Binding Letter of Intent of the Consortium NFDI4cat

This binding letter of intent serves as an advance notification for the proposal of the NFDI consortium NFDI4cat for 2019.

- **Planned name of the consortium**
- NFDI for Catalysis-Related Sciences

- **Acronym of the planned consortium**
- NFDI4cat

- **Spokesperson**
  - Prof. Dr. Matthias Beller
  - Leibniz-Institut für Katalyse e.V. (LIKAT),

- **Applicant institution**
- DECHEMA e.V., Theodor-Heuss-Allee 25, 60486 Frankfurt
  Executive Director Prof. Dr. Kurt Wagemann
  Impersonating Spokesperson Prof. Dr. Kurt Wagemann kurt.wagemann@dechema.de

Objectives, work programme and research environment

Research area of the proposed consortium (according to the DFG classification system)

3 Natural Sciences; 31 Chemistry; 301 Molecular Chemistry; 302 Chemical Solid State and Surface Research; 303 Physical and Theoretical Chemistry; 304 Analytical Chemistry, Method Development (Chemistry); 305 Biological Chemistry and Food Chemistry; 33 Mathematics; 312 Mathematics

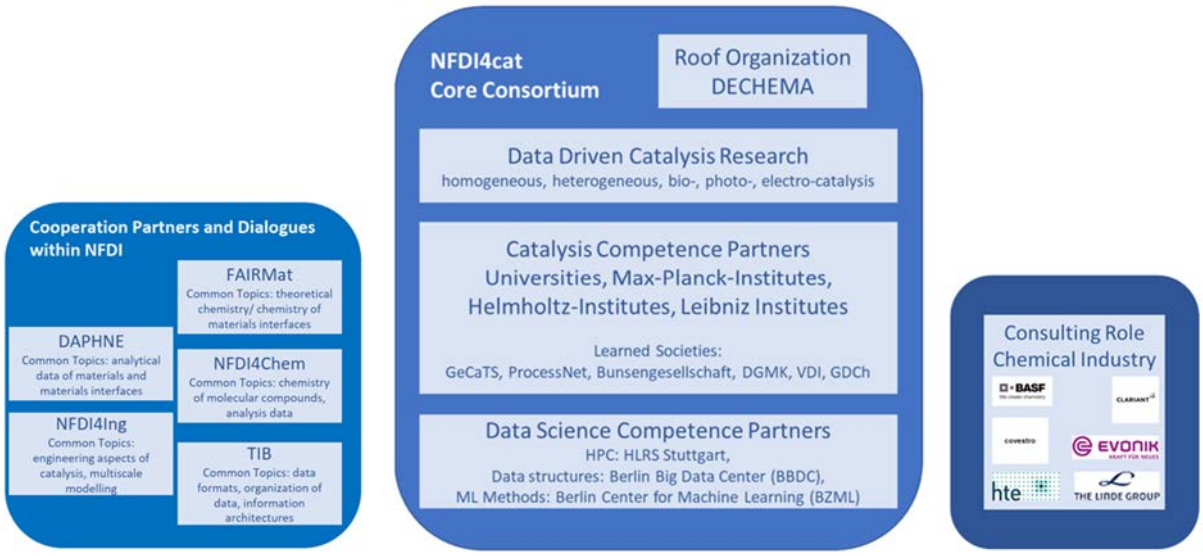
4 Engineering Sciences; 42 Thermal Engineering/ Process Engineering; Computer Science, Systems and Electrical Engineering; 403 Process Engineering, Technical Chemistry; 404 Heat Energy Technology, Thermal Machines, Fluid Mechanics; 406 Materials Science; 407 Systems Engineering; 409 Computer Science

Catalysis is an interdisciplinary scientific technology field of strategic importance for the economy and the society as a whole. It is one of the most important core technologies for solving pressing challenges concerning climate change, supply of sustainable energy and supply of sustainable materials at the same time. Concrete examples of such pressing issues are the reduction or complete avoidance of CO_2 emissions, the valorization of plastic waste and CO_2 in chemical production, sustainable hydrogen generation, fuel cell technology, or feeding more than 7 billion people sustainably on this planet – they all require break-through advances in catalysis science and technology.

