

Sonja Herres-Pawlis, Felix Bach, Theo Bender, Nicole Jung, Oliver Koepler, Johannes Liermann, Steffen Neumann, Jochen Ortmeier, Christoph Steinbeck

NFDI4Chem: Building a sustainable data infrastructure for chemical research

Since its foundation in autumn 2019, the NFDI4Chem consortium ('Chemistry Consortium in the NFDI') has been committed to developing and implementing a national research data infrastructure for chemical research in Germany. The vision of 'All chemists publish FAIR data' emphasises the aim of supporting all steps of chemical research – from data collection to reuse – in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) and providing a holistic infrastructure (Figure 1).

The interdisciplinary collaboration between universities, research institutions, infrastructure partners and professional associations such as the German Chemical Society (GDCh), the German Pharmaceutical Society (DPhG) and the Bunsen society forms the foundation of the project. NFDI4Chem represents all areas of chemistry as a part of the umbrella organisation National Research Data Infrastructure (NFDI) and works closely with international initiatives.

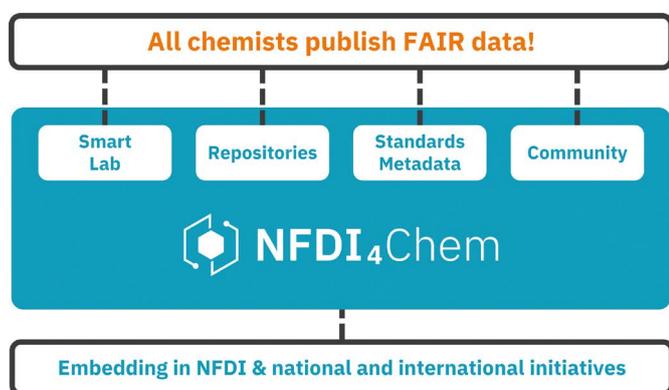


Fig. 1: Digitalising chemical research by offering a comprehensive infrastructure to generate, preserve and provide FAIR research data. CC-BY-SA 40

Digital laboratories and linked data collections

In the first funding phase, a key focus was on digitalising chemical laboratory work in order to collect experimental data in a structured and machine-readable format as early as possible in the research data life cycle. This is where the 'Smart Lab' concept comes into play, integrating digital tools such as elec-

tronic laboratory notebooks (ELNs). In particular, the open source ELN Chemotion was developed to collect chemical experiments, analysis data and metadata in a homogeneous, structured form, enabling seamless transfer of data to repositories – a crucial step towards making FAIR publications the norm (Figure 2). These tools have been gradually introduced to working groups and courses so that students and researchers can learn and disseminate the prerequisites, benefits and application of digital practices. This shift towards digital laboratory concepts not only increases efficiency, but also strengthens the basis for reproducible and collaborative research.

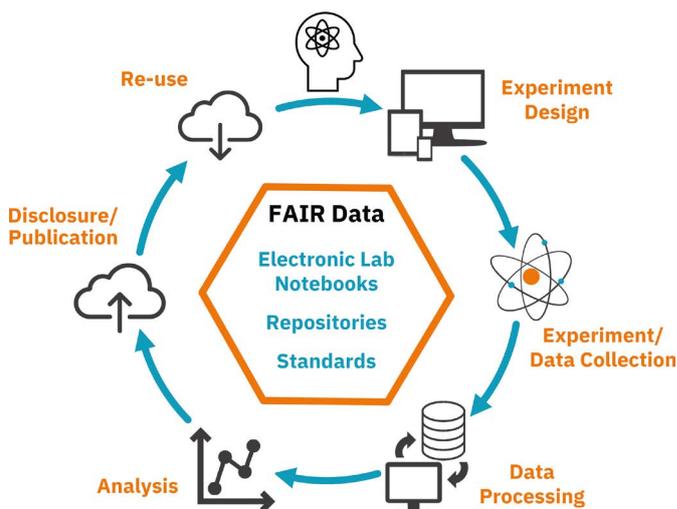


Fig. 2: Data life cycle: Workflow in the digital laboratory. CC-BY-SA 40

Repository landscape and data access

Another core area is the development and networking of chemistry-specific repositories that make data accessible and reusable in the long term. NFDI4Chem has further developed existing repositories such as Chemotion Repository, MassBank, and SupraBank and embedded them in a federated infrastructure (Figure 3). At the same time, new services have been created, such as RADAR4Chem, nmrXiv, and VibSpecDB which cover different data types and sub-disciplines. Especially RADAR4Chem is a repository using rich metadata and perfectly suits for physical chemistry and theoretical chemistry [1].

