NFDI4Chem
ENHANCE
YOUR
DATA.

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

NFDI4Chem
Consortium Meeting 4.0 2023

Normannenhaus Jena
22/23 November 2023

Contents

Sponsors

This conference would not be possible without the generous support of these sponsors:

Schedule

Wednesday, 22.11.2023

Day 1 : Analysis - Where are we now?

12:00	*Registration & Light Lunch*
13:00 - 13:10	**Welcome** Christoph Steinbeck, Oliver Koepler
13:10 - 14:05	**Consortium Summary of 2023 -Part I** Upcoming Dates & Duties (10 min) Chemotion ELN (15 min) Repositories (15 min) Metadata (15 min)
14:05	*Group Photo*
14:20 - 15:30	**Consortium Summary of 2023 -Part II** Community Engagement (20 min) Cross-Cutting Topics & Ontologies (15 min) Interconnections (20 min) General Discussion about Summaries (15 min)
15:30 -16:00	*Coffee break*
16:00 - 17:05	**Outside the box: Perspectives on the NFDI4Chem** A Data Repository Perspective (Jürgen Harter) (20 min) A Publisher Perspective (Wendy Pattersonn) (15 min) An Industry Perspective (Joachim Richert) (15 min) A (bench) Chemist Perspective (Ulrich Schatzschneider) (15 min)
17:05 - 18:30	**Poster Session, software demo, mingling** Feedback & Wishlist - Board: Identify challenges & potentials of NFDI4Chem, analyse previous work
18:30 - 18:45	*Walk to the Restaurant (Schwarzer Bär)*
19:00	**Conference Dinner (Schwarzer Bär)** Food paid by NFDI4Chem; Drinks self-paid

Thursday, 23 November 2023, Morning

Day 2: Ambition - Where do we want to go?

08:30	*Event-Admission/ Registration*
9:00 - 9:30	**Impulse Talk: Chemistry meets AI (Christoph Steinbeck)**
9:30 - 11:00	**The Future of Chemistry Data: Challenges and Potentials**
	Workshop with all Consortium Participants
	Opening statement with recap of yesterday's progress report and feedback
	Group Discussions (4 groups) Work
11:00 - 11:30	*Brunch break*
11:30 - 12:00	Results from the Workshop:
	Short Presentation 5- 10 min per Group
12:00 - 13:15	**Panel Discussion with the Advisory Boards and the Steering Committee**
	-Assessment of the current Progress
	-Identification of realistic and relevant Goals to be achieved in the second Funding phase
13:15 - 13:30	Final Statement (Christoph Steinbeck, Oliver Koepler)
13:30	**Close of the meeting**

Thursday, 23 November 2023, Afternoon

Optional Workshops

14:30 - 16:30 Workshop 1:

Metadata and Standards across all technical Task Areas (TA15)

(moderation: Oliver Koepler)

Bringing together developers and experts from Data Repositories, ELNs, Tools, Standards and Terminologies the workshops aims to identify commonalities in the development and implementation of data and metadata standards. We seek to examine issues related to data annotation, provision, as well as their reusability by downstream services. Furthermore, we will explore the use of terminologies within metadata schemas.

14:30 - 16:30 Workshop 2:

Knowledge Base, Social Media, Newsletter - What we do and how you can contribute

(moderation: Johannes Liermann, Theo Bender)

An important part of the NFDI4Chem strategy is the continuous communication with the chemical community. For this, we established several services and channels. The Knowledge Base, our Social Media channels, and our newsletter depend on input from the whole consortium. If you are interested in contributing, join us at this workshop and learn about the different possibilities for providing content.

Abstracts

1 Flexible and adaptable training program for research data management in chemistry

Ann-Christin Andres, Benjamin Golub, Daniela Hausen, Pascal Scherreiks, Sonja Herres-Pawlis, Johannes Liermann

In this poster, we would like to present the further development and adaptation of the chemistry-specific research data management (RDM) workshops. Through the numerous workshops since 2022 and the feedback from participants, we know where the community would like more support and which topics are relevant, such as: How to help chemists to get started with FAIR data? The demand for institutional workshops, including for larger groups of up to 100 participants, has increased significantly. Participants also would like to present best practices from their own research during the workshops. These developments require the continuous adaptation of the concept and the development of other workshop formats.