To solve these pressing issues, a fundamental change is required in catalysis research, chemical engineering and process technology. A key challenge is to bring together the different disciplines in catalysis science and technology. The consortium aims to redefine catalysis research in the digital age and add new facets of digital empowerment. Core challenge is a fundamentally improved understanding in catalysis sciences, the creation of workflows in catalysis that build a bridge between theory/simulation and experimental studies in design, characterization and kinetics of catalysts and the related engineering aspects. This challenge requires also a unified view on all catalysis disciplines to reveal universal guiding principles common to homogenous, bio, heterogeneous and electro-catalysis. This unification can only be successful if the full potential of the emerging data science technology is made accessible to each catalysis researcher. Vital elements of this strategy are the unification of data formats and the understanding of the requirements for the creation of high-performance information architectures that allow storage, exchange and analysis of data utilizing the latest state of the art tools of artificial intelligence. However, data science requires high quality data based on unified formats in the first place. The enabling of all members of the community to produce and work with high quality data using unified formats with common ontology and semantics will be a key aspect of the consortiums endeavor. Other critical issues that need to be solved concern e.g. reward models, legal aspects,

licenses and IP, and the organization of the community within the infrastructure. The consortium embarks on the endeavor of realizing “digital catalysis” along a value chain of data oriented from “molecules to chemical processes”. This data-based value chain has a non-virtual analogue and is therefore seen as ambitious but of high intrinsic value allowing for a high degree of sustainability.

The consortium NFDI4cat consists of experts from the field of homogeneous, heterogeneous, photo-, bio- and electro catalysis. The disciplines of reaction and process engineering are also represented in the consortium. The catalysis and engineering competences are complemented by expertise in the field of data science, high-performance computing and machine learning. NFDI4cat will collaborate with the relevant partners in the fields of chemistry, materials science and engineering. Consulting with respect to future use of the results will be provided by a panel of representatives of the chemical industry.



Concise summary of the planned consortium’s main objectives and task areas

The overall strategy for the transformation of catalysis research accross all relevant disciplines in catalysis and chemical engineering follows a stepwise approach as outlined in the Whitepaper: “The Digitalization of Catalysis-Related Sciences” published by GeCatS – German Catalysis Society in March 2019. It follows the path of transforming catalysis related sciences into a branch of digital sciences. In order to turn this strategy into a tangible plan a structure of projects and sub-projects will be pursued in working groups covering technical, organizational and educational aspects. The approach will be fostered and validated by a further element required: a branch which takes care of making experimental data libraries and data sets from theoretic assessments available, so that the data and metadata standards can be validated, as well as the performance of the infrastructure employed.

A roadmap towards institutionalization of these elements is also required assuring continuous existence. The whole process should be user-driven and follow the principles of stakeholder participation, from both academia and industry, in all its elements. In conjunction with this initiative it is also vital to work on models for an implementation plan in the national educational system, so that future generations of scientists working in the field of catalysis do have a smooth start regarding the skillsets acquired during their education. From a current point of view, it is also important to emphasize that the time scale, until a full implementation and the final goal of a fully digitalized national scene in catalysis can be reached, is expected to be on the order of a decade, possibly longer. It therefore is anticipated that ultimately the information architecture will become an indispensable tool of a digitalized community in catalysis on a national and international basis. The core topics that will be addressed by the consortium are addressed by the following workstreams, which all follow the value chain "from molecules to chemical processes":

Workstream (A) Data and Meta-Data Standards

Workstream (A) addresses core capabilities that the consortium intends to develop to enable data-based catalysis science and foster the development of a common semantics and ontology for catalysis, catalysis related chemical engineering and catalysis related aspects of industrial chemistry in general. All efforts are concentrated around data-based value chains and data valorization that follow the common guideline "from raw data to knowledge". Apart from vital efforts in standardization Workstream (A) will also follow the lifeline of data generated in catalysis - form their initial generation and documentation in electronic lab-notebooks, to their conversion in formats that can be uploaded on community available information infrastructures, till their use and re-use for other purposes. Software and software-related tools and their development along the line of the data value chain will be also treated as topics in Workstream (A). Catalysis science will rest on five important pillars that provide information in the form of data, i.e. synthesis data, performance data, characterization data, operando data as well theoretical computational data. One of the most important aspects will be respective cross-disciplinary data formats. These data formats should facilitate also a sustainable collaboration between academia and chemical industry.

