

Interim Report NFDI4Chem 2023

Steffen Neumann[‡], Ann-Christin Andres[§], Felix Bach^l, Theo Bender[§], Christian Bonatto Minella^l, Franziska Eber[¶], Tillmann G. Fischer[‡], Benjamin Golub[#], Shashank S. Hariyasi[□], Sonja Herres-Pawlis[«], Pei-Chi Huang[□], Johannes Hunold[»], John D. Jolliffe[§], Nicole Jung[□], Johannes C. Liermann[§], Venkata Chandrasekhar Nainala[¶], Matthias Razum^l, Oliver Koepler[»], Christoph Steinbeck[¶]

‡ Leibniz Institute of Plant Biochemistry, Halle, Germany

§ Johannes Gutenberg University Mainz, Mainz, Germany

l FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe, Germany

¶ Friedrich-Schiller-University, Jena, Germany

Technische Universität Braunschweig, Braunschweig, Germany

□ Karlsruhe Institute of Technology, Karlsruhe, Germany

« RWTH Aachen University, Aachen, Germany

» TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

Corresponding author: Steffen Neumann (sneumann@ipb-halle.de)

Abstract

The progress of the DFG-funded NFDI4Chem consortium (NFDI 4/1 - project number 441958208) in data management in chemistry is outlined in our latest report, highlighting the steps we have taken to integrate a data-centric approach within the chemistry community. This interim report offers a comprehensive overview of our data management activities, covering the reporting period from October 2020 to August 2023.

The shift to digital tools in research documentation is driven by our work with Electronic Laboratory Notebooks (ELNs), such as **Chemotion** ELN, offering systematic data storage for easy retrieval and sharing. Additionally, we focus on developing repositories, such as Chemotion repository and **RADAR4Chem**, which fulfil the needs for the storage of chemical data. The NFDI4Chem **Search Service** ensures easy data access from our repositories. Our efforts extend to **community engagement** through conference visits and online presence, aimed at creating awareness for (digital) research data management and connecting to chemistry students and researchers. Our training programs have reached over 600 participants to date. Initiatives like the **FAIR4Chem** award and the Chemistry Data Days promote cultural change towards FAIR data. Our **Editors4Chem** initiative collaborates with publishers for standardised data management and the **Ontologies4Chem** workshops organised by our consortium promote the ontology development in the field.

Apart from the consortium's engagement for chemists, NFDI4Chem members played key roles in the development of the **NFDI** as a whole. Being actively involved in the sections and task forces, NFDI4Chem promotes collaborative solutions across NFDI consortia.

Keywords

Chemistry, Research Data Management, Electronic Laboratory Notebooks, Repositories, Metadata, NFDI, Research Data Infrastructure, Digitalisation

Introduction

This interim report of the progress in the DFG-funded project NFDI4Chem (NFDI 4/1 - project number 441958208) covers the reporting period from October 2020 to August 2023.

The NFDI4Chem consortium and its embedding in the chemistry community

NFDI4Chem started in October 2020 with 27 partners from 21 organisations. Over time, some participants changed their affiliation and two new partners joined. In July 2022, the University of Stuttgart joined the consortium, bringing expertise in enzymology and biocatalysis and supporting the development of standards and smart laboratories. In June 2023, the Federal Institute for Materials Research and Testing (BAM) joined the consortium, strengthening its international integration and expertise in Research Data Management (RDM) solutions for materials science and chemical analytical techniques. As of August 2023, the NFDI4Chem consortium comprises 22 organisations, of which are seven (co-) applicant institutions and 15 participating institutions.

From the beginning, NFDI4Chem has made efforts to systematically identify the needs of the community with our survey (Herres-Pawlis et al. 2020) in 2019 for the NFDI4Chem project proposal (Steinbeck et al. 2020). Four years later, in spring 2023, we conducted a second survey with more than 800 responses (680 from Germany). Participants included professors (22%), senior researchers (24%) and PhD students (33%) from different areas of chemistry (27% inorganic chemistry, 34% organic chemistry, 21% physical chemistry, plus theoretical chemistry, pharmaceutical chemistry, biochemistry, chemical engineering and materials science). Digital data analysis improved (45% seamless, 65% some non-digital steps, such as analysis of NMR spectra). Metadata description was carried out by 56% (42% in 2019). ELN use increased to 30% (18% in 2019) and is slightly above average in organic chemistry (34%) and materials chemistry (33%). Chemotion (26%), ELabFTW (20%) and Sciformation/OpenInventory (9%) were frequently used amongst the 20 Electronic Laboratory Notebooks (ELNs) mentioned - including Word and Excel. Data repositories were used by 20% (up from 13%). In summary, 86% felt that research data management in the curriculum would benefit future students and groups. We therefore have direct feedback on the impact of NFDI4Chem's work on the chemistry community.

NFDI4Chem has strong links with the major sub-disciplines of chemistry (organic, inorganic, pharmaceutical, physical chemistry) through learned societies as consortium members (GDCh, DPhG, Bunsen Society, FID Pharmazie) and outreach activities. Collaborative development with specialised communities, such as chemical ontologies, NMR, EPR standards, electrochemistry, coordination chemistry, macromolecules and enzymes, includes regular meetings and workshops to gather feedback. Users can provide feedback via the Helpdesk and GitHub repositories. Interaction with the ontology community is particularly strong, as evidenced by the annual Ontologies4Chem workshop with ~ 50 participants from major chemistry ontology projects. NFDI4Chem actively participates in working groups on standards for chemical analytical methods.

The NFDI4Chem training programme is critical in promoting RDM awareness. We have developed a basic two-day interactive course specifically for chemists. It covers the basics of RDM and specific chemistry-related scenarios. The workshop, based on the FD Mentor concept (Biernacka et al. 2020) and materials from the research data teams at FSU, RWTH and JGU, started in January 2022 and is open to all chemists in Germany, with growing international participation. The online workshop took place every two months in 2022 and was consistently well booked. As of 2023, we have been focusing more on institution-specific workshops for researchers and data stewards in collaboration with local RDM teams to bridge the gap between local support and chemistry-specific needs. The Chemotion workshop series introduces the electronic lab notebook and its features to beginners in a "learning by doing" approach. The interactive hands-on workshop can be adapted to the sub-discipline and, most importantly, to the needs of the participants. In total, by the end of 2023, we will have organised 30 workshops at institutions in Germany (Fig. 1) and one in Liverpool (online) due to high international demand. All workshop series together, both institutional and public, have reached more than 600 active participants. To meet the increasing demand and to contribute to EduTrain, we are currently expanding the programme to include workshops for data stewards and a 1-day Lead-by-example workshop to provide practical support to researchers and data stewards.

To promote RDM awareness, we presented the FAIR4Chem award at the GDCh JungChemikerForum (JCF) in 2022 and 2023. In June 2023, we organised the first Chemistry Data Days, a two-day conference on data management in chemistry for the non-RDM expert, with around 100 participants, inspiring chemists about the potential of research data. We engage with the community through monthly 'Stammtisch' discussions on RDM, ELNs, repositories, ontologies and molecular representations, along with Chemotion Q&A sessions and various virtual and physical RDM and Chemotion workshops.

NFDI4Chem is active on several social media channels. Our X (formerly Twitter) account has 1062 followers (15/09/2023) and has published 412 tweets. On LinkedIn, we have published > 150 articles for 442 followers and we have launched an Instagram account in March 2023.

NFDI4Chem within the NFDI and its engagement in cross-cutting topics

NFDI4Chem members have been active in shaping the NFDI and its initiatives, with our spokesperson leading the Consortium Assembly in 2021/22. NFDI4Chem spokespersons and co-spokespersons played key roles in NFDI strategy workshops (Glöckner et al. 2019, Bierwirth et al. 2020, Ebert et al. 2021, Konsortialversammlung, des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. 2022), resulting in the election of three consortium members as spokespersons for NFDI e.V. sections, i.e. Section “Education and Training” (Herres-Pawlis et al. 2021), “(Meta)data, Terminologies and Provenance” (Koepler et al. 2021) and “Ethical, Legal and Social Aspects” (Boehm et al. 2021). Other NFDI4Chem representatives contribute to various section working groups. This way, NFDI4Chem members, including section spokespersons, play an important role in addressing key aspects of the NFDI4Chem agenda, such as (meta)data standardisation, terminology, AAI infrastructure and legal aspects of RDM (Hunold et al. 2023). Furthermore, more NFDI4Chem representatives are actively participating in NFDI e.V. task forces, especially in the task force “Evaluation and Reporting” where they have contributed to two white papers (Amelung et al. 2023b, Amelung et al. 2023c) and the [Co laborative Work Documentation](#) (Amelung et al. 2023a).

The work of the sections highlighted the need for collaborative solutions across NFDI consortia. As a result, a collaborative grant proposal called “Base4NFDI” (Bernard et al. 2022) was developed and submitted in 2022. NFDI4Chem is involved in the development of the Terminology Service Basic Service, leading to increased coordination efforts, which are expected to create synergies in the medium term. Funding programmes such as the BMBF Data Competence Centres (Bundesministerium für Bildung und Forschung (BMBF) 2022) or the Helmholtz Metadata Collaboration have created new points of contact and cooperation.

NFDI4Chem strongly supports the Code of Conduct adopted by the NFDI (NFDI e.V. 2023). In addition, we have identified additional points of particular importance in terms of equal opportunities and have published these as an addendum (NFDI4Chem 2023).

NFDI4Chem and community members have also contributed to the NFDI infra-talk series and co-organise the Physical Sciences Consortia Joint Colloquium with five other consortia. One of our co-spokespersons leads the ELN for Experimental Sciences interest group with three other natural sciences consortia. The interdisciplinary Joint Terminology Service is led by another NFDI4Chem co-spokesperson and a member of NFDI4Ing, with contributions from NFDI4Cat and NFDI4Culture. In addition, NFDI4Chem is strengthening local links with other consortia, having (co-)organised six local networking events in the past.

A complete list of all collaborations within the NFDI Association, as well as bi- and multilateral collaborations with other consortia, can be found in Amelung et al. (2023a). As documented there, NFDI4Chem has taken the lead in 17 out of 75 NFDI-related

activities and is involved in more than one quarter of all cross-consortium collaborations within its domain and beyond.

While collaborations offer opportunities to exploit synergies and fasten or optimise the output for the community, it also creates additional workload that was not foreseen in the consortium's work program. Nevertheless, we consider the engagement in cross-cutting topics within the NFDI as absolutely valuable and necessary to create an overarching research data infrastructure across all disciplines.