2 New Features in Chemotion ELN: Enhanced Data Search, Spectra Viewer, Inventory Data Management and Versioning

Adam Basha, Lan B. Q. Le and Mehreen Mansur

The Chemotion electronic lab notebook (ELN) is well known for being an open-source, web-based application designed to provide FAIR solutions for digitalizing chemical data, with a primary focus on organic chemistry. This work presents the recent updates that have been incorporated along with upcoming features nearing completion. The upgraded spectra editor now enables the editing and viewing of more data file formats. The new search feature in Chemotion ELN streamlines the retrieval of information, optimizing the user experience. Researchers can quickly locate specific data, experiments, or relevant content within the ELN, optimizing the research workflow. The inventory data management feature within Chemotion ELN facilitates the storage and administration of chemical inventory data. This feature provides users with a robust tool to promptly access relevant information, including safety data sheets and safety phrases, directly from designated vendors in real-time, subsequently archiving this information within the ELN. Ongoing developments aim to extend the capability to establish connections between the Chemotion inventory management system and external inventory systems, enabling seamless data transfer between these platforms.
The next version of the ELN will also incorporate a versioning feature, allowing users to track and manage changes made to entries over time. This feature enhances data integrity, facilitating a more organized and transparent record of the evolution of experiments and projects within the electronic lab notebook. These innovative additions to Chemotion ELN significantly contribute to its adaptability and usability across a spectrum of chemical research applications, offering researchers advanced tools for efficient data management and analysis.

3 NFDI4Chem: the federated repositories

Christian Bonatto Minella, Felix Bach, Kerstin Soltau,

e-Research Institute, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen

Within the framework of the German NFDI initiative and in particular the NFDI4Chem consortium, Task 3 (Repositories) aims to establish a federation of repositories for the collection, storage, processing, analysis, publication and re-use of research data as part of its infrastructure. The federation will be realised through the implementation of metadata and application programming interface (API) standards as well as vocabularies and ontologies addressed within the consortium. In order to identify those repositories that could form the nucleus of the envisaged virtual federation, the TA3 team developed a list of selection criteria. A number of relevant repositories were identified. Although some of them do not yet meet all the criteria required for integration into the NFDI4Chem service platform, they already play an important role in the chemistry research community and have committed to work towards the standards set by the consortium. Task Area 3 will support other databases and data repositories on interoperability issues and encourage them to participate in the development of NFDI4Chem standards and interfaces. With the aim of both identifying gaps in coverage in terms of data types or disciplines and addressing these through adaptation or, where necessary, new development, the TA3 team carried out a gap analysis of existing relevant repositories through individual interviews with repository leaders. The interview consisted of a series of questions ranging from general information to metadata standards and ontology, data content, technical information about the Authorisation and Authentication Infrastructure (AAI), API, services and functionality, operating environment, software architecture and workflows. The interviews were used to determine the current maturity and operational readiness of the selected repositories and to derive appropriate recommendations to address missing requirements. Future activities will focus on community needs to avoid the development of undesired functionalities. The poster will highlight the main characteristics and features of the repositories relevant to the chemistry community that have been selected to implement the NFDI4Chem federation.

4 Advantages of Packaging Scientific Data for Reusability - An Illustration with Electrochemistry Data

Albert K. Engstfeld 1*, Johannes Hermann 1, Nicolas Hörmann 2, and Julian Rüth 3

1 Institute of Electrochemistry, Ulm University, Ulm, 89069 (Germany)
2 Theory Department, Fritz-Haber-Institute, Berlin, 14195 (Germany)
3 Julian Rüth GmbH, Güntersleben, 97261 (Germany)

*E-Mail: * albert.engstfeld@uni-ulm.de*

Storing scientific data according to the FAIR principles[1] requires annotating such data with additional information or metadata. Research data consists, for example, of single values (the temperature at a specific time), time series data (observation of temperature with time), or images, all providing information on the state of a system at a particular time. The additional information on the system, in turn, usually must be stored separately, such as in a laboratory notebook. While text-based descriptions are valuable resources, data-interchange formats, such as JSON, YAML, or XML, are desirable for further data processing by a machine. Annotating research data is usually challenging, especially when many parameters should be stored, such as information on the users, equipment, software, involved materials, or type of measurement. In that sense, automation is highly desirable. This is the topic of this work, where we[2] present solutions to annotate time series data, the automated generation of related metadata, and show use-case examples of such annotated files in the research area of electrochemistry.
To annotate research data, we developed “autotag-metadata”,[3] which observes a local folder for file changes and writes a file with metadata provided in a template YAML file in the same folder. In a first step we developed a simplified metadata schema containing general objects for any research area and objects specific to the electrochemical community.[4] From the raw CSV (datafile) and the YAML (metadata file) one can create a Frictionless datapackage (JSON file).[5] We augmented the standardized frictionless datapackage such that it also contains information on the columns in the CSV, specifically the units. To compared and validated local data with those from scientific publications which are no longer available, we developed “svgdigitizer”, which extracts the data from carefully prepared scalable vector graphics and generates a similar datapackage as for the experimental data.[6]
Both the experimental as well as extracted datapackages can now be used with software solutions provided by Frictionless to do data validation or create static websites.[5] To interact with our datapackages, including units associated with the data in the CSV and additional metadata, we augmented the Frictionless Python API with the “unitpackage” module.[7] It provides simple access to the metadata associated with a CSV resource and allows unit

conversion or visualizing of the underlying data. On the other hand, it can also be used to generate a collection of unitpackages, which can be pictured appealingly on a website.[2] Finally, possible further applications can be discussed, such as the automated generation of electronic lab notebook entries. In total, storing scientific data with associated metadata as datapackages in the local file system or a remote repository provides a highly interoperable approach to working with scientific data in general.