Workstream (B) Data Science and Information Infrastructure Design

Workstream (B) will focus on questions of Data Science and Information Infrastructure Design. Both elements are seen as crucial elements that act as enablers in order to arrive at highly performing and community accepted systems and procedures. Data Science will serve as an important tool along the data-value chain starting from automatic readout of electronic lab-notebooks until the final analysis of data stored in repositories; in all cases artificial intelligence will be an indispensable core enabling technology. Apart from artificial

intelligence, the choice of standard procedures and suitable software tools that allow fast and efficient transfer and transformation as well as data curation is seen as core topic within this workstream. As more than one physical location is foreseen to be used as repository, performance related aspects play a major role in order to ensure aspects of system integrity and user friendliness. The development of suitable overarching software tools that create a work environment unifying a heterogeneous repository infrastructure will ensure user friendliness and system acceptance. Another topic that will also be addressed within this workstream are questions of raw data transfer from various heterogeneous sources and the respective transformation of such data formats into standardized formats; ideally this workstream seeks discussion with providers of tools for data generation (typically analysis instrumentation).

Workstream (C) Community and User related Aspects

Workstream (C) focusses largely on non-technical and community related aspects of the digitalization approach under the general motto “enabling scientists” and is therefore orthogonal to the Workstreams (A) and (B). Within this workstream legal topics like intellectual property and confidentiality will be treated. Although not fundable in the NFDI-format the active participation and facilitation of the industrial partners is seen as vital in this Workstream. Other aspects that will be covered within this workstream are aspects of education and digital skill development as well as community building efforts in order to ensure the development of a collaborative culture.

Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium’s objectives

NFDI4cat will connect existing infrastructure and combine local data hosting and to develop a virtual national data repository. Hence, tools will be developed to generate, import, store, clean, annotate, manage and analyze data locally. Additional tools will enable the transfer (publication) of selected data into national repositories, and comprehensive analysis within this national pool of data. Services will include data hosting along with ways of analyzing data with statistical and machine learning tools. On a local level, the integration with existing data-generating equipment (synthesis and analysis tools, reactor setups, analytics from large scale facilities etc.) will need to be solved.

Within the NFDI4cat consortium the HLRS will take over the role of hosting. For the initial phase a single provider is chosen in this role, in order to simplify procedures and ensure governance of the initial repository.

The final target of the consortium regarding a repository structure relies on a heterogeneously structured delocalized approach. Large central repositories like the initial repository of HLRS will have to be connected with smaller local ones provided by other individual partnering organizations.

The role of system integration and architectural design will be also taken over by HLRS;in this role it will provide services and render access to existing infrastructure e.g. a tape based hierarchical storage management system. Within the consortium this task of structuring an infrastructure based on a heterogeneous repository landscape using different back-end storage devices and storage systems and transforming it into a homogeneously accessible high-performance repository, that can serve as a work-platform, is seen as core capability and mission for the NFDI4cat consortium.

Based on HLRS´s experiences in meta-data standard development for engineering sciences, HLRS will also be active repurposing or potentially adapting existing workflows and tools like the tool for automatic metadata capturing.

From the side of KIT, particularly the Deutschmann group, existing tools can be contributed from the field of multiscale modelling and simulation in combination with an archiving function for local repositories. The tools contributed here are based on the DETCHEM code - CARMEN is the furthest developed tool that will be contributed and can be adapted within the work of the consortium.

The Berlin Big Data Center (BBDC) and Berlin Center for Machine Learning (BZML), represented by K.R. Müller, will address challenges in key aspects related to data evaluation and machine learning. These challenges are most of all (1) heterogeneity in data sources and experimental methods, (2) data quality assessment and outlier detection, (3) explainable ML models and (4) integration of a priori domain knowledge in catalysis into ML models.

The industrial panel of NFDI4cat will not be eligible to receive funds within the NFDI. Nevertheless, industry has a clear commitment towards the goals of NFDI4cat and will consult in their role and help the consortium in the choice of open-source tools and organizational issues whenever necessary and helpful.

Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

Networking is facilitated through a task force within GeCatS that is represented by members of universities, Max Planck Institutes, Helmholtz Institutes, LIKAT as well as industry representatives. Institutions experienced in hosting and access providing are part or partners of the consortium, i.e. HLRS, MPG as well as KIT. Current and planned collaborations include NFDI initiatives that focus on more general aspects of material science, chemistry

and engineering. The chemical physics of surfaces of materials, chemical reactions at surfaces, and multi-scale modeling of heterogeneous catalysis will be explored in collaboration with FAIRmat. Together with NFDI4chem, the digital approaches towards the digital representation of the chemistry of molecular compounds as well as standards and formats of analytical data will be discussed. Engineering aspects of catalysis and multiscale modelling will be approached within NFDI4cat in close dialogue with NFDI4ing. Building strong connections in respect to the catalysis-related aspects of these consortia will further the success of NFDI4cat. TIB as institutional library for science and technology can act as mediator for several planned initiatives in the area of chemistry related sciences. Together with NFDI4chem a harmonization between the initiatives will be pursued, mutual learning through exchange will be fostered. In addition, the TIB within the consortium NFDI4chem, and in the same role FAIRmat are regarded as important discussion partners for ontology engineering, vocabulary development and conceptual development for association of meta-data in suitable formats. As characterization data are vital to scientific approaches in catalysis for the development of structure property relationships, X-ray and neutron data conducted operando are relevant for NFDI4cat, a close collaboration with DAPHNE is part of the planned mode of interaction.

