

Bausteine Forschungsdatenmanagement

Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern

NFDI4Chem – Fachkonsortium für die Chemieⁱ

Jochen Ortmeyer Florian Schön Sonja Herres-Pawlis Nicole Jung
Felix Bach Johannes Liermann Steffen Neumann Christian Popp
Matthias Razum Oliver Koepler Christoph Steinbeck

2021

Zitiervorschlag

Ortmeyer, Jochen et al. 2021. NFDI4Chem – Fachkonsortium für die Chemie. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 34-45. DOI: [10.17192/bfdm.2021.2.8340](https://doi.org/10.17192/bfdm.2021.2.8340).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱJochen Ortmeyer (ORCID: [0000-0003-2074-8027](https://orcid.org/0000-0003-2074-8027)), Florian Schön (ORCID: [0000-0001-6197-2741](https://orcid.org/0000-0001-6197-2741)), Sonja Herres-Pawlis (ORCID: [0000-0002-4354-4353](https://orcid.org/0000-0002-4354-4353)), Nicole Jung (ORCID: [0000-0001-9513-2468](https://orcid.org/0000-0001-9513-2468)), Felix Bach (ORCID: [0000-0002-5035-7978](https://orcid.org/0000-0002-5035-7978)), Johannes Liermann (ORCID: [0000-0003-2060-842X](https://orcid.org/0000-0003-2060-842X)), Steffen Neumann (ORCID: [0000-0002-7899-7192](https://orcid.org/0000-0002-7899-7192)), Christian Popp (ORCID: [0000-0003-0691-2927](https://orcid.org/0000-0003-0691-2927)), Matthias Razum (ORCID: [0000-0002-5139-5511](https://orcid.org/0000-0002-5139-5511)), Oliver Koepler (ORCID: [0000-0003-3385-4232](https://orcid.org/0000-0003-3385-4232)), Christoph Steinbeck (ORCID: [0000-0001-6966-0814](https://orcid.org/0000-0001-6966-0814)).

1 Zusammenfassung

Als Fachkonsortium für die Chemie hat sich NFDI4Chem innerhalb der Nationalen Forschungsdateninfrastruktur (NFDI) gebildet. In diesem Beitrag stellt sich das Konsortium kurz vor und legt seine zentralen Ziele und wichtigsten Verbesserungen für das Forschungsdatenmanagement (FDM) in der Chemie sowie die praktischen Herausforderungen dar. Die Vision von NFDI4Chem ist die umfassende Digitalisierung und Vernetzung aller Prozesse im Umgang mit Forschungsdaten in der chemischen Forschung. Beginnend mit der Erzeugung der Daten, über deren Verarbeitung und Analyse bis hin zur Publikation wird eine modulare, vernetzte Infrastruktur aus Software-Tools, elektronischen Laborjournalen und Datenrepositorien entwickelt und bereitgestellt, die Forschende im Laboralltag unterstützt. Die Digitalisierung wird begleitet durch die Entwicklung von Minimalinformationen für Datenpublikationen, bestehend unter anderem aus Standards für Daten- und Metadatenformate sowie Ontologien zur semantischen Beschreibung. Seine Aufgaben verfolgt das NFDI4Chem-Konsortium wissenschaftsgeleitet und mit dem klaren Ziel, eine intuitiv und effizient nutzbare Infrastruktur zu entwickeln. Das Gestalten eines kulturellen Wandels, gemeinsam mit der wissenschaftlichen Community, zur Etablierung und Akzeptanz eines FAIRen Umgangs mit Daten ist daher ein weiteres wichtiges Element der NFDI4Chem-Aktivitäten.

