**The following pages are an excerpt from the proposal of the chemistry consortium NFDI4Chem submitted on 15.10.2019 to the DFG for the first funding period. They are intended for an open, tranparent inter-consortia discussion of cross-cutting topics.**

## 2.3. The consortium within the NFDI

### 2.3.1. Thematic Embedding

In preparation for this proposal, NFDI4Chem has extensively cooperated via joint workshops with FairMat, NFDI4Ing, NFDI4Cat, DAPHNE, PAHN-Pan, NFDI4Phys, NFDI4BioDiversity, NFDI4Agri, NFDI4Health, NFDI4Microbiota, and further neighbouring consortia. The cooperation has covered topics like interdisciplinary (meta)data standards, cross-domain search, legal aspects, and access to repositories. Networking with other consortia has been facilitated by the fact that many members of NFDI4Chem are also active in other consortia. These are: NFDI4Biodiversity, NFDI4Medicine, PAHN-PaN, NFDI4Culture, MaRDI, NFDI4MobilTech, NFDI4Agri, NFDI4MSE, FAIRMat, NFDI4Ing, NFDI4Phys, NFDI4Earth, GHGA, DAPHNE.

To emphasise the importance of cross-cutting topics in the NFDI as a whole, 21 NFDI consortia signed the Berlin Declaration, co-authored by the NFDI4Chem leadership, identifying central topics of general interest for all consortia. We will participate in preliminary workshops to further foster cooperation in preparation of interconsortial working groups. Particularly relevant are our interactions with thematically related consortia where we want to contribute our expertise. We have been in close coordination with the neighbouring consortia from the material and engineering sciences from the very beginning. Discussions with NFDI4Cat showed, that a community-tailored approach is best implemented through agreements on shared tasks in the areas of standards and cross-cutting topics like ontologies, metadata formats and the cross-linking of data repositories. While NFDI4Chem focuses on molecules and their characterisation data, NFDI4Cat covers the areas of technical chemistry and chemical engineering sciences. FAIRmat embraces condensed matter physics which includes soft matter and the (chemical) physics of solids and liquids, addressing a distinct community and research area. Nevertheless, there are partially overlapping areas where NFDI4Chem and FAIRmat will collaborate, for example in the definition and implementation of extended, new metadata standards for quantum chemistry in the NOMAD Repository. With NFDI4Ing, we have discussed options to model metadata and exchanged experiences on digitalisation of workflows for scientific data in chemistry and material science. In life sciences, molecule characterisation data like physicochemical, target engagement, bioactivity, pharmacokinetic, toxicology or safety and regulatory data have

been identified as linking elements. With NFDI4BioDiversity we will collaborate on integrated data access across the consortia, and development of data management tools for smart Lab environments. Here, metabolomics data is of particular interest for the biodiversity community. Together with NFDI4Health and NFDI4Microbiota we will discuss metadata standardisation and cross-sectional mapping for chemical compound characterisation data in contexts such as medication, dietary factors or metabolome data. Synergies between NFDI4Chem and NFDI4BIMP have been identified for spectral and spectrometric imaging data along with 'pure' image data like atomic force microscopic (AFM) imaging data and will be further investigated. With DataPlant we share a common interest in developing training material for data literacy with a special focus on molecule-specific aspects. Together with MaRDI we have identified the potential of new research insights by providing suitable and easy-to-use interfaces to apply mathematical methods of MaRDI on chemical data retrieved from the NFDI4Chem repositories. In the discussions with all consortia mentioned above, the consensus for collaborative measures in dealing with molecule data became apparent. NFDI4Chem aims to coordinate these efforts with the NFDI.

We see the following topics as areas where NFDI4Chem would specifically invest effort to coordinate across consortia with the whole of the NFDI:

**General principles of FAIR data management, international networking and awareness-raising:** Key personnell of NFDI4Chem are active in a number of international efforts, such as GO FAIR, RDA interest groups, ELIXIR implementation networks, the European Open Science Cloud (EOSC) and more, which promote FAIR data in both the chemical as well as biomedical domain. We will aim to harmonize those existing efforts with FAIR data aspects across the whole of NFDI and engage in international networking with generic and specialized bodies promoting RDM and standards. As leaders and participants in collaborative research and excellence clusters in Germany, we will help to promote and implement the principles of FAIR data management in our local community, gather requirements and promote the adoption of the NFDI.

**Repository technology and customisation toward individual domains:** Repository technology will be at the heart of virtually any NFDI consortium's implementation plan. To foster the interoperability of a potentially diverse portfolio of repository technologies, NFDI4Chem wants to promote standardisation of interfaces and technological platforms across the NFDI which can be customised to individual research domains and application scenarios.

**Catalogue of all services developed by the NFDI:** Following the model of the European Open Science Cloud (EOSC) and enabling easier integration into the same, we suggest the cross-cutting service catalogue enabling the keyword based discovery of services by users. Individual catalogues for exposure on individual consortium portals can then be generated on

the fly from the central catalogue. The central catalogue will be designed to feed automatically, if desired, into the EOSC service catalogue.

**Mechanisms and instruments for agreeing on international standards:** Research data can only be re-used when annotated with sufficient metadata adhering to community agreed standards. New standards required for the NFDI cannot be negotiated at a national level but require extensive and long-term international consultations. The NFDI4Chem leadership has been engaged in such efforts for the past 10 years and we want to contribute to agreeing on common best practises for international development of standards within the NFDI.

**Ontologies, terminology services:** Once agreed, controlled vocabularies and ontologies will ideally be managed through lookup terminology services used across the entire NFDI.

**Machine-readable data, data validation:** Especially for cross-domain applications data needs to be unambiguously semantically annotated, both for humans and machines. Using discipline-specific terminologies we will describe research data in machine-readable form and adopt and develop research data semantics for properties, methods, units.

**Efficient and harmonised materials and measures for outreach and training across NFDI:** Established outreach instruments such as workshops, conferences, tutorials and training material, feedback mechanisms ranging from electronic surveys via issue trackers to social media elements will be explored throughout the NFDI. We further expect public policy, funders and learned societies to increase their demand for FAIR and open data management which will naturally increase the incentive for users to engage with these ideas. NFDI4Chem would like to promote concerted efforts with the NFDI towards those goals.

**Legal aspects of research data management, data sharing:** NFDI4Chem participants have expertise to address legal aspects of RDM and provide support for the NFDI community on e.g. legal questions about data ownership, legally compliant operation of the NFDI infrastructures, and the development of science-friendly guidelines for RDM. We assume that there will be similar legal issues in other consortia at a higher level and propose a joint approach to those fundamental issues.

**Unified and interoperable governance models across NFDI:** NFDI4Chem leadership and participants have extensive experience in building international research data infrastructures in the biomolecular and chemical domain and beyond and will happily share this knowledge during discussion across NFDI domains.