This infrastructure serves not only as a storage location, but also as a linked network in which data is accessed via services such as a central search platform [2]. Moreover, versioning, cross-linking to the classical text publication and embargos are



Fig. 3: Overview of federated NFDI4Chem repositories. CC-BY-SA 4.0

possible. The harmonisation of interfaces, terminology services and single sign-on mechanisms creates a coherent environment in which research results can be efficiently found, linked and used – a significant advance for chemical research. The Search Service offers a search function across five different repositories, providing 153,000 datasets at a glance.

Standards, metadata and semantic enrichment

In addition to technical components, NFDI4Chem develops and curates standards to harmonize data annotation with metadata and ontologies, creating the foundation of data interoperability. In cooperation with international partners such as the International Union of Pure and Applied Chemistry (IUPAC) and the Research Data Alliance (RDA), minimum information standards and ontological components are being developed and promoted that enable semantic annotation and machine-interpretable data. These include the VIBSO ontology [3], chemDCAT-AP [4] as a chemistry specific extension of a well-established metadata model, reaction descriptions and other core elements of chemical research. The consolidation of ontologies and the integration of chemistry-specific terminology contribute significantly to not only storing data, but also linking it in a semantically rich way, thus making it usable across disciplinary boundaries.

Community involvement and training

A key component for the sustainability of the NFDI4Chem infrastructure is the active involvement of the chemical community. Through nationwide workshops, hands-on training, continuous public relations work and topic-specific events, researchers at all career levels have been sensitised to and trained in FAIR data management. In addition to general FAIR research data management (RDM) training courses, NFDI4Chem has also established specific offerings such as train-the-trainer programmes and introduced formats such as the FAIR4Chem Award to highlight outstanding practical examples. At the same time, research data management content has been integrated into curricular and subcurricular teaching so that students can acquire the digital skills that are essential for future scientific practices. To this end, experience reports have been and continue to be communicated in order to provide best practice examples for other universities.

Networking, synergies and international cooperation

NFDI4Chem positions itself not only nationally, but also in an international context. Through cooperation with international standardisation initiatives such as CODATA, Pistoia, RDA, IUPAC and PSDI, the development of standards and joint services is being driven forward and the interoperable use of data across disciplinary boundaries is being promoted. NFDI4Chem is actively involved in the initiative to build a coalition for digital standards in chemistry to pave the way to sustainable and interoperable chemical data exchange standards [5]. NFDI4Chem is also participating in the CODATA Task Force on improving the quality of research data [6]. These activities are important cornerstones to create knowledge graphs and AI ready data in the next steps.

Sustainability and successes in the first funding phase

The first funding phase of NFDI4Chem has laid a robust digital foundation. The widespread adoption of digital tools such as ELNs, the establishment of a federated repository landscape, and the implementation of standards have already significantly changed the culture of data management in chemistry. The chemical community is increasingly embracing the services, as evidenced, for example, by the rising demand for support requests, the number of data sets in the federated network, and active participation in workshops. These results demonstrate that NFDI4Chem has not only understood the vision of 'All chemists publish FAIR data' as a guiding principle, but has also translated it into concrete services and collaborative practice.

Outlook: Focus and objectives of the second funding phase (2025–2030)

The second phase of NFDI4Chem comprises a strategic programme that builds on previous achievements to implement sustainable operating models for infrastructure components and anchor the cultural shift towards a fully digital research landscape [7]. A key concern is the expansion of ELNs and repositories to cover more sub-disciplines and data types and to make them more interoperable, in particular through deeper integration into federated systems. Especially, physical chemistry and theoretical chemistry with their vast multitude of

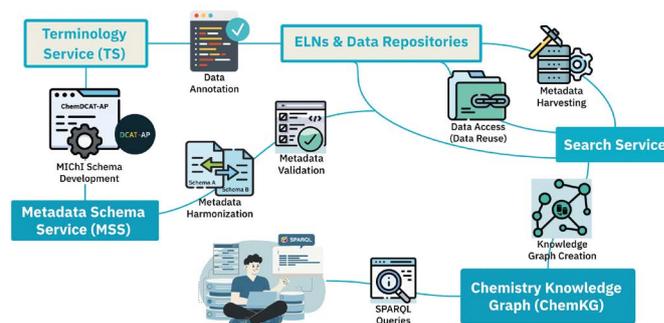


Fig. 4: 2nd Funding phase: Towards a semantic data hub. CC-BY-SA 4.0

methodologies come into the focus and we invite all interested colleagues to approach us for further standardisation efforts. Moreover, the working group TheoChem4ELN implements theoretical workflows into ELNs.