2 Kurzvorstellung des Konsortiums mit dessen zentralen Zielen

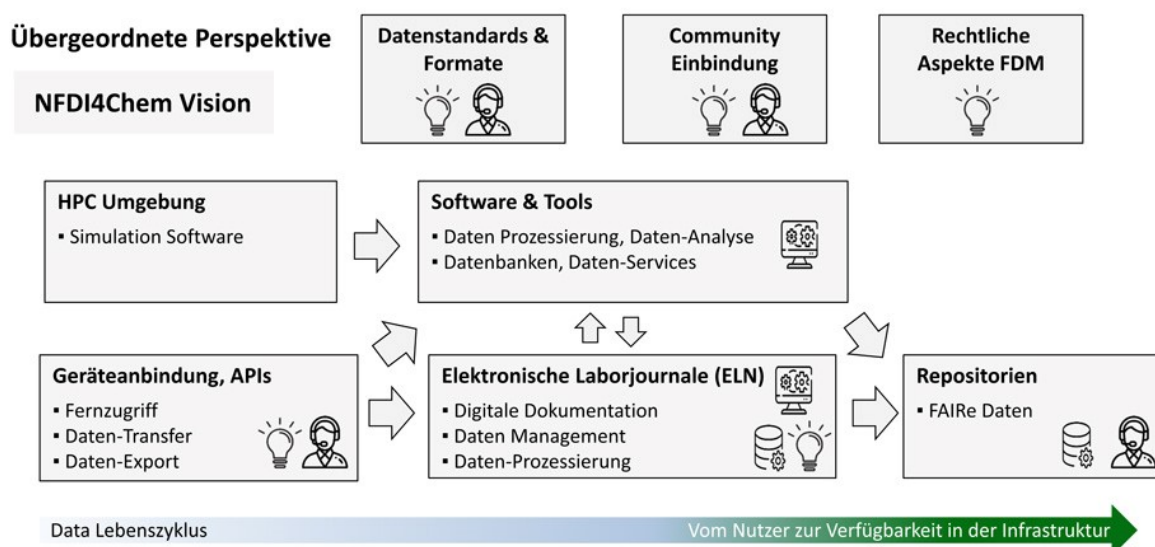
NFDI4Chem vertritt innerhalb der NFDI die Interessen und Bedarfe der Chemie. Hierzu schlossen sich im Jahr 2018 engagierte Vertreterinnen und Vertreter aus universitärer und außeruniversitärer Forschung, Infrastruktureinrichtungen, Rechenzentren und nationalen Fachgesellschaften, wie der Gesellschaft Deutscher Chemiker (GDCh), der Deutschen Bunsen-Gesellschaft (DBG) und der Deutschen Pharmazeutischen Gesellschaft (DPHG), zusammen. Sie repräsentieren damit mehr als 40.000 Mitglieder der Chemie-Gemeinschaft. Aus dieser Initiative bildete sich das Fachkonsortium Chemie NFDI4Chem unter der Leitung von Christoph Steinbeck (Friedrich-Schiller-Universität, Jena) und Oliver Koepler (TIB - Leibniz Informationszentrum Technik und Naturwissenschaften, Hannover). NFDI4Chem wird seit Oktober 2020 von der DFG im Rahmen der Initiative zum Aufbau einer Nationalen Forschungsdateninfrastruktur gefördert.¹ Das Arbeitsprogramm der NFDI4Chem verteilt sich auf die Arbeitsbereiche (**Task Areas**) TA1 Management, TA2 Smart Lab, TA3 Repositorien, TA4 Standards, TA5 Community-Einbindung und Training und TA6 Synergien. Im Sinn eines intensiven Austauschs und einer weitreichenden Vernetzung kooperiert NFDI4Chem bei wichtigen und vor allem

¹Förderantrag: Steinbeck, Christoph, Oliver Koepler, Felix Bach et al. „NFDI4Chem - Towards a National Research Data Infrastructure for Chemistry in Germany“. *Research Ideas and Outcomes* 6 (2020): e55852, <https://doi.org/10.3897/rio.6.e55852>.

fachübergreifenden Themen mit anderen Fachkonsortien aus der NFDI (z. B. mit den bereits geförderten Konsortien NFDI4Cat und NFDI4Ing sowie den vorgeschlagenen Konsortien FAIRmat oder NFDI-MatWerk). Auf internationaler Ebene werden die Aktivitäten zur Standardisierung des FDMs mit wichtigen Gremien, wie der International Union of Pure and Applied Chemistry (IUPAC), dem GO FAIR Chemistry Implementation Network, der European Open Science Cloud (EOSC) oder der Chemistry Research Data Interest Group der Research Data Alliance (RDA), abgestimmt.