1 https://www.go-fair.org/fair-principles/

2 https://www.echemdb.org,https://github.com/echemdb

3 https://echemdb.github.io/autotag-metadata/

4 https://github.com/echemdb/metadata-schema

5 https://framework.frictionlessdata.io/,https://frictionlessdata.io/

6 A.K. Engstfeld, J. Hermann, N. Hörmann, J.Rüth (2021) svgdigitizer, Zenodo: doi: 10.5281/zenodo.5874747, https://echemdb.github.io/svgdigitizer/

8 A.K.Engstfeld, J.Hermann, N.Hörmann, J.Rüth (2023) unitpackage, doi: 10.5281/zenodo.794320, https://echemdb.github.io/unitpackage/

5 Tracking the Integration of Research Data Management and the Electronic Laboratory Notebook Chemotion into Chemistry Curricula

Fabian Fink[a], Alexander Hoffmann[a], and Sonja Herres-Pawlis[a]*

[a] RWTH Aachen University, Institute of Inorganic Chemistry

At last year's meeting, I presented a poster on the integration of RDM and the Chemotion ELN in an upper-devision undergraduate student lab course at RWTH Aachen University (see title and abstract of last year's registration). This year, I would like to present an updated version of this poster extended with this year's results as used for and presented in the corresponding publication. The title thus changes to "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".

6 Teaching RDM and the Usage of the Chemotion ELN with Training Videos

Fabian Fink[a], Salim Benjamaa[a], Fabian Schüler[a], Nicole Parks[b], Florian Schön[a], Alexander Hoffmann[a], and Sonja Herres-Pawlis[a]*

[a] RWTH Aachen University, Institute of Inorganic Chemistry, 52074 Aachen (Germany)
[b] RWTH Aachen University, IT Center, 52074 Aachen (Germany)

There are many ways to teach RDM and how to use specific software. Training videos are an excellent option because they offer the advantages of easy production, clear visualization, and on-demand availability. Our poster shows our general approach to video production and our recently released training videos. The video production is summarized in a video production (life) cycle and key issues are discussed, starting from conception, through production and release, to improvement and updating. The outcome of this process are chemistry-specific RDM training videos, general instructional videos for the Chemotion ELN to help beginners get started, and application-specific Chemotion ELN videos to meet the needs of specific use cases.

7 Electronic Lab Notebooks in Pharmaceutical Chemistry

Benjamin Golub 1, Christina Draheim 1, Stefan Wulle 1, Christoph W. Grathwol 2

[1] Technische Universität Braunschweig, Universitätsbibliothek, Fachinformationsdienst Pharmazie.
[2] Karlsruhe Institute of Technology, Institute of Biological and Chemical Systems – Functional Molecular Systems (IBCS-FMS), Eggenstein-Leopoldshafen

In this poster, we would like to show which electronic lab notebooks (ELNs) are used in pharmaceutical chemistry. We will go into more detail about the needs of pharmaceutical chemists in an ELN. Furthermore, we will take a look Chemotion ELN and how well it fits these needs.

8 EnzymeML-based workflow for FAIR enzyme kinetics

Max Häußler, Jürgen Pleiss

Institute of Biochemistry and Technical Biochemistry, University of Stuttgart

Data management according to FAIR data principles is an emerging best practice in science.[1] In biocatalysis, the EnzymeML format implements these principles, enabling storing and working with biocatalytical (meta)data.[2] Furthermore, modeling results can be stored together with data on which the results are based on. Together with PyEnzyme, a Python interface for EnzymeML, powerful analysis workflows can be implemented, which exceed the capabilities of analysis in spreadsheet applications. Thus, the EnzymeML infrastructure supports the field by providing structured documentation to experimental data. Hence, enabling the implementation of FAIR analysis pipelines. Here we present an EnzymeML-based workflow for kinetic parameter estimation, enabling a continuous data flow from raw data to kinetic parameters. Thereby, analytical raw data together with calibration data is used for accurate concentration calculation, enabling description of linear and non-linear relationships between the analytical signal and the analyte concentration. After calibration, the respective information is directly applied to the EnzymeML data model containing analytical raw data and yielding an EnzymeML data model with concentration data. Thereafter, concentration data from the EnzymeML data model is mapped to EnzymePynetics, a tool for kinetic parameter estimation based on time-course measurement data. EnzymePynetics enables timecourse analysis of enzymatic reactions, assessing the inhibition constant Ki for potential substrate and product inhibition or an enzyme inhibitor apart from substrate and product. Furthermore, different enzyme inactivation models are applied. After fitting of different model combinations, the best fitting model is selected based on different statistical criteria and thus provides an insight into the kinetic mechanism of the enzymatic catalysis. Finally, the modeling results are written back to the EnzymeML data model and serialized to the standardized EnzymeML format.