The proposed initiative aims primarily at improving catalysis science on a national level within the framework of NFDI; nevertheless, through active engagement within the Research Data Alliance it will be ensured that developed standards within the consortium will be compatible with international standards. It will therefore create a lasting impact and assure the ability to compete in both science and applications also internationally. International collaborations within and beyond the catalysis community will be pursued. Collaboration and discussion between parties with experience in infrastructure management and “users” within the catalysis community will be a core element. Beyond the expertise assembled within the consortium in the field of high-performance computing, data science and machine learning, further collaborations are aimed for, e.g. Research Data Alliance and others. New forms of networking will be required to facilitate the collaboration between experimentalists, application and theory.

Cross-cutting topics

The following cross cutting topics have been identified as relevant for NFDI4cat:

- Development of ontologies and semantics in the field of chemistry, materials sciences and chemical engineering
- Development on suitable Metadata Standards in the field of chemistry, materials sciences and chemical engineering
- Information Infrastructure Design: Development of tools that allow the connection of a variety of physically separated repositories into one highly functional virtual repository

- Raw data transfer from various heterogeneous sources and the respective transformation of such data formats into standardized formats
- Tools that allow fast and efficient transfer and transformation as well as data curation
- Development of standard and legally sound practices concerning the treatment of intellectual property and topics around confidentiality
- Education and digital skill development and community building
- Fostering of platforms for sustainable collaboration between academia and chemical industry

Please indicate which of these cross-cutting topics your consortium could contribute to and how.

- Development of ontologies and semantics in the field of chemistry, materials sciences and chemical engineering.

  This topic will be addressed in collaboration with NFDI4chem, FAIRmat and NFDI4ing. NFDI4cat will be the competent discussion and development partner for the field of catalysis and chemical engineering. The industrial partners will aid in the process.

- Development on suitable Metadata Standards in the field of chemistry, materials sciences and chemical engineering

  This topic will be addressed in collaboration with NFDI4chem, FAIRmat and NFDI4ing. NFDI4cat will be the competent discussion and development partner for the field of catalysis and chemical engineering. The industrial partners will aid in the process.

- Information Infrastructure Design: Development of tools that allow the connection of a variety of physically separated repositories into one highly functional virtual repository - a topic of high relevance for all NFDI consortia. HLRS will address the topic within NFDI4cat; solution approaches and experiences will be shared with other NFDI´s.
- Raw data transfer from various heterogeneous sources and the respective transformation of such data formats into standardized formats- a topic of high relevance for all NFDI consortia. The NFDI4cat data-science partners HLRS and BBDC will address the topic within NFDI4cat; solution approaches and experiences will be shared with other NFDI´s.
- Tools that allow fast and efficient transfer and transformation as well as data curation - a topic of high relevance for all NFDI consortia. The NFDI4cat data-science partners HLRS and BBDC will address the topic within NFDI4cat; solution approaches and experiences will be shared with other NFDI´s. hte GmbH will act as an industrial consultant for the data science partners.
- Development of standards and legally sound practices concerning the treatment of intellectual property and topics around confidentiality

This topic is of high relevance for most of the other NFDI consortia. A joint work group within NFDI4cat will address the topic, the industrial panel will be involved in the discussion and help seek sustainable solutions which are legally sound and attractive also to industrial users.

- Education and digital skill development and community building

  This topic is of high relevance for all other consortia as well. NFDI4cat will pursue a close dialogue with NFDI4chem, FAIRmat and NFDI4ing to come up with measures that are appropriate to develop formats which ensure sustainability in digital skill development and community building.

- Fostering of platforms for sustainable collaboration between academia and chemical industry

NFDI4cat is one of the few, if not the only consortium with intimate links to industry. In order to bring scientific results to a level that has a socio-economic impact, the German Catalysis Society and Dechema have built solid networks and communities. NFDI4cat will continue the tradition of academic industrial collaboration and seek measures on how both groups can do better science and engineering and also maximize socio-economic impact in a digitalized form.