Das Arbeitsprogramm von NFDI4Chem unterstützt und fördert FDM und Open Science gemäß der FAIR-Prinzipien (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable)². Das Konsortium will der Chemie-Gemeinschaft ein ganzheitliches Konzept für den Umgang mit Forschungsdaten zur Verfügung stellen. Dabei stehen zunächst Daten zu Molekülen und Reaktionen und Daten zur Molekül- bzw. Materialcharakterisierung im Mittelpunkt, die sowohl experimentell als auch theoretisch erhoben worden sein können. Somit werden Forschende bei der digitalen Erhebung, Speicherung, Verarbeitung, Analyse, Publikation und Wiederverwendung von Forschungsdaten unterstützt und gefördert. Diese Vorgehensweise wird den digitalen und kulturellen Wandel im Umgang mit Forschungsdaten in der Chemie (**Abbildung 1**) vorantreiben.

Digitaler und kultureller Wandel in der Chemie



Perspektive Forschende

Idee – Vorbereitung Experiment – Datengenerierung – Dokumentation Experiment – Publikation Ergebnisse

Abbildung 1: Auf dem Weg ins digitale Zeitalter: Komponenten eines digitalen Forschungsprozesses.

²Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg et al. „The FAIR Guiding Principles for scientific data management and stewardship“. *Scientific Data* 3, Nr. 160018 (2016): <https://doi.org/10.1038/sdata.2016.18>.

Die Digitalisierung der Forschungsprozesse erfolgt durch modulare Software-Lösungen, die über Schnittstellen vernetzt werden und auf offenen Datenstandards beruhen. Ein Schwerpunkt wird die Entwicklung intuitiver Bedienkonzepte sein, um niedrigschwellige Einstiege für Nutzerinnen und Nutzer für die Dienste der NFDI4Chem zu ermöglichen. Benutzungsfreundlichkeit wird ein Kernaspekt für die Akzeptanz und den Erfolg der Dienste und damit der zukünftigen Nachnutzung von Forschungsdaten sein.

Die Schlüsselziele von NFDI4Chem³ sind:

Schlüsselziel 1: Aufbau einer virtuellen Föderation von Datenrepositorien für die Speicherung, Veröffentlichung, Suche und Nachnutzung von Forschungsdaten. Verbindung bestehender Datenrepositorien und – basierend auf einer Bedarfsanalyse – Aufbau eines oder mehrerer domänenspezifischer Forschungsdatenrepositorien für die nationale Forschungsgemeinschaft sowie deren Verknüpfung mit internationalen Repositorien.

Schlüsselziel 2: Einleitung von Prozessen der internationalen Gemeinschaft zur Festlegung von Standards für Minimalinformationen (MI) für Daten und maschinenlesbare Metadaten in Schlüsselbereichen der Chemie. In Fällen, in denen offene Datenstandards vorhanden sind, sollen diese systematisch unterstützt werden, um die Umsetzung der FAIR-Prinzipien für Forschungsdaten zu ermöglichen.

Schlüsselziel 3: Förderung des kulturellen und digitalen Wandels hin zur Nutzung intelligenter Laborumgebungen durch Etablierung digitaler Instrumente in allen Phasen der Forschung. FAIRes Forschungsdatenmanagement soll auf allen Ebenen der Wissenschaft, beginnend mit den Lehrplänen des Bachelorstudiums, eingebracht werden.