International networking of NFDI4Chem

NFDI4Chem believes that the development of standards and best practices for research data management should be undertaken with a global perspective. Therefore, NFDI4Chem is active in the global community of chemists and research data infrastructure experts. NFDI4Chem has strengthened its collaboration with the International Union of Pure and Applied Chemistry (IUPAC), linking measures from its task areas (TAs) to several IUPAC projects. TA4 and TA6 are collaborating with the FairSpec project (IUPAC Committee on Publications and Cheminformatics Data Standards 2020). In the WorldFAIR: Global cooperation on FAIR data policy and practice project, NFDI4Chem contributed to the IUPAC WorldFAIR Chemistry deliverable 3.2 "Chemistry Training Package" (IUPAC Committee on Publications and Cheminformatics Data Standards 2022a). TA6 and IUPAC are working together to develop a recommendation on how to use the Compendium of Chemical Terminology ("GoldBook") as a source of definitions for new ontology terms in chemistry-specific ontologies. NFDI4Chem representatives are active in various Research Data Alliance (RDA) working and interest groups. NFDI4Chem regularly contributes to the Chemistry Research Data Interest Group (CRDIG) sessions during RDA plenary meetings, for example, co-organising the ["Describing diverse chemistry datasets across distributed data resources"](#) session in 2022. The interoperable federation of repositories and other NFDI4Chem services, as well as RADAR4Chem, were presented and discussed with RDM experts at several international conferences, including [FDO2022 in Leiden](#) and [RDA plenary 20 in Gothenburg](#). Members of NFDI4Chem are founding members of the potential RDA working group "Data representation in materials and chemicals based on harmonised domain ontologies", which has received RDA-Tiger funding (Goldbeck 2022, Rettberg 2023). Furthermore, one NFDI4Chem co-spokesperson is a member of the [Board of the InChI Trust](#), another consortium participant is chair of the [Sub-Committee on Polymer Terminology](#) and the NFDI4Chem spokesperson is a member of the [Committee on Publications and Cheminformatics Data Standards](#).

Additionally, NFDI4Chem has participated in two [IUPAC Global women's breakfasts](#), showcasing the achievements of women in NFDI4Chem. We also recently participated in a panel discussion on critical reflections on RDM at the [SDG Graduate Schools Alliance midterm conference](#), focusing on equal opportunities for researchers in the global south.

NFDI4Chem has established collaborations and joint activities with international learned societies. An important forum for data management in chemistry is the Division of

Chemical Information (CINF) of the American Chemical Society (ACS). During ACS CINF events, NFDI4Chem has strengthened links with international collaborators, including the UK's Physical Sciences Data Infrastructure (PSDI). At the 2023 Fall Meeting of the American Chemical Society, NFDI4Chem co-organised two sessions on "Helping Chemists manage their Data" and "Metadata to Knowledge Graphs".

NFDI4Chem regularly collaborates with the European Chemical Society (EuChemS). During the EuChemS congress 2022 in Lisbon, Chemotion ELN attracted considerable interest. We are also working with IYCN and EYCN, which are key stakeholders for RDM implementation in universities, while GDCh-JCF is acting as a multiplier for NFDI4Chem in the German community. NFDI4Chem collaborates with the Royal Society for Chemistry (RSC) in the development and curation of ontologies.

NFDI4Chem partners are active in ELIXIR, where we contribute to Bioschemas developments and organise projects at the 2022 and 2023 ELIXIR Hackathons. The NFDI4Chem knowledge base (TA5) is modelled on the ELIXIR Converge RDMkit and the metadata standards (TA4) are aligned with Bioschemas.

The European Chemistry Thematic Network (ECTN) includes over 20 members from 30 European countries, promoting science and engineering education across borders and shaping chemistry degrees. The ECTN is revising the recommendations for "Bachelor core chemistry content" to integrate modern RDM into European curricula, with the involvement of NFDI4Chem partners. Embedded in the NFDI activities, NFDI4Chem has presented its work at European Open Science Cloud (EOSC) events, i.e. the presentation of the Terminology Service at the EOSC Symposium 2021 in the session "Metadata and Data Quality".

As we attend more international conferences and become more known internationally, we expect the number of Chemotion installations to increase significantly (Fig. 2). It should be noted that these are only installations known to us through the helpdesk (Chemotion does not call home; therefore, we cannot track all installations). To increase visibility and engagement with users, we plan to create dedicated social media accounts for Chemotion.

Sustainability of services

The NFDI4Chem services are located at reliable institutions supported by robust data centres, including KIT, TIB, FIZ and FSU. These centres adhere to state-of-the-art standards and provide fail-safe operations, data security, fast networks and skilled IT staff to ensure high availability. An [uptime tracking service](#) monitors repository and service availability and performance, enabling problem identification and smooth operation. In addition to technical reliability, NFDI4Chem's services prioritise sustainability. A key aspect is the use of open source software with modular code design. This approach facilitates flexible software reuse, customisation, updates, security enhancements and integration with other services and tools (e.g. NMRium in nmrXiv, Chemotion ELN and Chemotion Repo). This sustainability focus extends to the Chemotion ELN, core data

repository software, [FIZ-OAI provider](#), ontology and terminology services, search services, research software and all other reusable tools.

Another aspect of sustainability is Docker containerisation, which provides easy deployment and scalability and allows services to be easily moved between data centres and the cloud. Services become adaptive and not tied to specific technical setups. However, moving large amounts of archived data remains a challenge. We are developing exit strategies, including the use of BagIt containers (Kunze et al. 2018) for datasets. This preserves context during repository moves by embedding data and metadata in standardised containers. Services are optimised for interoperability, using microservice architecture where possible and modular code design for rapid replacement of components and frameworks. To increase interoperability of datasets, we follow standards for data formats and metadata developed in TA4 and with IUPAC and endorsed by the international chemistry community.

Operating model of the NFDI4Chem services

Under the current operational and financial model, all NFDI4Chem services (see section on [Services provided by NFDI4Chem](#)) are free to users to encourage widespread use regardless of financial constraints. We believe this is currently the only successful model for the efficient adoption of RDM, as it is a relatively new requirement for many chemists and free services to lower the barriers. Partner contributions have adequately supported the establishment and operation of the services. While free services are an essential part of the overall concept, the infrastructure and human resources required to deliver the services cannot be provided entirely as in-kind contributions and need to be supported by public funding. Currently, the NFDI4Chem operating model is based on a combination of various in-kind contributions of infrastructure and staff, public funding through the NFDI4Chem project and other third party funding. We believe that this operating model is the key to success and must continue at least until FAIR RDM practices are established as a natural part of scientists' work. However, should political decisions require a change in the operational model, NFDI4Chem can draw on concepts and expertise developed with consortium partners at FIZ Karlsruhe. While the [RADAR4Chem](#) service is offered as a free service to German scientists, the widely available RADAR service provides repository services to institutions based on a sustainable business model (Bein et al. 2016). This model could be used as a blueprint for similar services (i.e. data repositories) that rely on covering the high costs of operation and data storage.

Research data management strategy

Achievements of the Task Areas and their relevance for the scientific community

NFDI4Chem's vision is that all chemists publish FAIR data. To achieve this, we are developing the infrastructure and services for research data management and training

the chemistry community to use them. The use of our services and infrastructure greatly facilitates and accelerates the daily research routine of our community, for example, by making data more findable and reusable. At the same time, the use of our infrastructure improves the quality of data by promoting best practices, standards and adherence to good scientific practice. The actions we will take to achieve our goal are divided into six task areas (TAs), which will work closely together and are described in the following sections.

TA1 (Management), based at FSU as an applicant institution, manages the technical, financial and administrative processes with the support of the partner institutions. The cooperation agreement of November 2020 serves as the legal basis for the cooperation and the transfer of funds. OpenProject, for which NFDI4Chem, together with other consortia, has purchased an enterprise licence, is being used to monitor the progress of work at consortium level, in addition to a meeting and reporting system that ensures project control. TA1 organises two consortium meetings per year. Additionally, smaller retreats within and between TAs have been effective in driving developments. Four [Advisory Boards](#) (National, International, Industry, Publishers) have been established to provide advice and feedback at the consortium meetings. Communication within the consortium is based on regular meetings, mailing lists and the chat tool rocket.chat. A strategic communications concept was initiated early on, including a corporate design and the [website](#), which was launched in Q4 2021 and is frequently updated and expanded. It serves as a single point of information for the community and provides an overview of the consortium and all its services.

TA2 (Smart laboratory) is developing open source software to create a digital infrastructure for FAIR data management. This Smart Lab environment includes instrument integration, electronic lab notebooks and additional scientific digitalisation tools. Seamless data transfer is a key focus to ensure interoperability with NFDI4Chem components.

Over 39 local instances (at 37 different locations) of the Chemotion ELN have already been installed with the support of NFDI4Chem (see Fig. 2), spanning academia and industry. They are being integrated as a core component of data management in German universities, fostering interaction between users and developers for various scientific processes. Chemotion ELN for individual scientists and small groups as Software as a Service is currently being tested by pilot users and is planned to be rolled out in early 2024.

During the early years of NFDI4Chem, the Chemotion ELN underwent continuous and substantial improvements in functionality, development approaches and deployment methods. This facilitated rapid [Docker container](#) installations for a diverse chemistry user base (Docker 2023). Adapting the ELN to different sub-discipline needs was challenging due to different documentation, data structure, data processing and workflow requirements. TA2 tackled this through sub-discipline specific work packages, led by domain experts, which were continuously merged into the main ELN code. Highlights include new entities for [chemical biologists](#) (Chemotion ELN contributors 2022),

inorganic chemists, [polymer chemists](#) and [image annotation](#) (Chemotion ELN contributors 2022), as well as flexible module integration ([LabIMotion](#)) (Huang and Lin 2023) to support the development of new documentation standards across sub-disciplines.

Chemotion development and feature integration follows a defined workflow: planning, coding, testing, community consultation in TA team meetings and frequent releases (2020: 0, 2021: 4, 2022: 4, 2023: 5 until August). Feedback via helpdesk and [GitHub](#) (resolved issues 2020-2023: 359) allows transparent activity tracking, requirement discussions and task allocation amongst the geographically distributed German team (Karlsruhe, Halle, Braunschweig, Aachen, Jena).

TA2 improved device integration into the ELN, enabling remote device control via the user interface (UI) and automated data transfer (Starman 2023). A growing number of supported devices and protocols support the Smart Lab strategy. In parallel, TA2 has improved data conversion from device-generated files to open, standardised formats. Converter software with readers and routines for common file types has been developed and integrated into the ELN. TA2 is developing and harmonising software for reading, processing, visualising and analysing data, enabling the use of the open source tools ChemSpectra (Huang et al. 2021) and NMRium (Patiny et al. 2023). These tools cover multiple measurement types, minimising reliance on commercial software. Comprehensive online documentation describes features and methods and provides examples, videos and SOPs for easy reuse of components by administrators and users.

Within TA2, work on a smart lab environment will enable early digitisation and a digital workflow for deposition in NFDI4Chem repositories. Data transfer and publication from Chemotion ELN to Chemotion repository (Chemotion ELN 2021, Tremouilhac et al. 2020) and RADAR4Chem have been established. Similar processes for other NFDI4Chem repositories are being designed ("[repotracker](#)", more information below) and will be implemented soon, in close collaboration with TA3. The ELN customisation and data transfer developments will be used by scientists in NFDI4Cat, Daphne4NFDI and FAIRmat. This collaboration with other consortia and communities enables even more sophisticated interdisciplinary data management. ELN interoperability and data exchange mechanisms will be explored through participation in workshops and cross-ELN projects. TA4, TA5 and TA6 strongly support the successful activities of TA2 by contributing to standards, metadata schemes and community training on digital working environments.

TA3 (Repositories) establishes a virtual environment of federated repositories for molecule-related data and ensures the integration of existing repositories selected on the basis of suitability, open source accessibility, adherence to standards and funding requirements.