1 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9 (2016).

2 Lauterbach, S., Dienhart, H., Range, J. et al. EnzymeML: seamless data flow and modeling of enzymatic data. Nat Methods 20, 400-402 (2023).

9 Research Data Management and Data Publication with Chemotion Repository

Pei-Chi Huang, Chia-Lin Lin

The Chemotion Repository addresses contemporary challenges in research data management, providing effective solutions for storing data and preserving domain-specific information in a machine-readable format. With its comprehensive functionality, the repository offers avenues for collecting, preparing, and reusing data through discipline-specific methods and data processing tools. Automated procedures are implemented for selected analytical data to streamline the data curation process. It facilitates data publication and citation, supporting automated Digital Object Identifier (DOI) generation, comparison with PubChem instances, and workflows for peer review, including embargo settings.

10 Metadata Handling and FAIR Compliance in Analytical Chemistry

Fathia Idiris and Martin Starman

The Chemical Methods Ontology establishes a structured framework for organizing chemistry methods, enhancing communication and data integration. Chemotion utilizes this ontology to improve data representation and establish a standardized language for describing chemical methods, fostering systematic and interoperable approaches in electronic lab notebooks. Simultaneously, LabIMotion Hub serves as a platform for resource sharing and accessing templates contributed by researchers. In various disciplines, where analytical data lacks consistent standards, our outlined workflow transforms diverse cyclic voltammetry data formats into open, standardized data with rich metadata. This comprehensive workflow spans the entire data life cycle, from measurement to publication in repositories.

11 News from the InChI

Frank Lange 1, Gerd Blanke 2,3, Djordje Baljozovic 1, Nauman Khan 1, Jan C. Brammer 1, Sonja Herres-Pawlis 1*

1 Institut für Anorganische Chemie, RWTH Aachen University, Germany
2 InChI Trust, Cambridge, UK
3 StructurePendium Technologies GmbH, Essen, Germany

The InChI (IUPAC International Chemical Identifier) is an algorithmic and unique identifier for chemical compounds that has been widely adopted by science and industry within the last two decades. It serves to connect islands of data on chemical entities making them interoperable and forsters information retrieval in the web with great precision by means of the derived InChIKey. InChI has become a key element of FAIR data handling within cheminformatics as it is the glue between all steps in the research data life cycle of chemical compounds. After its inception, efforts by various working groups started in extending the application of InChI within the different branches of chemistry and to describe and identify other chemical entities such as reactions (RInChI), mixtures (MInChI), polymers and nanomaterials (NInChI).
This poster presentation will provide an overview of the current developments and functionality extensions in InChI's codebase towards the release of version 1.07 and of the previously mentioned extensions within InChI's application framework.
We will also showcase a new web application for the generation of InChIs and RInChIs from drawn chemical structures and reactions and vice versa (https://iupac-inchi.github.io/InChI-Web-Demo).
It opens the InChI ecosystem to new target groups without the burden of installing the InChI software suite or any other (proprietary) software as the application runs on any device with a modern web browser. In particular, we encourage its use for educational purposes. For the audience interested in the technical details behind the application, we will show how to port the InChI library via the WebAssembly technology stack and like to motivate other projects to bring their existing cheminformatics tools to the user's web browser.

12 LabIMotion: A software that provides the option to design new modules that can be adapted to the needs of the scientists

Chia-Lin Lin, Pei-Chi Huang

LabIMotion ELN is a powerful software designed to meet the diverse needs of scientists. This platform extends the capabilities of Chemotion ELN by introducing customizable modules, providing flexibility in utilizing the software with or without chemistry-specific features. Notably, LabIMotion ELN empowers scientists to create tailored modules that seamlessly integrate, streamlining the assignment of materials and samples to various processes and workflows.
In addition to its customizable functionality, LabIMotion ELN enhances collaboration with its user-friendly center. This feature allows scientists to effortlessly share templates across different applications and synchronize them through the LabIMotion Template Hub, promoting efficiency in collaborative research efforts.