Schlüsselziel 4: Zusammenarbeit mit der Chemie-Gemeinschaft in Deutschland durch eine Vielzahl von Maßnahmen zur Sensibilisierung und Förderung der Akzeptanz von FAIRem Datenmanagement. Einleitung von Prozessen zur Integration von FDM und Datenwissenschaften in die Lehrpläne. Schaffung eines breiten Spektrums an Ausbildungsmöglichkeiten für Forschende.

Schlüsselziel 5: Ermittlung von Synergien mit anderen Konsortien und Mitgestaltung von Querschnittsthemen innerhalb der NFDI.

Schlüsselziel 6: Schaffung und Bereitstellung eines rechtssicheren Rahmens von Richtlinien und Empfehlungen für FAIRes und offenes Forschungsdatenmanagement.

Diese Ziele adressiert das NFDI4Chem-Konsortium im intensiven Austausch und unter Beteiligung der wissenschaftlichen Chemie-Gemeinschaft.

³Weiterführende Informationen zur Arbeit des Konsortiums sowie zu Veranstaltungen: <https://nfdi4chem.de/> (letzter Zugriff: 2021-01-31).

3 Was sind die drei wichtigsten Dinge, die sich durch die Arbeit des Konsortiums für die Chemie-Gemeinschaft verändern/verbessern werden?

3.1 Durchgängige Digitalisierung im Labor

Die Umsetzung eines Forschungsprojektes in der Chemie basiert meist auf einer Vielzahl von Einzelexperimenten. Diese beginnen grundsätzlich mit Überlegungen zu bzw. der Planung der jeweiligen Synthese, der entsprechenden Probenvorbereitung oder der theoretischen Berechnung. Dazu wird heutzutage – im digitalen Zeitalter – immer noch mehrheitlich das klassische, handgeschriebene Laborbuch verwendet, obwohl bereits einige elektronische Lösungen existieren. Diese bestehen in der Möglichkeit, frei zugängliche oder kommerzielle elektronische Lab Notebooks (ELN) zu nutzen. Manche Chemikerinnen und Chemiker verstehen bereits unter der Verwendung von Word oder Excel, als eine Art ELN, einen digitalen Fortschritt. Allerdings können diese Office-Programme bei weitem nicht alle Vorteile, wie beispielsweise das Speichern von FAIRen Daten, eines ELNs bieten. In diesem Zusammenhang kam eine kürzlich veröffentlichte Umfrage zum Thema FDM in der Chemie mit 541 Teilnehmenden aller Karrierestufen aus den klassischen Chemie-Fachrichtungen in Deutschland (hauptsächlich Professoren, Postdocs und Doktoranden) zu dem Ergebnis, dass nur 17 Prozent der befragten Personen ihre Daten zu Versuchsdurchführung, Beobachtung etc. in ELNs (Word und Excel zählten in dieser Umfrage ebenfalls zu den ELNs) digital verarbeiten.^{4,5} Dies mag an der eingeschränkten Nutzerfreundlichkeit der bestehenden ELNs liegen. Nicht zu unterschätzen ist jedoch auch die Tatsache, dass eine Vielzahl der Chemikerinnen und Chemiker einen zeitlichen Mehraufwand befürchtet. Um also die generierten Daten während des Forschungsprozesses direkt, schnell und digital erfassen zu können, sind nutzerfreundliche ELNs dringend erforderlich.