The repositories that are currently part of the NFDI4Chem federation are listed below in "Services provided by NFDI4Chem for the community". For each of them, detailed information has been gathered through interviews and workshops with providers (Bach et

al. 2023) and summarised as [profiles on our website](#). A knowledge base article "How to Choose the Right Repository" guides researchers in choosing the right repository for their data and discipline. In addition, we evaluated other international repositories researched through [re3data](#) using transparent criteria (Minella et al. 2023) selected by TA3 in collaboration with other TAs, which proved valuable even if they did not fully meet our criteria. We have signed letters of intent with CSD, ICSD and CCDC Access Structures Service for crystallography offerings and are exploring integration with NFDI4Chem.

All NFDI4Chem repositories were supported to optimise their operational fitness, interoperability, metadata standards (developed by TA4) and harvesting, dataset landing pages and interfaces to other services such as the NFDI4Chem search service. Concepts have been developed to link the repositories to the scientists' workspace by allowing data to be transferred from the Chemotion ELN to the repositories. The "[reptracker](#)" software was developed to record and monitor the various data transfer processes. In collaboration with TA4, we decided on minimum information standards (Herres-Pawlis et al. 2022) for the metadata of datasets in the repositories of our federation, based on DataCite. The FIZ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) provider and backend were published on GitHub and integration help was made available to repositories.

The relevance of TA3's work can be seen in the author instructions of journals, including *Angewandte Chemie International Edition*, which recommend the NFDI4Chem repositories for data deposition (Journal Editor 2023, Proteau 2023).

TA4 (Metadata, data standards and publication standards) focuses on the development and harmonisation of minimum information (MI) standards and metadata for chemical research data, as well as data standards for molecules and reactions, including experimental and theoretical characterisations (also see the following section). Working closely with TA6 (Terminology Services), TA4 contributes relevant chemistry file formats to the EDAM ontology and adds repositories, data types and formats to the FAIRsharing catalogue.

Internationally, we work with IUPAC and the InChI Trust and with TA5 on the [WorldFAIR initiative](#) to develop a FAIR data cookbook for chemists using CODATA. Our involvement in the InChI Trust Organometallics Working Group has led to important developments in the international open molecular representation standard, including a non-disconnection approach. An NFDI4Chem co-spokesperson now leads the unified InChI organometallics/inorganics working group and sits on the InChI Trust Board. In addition, the recently developed TUCAN, a graph-based molecular representation (Brammer et al. 2022), improves the handling of inorganic compounds. A review of data standards and file formats in analytical chemistry has been published (Brammer et al. 2022, Rauh et al. 2022). We are also working with NFDI4Phys initiative and NFDI4Ing on a framework for file conversion services presented at the Data Formats Workshop in October 2022.

TA4 and TA2 are working together to develop documentation and reporting standards at all scientific levels (processes, entities and data files). TA2's LabIMotion extension guides

the creation of comprehensive documentation templates through engagement with the chemistry community, both within and beyond NFDI4Chem. These community-driven templates are iteratively refined and made available on GitHub for scientists to use directly in their ELN instances. After refinement and incorporation of feedback, these [templates](#) may become documentation standards or serve as the basis for modules in the next ELN release. These ELN standards can also be applied to NFDI4Chem repositories. For the Chemotion repository, processes for the distribution of standardised templates and the management of future versions have already been established in the form of a [template hub](#), built from the versioned templates on GitHub.

Other activities focus on monitoring and improving the FAIRness of data publications. We analysed DataCite "Dataset" records using F-UJI (Devaraju and Huber 2021), comparing FAIR scores in chemistry and other fields to identify (anti)patterns. The report (Fischer et al. 2023) highlights the results.

The development of MI guidelines is an ongoing process involving discipline-specific workshops, consensus on journal reporting standards and technical interoperability. Workshops on FAIR NMR RDM and MI standards in polymer chemistry discussed specific metadata and reporting rules. Reports of the workshops are available on the [NFDI4Chem website](#). NFDI4Chem assists the chemistry community and partners in preparing standards-compliant data publications. Key datasets from Lead-by-Example initiatives are described in the [Knowledge Base](#) and serve as practical test data for repository and service development (Fischer et al. 2023). Proposals for revised reporting standards in analytical chemistry have been initiated by scientists and stakeholders, such as BI, as a draft for further development in TA2 and TA3. The first Editors4Chem workshop (Fischer and Neumann 2021) with IUPAC in November 2021 involved 18 editors, including many editors-in-chief, working with publishers to integrate FAIR data recommendations into journal author guidelines. Survey results on author guidelines in chemistry journals have been published (Parks et al. 2023) and presented at workshops and conferences. A second workshop is planned for 2 November 2023 (Note that this report reflects the status as of September 2023. The second Editors4Chem was successfully held in November 2023).

In **TA5 (Community involvement and training)**, we aim to drive cultural change towards digital research data in chemistry. We provide resources, training and support to researchers. A major milestone was the release of the [NFDI4Chem Knowledge Base](#). Our [Youtube channel](#) provides RDM training material on general RDM and Chemotion from basic to advanced topics. We run regular online and on-site RDM training for chemists and a dedicated Chemotion course. In collaboration with TA2, we organise online Chemotion Q&A sessions with 10-20 participants. Our monthly online "Chemotion/NFDI4Chem Stammtisch" features discussions on RDM and various ELNs, as well as advanced topics, such as molecular machine-learning or recent InChI and SMILES developments, with 25-45 participants.

Outside the consortium, we play a leading role in the [NFDI section Training & Education](#). Our involvement in the BMBF project DALIA (Data Literacy Alliance) aims to develop a

semantic platform for NFDI teaching materials. We are working closely with IUPAC on WorldFAIR, a FAIR data cookbook and improving digital molecular representation with InChI. We are working with major chemistry publishers on new publishing standards. Our collaboration with TA4 includes best practice examples from the consortium, showcasing research papers with good data management (Heck et al. 2022, Hermann et al. 2022, Raßpe-Lange et al. 2023, Fuchs et al. 2023a, Fuchs et al. 2023b). The Chemotion ELN serves as a working group database for several TA5 members, demonstrating the benefits of RDM within an active research group. A best practice working group policy, published jointly with TA2, is available in Chemistry Methods (Fink et al. 2022b).

Communication is key to spreading cultural change and TA5 actively communicates with the scientific community to this end. We promote the latest developments through presentations at conferences, institute seminars and booths at national and international chemistry conferences. During the reporting period, we participated in around 30 conferences, including ACS, Analytica, EuChemS, ORCHEM, MACRO, Coordination Chemistry Meetings, IUPAC Conference, Chemistry Teachers' Conference and others. We maintain an [active presence on Zenodo](#) and distribute information material about the work of NFDI4Chem. Colleagues often seek our guidance in transforming chemistry departments into FAIR departments and we provide support at all levels.

We believe in introducing RDM to the next generation at an early stage and integrated it into a 5th semester inorganic laboratory course at RWTH in 2020. More than 350 students learned the basics of chemical RDM and used the ELN Chemotion in their lab work. The digital change was well received and students expressed interest in further studies on digital chemistry (Fink et al. 2023a). We created Chemotion training videos in German (Fink et al. 2022a) and English (Fink et al. 2023b) to support self-learning. In addition, specific RDM modules were developed for inclusion in Master's level chemistry courses as a meta-competence to promote long-term cultural change. Chemistry students seem to prefer learning RDM with case studies from real research (Hermann and Herres-Pawlis 2020, Hermann et al. 2020, Conen et al. 2022a, Conen et al. 2022b, Conen et al. 2023).

TA6 (Synergies and Cross-Cutting Topics) aims at a holistic use of the NFDI4Chem infrastructure and services. It facilitates and promotes the harmonisation of existing and new components and adds cross-cutting infrastructure services. TA6 coordinates NFDI4Chem participation and contributions to NFDI measures (see section on [NFDI4Chem within the NFDI](#)) and international networking (see section on [International networking of NFDI4Chem](#)). TA6 contributes to the process of establishing a NFDI IAM (Identity and Access Management), initially in a NFDI task force. This joint initiative has resulted in the NFDI basic service project IAM4NFDI. TA6 ensures that these developments can be integrated into the NFDI4Chem services. TA6 is also contributing to the discussion and development of legal guidelines for a legally sound RDM, integrating our perspective in the work of the NFDI section ELSA.

TA6 works closely with TA2, TA3 and TA4 to develop ontologies and metadata standards and implement them in services for semantically annotated data. In the Ontologies4Chem activities, TA6 is contributing to the implementation of ontologies in the chemistry

community, with a focus on the creation of machine-actionable data. We have conducted a thorough evaluation of ontologies in the chemistry domain and defined selection criteria (Strömert et al. 2022b). We are continuously curating ontologies, such as CHMO, MOP, RXNO, CHEMINF or the IUPAC Goldbook in collaboration with RSC, IUPAC and the OBO Foundry. To address gaps in the chemistry ontology landscape, we collaboratively developed the new Vibrational Spectroscopy Ontology (VIBSO), initially focusing on Raman spectroscopy (Strömert et al. 2023). We organised the first Ontologies4Chem workshop (Strömert et al. 2022a) in September 2022, with over 50 international participants, to foster discourse on chemistry ontologies and their application in RDM. TA6 enhances the use of ontologies by developing the NFDI4Chem Terminology Service (TS), which provides rich search and exploration capabilities within ontologies. We have built the TS backend on top of the OLS software (Jupp et al. 2015), extending its features and adapting it to our needs, while contributing back to the community. TIB is developing the backend for NFDI4Chem and NFDI4Ing collaboratively and it is part of the effort to establish the TS4NFDI terminology basic service. The development of the Search Service involves collaboration amongst TA6, TA3 and TA4 to ensure seamless integration of repositories into the NFDI4Chem federation. Metadata from each repository will be collected and metadata formats and protocols will be harmonised with TA3. Efforts have started with the DataCite format and processing Bioschemas.org as JSON-LD. TA4 is working with the Bioschemas community to extend the schema to include additional chemical information.

In the development of the Search Service (TIB-Leibniz Information Centre for Science and Technology 2022), TA6, TA3 and TA4 jointly address critical steps to ensure seamless integration of repositories within the NFDI4Chem federation. The service harvester collects metadata from individual repositories. Together with TA3, we are standardising metadata formats and protocols for their delivery. Initial progress includes implementation of the DataCite format and OAI-PMH APIs used in Chemotion Repo, DaRUS and RADAR4Chem. For chemical information integration, we are integrating Bioschemas as JSON-LD in the current development phase (MassBank, nmrXiv).

Development of metadata standards

NFDI4Chem adopts and develops metadata standards associated with chemical research data, including information about molecules and reactions, as well as data for their experimental and theoretical characterisation. Datasets may either directly embed enriched metadata and/or metadata is available through APIs (e.g. OAI-PMH), registries (e.g. DataCite commons) and search engines (Google dataset search or NFDI4Chem Search Service).