13 Device Integration & Data Conversion for Chemotion ELN with Cyclic Voltammetry (CV) as an Example

Tasnim Mehzabin, David Herrmann

In this digital era, scientists face challenges in accessing and managing diverse data, which highlights the necessity for efficient methods. This obstacle can be overcome by adopting particularly electronic lab notebooks (ELNs). By transferring data to a centralized storage server, electronic lab notebooks (ELNs) enable seamless data conversion and visualization across multiple devices such as Gas Chromatography (GC), Size Exclusion Chromatography (SEC), Nuclear Magnetic Resonance (NMR), Infrared Spectroscopy (IR), and potentiostats used for electrochemistry. The data transfer workflow includes a manual process for devices without a network connection in addition to automatic processes like email collection and designated procedures for LAN/WLAN-connected devices. On the other hand, one prevalent issue encountered in data management pertains to the diverse file formats generated by devices and software, like the ones used for Cyclic Voltammetry (CV). While some formats are easily readable, others are proprietary or formatted in unconventional ways. Addressing the FAIR-data principles necessitates obtaining an open, readable output, making .jdx and/or .json formats ideal choices due to their accessibility and popularity in Chemistry. The software ChemConverter provides a useful workflow to meet this objective by transforming data files into open, standardized formats like jcamp-dx. This conversion process not only facilitates accessibility but also aligns with the FAIR-data concept. ChemConverter also extracts and saves useful metadata from the original files, saving it in both .json and .jdx formats, if present. Finally, with LabIMotion's assistance, this converted data can be easily incorporated into the electronic lab notebook, giving researchers a convenient and efficient method for managing and visualizing their data.

14 Ontology Elements: Enabling data interoperability across the web

Venkata Chandrasekhar Nainala 1, N. Sharma 1, H. Musallam 1, Oliver Koepler 2, Steffen Neumann 3, C. Steinbeck 1

1 FSU Jena
2 TIB Hannover
3 IPB Halle

Achieving data interoperability across applications such as Electronic Lab Notebooks, repositories, and other search platforms continues to be a significant challenge in the expanding digital landscape. To address this issue, ontologies and controlled vocabularies provide a formal and explicit specification of concepts and their relationships within a given domain. They establish a shared vocabulary for that area, which includes definitions of terms and the relationships between them.
Ontology Elements (OE) offers a collection of custom, rich ontology-driven HTML elements or web components. They are specifically designed to capture rich annotations that enhance the structure and semantic meaning of biological data. By leveraging the power of ontologies, OE enables a standardised representation of concepts and their relationships within the web interfaces of electronic lab notebooks and repositories. With OE, data interchange and understanding between web applications become more seamless, fostering improved data interoperability on the World Wide Web.
The Ontology Elements store the ontology-driven annotations in the standoff format. However, it deviates in a few ways to facilitate the storage of ontology IRI and classes. This allows entity recognition training/classification to proceed seamlessly without additional processing. When closely integrated with the terminology services via OE, electronic lab notebooks and repositories will enable a better standardised (machine-readable) contextual metadata annotation and create powerful semantically rich knowledge graphs.

[1] https://nfdi4chem.github.io/ontology-elements/

[2] Code: https://github.com/NFDI4Chem/ontology-elements

15 The current landscape of author guidelines in chemistry through the lens of research data sharing

N. A. Parks 1*, T. G. Fischer 2*, C. Blankenburg 2, V. F. Scalfani 3, L. R. McEwen 3, S. Herres-Pawlis 1, S. Neumann 2.

1 RWTH Aachen University, Aachen/Germany (04xfq0f34)
2 Leibniz Institute of Plant Biochemistry, Halle (Saale)/Germany (01mzk5576)
3 Rodgers Library for Science and Engineering, Tuscaloosa/USA (03xrrjk67)
4 Cornell University, Ithaca/USA (05bnh6r87)

E-Mail: parks@itc.rwth-aachen.de
E-Mail: tfischer@ipb-halle.de

NFDI4Chem's vision is the digitalization of all key steps in chemical research to support scientists in their efforts to collect, store, process, analyze, publish, and re-use research data [1]. For that goal, we aim at making chemical research data FAIR (Findable, Accessible, Interoperable, Reusable) [2].
As the primary method of communicating research results, journals and their author guidelines have a tremendous impact on community behavior. To work with scientific journal editors to enhance recommendations on data publication, we organized the Editors4Chem workshop in 2021 [3]. The 2nd Editors4Chem workshop was held in November 2nd, 2023. In addition to these workshops, we surveyed author guidelines from several publishers and journals active in chemistry research. The results not only impart important information for journals and publishers, but also for authors in terms of adjusting to new requirements for research data underlying a journal publication.
With this poster, we will present the results of a large-scale analysis [3] of author guidelines: To which extent is the publishing landscape supporting FAIR Data and Open Science practices? In which areas is this support lacking and what might be underlying reasons? How should authors navigate these changes?