Ein großer Vorteil von ELNs ist die Vernetzung aller Aktivitäten im Labor von der Planung über die Dokumentation, das Datenmanagement und die Datenanalyse bis hin zur Datenaufbereitung für Publikationen. So kann neben der eigentlichen Dokumentation der Versuchsdurchführung inklusive der Beobachtung und des Ergebnisses eine Verknüpfung mit erhaltenen analytischen Daten erfolgen. Dabei werden die Resultate des jeweiligen Verfahrens (Messung, quantenchemische Rechnung o. ä.) direkt mit dem Eintrag im ELN verknüpft, sodass die Ergebnisse bei Ansicht unmittelbar zur Verfügung stehen. Dadurch wird nicht nur die sofortige Sicherung und schnelle Wiederauffindbarkeit der Daten für die Einzelperson garantiert, sondern auch ein arbeitsgruppen-internes Archiv generiert. Dieses kann bei Bedarf auch noch Jahre später nach benötigten Versuchen und Ergebnissen durchsucht werden. Somit „ver-

⁴Herres-Pawlis, Sonja, Johannes C. Liermann und Oliver Koepler. „Research Data in Chemistry - Results of the first NFDI4Chem Community Survey“. *Zeitschrift für anorganische und allgemeine Chemie* 646, Nr. 21 (2020): 1748-57, <https://doi.org/10.1002/zaac.202000339>.

⁵Herres-Pawlis, Sonja, Oliver Koepler und Johannes Liermann. „First NFDI4Chem User Survey“. Dataset (2020): <https://doi.org/10.25835/0077933>.

staubt“ das zusammengetragene Wissen (positive und negative Resultate) einer Arbeitsgruppe nicht in der Sammlung handgeschriebener Laborbücher. Es steht vielmehr den nachfolgenden Teammitgliedern zur vollumfänglichen Nutzung und Steigerung ihres Forschungsfortschrittes zur Verfügung. Die gesammelten Daten werden zusätzlich mit maschinenlesbaren Metadaten versehen (dies kann je nach Arbeitsumgebung händisch, semi-automatisch oder automatisch erfolgen), wodurch die Ergebnisse später wiederverwendbar werden. Innerhalb der NFDI4Chem sollen in Zukunft die digitalen Prozesse weiter verbessert werden, indem Standardisierung (siehe Punkt 3.3) gefördert und proprietäre Datenformate soweit als möglich ersetzt werden.⁶ Idealerweise können somit in Zukunft im Falle einer Publikation die Daten ohne technische Einschränkungen direkt aus dem ELN in ein Repositorium übernommen und mit der Publikation verknüpft werden (siehe Punkt 3.2).

Die Einführung und Verbreitung eines solchen multifunktionalen ELNs ist von entscheidender Bedeutung. Dies gilt insbesondere, da Wissenschaftlerinnen und Wissenschaftler mehr und mehr ihre Forschungsdaten nicht nur erfassen und dokumentieren, sondern auch bereitstellen müssen, um die Anforderungen der fördernden Institutionen oder Verlage zu erfüllen.^{7,8} Die Erfassung digitaler Daten ist diesbezüglich sehr wichtig, da die Bereitstellung von Forschungsdaten nicht nur mit dem Ziel der Reproduzierbarkeit, sondern auch der Nachnutzbarkeit erfolgen sollte. Diese Anforderungen erfüllen die von der NFDI4Chem ausgewählten Referenzimplementierungen Chemotion ELN und Repositorium Chemotion in weiten Teilen.^{9,10,11} Deren Weiterentwicklung und Optimierung für eine vollständige Abdeckung der Anforderungen an FAIRes FDM stellen elementare Ziele des NFDI4Chem-Konsortiums dar. Durch gezielte Umfragen (erste bereits erfolgt^{12,13}) innerhalb der Community sollen noch bestehende Herausforderungen systematisch erfasst und gelöst werden. Die modulare Architektur der NFDI4Chem-Infrastruktur ist offen für die Einbindung alternativer ELN Systeme.

⁶Koepler, Oliver, Johannes Liermann, Florian Schön und Sonja Herres-Pawlis. „Forschungsdatenmanagement - Zeit für den Abschied vom analogen Laborbuch“. *Nachrichten aus der Chemie* 68, Nr. 12 (2020): 20-3, <https://doi.org/10.1002/nadc.20204095910>.