For domain-independent metadata about datasets, such as title, keywords and creators, we use DataCite in the nmrXiv, Chemotion and RADAR4Chem repositories, which also provide DOIs registered through DataCite. For truly rich annotation of chemical information, we use, adapt and develop specific metadata using the JSON-LD-based Schema.org framework. All repositories provide metadata via [Schema.org](#) compliant

JSON-LD. We are active in the [ELIXIR](#) Bioschemas community, where one of our co-spokespersons has become co-leader of the chemicals working group. New Bioschemas profiles, such as reactions, are under discussion to be specified and submitted to become an accepted standard. One NFDI4Chem co-spokesperson is a member of the Technical Specification & Implementation Group of the FAIR Digital Objects Forum.

The RADAR4Chem and Chemotion repositories provide an [OAI-PMH](#) endpoint for metadata harvesting by other services. In addition, we have extended the XML-based OAI-PMH architecture (using the [FIZ-OAI provider](#)), so that JSON-LD data can also be submitted to an OAI-PMH service, where it can be queried both as JSON-LD and via a crosswalk as OAI Dublin Core (Castro et al. 2023). The overarching search (TA6) can perform incremental harvesting of domain-specific metadata via OAI-PMH.

Schema.org allows the use of defined terms from ontologies and we have started to connect the repositories to the NFDI4Chem Terminology Service and embed the terms in the exported metadata. Services established within the NFDI4Chem are further described below in "Services provided by NFDI4Chem for the community".

Implementation of the FAIR principles

NFDI4Chem strongly endorses FAIR (meta)data. Several Task Areas (TA2, TA3, TA4, TA6) and service providers are contributing to a consistent implementation across services through the adoption of standards and technical requirements. In parallel, education and training will support these developments and the acceptance and understanding of FAIR data by the community.

To ensure **Findability**, all our resources provide rich, machine-readable (meta)data linked to domain-specific and cross-domain vocabularies. Our ELN and instrument integration strategy, as outlined in the section [Achievements of the Task Areas](#) (TA2) encourages the early collection of rich metadata during data generation. While (meta)data in ELNs are typically private to research groups and not globally accessible, they are prepared for transfer to repositories, making them findable while preserving domain-specific information. Most NFDI4Chem repositories register their datasets with DataCite and assign globally unique and persistent DOIs. This facilitates indexing in our Search Service. Dataset landing pages in repositories are optimised for human use. To improve machine actionability and interpretability, several NFDI4Chem services (including RADAR4Chem, Chemotion repository, MassBank and nmrXiv) adopt unified metadata via JSON-LD, following [Schema.org](#) and relevant chemistry types from [Bioschemas](#) (Neumann et al. 2023). This approach ensures data findability via search engines, such as Google Dataset Search and the NFDI4Chem Search Service.

To be **Accessible**, all data provided by NFDI4Chem services are retrievable by their persistent identifier using HTTPS as a standardised communication protocol that is secure, open, free and universally implementable. All NFDI4Chem components are available under open-access models. For some functions, such as data upload, download and editing, where registration and login are required, NFDI4Chem services

facilitate and manage access using established authentication protocols and identity providers, such as OpenID and Shibboleth. NFDI4Chem will use the NFDI-All service for authentication and authorisation procedures when available. To allow programmatic access to data and metadata, NFDI4Chem repositories support or will support the OAI-PMH protocol. An OAI-PMH provider is provided by FIZ. In addition, the components of the infrastructure already provide standardised APIs or will implement them as part of their work programme (see TA3 in section on [Achievements of the Task Areas](#)).

To be **Interoperable**, our metadata and data are stored and made available according to existing standards. We address gaps in standards, routines and representations by developing our own solutions, which are discussed and negotiated with the chemistry community. For example, we are developing MI metadata standards to semantically describe experiments, simulations, molecule characterisations and more. At the same time, NFDI4Chem is promoting open data formats (Rauh et al. 2022), extending them to cover currently unestablished data types and systematically implementing them in our services. All NFDI4Chem components, from ELN to repository services, include software and toolkits that automate the collection of standardised data and metadata and ensure interoperability using these standards. We use community-accepted toolkits, such as RDKit, CDK and OpenBabel, as well as terms from established ontologies, such as CHMO and RXNO. Automated data conversion processes allow storage in the jcamp-dx file format, originally developed as an IUPAC standard for IR, UV-Vis and NMR spectroscopy and adaptable to other measurement types. We use integrated data editors that store data in standardised, readable file formats or containers, even for proprietary inputs. This approach produces interoperable information that can be easily reused in other systems, including repositories.

To be **Reusable**, our (meta)data have accurate, relevant attributes that conform to domain-specific community standards. We use established ontologies, such as CHMO and RXNO and develop our own, such as VIBSO, where appropriate, to facilitate understanding by both humans and machines. These ontologies adhere to FAIR principles to ensure that annotated data remains FAIR. We prioritise the use of openly-licensed and well-maintained ontologies. We integrate metadata annotation standards, such as ROR ID (Research Organisation Registry) and GND (Integrated Authority File) into NFDI4Chem services, with plans for their universal adoption. Data are released with clear and accessible usage licences, supported by legal policies and guidelines. NFDI4Chem repositories curate data and metadata to ensure reusability, with the level of curation varying according to their role in the federation. For example, Chemotion is highly curated, while RADAR4Chem is mainly automatically checked. We also facilitate data reuse by providing targeted datasets for machine-learning and other support.

Services provided by NFDI4Chem for the community

Following the definitions in Amelung et al. (2023c), these services were developed and provided by NFDI4Chem:

The [Knowledge Base](#) (N4C-KB), launched in late 2021, involves 21 contributors and offers various entry points based on the viewer's discipline, role or specific interests. It covers a wide range of topics, from basic RDM concepts to more in-depth articles. The N4C-KB assists users in selecting the right data repository for their research data needs. Hosted at JGU and built using the open-source framework [DocuSaurus](#), the platform allows all content to be stored in a [GitHub repository](#), using simple Markdown syntax. This approach makes it easy for authors to contribute without requiring web programming skills. The website is automatically updated with every change to the repository. The N4C-KB team actively supports contributors and accepts content in a variety of formats.

The [Terminology Service](#) (TS, TIB-Leibniz Information Centre for Science and Technology (2023b)), hosted by the TIB, is a comprehensive repository and search service for ontologies, terminologies and vocabularies in chemistry and related disciplines. As of August 2023, it contains 39 terminologies, selected, based on criteria established by our Ontologies4Chem overview (Strömert et al. 2022b) and community approval (Strömert et al. 2022a). The TS provides advanced search, browse and access capabilities within these terminologies, providing rich metadata and information. Users can explore terminologies through graph or tree visualisations and access development and curation features. It facilitates connections to original terminology repositories, enabling term requests and comments as a step towards a comprehensive terminology curation platform. The TS plays a pivotal role in generating semantically annotated, machine-actionable data and provides a comprehensive API (TIB-Leibniz Information Centre for Science and Technology 2023a) for other NFDI4Chem services to integrate terminologies into their data annotation workflows, such as ELNs or data repositories. In addition, NFDI4Chem's Ontology Elements web components (Venkata et al. 2023) provide an easy way to implement semantic annotation widgets using these terminologies.

NFDI4Chem drives the development and establishment of **ELNs** as a key requirement to achieve systematic digitalisation. While the developed ELN software is offered to users as source code to be hosted locally, there are three additional ELN-based services within NFDI4Chem:

1. For IT staff and admins, a [Docker container](#) service is provided [as image](#) to easily install the ELN with all required dependencies. Management of single instances or multiple instances is supported by providing a [command line interface](#).
2. NFDI4Chem offers the hosting of four (five by the end of 2023) ELNs as [test instances](#) that can be easily used by scientists looking for the right ELN solution. From more than 60 available ELN software tools, the four (five) most important OS solutions for chemists have been selected and are hosted for testing, comparison and educational purposes.
3. Chemotion ELN will be hosted as a service for individual researchers and small groups. The service is currently being set up. Initial pilot use cases with individual users and small groups are already underway in preparation for the full service. The service includes the migration of content to a local instance when a

sustainable user group size is reached - which could be successfully completed with a first pilot user group in 2023.

The [federation of core repositories](#) comprises seven German-hosted repositories, each covering essential content in key sub-disciplines of the chemical community. These repositories are developed and provided as individual services, tailored to specific discipline-specific processes and functionalities driven by their respective communities. Within the NFDI4Chem federation, existing repositories adapt and emerging repositories develop workflows, standards and functionalities to create a harmonised data infrastructure. This infrastructure aims not only at data interoperability, but also at collaborative interaction between repositories and other NFDI4Chem services. The selected core repositories can handle different data types, chemical processes, analytical data and specialised methods, thus supporting the entire data landscape. The roles and requirements of these core repositories vary, with the first funding period focusing on strengthening existing repositories through strategic source code improvements for efficient development and stable hosting.

To date, five NFDI4Chem repositories are in operational use. Of these, we describe RADAR4Chem, the Chemotion repository, nmrXiv and massbanKEU in more detail, as they are currently relevant to the widest user community and major changes have been released.

RADAR4Chem is a cross-domain repository, launched in March 2022, that provides flexible storage options for a wide range of chemistry-related data, with no restrictions on data types or content. It has been developed by adapting the existing [RADAR](#) service. Each registered scientist can publish up to 10 GB of research data by default, with the option to increase storage upon reasonable request. The hosting and integration of RADAR4Chem into the federation was crucial to fill the gap in discipline-specific repositories and ensure data preservation while discipline-specific solutions are still being developed. A high priority has also been given to enabling seamless data transfer from the Chemotion ELN to RADAR4Chem. This allows direct and effortless publication of data collected in the ELN with just a few clicks (enabled with ELN version v.1.5.0).

The [Chemotion Repository](#) deals with data related to chemical reactions and chemical substances and was established at KIT in 2015, initially serving a narrow range of scientific data (Tremouilhac et al. 2020). Since then, it has evolved within NFDI4Chem to meet different scientific needs. It serves as a pilot for data transfer from ELN to repositories and will be adapted as a service to manage data transfer to other suitable repositories for interoperability. Key milestones have been achieved in 2022 and 2023, including the development of tools, such as the [repo-tracker](#) and [repo-downloader](#) software. It provides enhanced reporting capabilities and supports the publication of different data types with discipline-specific templates. NFDI4Chem funds its development and the content of the repository is curated by in-kind contributions.

The [nmrXiv](#) repository, hosted at FSU, is a new NMR spectroscopy data repository and analysis platform built from the ground up. It builds on the experience of its predecessor,

nmrshiftdb2. nmrXiv is open, FAIR and consensus-driven, preserving both raw and processed NMR data. In its pre-release phase in early 2023, it already contained 14 projects, 81 compounds and 490 spectra. It provides DOIs, web UI and REST APIs (Open API, DataCite, Bioschemas, NMRium). nmrXiv follows the DataCite metadata schema, enhanced with InChI and SMILES. It uses two-factor authorisation and single sign-on with popular social network logins, including ORCID. Storage capacity is provided in-kind by FSU.

MassBank EU, hosted at the UFZ, is the first public repository of mass spectrometry data, facilitating its sharing with the scientific community. Since 2021, its compound dataset has grown to 15,075 (from 14,788) and its spectra to 90,190 (from 86,576) in 2023. MassBank uses GitHub for AAI (open read access, limited write access) and uses [GitHub issues](#) for curation tracking, managed by the [MassBank record validator](#). Spectral data and metadata are stored in a human-readable record format within a revision control system and continuous integration ensures record integrity with each change. NFDI4Chem funding has enabled a modern software overhaul, with a first development release using a JS-based front-end and a REST-based back-end.