At the same time, emphasis is placed on the further development and integration of semantic standards and ontologies into NFDI4Chem services, linking them more closely together forming the semantic data hub (Fig 4.) to promote global data interoperability. The semantification of data is key to building the Chemistry Knowledge Graph in the second phase. This way we provide AI-ready data models so that data is not only stored but can also be used for modern analysis methods, machine learning and knowledge acquisition – a significant step towards making chemical research compatible with data-driven methods in the modern age. Training and community programmes will be consistently continued and embedded in academic curricula to equip future generations of chemists with digital data literacy.

Concomitantly, the consortium aims to intensify cooperation with publishers and editorial offices in order to integrate FAIR data publications even more closely into everyday scientific life and thus bring about a lasting change in publication culture. Through this integrative approach, NFDI4Chem aims not only to provide a stable infrastructure, but also to establish a lasting paradigm shift in the handling of chemical research data – from everyday laboratory work to publication and global scientific use.

Summary

The development of NFDI4Chem over the first funding phase has shown how a subject-specific consortium can help a scientific community master the challenge of digital data practices. The coordinated combination of technical services, content standards, shared tools and actively designed community processes creates an ecosystem that puts chemical research on a new data and knowledge basis. The second funding phase now opens up the opportunity to stabilise and expand these achievements and integrate them deeply into scientific practice. The vision of consistently FAIR-based chemistry is thus understood not as an abstract goal, but as an immediately achievable strategy that strengthens scientific collaboration, reproducibility and innovation in chemistry in the long term. Physical and theoretical chemistry with their rich methodology are now in the focus of the second funding phase and adapted services for FAIR data are at the fingertips of the physicochemical community soon.

References

- [1] “Radar4Chem Forschungsdaten.” Accessed: Feb. 02, 2026. [Online]. Available: <https://radar4chem.radar-service.eu/radar/de/home>
- [2] TIB-Leibniz Information Centre for Science and Technology, “NFDI4Chem Search Service.” Accessed: Feb. 02, 2026. [Online]. Available: <https://search.nfdi4chem.de/>
- [3] P. Strömert, J. Hunold, L. Giacomo, and R. Stosch, *VIBSO - Vibrational Spectroscopy Ontology*. Github. Accessed: Feb. 02, 2026. [Online]. Available: <https://github.com/NFDI4Chem/VibrationalSpectroscopyOntology>
- [4] S. Philip *et al.*, “ChemDCAT-AP: Enabling semantic interoperability with a contextual extension of DCAT-AP,” *arXiv [cs.DB]*, Feb. 02, 2026. doi: 10.48550/arXiv.2602.01822.
- [5] L. McEwen *et al.*, “Collaborative development and sustainability of digital chemical data standards DigSustain2: A community workshop, April 2025, Germany,” *Chem. Int.*, vol. **47**, no. **3**, pp. 45–48, Jul. 2025, doi: 10.1515/ci-2025-0317.
- [6] “Research Data Quality Management Across the Data Lifecycle,” CODATA, Committee on Data of the ISC. Accessed: Feb. 02, 2026. [Online]. Available: <https://codata.org/initiatives/task-groups/research-data-quality-management-across-the-data-lifecycle/>
- [7] C. Steinbeck *et al.*, “Proposal NFDI4Chem 2025-2030 in the national research data infrastructure (NFDI) — our vision: All chemists publish FAIR data,” *Res. Ideas Outcomes*, vol. **11**, no. e177037, p. e177037, Nov. 2025, doi: 10.3897/rio.11.e177037.



NFDI4Chem

The authors are the spokespersons of the NFDI4Chem consortium (Christoph Steinbeck, Oliver Koepler, Felix Bach, Nicole Jung, Sonja Herres-Pawlis, Johannes Liermann, Steffen Neumann), as well as two project managers (Jochen Ortmeier & Theo Bender). The consortium has set itself the goal of enabling all chemists to publish FAIR data. To this end, since 2020 it has been providing technical infrastructure such as Chemotion ELN, a federation of repositories, a terminology service and a search service, and is working internationally to improve standards.