needs to be considered.

The current status and future plans within FAIRmat's task area B3 –core level spectroscopy –with X-ray photoelectron spectroscopy (XPS) as its core experimental technique will be presented. Additionally, the connection to the other FAIRmat areas, and to the RDM related efforts and projects at the MPI-CEC will be discussed.

**Primary author:** HETABA, Walid (MPI für Chemische Energiekonversion)

**Co-author:** PIELSTICKER, Lukas (MPI für Chemische Energiekonversion)

**Presenter:** HETABA, Walid (MPI für Chemische Energiekonversion)

**Session Classification:** Session III



Contribution ID: 17

Type: **not specified**

## NFDI4BioDiv

*Friday, 21 April 2023 11:00 (30 minutes)*

NFDI4Biodiversity is one of the consortia under the umbrella of the National Research Data Infrastructure (NFDI) and dedicated to mobilize biodiversity and environmental data for collective use. The consortium includes close to 50 scientific institutions, museums, natural history societies, state offices, and other institutes and expert groups. They pool their scientific and technical expertise to provide a broad service portfolio for handling biodiversity and environmental data and to develop it further. The cooperation is guided by the knowledge that stakeholders in science, politics, nature conservation and landscape management need reliable data to be able to develop better contributions to the conservation of biodiversity.

To this end, the consortium partners offer added value to the professional community, specifically:

- Access to modern technologies and a comprehensive stock of biodiversity and environmental data
- Methods and tools for archiving, publishing, searching and analyzing data that are suitable for everyday use and have been tried and tested in practice &
- An expert forum for the safe and competent handling of data for broad and responsible use

**Primary author:** WIEDER, Philipp (GWDG)

**Co-author:** YAHYAPOUR, Ramin (GWDG)

**Presenter:** YAHYAPOUR, Ramin (GWDG)

**Session Classification:** Session V

Contribution ID: **18**Type: **not specified**

## Text+

*Thursday, 20 April 2023 12:00 (30 minutes)*

Language and text-based research data are of great importance in many disciplines in the humanities. Initially with the data domains Collections, Lexical Resources and Editions, Text+ addresses the requirements of a wide range of research fields. Therefore, it will systematically expand its data portfolio. This also includes tools to support researchers in the FAIR creation, use and provision of data throughout the entire research data life cycle.

The usability of the data and tools are central tasks of Text +. As a distributed infrastructure, Text+ relies on and incorporates a wide range of existing data and tools from the currently more than 30 partners in the consortium.

### Goals

- Development of a distributed infrastructure for speech language and text data
- Support of researchers in the creation, re-use and preservation of language and text data
- Enable innovative research through easy access to research data and tools
- Close cooperation with the communities involved: portfolio expansion, training, workshops

**Primary author:** WIEDER, Philipp (GWDG)

**Presenter:** WIEDER, Philipp (GWDG)

**Session Classification:** Session II