Two other databases, Suprabank and STREND A DB, are part of the NFDI4Chem repository federation. Both services are currently provided as a stable service with only minor adjustments in the production environment.

Suprabank is a specialised database, hosted at KIT since 2019, offering unique data on intermolecular and supramolecular interactions. It primarily addresses supramolecular and physical chemists, as well as biologists in organic chemistry, focusing on binding, assembly and interaction phenomena not found in other repositories.

STREND A DB, established in 2016 and operated by Beilstein Institute, is a well-established repository for enzymology data. It collaborates with over [55 international biochemistry journals](#) and has integrated the STREND A guidelines into its author instructions. The database ensures the completeness and validity of enzymology data prior to submission for publication. It primarily contains functional enzymology data, including kinetic and experimental data. STREND A DB is an in-kind contribution.

In addition to the services available in production mode, the VibSpecDB repository is under development. VibSpecDB will focus on Raman and IR spectra.

The **Search Service** by TIB, which was released in summer 2022, acts as a central hub for searching the federated repositories of NFDI4Chem. It currently includes 93,935 datasets from the Chemotion Repository, MassBank and Radar4Chem. The integration of the chemistry sub-collection of DaRUS marks the first integration of datasets from a generic data repository. The service regularly harvests and indexes metadata, handling different metadata models and protocols (Fig. 3). It enhances them with chemical metadata, enabling searches by chemical structure codes, molecules and analytical methods. The metadata and search capabilities are being developed in collaboration amongst TA3, TA4 and TA6, involving stakeholders, such as IUPAC, DataCite,

Bioschemas Community and NFDI sections. Agile development aims to integrate more repositories.

The NFDI4Chem [Helpdesk](#) serves as a central hub for community requests. It provides efficient support for all NFDI4Chem services and RDM topics. Basic issues and common questions are handled by first-level support, while complex cases are handled by specialised second-level teams of the corresponding services. The Helpdesk streamlines communication with our user community, collects common queries in the N4C-KB for proactive solutions and is hosted by TIB, supported by teams from JGU, FSU, TuBr, KIT and RWTH.

Appendix

List of outputs from the consortium

The following summary of NFDI4Chem outputs (Fig. 4) and the detailed listing in Table 1 have been limited to listing outputs in the following 15 of the 254 possible FaBiO classes: Computer Program, Conference Paper, Conference Poster, Dataset, Grant Application, Journal Article, Movie, Position Paper, Preprint, Presentation, Report, Repository, Review Article, Website, White Paper. In addition, only outputs where a digital output is available were listed, so workshops and conferences without reports or proceedings are not included. We also excluded blog posts and social media postings, which are summarised in the section on our [Embedding in the chemistry community](#).

Glossary

A glossary of abbreviations used in this report is available in Table 2.

Funding program

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure (NFDI4/1).

Grant title

NFDI4Chem – Chemistry Consortium in the NFDI (Project number 441958208)

Hosting institution

Friedrich Schiller University Jena

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Amelung L, Bodenschatz E, Danabalan R, Demandt É, Depping R, Eberl F, Ebert B, Espinoza S, Fuchsloch S, Goedicke M, Götz B, Hege C, Hennig C, Henzen C, Hunold J, Jansen L, Krieger U, Rodrigues CM, Miller B, Schatlowski N, Schleuß A, Schneide C, Schwarz A, Schwetje T, Seegert J, Trippel T, Wössner E, Zinke W (2023a) Collaborative work in NFDI. Zenodo <https://doi.org/10.5281/ZENODO.8296724>
- Amelung L, Bodenschatz E, Danabalan R, Demandt É, Depping R, Eberl F, Ebert B, Espinoza S, Fuchsloch S, Goedicke M, Götz B, Hege C, Hennig C, Henzen C, Hunold J, Jansen L, Krieger U, Rodrigues CM, Miller B, Nüst D, Schatlowski N, Schleuß A, Schneide C, Schwarz A, Schwetje T, Seegert J, Trippel T, Wössner E, Zinke W (2023b) White Paper: Umgang mit Zielen der BLV als Grundlage für die Strukturevaluation. Zenodo <https://doi.org/10.5281/ZENODO.8191842>
- Amelung L, Anthofer V, Danabalan R, Demandt É, Ebert B, Elschner E, Espinoza S, Eufinger J, Fuchsloch S, Götz B, Henzen C, Hunold J, Idda T, Jansen L, Krieger U, Rodrigues CM, Meister M, Miller B, Pitroff S, Popp C, Pushkina A, Schatlowski N, Schimmler S, Schneide C, Schörner T, Schwetje T, Seegert J, Trippel T, Weber J, Wössner E, Zinke W (2023c) White Paper: Interim Report Reference. Zenodo. <https://doi.org/10.5281/zenodo.7688729>
- Bach F, Binder K, Minella CB, Lutz B, Razum M (2023) NFDI4Chem - Deliverable D3.3.1: Gap analysis report for selected repositories. Zenodo <https://doi.org/10.5281/zenodo.7602101>
- Bein T, Helmes L, Neumair B, Wessjohann L (2016) RADAR Research Data Repository: Abschlussbericht. URL: https://radar.products.fiz-karlsruhe.de/sites/default/files/radar/docs/info/publications/Abschlussbericht_RADAR_DFG-Projekt_Publikation.pdf
- Bernard L, Althenöner R, Böhm F, Diepenbroek M, Ebert B, Fluck J, Herres-Pawlis S, Klinger A, Koepler O, Lorenz S, Mathiak B, Miller B, Pelz P, Rißler-Pipka N, Ritz R, Sax U, Schimmler S, Schörner T, Schrade T, Schubotz M, Sczyrba A, Stein R, Suchodoletz Dv (2022) Base4NFDI - Basic Services for NFDI. Zenodo <https://doi.org/10.5281/ZENODO.8329192>
- Biernacka K, Bierwirth M, Buchholz P, Dolzycka D, Helbig K, Neumann J, Odebrecht C, Wiljes C, Wuttke U (2020) Train-the-Trainer Concept on Research Data Management. Zenodo <https://doi.org/10.5281/ZENODO.4071471>
- Bierwirth M, Glöckner FO, Grimm C, Schimmler S, Boehm F, Busse C, Degkwitz A, Koepler O, Neuroth H (2020) Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. Zenodo <https://doi.org/10.5281/ZENODO.3895208>
- Boehm F, Buchner B, Kipker D, Kuntz A, Petri G, Sax U, Schaar K, von Suchodoletz D, Vettermann O (2021) Sektionskonzept "Ethical, Legal & Social Aspects" (section-ELSA). Zenodo. <https://doi.org/10.5281/zenodo.5646929>

- Brammer JC, Blanke G, Kellner C, Hoffmann A, Herres-Pawlis S, Schatzschneider U (2022) TUCAN: A molecular identifier and descriptor applicable to the whole periodic table from hydrogen to oganesson. *J. Cheminform.* 14 (1). <https://doi.org/10.1186/s13321-022-00640-5>
- Bundesministerium für Bildung und Forschung (BMBF) (2022) Bekanntmachung BMBF - Richtlinie zur Förderung von Projekten zum Aufbau von Datenkompetenzzentren in der Wissenschaft, Bundesanzeiger vom 21.06.2022. URL: <https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2022/06/2022-06-21-Bekanntmachung-Datenkompetenzzentren.html>
- Castro LJ, Fluck J, Arend D, Lange M, Martini D, Neumann S, Schimmeler S, Rebholz-Schuhmann D (2023) Schema.org as a Lightweight Harmonization Approach for NFDI. *Proc Conf Res Data Infrastr* 1 <https://doi.org/10.52825/cordi.v1i.280>
- Chemotion ELN (2021) Chemotion ELN and repository. Youtube. URL: <https://www.youtube.com/watch?v=tZHaP6DW-Dw>
- Chemotion ELN contributors (2022) Chemotion Electronic Lab Notebook (ELN). Zenodo. <https://doi.org/10.5281/zenodo.6797527>
- Conen N, Fuchs M, Hoffmann A, Herres-Pawlis S, Jupke A (2022a) Experimental data to the publication "Taking the next step – Model-based analysis of robust and non-toxic Zn catalysts for the ring opening polymerization of lactide in the polymer melt" - Part 2. RWTH Aachen University <https://doi.org/10.22000/808>
- Conen N, Fuchs M, Hoffmann A, Herres-Pawlis S, Jupke A (2022b) Experimental data to the publication "Taking the next step – Model-based analysis of robust and non-toxic Zn catalysts for the ring opening polymerization of lactide in the polymer melt". RWTH Aachen University <https://doi.org/10.22000/579>
- Conen N, Fuchs M, Hoffmann A, Herres-Pawlis S, Jupke A (2023) Taking the Next Step —Model-Based Analysis of Robust and Non-Toxic Zn Catalysts for the Ring Opening Polymerization of Lactide in the Polymer Melt. *Advanced Sustainable Systems* 7 (2). <https://doi.org/10.1002/advsu.202200359>
- Devaraju A, Huber R (2021) An automated solution for measuring the progress toward FAIR research data. *Patterns (N Y)* 2 (11). <https://doi.org/10.1016/j.patter.2021.100370>
- Docker (2023) Chemotion Docker image. cited Sept. 2023. URL: <https://hub.docker.com/r/ptrxyz/chemotion/tags>
- Ebert B, Fluck J, Glöckner FO, Koepler O, Miller B, Schmitt R, Schrade T, Stegle O, Steinbeck C, von Suchodoletz D, Wagemann K, Knebes J, Kraft S, Seitz-Moskaliuk H, Sure-Vetter Y, Wössner E (2021) NFDI cross-cutting topics workshop report. Zenodo <https://doi.org/10.5281/ZENODO.4593769>
- Fink F, Benjamaa S, Hoffmann A, Herres-Pawlis S (2022a) Chemotion ELN Erklärvideos. Zenodo. <https://doi.org/10.5281/zenodo.6356844>
- Fink F, Hüppe HM, Jung N, Hoffmann A, Herres-Pawlis S (2022b) Sharing is caring: Guidelines for sharing in the electronic laboratory notebook (ELN) chemotion as applied by a synthesis-oriented working group. *Chemistry Methods* <https://doi.org/10.1002/cmtd.202200026>
- Fink F, Hoffmann A, Herres-Pawlis S (2023a) Results of a Three-Year Survey on the Implementation of Research Data Management and the Electronic Laboratory Notebook (ELN) Chemotion in an Advanced Inorganic Lab Course. *ChemRxiv* <https://doi.org/10.26434/chemrxiv-2023-09ljg>

