Advancing the Federated Repositories in NFDI4Chem

Christian Bonatto Minella, Felix Bach
FIZ Karlsruhe – Leibniz Institute for Information Infrastructure
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen, Germany

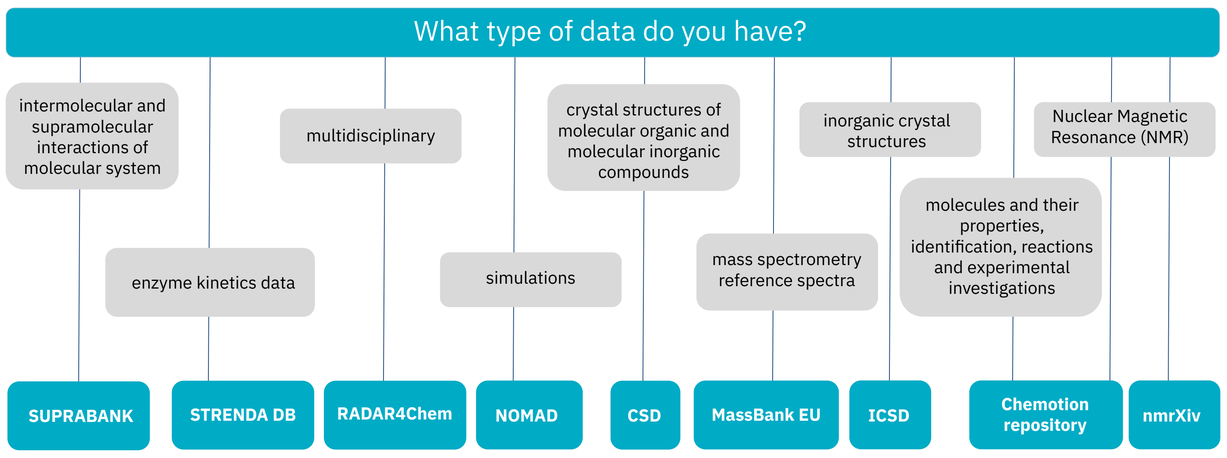
NFDI4Chem
ENHANCE YOUR DATA.
Find out more at www.nfdi4chem.de

nfdi Nationale Forschungsdaten Infrastruktur

2020 → 2025

Activities of Task Area 3 (Repositories)

- Enhances a federated system of chemistry repositories
- Ensures trusted, interoperable data exchange
- Supports publication of high-quality, open-access FAIR data
- Continuously evaluates new chemistry repositories
- Provides and maintains an accessible data network for the chemistry community
- Ensures sustainability of the data infrastructure



RADAR4Chem

- Launched in 2022, accepts all data types and formats
- FAIRness Assessment Challenge: enhanced **F-UJI Scores** up to 87%
- Unlimited retention period (minimum **25 years** / bitstream preservation)
- **DOI** registration (DataCite) incl. DOI reservation option
- Data upload via browser (https), via **WebDAV** protocol and via RADAR API
- RADAR metadata schema based on **DataCite** metadata schema (Kernel 4.4)
- Metadata **indexing** (RADAR, DataCite, harvestable via OAI-PMH)
- Metadata formats: addition of **JSON** and **Schema.org** (JSON-LD)
- **Authentication** after self registration or via single-sign-on (SSO) with **DFN-AAI**
- RADAR knowledge graph and support of **RADAR SPARQL** endpoint
- Chemotion ELN -> RADAR4Chem: data transfer and publication possible
- **Embargo** options: many enhancements for more comfort and flexibility
- **Large volume** data: resumable data downloads after interruptions
- **Recommended** by Angewandte Chemie - International Edition
- Soon available for EU countries, Switzerland and UK