1 C. Steinbeck, O. Koepler, F. Bach, S. Herres-Pawlis, N. Jung, J. Liermann, S. Neumann, et al., Research Ideas and Outcomes, 2020, 6: e55852. DOI: 10.3897/rio.6.e55852. Website: www.nfdi4chem.de..

2 M. Wilkinson, M. Dumontier, I. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. Sci Data 2016, 3, 160018, DOI: 10.1038/sdata.2016.18.

3 T. G. Fischer, 1st Editors4Chem Workshop, www.nfdi4chem.de, 2021, URL: www.nfdi4chem.de/index.php/2021/11/23/1st-editors4chem-workshop/

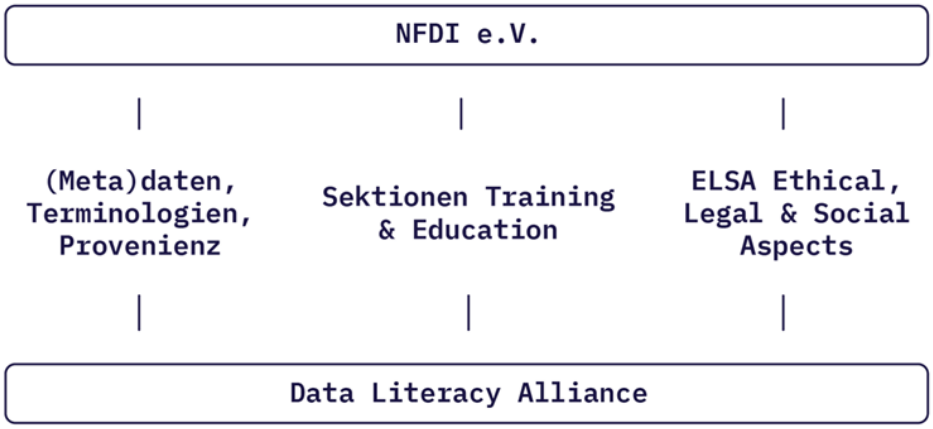
4 N. A. Parks, T. G. Fischer, C. Blankenburg, V. F. Scalfani, L. R. McEwen, S. Herres-Pawlis, S. Neumann, Pure Appl. Chem. 2023, ASAP. DOI: 10.1515/pac-2022-1001.

16 DALIA, Data Literacy for all from the very beginning

Prof. Dr.-Ing. Peter F. Pelz 1, Prof. Dr. rer. nat. Sonja Herres-Pawlis 2, Prof. Dr. rer. nat. Matthias Müller 2, Prof. Thorsten Schrade 3, Prof. Dr. Ulrich Sax 4, Prof. Dr. Thomas Stäcker 1, Dr. Gábor Kismihók 5, Dr. phil. Canan Hastik 1

1 Technische Universität Darmstadt
2 RWTH Aachen University
3 Akademie der Wissenschaften und der Literatur | Mainz
4 Universitätsmedizin Göttingen
5 Technische Informationsbibliothek

The NFDI founded a permanent digital knowledge network for research and innovation following the FAIR principles (findable, accessible, interoperable, reusable) and cooperates with international initiatives such as the European Open Science Cloud (EOSC). They cover a wide range of disciplines from culture to natural sciences. These different focal points of the consortia cooperate with the Data Literacy Alliance (DALIA), which is funded by the Federal Ministry of Education and Research (BMBF), via the Training & Education section in order to bring together and make accessible common educational materials to promote data literacy via a semantic knowledge graph. The platform connects educational materials on FAIR data use, improved metadata quality and promotes data culture change. For good user support, a question-answer interaction should be integrated and personalized learning paths should be created.



17 Recommendation Standards for Reporting Liquid-State NMR Experiments of Small Molecules

Noura Rayya 1, Stuart Chalk 2, Tillmann Fischer 3, Robert Hanson 4, Pavel Kessler 5, Stefan Kuhn 6, Johannes Liermann 7, Venkata Chandra Sekhar Nainala 1, Steffen Neumann 3, Luc Patiny 8, Nils Schlörer 1, Nisha Sharma 1, Christoph Steinbeck 1

1. Friedrich-Schiller-Universität, Jena
2. University of North Florida, Jacksonville
3. Leibniz-Institut für Pflanzenbiochemie, Halle
4. St. Olaf College, Northfield
5. Bruker BioSpin, Ettlingen
6. University of Tartu, Tartu
7. Johannes Gutenberg-Universität, Mainz
8. Swiss Federal Institute of Technology Lausanne, Lausanne

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical technique that plays a pivotal role in unraveling the structural and dynamic properties of small molecules. Its applications in various scientific disciplines, from chemistry to biology, are vast and continue to grow. However, despite its widespread usage, there exists a pressing need for standardized reporting practices in the field of small molecule liquid-state NMR experiments. This gap in reporting standards has led to inconsistencies in data interpretation, hindering scientific reproducibility and progress.