- Fink F, Benjamaa S, Parks N, Hoffmann A, Herres-Pawlis S (2023b) Chemotion ELN Instruction Videos. Zenodo. <https://doi.org/10.5281/ZENODO.7634480>
- Fischer T, Neumann S (2021) 1st Editors4Chem Workshop. <https://www.nfdi4chem.de/2021/11/23/1st-editors4chem-workshop/>. Accessed on: 2024-3-27.
- Fischer TG, Ditte T, Rauh D, Parks NA, Liermann J, Schulze T, Neumann S (2023) NFDI4Chem - Deliverable D4.4.1: 50 Experimental processes and data publications using NFDI4Chem infrastructure and beyond. Zenodo <https://doi.org/10.5281/ZENODO.8137599>
- Fuchs M, Schäfer PM, Wagner W, Krumm I, Walbeck M, Dietrich R, Hoffmann A, Herres-Pawlis S (2023a) Experimental data to the publication "A multitool for circular economy – fast ring-opening polymerization and chemical recycling of (bio)polyesters using a single aliphatic guanidine carboxy zinc catalyst". RWTH Aachen University. <https://doi.org/10.22000/923>
- Fuchs M, Schäfer PM, Wagner W, Krumm I, Walbeck M, Dietrich R, Hoffmann A, Herres-Pawlis S (2023b) A Multitool for Circular Economy: Fast Ring-Opening Polymerization and Chemical Recycling of (Bio)polyesters Using a Single Aliphatic Guanidine Carboxy Zinc Catalyst. ChemSusChem 16 (12). <https://doi.org/10.1002/cssc.202300779>
- Glöckner FO, Diepenbroek M, Felden J, Overmann J, Bonn A, Gemeinholzer B, Güntsch A, König-Ries B, Seeger B, Pollex-Krüger A, Fluck J, Pigeot I, Toralf K, Mühlhaus T, Wolf C, Heinrich U, Steinbeck C, Koepler O, Stegle O, Weimann J, Schörner-Sadenius T, Gutt C, Stahl F, Wagemann K, Schrade T, Schmitt R, Eberl C, Gauterin F, Schultz M, Bernard L (2019) Berlin declaration on NFDI cross-cutting topics. Zenodo <https://doi.org/10.5281/ZENODO.3457212>
- Goldbeck G (2022) Data representation in materials and chemicals based on harmonised domain ontologies. URL: <https://www.rd-alliance.org/data-representation-materials-and-chemicals-based-harmonised-domain-ontologies>
- Heck J, Metz F, Buchenau S, Teubner M, Grimm-Lebsanft B, Spaniol TP, Hoffmann A, Rübhausen MA, Herres-Pawlis S (2022) Manipulating electron transfer - the influence of substituents on novel copper guanidine quinoliny complexes. Chem. Sci. 13 (28): 8274-8288. <https://doi.org/10.1039/d2sc02910c>
- Hermann A, Herres-Pawlis S (2020) C38H68F6N12O7S2Zn. chemotion.net <https://doi.org/10.14272/rvbyfqzwfrudca-wafkzfkjsa-l>
- Hermann A, Hill S, Metz A, Heck J, Hoffmann A, Hartmann L, Herres-Pawlis S (2020) Next Generation of Zinc Bisguanidine Polymerization Catalysts towards Highly Crystalline, Biodegradable Polyesters. Angew. Chem. Int. Ed. 59 (48): 21778-21784. <https://doi.org/10.1002/anie.202008473>
- Hermann A, Becker T, Schäfer MA, Hoffmann A, Herres-Pawlis S (2022) Effective Ligand Design: Zinc Complexes with Guanidine Hydroquinoline Ligands for Fast Lactide Polymerization and Chemical Recycling. ChemSusChem 15 (18). <https://doi.org/10.1002/cssc.202201075>
- Herres-Pawlis S, Liermann JC, Koepler O (2020) Research Data in Chemistry – Results of the first NFDI4Chem Community Survey. Z. Anorg. Allg. Chem. 646 (21): 1748-1757. <https://doi.org/10.1002/zaac.202000339>
- Herres-Pawlis S, Pelz P, Kockmann N, Gläser R, Richter M, Liermann J, Ortmeier J, Heine I, Metzmacher A, Andres A, Münzmay A, Heuer J, Hagener M, Dierkes J, Wiljes C, Lindstädt B (2021) Sektionskonzept Training & Education zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. Zenodo. <https://doi.org/10.5281/zenodo.5619089>

- Herres-Pawlis S, Bach F, Bruno I, Chalk S, Jung N, Liermann J, McEwen L, Neumann S, Steinbeck C, Razum M, Koepler O (2022) Minimum Information Standards in Chemistry: A Call for Better Research Data Management Practices. *Angewandte Chemie International Edition* 61 (51). <https://doi.org/10.1002/anie.202203038>
- Huang P, Lin C (2023) LabIMotion. Zenodo. <https://doi.org/10.5281/ZENODO.8305411>
- Huang Y, Tremouilhac P, Nguyen A, Jung N, Bräse S (2021) ChemSpectra: A Web-based Spectra Editor for Analytical data. *Journal of Cheminformatics*. <https://doi.org/10.1186/s13321-020-00481-0>
- Hunold J, Koepler O, von Maltzan S, Klix M (2023) Activity Update - Report of relevant cross-cutting topics for NFDI4Chem. Zenodo <https://doi.org/10.5281/ZENODO.8334137>
- IUPAC Committee on Publications and Cheminformatics Data Standards (2020) Development of a Standard for FAIR Data Management of Spectroscopic Data. URL: <https://iupac.org/project/2019-031-1-024/>
- IUPAC Committee on Publications and Cheminformatics Data Standards (2022) WorldFAIR Chemistry - 2. Training Cookbook: Digital recipes for managing chemical data. URL: <https://iupac.org/project/2022-028-1-024/>
- Journal Editor (2023) *Angewandte Chemie International Edition* - Notice to Authors - Wiley Online Library. Wiley. [https://doi.org/10.1002/\(ISSN\)1521-3773](https://doi.org/10.1002/(ISSN)1521-3773)
- Jupp S, Burdett T, Leroy C, Parkinson H (2015) A new Ontology Lookup Service at EMBL-EBI. Workshop on Semantic Web Applications and Tools for Life Sciences URL: <https://www.semanticscholar.org/paper/b83bfbfc1f2f08e5b88af5ef65ef2a8687ac4112>
- Koepler O, Schrade T, Neumann S, Stotzka R, Wiljes C, Blümel I, Bracht C, Hamann T, Arndt S, Hunold J (2021) Sektionskonzept Meta(daten), Terminologien und Provenienz zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. Zenodo. <https://doi.org/10.5281/zenodo.5619089>
- Konsortialversammlung, des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. (2022) Stellungnahme der NFDI-Konsortien zu Basisdiensten. Zenodo <https://doi.org/10.5281/ZENODO.6091656>
- Kunze J, Littmann J, Madden L, Scancelli J, Adams C (2018) The BagIt File Packaging Format (V1.0). URL: <https://datatracker.ietf.org/doc/html/rfc8493>
- Minella CB, Bach F, Jolliffe J, Fischer T (2023) Repos4Chem - criteria for acquisition - for suggestion by NFDI4Chem for data providers. Zenodo <https://doi.org/10.5281/zenodo.8199755>
- Neumann S, Bach F, Castro LJ, Fischer T, Hofmann S, Huang P, Jung N, Katabathuni B, Mauz F, Meier R, Nainala VCS, Rayya N, Steinbeck C, Koepler O (2023) Harmonising, Harvesting, and Searching Metadata Across a Repository Federation. *Proc Conf Res Data Infrastr* 1. <https://doi.org/10.52825/cordi.v1i.202>
- NFDI4Chem (2023) NFDI4Chem Addendum to the Code of Conduct. URL: <https://www.nfdi4chem.de/wp-content/uploads/2023/09/NFDI4Chem-Addendum-Verhaltenskodex.pdf>
- NFDI e.V. (2023) Code of Conduct of the association Nationale Forschungsdateninfrastruktur (NFDI) e.V. URL: https://www.nfdi.de/wp-content/uploads/2023/08/NFDI-Verhaltenskodex_NFDI-code-of-conduct.pdf
- Parks NA, Fischer TG, Blankenburg C, Scalfani VF, McEwen LR, Herres-Pawlis S, Neumann S (2023) The current landscape of author guidelines in chemistry through the lens of research data sharing. *J. Macromol. Sci. Part A Pure Appl. Chem.* 95 (4): 439-450. <https://doi.org/10.1515/pac-2022-1001>

- Patiny L, Musallam H, Zasso M, Bolaños A, Kostro D, Wenk M, Ahkrin S, Jeannerat D, Ziegler E, Liermann J, Schloerer N, Wist J (2023) NMRium: ReactJS component to display and process NMR spectra. Zenodo <https://doi.org/10.5281/zenodo.8345822>
- Proteau PJ (2023) 2023 - New NMR Data Requirements and Editor Changes. *J. Nat. Prod.* 86 (4): 653-654. <https://doi.org/10.1021/acs.jnatprod.3c00281>
- Raßpe-Lange L, Hoffmann A, Gertig C, Heck J, Leonhard K, Herres-Pawlis S (2023) Geometrical benchmarking and analysis of redox potentials of copper(I/II) guanidine-quinoline complexes: Comparison of semi-empirical tight-binding and DFT methods and the challenge of describing the entatic state (part III). *J. Comput. Chem.* 44 (3): 319-328. <https://doi.org/10.1002/jcc.26927>
- Rauh D, Blankenburg C, Fischer TG, Jung N, Kuhn S, Schatzschneider U, Schulze T, Neumann S (2022) Data format standards in analytical chemistry. *Pure Appl. Chem.* 94 (6): 725-736. <https://doi.org/10.1515/pac-2021-3101>
- Rettberg N (2023) RDA TIGER is pouncing along. URL: <https://rd-alliance.org/blogs/rda-tiger-pouncing-along%C2%A0.html>
- Starman M (2023) StarmanMartin/EfwBuilder: v1.0. Zenodo <https://doi.org/10.5281/zenodo.8337272>
- Steinbeck C, Koepler O, Bach F, Herres-Pawlis S, Jung N, Liermann J, Neumann S, Razum M, Baldauf C, Biedermann F, Bocklitz T, Boehm F, Broda F, Czodrowski P, Engel T, Hicks M, Kast S, Kettner C, Koch W, Lanza G, Link A, Mata R, Nagel W, Porzel A, Schlörer N, Schulze T, Weinig H, Wenzel W, Wessjohann L, Wulle S (2020) NFDI4Chem - Towards a National Research Data Infrastructure for Chemistry in Germany. *Research Ideas and Outcomes* 6 <https://doi.org/10.3897/rio.6.e55852>
- Strömert P, Hunold J, Koepler O (2022a) 1st Ontologies4Chem Workshop – Ontologies for chemistry. <https://doi.org/10.25798/FRNP-SN04>
- Strömert P, Hunold J, Castro A, Neumann S, Koepler O (2022b) Ontologies4Chem: the landscape of ontologies in chemistry. *Pure Appl. Chem.* <https://doi.org/10.1515/pac-2021-2007>
- Strömert P, Hunold J, Giacomo L, Stosch R (2023) VIBSO - Vibrational Spectroscopy Ontology. URL: <https://github.com/NFDI4Chem/VibrationalSpectroscopyOntology>
- TIB-Leibniz Information Centre for Science and Technology (2022) NFDI4Chem Search Service. URL: <https://search.nfdi4chem.de/>
- TIB-Leibniz Information Centre for Science and Technology (2023a) Terminology Service API - Swagger UI. URL: <https://service.tib.eu/ts4tib/swagger-ui.html>
- TIB-Leibniz Information Centre for Science and Technology (2023b) NFDI4Chem Terminology Service. URL: <https://terminology.nfdi4chem.de/>
- Tremouilhac P, Lin C, Huang P, Huang Y, Nguyen A, Jung N, Bach F, Ulrich R, Neumair B, Streit A, Bräse S (2020) The Repository Chemotion: Infrastructure for Sustainable Research in Chemistry. *Angew. Chem. Int. Ed Engl.* 59 (50): 22771-22778. <https://doi.org/10.1002/anie.202007702>
- Venkata C, Sharma N, Koepler O, Neumann S, Steinbeck C (2023) Ontology Elements. Zenodo. <https://doi.org/10.5281/zenodo.8212134>



Figure 1.