Chemotion repository

- Chemotion ELN -> Chemotion repository: data transfer and publication possible
- **Repotracker** software to record and monitor data transfer processes
- **OAI-PMH endpoint** for metadata harvesting by other services
- Provides **DOIs** registered through DataCite
- Specific metadata using the **JSON-LD** based **Schema.org**
- One-Click Metadata and Data Downloads
- Global **Label** and Review Label to standardise data classification and curation
- Repository **search tool** and **structure search** function
- **Embargo** function that enables controlled access before public release
- Configured and enabled Matomo for Chemotion Repository **analytics**
- Ensures high-quality data management through **data curation** workflow
- **Recommended** by Angewandte Chemie - International Edition and further Wiley journals

nmrXiv

- Launched in 2023
- Ready-to-use **NMR data** management platform to host your raw instrument data and processed files for free and provides you with tools and services to analyse data
- **Recommended** by Angewandte Chemie and Journal of Natural Products as the community-trusted repository for NMR data deposition
- Open-source NMR spectra processing tool - **NMRium integration**
- Publication friendly with spectra level **DOIs**, **embargo** mode and reviewer's access links
- Metadata exports in various formats (e.g. **JSON-LD** based on **Schema.org**)

The Federated Repositories

Core repositories:

- Chemotion
- RADAR4Chem
- nmrXiv
- MassBank EU
- Suprabank
- STRENDA DB

Associated repositories:

- CSD and ICSD

Other relevant repositories:

- NOMAD

Further repositories (testing phase):

- VibSpecDB

FAIRification of the federation

- **(F)** Rich, machine-readable (meta)data linked to domain-specific and cross-domain vocabularies
- **(F)** Registration of datasets with **DataCite** and assignment of globally unique and persistent **DOIs**
- **(F)** Optimisation of repositories dataset landing pages for human use
- **(F)** RADAR4Chem, Chemotion repository, MassBank and nmrXiv adopt unified metadata via JSON-LD, following **Schema.org** and relevant chemistry types from **Bioschemas** => **(F)** through **Google Dataset Search** and **NFDI4Chem Search Service**
- **(A)** All data provided are **retrievable by their persistent identifier** using HTTPS as a standardised communication protocol
- **(A)** All components are available under **open access** models (in some cases registration and login are required)
- **(A) Authentication** protocols and identity providers (e.g. OpenID and Shibboleth)
- **(A)** Supporting the **OAI-PMH** protocol (FIZ offers an OAI-PMH provider)
- **(I) MiChI**s: Development of standards and solutions to address gaps within the chemistry community (e.g. MI metadata standards to semantically describe experiments, simulations, molecule characterisations)
- **(I)** Promoting **open data** formats, their extension to cover currently unestablished data types and their implementation
- **(I) Automating** standardised data and metadata collection for interoperability
- **(I)** Usage of community-accepted **toolkits** (RDKit, CDK, OpenBabel) and established **ontologies** (CHMO, RXNO) to ensure compatibility
- **(I)** Automated data conversion for data storage in **jcamp-dx** (IUPAC standard for spectroscopy)
- **(I) Data editors** to ensure standardised, readable formats for interoperability and reuse across systems
- **(R)** Using openly licensed and well-maintained **ontologies** (CHMO, RXNO, and own development (e.g. VIBSO) to ease humans and machines readability
- **(R)** Integrating metadata annotation standards e.g. **ROR** ID (Research Organisation Registry) and **GND** (Integrated Authority File)
- **(R)** Clear and accessible usage **data licences**, supported by legal policies and guidelines
- **(R) Curating** data and metadata (the level of curation varying according to their role in the federation)
- **(R)** Facilitating **data reuse** by providing targeted datasets for machine learning and other support

In the next funding period

- Optimising and consolidating existing repositories
- Improving user interfaces for better usability
- Enhancing metadata quality and standardisation
- Supporting interoperability across repositories
- Onboarding new repositories into the network
- Developing sustainable operating models for long-term maintenance
- Expanding the federation of chemistry repositories
- Enabling seamless data reuse within the community
- Integrating AI tools for advanced data processing
- Ensuring long-term accessibility of chemistry data and repositories

linkedin.com/company/nfdi4chem
instagram.com/nfdi4chem/

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation
NFDI4Chem is supported by DFG Under project number 441958208