NFDI4Chem aims to provide a sustainable research data management infrastructure by applying digitalisation to all key steps in chemistry research. Thereby fundamentally promoting Open Science that adheres to FAIR data principles. To this end, NFDI4Chem aims to develop domain-specific Minimum Information (MI) and reporting standards in a variety of subdisciplines of chemistry. Here we present the initial version of the Standards for Reporting Liquid-State NMR Experiments (Small Molecules). The NMR task force group at NFD4Chem have gone through an extensive review of existing NMR community standards, literature, data reported in repositories, and applications to improve FAIR reporting in collaboration with IUPAC and community members.

The proposed standards cover sample preparation, NMR acquisition, and spectral processing. Emphasis is laid on aiming NMR repositories and the supplementary materials of publications towards the reproducibility of NMR experiments. Ontologies and controlled vocabularies, mainly nmrCV, are utilized to define and specify accepted values, along with contributions to the development of existing controlled vocabulary to cover the needs of the NMR data reporting.

The link to the tabular recommendations: https://docs.google.com/spreadsheets/d/

1MxCceGO3UUAvIn-GWxxgeOUFnR34A3ileqIW3sYgZNU/edit#gid=0

18 nmrXiv: A highly visible and consensus-driven NMR data repository and computational platform

N. Sharma 1, J. Liermann 2, V. Nainala 1, H. Musallam 1, N. Rayya 1, N. Schlörer 1, S. Neumann 3, Tillmann G. Fischer 3, S. Kuhn 4, C. Steinbeck 1.

1 FSU Jena
2 JGU Mainz
3 IPB Halle
4 Tartu/EE

Nuclear magnetic resonance (NMR) allows researchers to obtain rich structural information from the vibrations of the molecules in their natural environment while they are still intact. It is the standard method of determining the absolute chemical structure (chirality and, more commonly, to determine the relative configuration of diastereomers). NMR is also widely used in natural product dereplication, unknown identification and metabolomics [1]. Owing to the importance of NMR data, several databases have been developed to host literature-derived and experimental NMR data of synthetic molecules and natural products over the past few decades. However, most of these databases cannot accept NMR data depositions from community members. They also do not meet the needs of the natural products/metabolomics communities in several ways. That includes closed licenses, lack of experimental data, limited search capabilities and, more importantly, they are not FAIR compliant.
In today's world of data-driven research, it is critical to have easy access, interoperability, and reusability of NMR data. This is essential for advancing scientific knowledge and promoting collaboration among researchers. nmrXiv (pronounced nm-archive) is a data repository and computational platform for Nuclear Magnetic Resonance (NMR) data that is both FAIR and Open and driven by consensus. Its ultimate goal is to accelerate the coordination and sharing of data among researchers by creating a computational platform for managing, sharing, and analyzing NMR data. The cloud-based platform uses open-source code and deployment methods to ensure sustainability. Curation and analysis standards for NMR data and metadata on nmrXiv are non-prescriptive. The standards are developed with input and contributions from the community and are compatible with existing and potential newly developed formats and approaches. nmrXiv also provides users and curators with an intuitive and easy-to-use interface to create, edit, and annotate their NMR data online. By maintaining regular engagement with analytical chemists and adjacent communities through webinars, workshops, tutorials, and community calls, nmrXiv aims to meet the needs of researchers and implements effective incentive mechanisms for data contributions and curation.

1 Nuzillard, Jean-Marc. (2021). Taxonomy-focused Natural Product Databases for Carbon-13 NMR-based Dereplication. 10.20944/preprints202105.0701.v1.

19 On the right track: Demand-oriented development of the research data repository RADAR

Kerstin Soltau, Felix Bach, Stefan Hofmann, Sandra Göller

e-Research Institute, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen

The research data repository RADAR, launched in 2017, supports the secure management, archiving, publication and dissemination of digital research data. The system is operated by FIZ Karlsruhe - Leibniz Institute for Information Infrastructure and is currently used as a generic cloud service by 20 universities and non-university research institutions in Germany. Since its launch, the environment for research data repositories and also the requirements of researchers and using institutions have developed dynamically. Using a timeline, our poster illustrates RADAR's feature set and how the system has responded to these dynamic changes until now. It also presents our future plans. The poster may therefore provide solutions and potential synergies for RDA activities and other service providers.