Institutional Workshops on Research Data Management (RDM) and Chemotion Electronic Lab Notebook (ELN) conducted by NFDI4Chem until summer 2023 and still to be held until the end of 2023.

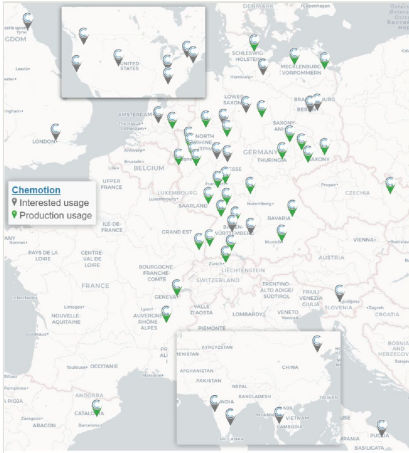


Figure 2. Chemotion ELN Instances in Germany and abroad. Shown are organisations interested or testing Chemotion ELN and organisations known to have rolled out Chemotion ELN instances.

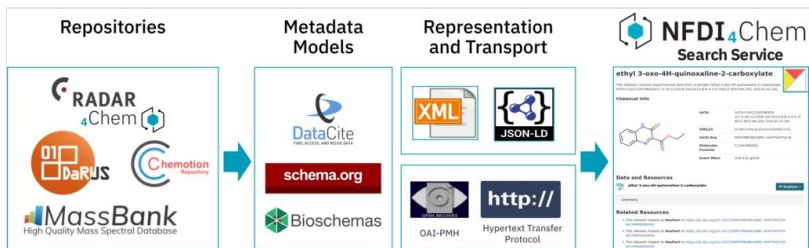


Figure 3.

Metadata architecture in NFDI4Chem. All repositories provide metadata in at least Data-Cite, Schema.org in XML or JSON-LD and are harvested in the central metadata search service.

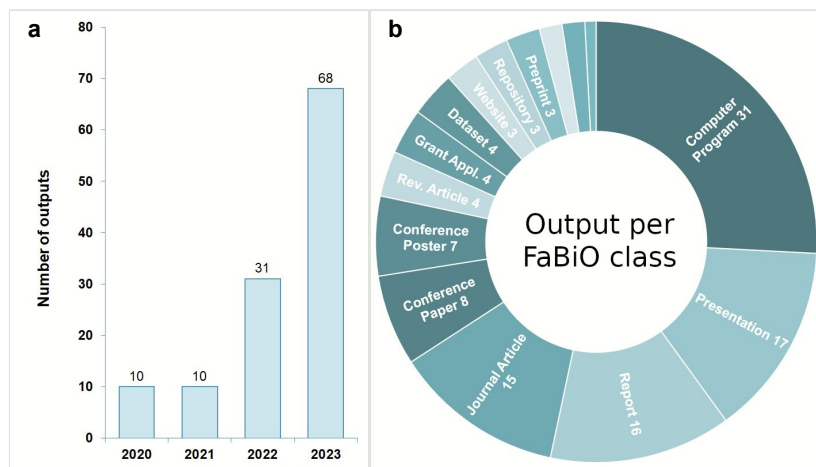


Figure 4.

Number of outputs by NFDI4Chem from 2020 to 09/2023 per year (a) and per FaBiO class (b).

Table 1.

Outputs by NFDI4Chem sorted by FaBiO class (alphabetic order).

| Year | Title | DOI or Link |
|-------------------------|--|---|
| Computer Program | | |
| 2023 | ChemCLI is a tool to help you manage Chemotion ELN on a machine. | github.com/Chemotion/ChemOrc |
| 2023 | ChemConverter app v0.10.0 | doi:10.5281/zenodo.8033808 |
| 2023 | ChemConverter app v1.0.0 (released 03.07.2023) | doi:10.5281/zenodo.8109589 |
| 2023 | cheminformatics-python-microservice v1.0.0 | github.com/Steinbeck-Lab/cheminformatics-python-microservice |
| 2023 | nmrium-react-wrapper v0.1.0 | github.com/NFDI4Chem/nmrium-react-wrapper |
| 2023 | Chemotion ELN Release v1.5.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.5.0 |
| 2023 | Chemotion ELN Release v1.6.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.6.0 |
| 2023 | Chemotion ELN Release v1.7.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.7.0 |
| 2023 | ChemSpectra: Chem Spectra app (26 releases) | github.com/ComPlat/chem-spectra-app |
| 2023 | InChI Webdemo | iupac-inchi.github.io/InChI-Web-Demo/ |
| 2023 | LabIMotion - a Ruby Gem extension to Chemotion ELN | doi.org/10.5281/zenodo.8305412 |
| 2023 | LabIMotion/dataset/cyclic voltammetry | doi.org/10.5281/zenodo.8038708 |
| 2023 | nmrium-react-wrapper v0.1.0 | github.com/NFDI4Chem/nmrium-react-wrapper |
| 2023 | nmrium-react-wrapper v0.2.0 | github.com/NFDI4Chem/nmrium-react-wrapper |
| 2023 | nmrium-react-wrapper v0.3.0 | github.com/NFDI4Chem/nmrium-react-wrapper |
| 2023 | ontology-elements - pre-release | github.com/NFDI4Chem/ontology-elements |
| 2023 | repo-helm-charts | github.com/NFDI4Chem/repo-helm-charts |
| 2023 | Repository downloader | github.com/ComPlat/Repository-Downloader |
| 2023 | Repository tracker | github.com/ComPlat/Repository-Tracker |
| 2023 | Shiny App - Implementation for ELN | github.com/ComPlat/shinychem |
| 2023 | SVG composer: software enabling the composition and rendering of reactions SVGs based on molecule SVGs | github.com/ComPlat/reaction-svg-composer |
| 2023 | Vibrational Spectroscopy Ontology | github.com/NFDI4Chem/VibrationalSpectroscopyOntology |
| 2022 | ChemConverter client | github.com/ComPlat/chemotion-converter-client |

| | | |
|--------------------------|---|---|
| 2022 | Chemotion ELN Release v1.1.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.1.0 |
| 2022 | Chemotion ELN Release v1.2.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.2.0 |
| 2022 | Chemotion ELN Release v1.3.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.3.0 |
| 2022 | Chemotion ELN Release v1.4.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.4.0 |
| 2022 | ChemSpectra: Chem Spectra Client (17 releases) | github.com/ComPlat/chem-spectra-client |
| 2022 | ChemSpectra: react spectra editor (22 releases) | github.com/ComPlat/react-spectra-editor |
| 2022 | nmrXiv - pre-release | github.com/NFDI4Chem/nmrxiv |
| 2022 | TUCAN - a molecular identifier and descriptor for all domains of chemistry | tucan-nest.github.io |
| 2021 | Chemotion ELN Release v1.0.0 | github.com/ComPlat/Chemotion_ELN/tree/v1.0.0 |
| Conference Paper | | |
| 2023 | Digitalizing the Chemical Landscape: A Comprehensive Overview and Progress Report of NFDI4Chem | doi:10.52825/cordi.v1i.213 |
| 2023 | Finding a Common Ground for NFDI Terminologies: Proposing I-ADOPT as a NFDI Wide Semantic Layer | doi:10.52825/CoRDI.v1i.366 |
| 2023 | LabIMotion ElectronicLab Notebook as Research Data Management tool in Catalysis | doi.org/10.52825/CoRDI.v1i.334 |
| 2023 | Leveraging Terminology Services for FAIR Semantic Data Integration across NFDI Domains: How to Integrate Terminology Services Into Other Service Applications | doi:10.52825/CoRDI.v1i.356 |
| 2023 | RADAR: building a FAIR and community tailored Research Data Repository | doi:10.52825/CoRDI.v1i.295 |
| 2023 | RDM in Chemistry: How to Educate and Train Future Researchers to Manage Their Data | doi:10.52825/cordi.v1i.408 |
| 2023 | Schema.org as a Lightweight Harmonization Approach for NFDI | doi:10.52825/cordi.v1i.280 |
| Conference Poster | | |
| 2023 | A Practical Guide to FAIR Research Data Management in Medicinal Chemistry | doi:10.57747/pharmrxiv-2023041040405-000 |
| 2023 | Harmonising, Harvesting, and Searching Metadata across a Repository Federation | doi:10.5281/zenodo.8328199 |
| 2023 | nmrXiv: A FAIR and Open, Consensus-Driven NMR Data Repository and Computational Platform | doi:10.5281/zenodo.6542508 |
| 2023 | PIDs in the Natural Sciences | doi:10.52825/CoRDI.v1i.361 |
| 2023 | www.nmrium.org: Revolutionizing NMR Spectra Processing with a Free Web-Based Application | doi.org/10.5281/zenodo.8171433 |
| 2022 | Metadata, Data Standards and Publication Standards: NFDI4Chem | doi:10.5281/zenodo.6556915 |
| 2022 | MICHI Workshop Series | doi:10.5281/zenodo.8171885 |

| Dataset | | |
|--------------------------|--|---|
| 2023 | Collaborative work in NFDI | doi:10.5281/zenodo.8296724 |
| 2023 | Dataset: Chemotion Repository - Data collection: mass spectrometry data | doi:10.35097/1663 |
| 2023 | Dataset: The current landscape of author guidelines in chemistry through the lens of research data sharing | doi:10.22000/702 |
| 2021 | Collection of SOPs for extensions of chemotion repository and chemotion ELN | github.com/ComPlat/Chemotion-Templates |
| Grant Application | | |
| 2023 | Base4NFDI - Basic Services for NFDI | doi:10.5281/zenodo.8329192 |
| 2022 | LabIMotion4Catalysis (KIT RDM grant 2022) | www.chemotion.net/docs/labimotion |
| 2020 | NFDI4Chem - Towards a National Research Data Infrastructure for Chemistry in Germany | doi:10.3897/rio.6.e55852 |
| 2020 | SciMotion ELN (KIT RDM grant 2020) | fms.ibcs.kit.edu/LabIMotion.php |
| Journal Article | | |
| 2023 | Integrative analysis of multimodal mass spectrometry data in MZmine 3 | doi:10.1038/s41587-023-01690-2 |
| 2022 | Minimum Information Standards in Chemistry: A Call for Better Research Data Management Practices | doi:10.1002/anie.202203038 |
| 2022 | SELFIES and the future of molecular string representations | doi:10.1016/j.patter.2022.100588 |
| 2022 | Sharing is Caring: Guidelines for Sharing in the Electronic Laboratory Notebook (ELN) Chemotion as applied by a Synthesis-oriented Working Group | doi:10.1002/cmt.202200026 |
| 2022 | Treatment of research data | doi:10.1002/nadc.20224131398 |
| 2022 | TUCAN: A molecular identifier and descriptor applicable to the whole periodic table from hydrogen to oganesson | doi:10.1186/s13321-022-00640-5 |
| 2021 | Den Datenschatz endlich heben | doi:10.1002/nadc.20214117508 |
| 2021 | FAIR and Open Data in Science: The Opportunity for IUPAC | doi:10.1515/ci-2021-0304 |
| 2021 | NFDI4Chem – Fachkonsortium für die Chemie | doi:10.17192/bfdm.2021.2.8340 |
| 2021 | NFDI4Chem – Infrastruktur für den digitalen Wandel in der Chemischen Forschung | doi:10.26125/r978-6f93 |
| 2020 | Chemotion Repository, a Curated Repository for Reaction Information and Analytical Data | doi:10.1002/cmt.202000034 |
| 2020 | Comparability of Raman Spectroscopic Configurations: A Large Scale Cross-Laboratory Study | doi:10.1021/acs.analchem.0c02696 |
| 2020 | Forschungsdatenmanagement - Zeit für den Abschied vom analogen Laborbuch | doi:10.1002/nadc.20204095910 |
| 2020 | Research Data in Chemistry - Results of the first NFDI4Chem Community Survey | doi:10.1002/zaac.202000339 |
| 2020 | The Repository Chemotion: Infrastructure for Sustainable Research in Chemistry | doi:10.1002/anie.202007702 |
| Movie | | |
| 2023 | Chemotion ELN Instruction Videos | doi:10.5281/zenodo.7634481 |