20 VibSpecDB: a repository for vibrational spectroscopy data

Darina Storozhuk 1,2 Thomas Bocklitz 1,2,3

1 Leibniz Institute of Photonic Technology, Member of Leibniz Health Technologies, Member of the Leibniz Centre for Photonics in Infection Research (LPI)
2 Institute of Physical Chemistry (IPC) and Abbe Center of Photonics (ACP), Friedrich Schiller University Jena, Member of the Leibniz Centre for Photonics in Infection Research (LPI)
3 Institute of Computer Science, Faculty of Mathematics, Physics & Computer Science, University Bayreuth

According to the standards of data management in research, data storage needs to comply with FAIR (findability, accessibility, interoperability, and reusability) principles. To assist researcher with that, different systems were developed and will be created for some types of data [1]. Unfortunately, until now there were no convenient systems for storing and analysing vibrational spectroscopic data like Raman and Infrared spectra according to the FAIR principles.
To fill this gap, we started working on a web-based application with a modern and convenient user interface. It is called VibSpecDB and already has a basic functionality, such as: file upload; user's profile pages; single sign on and registration; admin console; spectral and file tree viewer. Uploaded files are structured into 'Projects' and 'Studies'. User can set up minimal required or custom metadata information and add tags for data traceability. A separate python microservice can parse input data and metadata of some common file formats, which simplifies integration of existing data into VibSpecDB.
To ensure uninterrupted operation of services for a large number of users, we have developed a cloud architecture with an automatic scaling system. Pre-released version of the web application and few supportive servers, such as file storage, cashing system, and database can be automatically deployed and updated on local Kubernetes cluster. The high speed of data transfer between web platform services allows to minimize the time of loading and unloading data, as well as visualizing spectra.
VibSpecDB is under continuous development and a first pre-release version will be presented and demonstrated at the NFDI4Chem Consortium Meeting 4.0.

1 C. Steinbeck, C. et al. 2020-06-26. "NFDI4Chem – Towards a National Research Data Infrastructure for Chemistry in Germany," Research Ideas and Outcomes, vol. 6, p. e55852. doi: 10.3897/rio.6.e55852.

21 The EuroSAMPL blind prediction challenge - A community effort towards best practices in RDM

N. Tielker 1, P. Kibies 1, M. Urban 1, B. Hein-Janke 2, M. Lim 3, J. Gretz 4, R. A. Mata 2, P. Czodrowski 3, S. M. Kast 1

1 TU Dortmund University
2 Georg August University of Göttingen
3 Johannes Gutenberg University Mainz
4 Ruhr University Bochum

Making research data FAIR (Findable, Accessible, Interoperable, and Reusable) is increasingly a requirement for research groups, scientific journals, and funding organizations, and significant progress has been made by taking advantage of the increasing digitalization of research data. However, despite computational chemistry being well-positioned to establish modern and sustainable research data management (RDM) techniques in chemistry, they are not yet used to the fullest extent.
Furthermore, good scientific practice demands that research data is published in a way that makes it reproducible. The reproducibility of computational chemistry data using only the information in a given journal article and its supporting information is vital for other researchers to easily verify and use newly developed methods. For the combination of FAIR data with data reproducibility standards to make RDM even "fairer", we choose the acronym FAIR+R. This includes methods such as the automated or manual annotation of generated research data with relevant author- and domain-specific metadata, persistent storage in suitable repositories accessible to other researchers, such as ioChem-BD or RADAR4Chem, and the transparent and automated analysis of raw data to generate the chemically relevant information which usually encompasses the published research data.
A primary goal of this challenge is to combine a blind prediction community task for molecular properties in the spirit of the SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) challenges [1] with the implementation of such FAIR+R RDM methods. Such a challenge involves experimental measurements done in-house that will be revealed only after the community has submitted predicted values. Participants' data will be automatically processed from submission up to and including the potential publication in a scientific journal. We chose small molecule acidity constants as the prediction goal, covering a wide range of computational methods in order to address diverse RDM requirements. While it will not be possible to fulfill the full FAIR standards in this first iteration of the EuroSAMPL challenge, it will serve as a first step to identify the most common obstacles encountered by different computational methods and take action to overcome them.
The ideal submission would seamlessly connect raw data, such as log files, with information-

bearing derived data, such as molecular energies and final prediction quantities, which are usually supplied as supporting information in scientific publications, by an automated approach. This way the integrity of research data, encompassing both raw and derived data, as well as the conclusions drawn from them can be maximally preserved.
The FAIR+R quality of the submissions will be assessed by peer evaluations upon conclusion of the challenge. This is carried out with a peer system to develop suggestions for metadata, data deposition, workflow reporting and other standards. One particular measure is to strongly encourage all participants in the use of metadata to improve reproducibility. The challenge will be rolled out in November 2023 [2].

1 https://www.samplchallenges.org

2 https://qmbench.net/challenges/eurosampl/ (available at challenge start)