| | | |
|-----------------------|---|---|
| 2022 | Chemotion ELN Erklärvideos | doi:10.5281/zenodo.6356844 |
| Position Paper | | |
| 2020 | Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung | doi:10.5281/zenodo.3895209 |
| Preprint | | |
| 2023 | Cheminformatics Python Microservice (CPM): unifying access to open cheminformatics toolkits | doi:10.26434/chemrxiv-2023-hk8zn |
| 2023 | Results of a Three-Year Survey on the Implementation of Research Data Management and the Electronic Laboratory Notebook (ELN) Chemotion in an Advanced Inorganic Lab Course | doi.org/10.26434/chemrxiv-2023-09ljg |
| 2023 | Supporting Sustainability of Chemistry by Linking Research Data with Physically Preserved Research Materials | doi:10.26434/chemrxiv-2023-2dd4c |
| Presentation | | |
| 2023 | FAIR Research Data Management: Basics for Chemists | doi:10.5281/zenodo.8238499 |
| 2023 | HeFDI Data Talk "Chemotion. An Introduction to an Open-Source ELN for FAIR Data" | doi:10.5281/zenodo.8307691 |
| 2023 | HeFDI Data Week 2023: Chemotion. An Introduction to an Open-Source ELN for FAIR Data | doi:10.5281/zenodo.8252090 |
| 2023 | NFDI4C* Workshop on synergy & cooperation | doi:10.5281/zenodo.7839663 |
| 2023 | Overview of Research Data Management in Chemistry | doi:10.5281/zenodo.7767144 |
| 2023 | Schema.org as a Lightweight Harmonization Approach for NFDI | doi:10.5281/zenodo.8331237 |
| 2023 | Setting up your own ODK ontology repository | doi:10.5281/zenodo.7623877 |
| 2023 | NFDI4Chem bei der SaxFDM Digital Kitchen am 11.05.2023 | doi:10.5281/zenodo.7961306 |
| 2023 | NFDI4Chem: from chemical research data management to digital chemistry | doi:10.5281/zenodo.8340453 |
| 2022 | Breakout Session II: Hands on Data Annotation using Ontologies - Creating a prototype knowledge graph from NMR spectroscopy research data | doi:10.5281/zenodo.7050762 |
| 2022 | Chemotion ELN and Chemotion Repository as tools for the digitalization in chemical research within the framework of NFDI4Chem | doi:10.5281/zenodo.6772579 |
| 2022 | Chemotion & Research Data Infrastructure NFDI4Chem | doi:10.5281/zenodo.6985033 |
| 2022 | NFDI4Chem Knowledge Base | doi:10.5281/zenodo.6685262 |
| 2022 | NFDI4Chem Terminology Service: Enabling semantic research data interoperability, discovery and exploitation in chemistry | doi:10.5281/zenodo.6006729 |
| 2022 | Ontologies4Chem: Current chemical ontologies 4 research data management | doi:10.5281/zenodo.7049723 |
| 2021 | NFDI4Chem - Digitising Research Workflows in Chemistry | doi:10.5281/zenodo.5764092 |
| Report | | |
| 2023 | 50 Experimental processes and data publications using NFDI4Chem infrastructure | doi:10.5281/zenodo.8137599 |

| | | |
|-----------------------|---|---|
| 2023 | Accessible Documentation on NFDI4Chem portal | doi:10.5281/zenodo.8246684 |
| 2023 | Analysis of the Landscape of Repositories for Chemistry in re3data | doi:10.5281/zenodo.8347993 |
| 2023 | Continuously updated protocols and minutes of consortium and TA meetings | doi:10.5281/zenodo.8228231 |
| 2023 | Gap analysis report for selected repositories | doi:10.5281/zenodo.7602102 |
| 2023 | Minutes of Advisory Board meetings published on portal | doi:10.5281/zenodo.8228629 |
| 2023 | NMR Task Force Meeting | nfdi4chem.github.io/workshops/docs/workshops/nmr-michi/overview |
| 2023 | Report of relevant cross-cutting topics for NFDI4Chem | doi:10.5281/zenodo.8334137 |
| 2023 | Report on FAIRness of data standards and datasets published by the community | doi:10.5281/zenodo.8137711 |
| 2023 | Repos4Chem - criteria for acquisition - for suggestion by NFDI4Chem for data providers | doi:10.5281/zenodo.8199754 |
| 2023 | The NFDI4Chem portal | doi:10.5281/zenodo.8228654 |
| 2022 | Data Formats | nfdi4chem.github.io/workshops/docs/workshops/standard-formats/overview |
| 2022 | FAIR NMR Research Data Management | nfdi4chem.github.io/workshops/docs/workshops/fair-nmr/overview |
| 2022 | Minimum Information Standards in Polymer Chemistry | nfdi4chem.github.io/workshops/docs/workshops/polymer/overview |
| 2022 | NFDI4Trackact (Organised by Daphne4NFDI) | www.daphne4nfdi.de/downloads/10272022_DAPHNE_TrackACT.pdf |
| 2021 | NFDI Cross-cutting Topics Workshop Report | doi:10.5281/zenodo.4593770 |
| Repository | | |
| 2023 | Chemotion Repository Release v1.1.0 (released on 12.06.2023) | doi:10.5281/zenodo.8028033 |
| 2023 | Chemotion Repository Release v1.2.0 (released on 29.06.2023) | doi:10.5281/zenodo.8093570 |
| 2020 | Chemotion repository | chemotion-repository.net/welcome |
| Review Article | | |
| 2023 | The current landscape of author guidelines in chemistry through the lens of research data sharing | doi:10.1515/pac-2022-1001 |
| 2023 | The Impact of Digitalized Data Management on Material System Workflows | doi:10.1002/adfm.202303615 |
| 2022 | Data format standards in analytical chemistry | doi:10.1515/pac-2021-3101 |
| 2022 | Ontologies4Chem: the landscape of ontologies in chemistry | doi:10.1515/pac-2021-2007 |
| Website | | |
| 2021 | Chemotionsaurus: Dokumentation for Chemotion ELN and Chemotion repository | github.com/ComPlat/chemotion_saurus |
| 2021 | Knowledge Base | knowledgebase.nfdi4chem.de/knowledge_base/ |
| 2020 | NFDI4Chem website | www.nfdi4chem.de |

| White Paper | | |
|--------------------|---|---|
| 2023 | Interim Report Reference | doi.org/10.5281/zenodo.7688728 |
| 2023 | Umgang mit Zielen der BLV als Grundlage für die Strukturrevaluation | doi.org/10.5281/zenodo.8191842 |

Table 2.
Index of Abbreviations

| | |
|---------|--|
| ACS | American Chemical Society |
| API | Application Programming Interface |
| BMBF | Bundesministerium für Bildung und Forschung |
| CCDC | Cambridge Crystallographic Data Centre |
| CDK | Cloud Development Kit |
| CHEMINF | Chemical Information Ontology |
| CHMO | Chemical Methods Ontology |
| CINF | Division of Chemical Information |
| CODATA | Committee on Data of the International Science Council |
| CPU | Central Processing Unit |
| CRDIG | Chemistry Research Data Interest Group |
| CSD | Cambridge Structural Database |
| DALIA | Data Literacy Alliance |
| DaRUS | Data Repository of the University of Stuttgart |
| DB | Data Base |
| DNS | Domain Name System |
| DOI | Digital Object Identifier |
| ECTN | European Chemistry Thematic Network |
| EDAM | Bioinformatics operations, data types, formats, identifiers and topics |
| ELIXIR | European life sciences infrastructure |
| ELN | Electronic Laboratory Notebook |
| EOSC | European Open Science Cloud |
| EPR | Electron Paramagnetic Resonance |
| EuChemS | European Chemical Society |
| EYCN | European Young Chemist Network |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| FID | Forschungsinformationsdienste |
| FTE | Full-Time Equivalent |
| GND | Integrated Authority File |
| HPC | High-Performance Computing |
| HTTPS | Hypertext Transfer Protocol Secure |
| IAM | Identity and Access Management |
| ICSD | Inorganic Crystal Structure Database |
| InChI | International Chemical Identifier |
| IR | Infrared |

| | |
|-----------|---|
| IUPAC | International Union of Pure and Applied Chemistry |
| IYCN | International Younger Chemist Network |
| JCF | JungChemikerForum |
| JS | JavaScript |
| JSON-LD | JavaScript Object Notation for Linked Data |
| MI | Minimum Information |
| MICHI | Minimum Information for Chemical Investigations |
| MOP | Molecular Process Ontology |
| N4C-KB | NFDI4Chem Knowledge Base |
| NFDI | Nationale Forschungsdateninfrastruktur |
| NMR | Nuclear Magnetic Resonance |
| OAI-PMH | Open Archives Initiative Protocol for Metadata Harvesting |
| OBO | Open Biological and Biomedical Ontologies |
| OLS | Ontology Lookup Service |
| ORCID | Open Researcher and Contributor Identifier |
| OS | Open Source |
| PM | Personenmonat |
| PSDI | Physical Sciences Data Infrastructure |
| Q&A | Questions and Answers |
| RDA | Research Data Alliance |
| RDM | Research Data Management |
| REST | Representational State Transfer |
| ROR | Research Organisation Registry |
| RSC | Royal Society for Chemistry |
| RXNO | Name Reaction Ontology |
| SC | Steering Committee |
| SMILES | Simplified Molecular-Input Line-Entry System |
| STRENDATA | Standards for Reporting Enzymology Data |
| TA | Task Area |
| TS | Terminology Service |
| UI | User Interface |
| UV-Vis | Ultraviolet-visible |
| VIBSO | Vibrational Spectroscopy Ontology |