⁷Am Beispiel der DFG als fördernde Institution: Deutsche Forschungsgemeinschaft. „Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex.“ „Guidelines for Safeguarding Good Research Practice. Code of Conduct.“ (2019): <http://doi.org/10.5281/zenodo.3923602>.

⁸Am Beispiel der DFG als fördernde Institution: https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html (letzter Zugriff: 2021-01-31).

⁹Tremouilhac, Pierre, Chia-Lin Lin, Pei-Chi Huang et al. „Das Repositorium Chemotion: Infrastruktur für nachhaltige Wissenschaft in der Chemie“. *Angewandte Chemie* 132, Nr. 50 (2020): 22960-8, <https://doi.org/10.1002/ange.202007702>. „The Repository Chemotion: Infrastructure for Sustainable Research in Chemistry“. *Angewandte Chemie International Edition* 59, Nr. 50 (2020): 22771-8, <https://doi.org/10.1002/anie.202007702>.

¹⁰Tremouilhac, Pierre, Pei-Chi Huang, Chia-Lin Lin et al. „Chemotion Repository, a Curated Repository for Reaction Information and Analytical Data“. *Chemistry-Methods* 1, Nr. 1 (2021): 8-11, <https://doi.org/10.1002/cmt.d.202000034>.

¹¹Web-Zugang zu Chemotion: <https://www.chemotion-repository.net/home/welcome> (letzter Zugriff: 2021-01-31).

¹²Herres-Pawlis, Sonja, Johannes C. Liermann und Oliver Koepler. „Research Data in Chemistry“.

¹³Herres-Pawlis, Sonja, Oliver Koepler und Johannes Liermann. „First NFDI4Chem User Survey“.

Regelmäßige Newsletter, Webinare, Videotutorials und ein Stammtisch unterstützen die Anwendenden durch kontinuierliche Fortbildung, um die Einführung und Nutzung von ELNs zur digitalen Datenerfassung in der Chemie bestmöglich zu erreichen. Als NFDI4Chem stehen wir dazu als zentrale Anlaufstelle zur Verfügung.

3.2 Etablierung von Datenrepositorien in der Chemie

Datenrepositorien ermöglichen die Publikation und den nachhaltigen Zugang zu originalen Daten aus dem Forschungsprozess und sind ein wichtiges Element in der Umsetzung der Richtlinien zur guten wissenschaftlichen Praxis^{14,15}. Daten in Repositorien erhöhen die Nachvollziehbarkeit von wissenschaftlichen Publikationen, z. B. in Form einer Zitation oder Verlinkung der Datenpublikation in Zeitschriftenartikeln, in denen auf Grundlage dieser Daten wissenschaftliche Erkenntnisse abgeleitet werden. Für die Umsetzung der FAIR-Prinzipien sind in der Chemie eine Reihe von Herausforderungen zu meistern.

Bei der Charakterisierung von Molekülen wird ein großes Spektrum von experimentellen und theoretischen Methoden, wie z. B. NMR- (Nuclear Magnetic Resonance), IR- (Infrarot) und UV/VIS-Spektroskopie (ultraviolettes und sichtbares (visible) Licht) sowie Massenspektrometrie (MS), High Performance Liquid Chromatography (HPLC), Elektronenmikroskopie, Quanten- und Kraftfeldberechnungen, angewendet, wodurch eine ebenso große Anzahl an heterogenen Datentypen und -formaten erzeugt wird, ein Großteil davon in proprietären Dateiformaten. Zwar gibt es einige offene Datenaustauschformate wie mzML, JCAMP oder AniML, jedoch werden diese nicht durchgängig von den Geräteherstellern unterstützt.

Relevant für die Veröffentlichung von Daten in Repositorien ist ferner die Annotation mit beschreibenden Metadaten. Neben etablierten, generischen Kern-Metadaten sind disziplinspezifische Metadaten zu berücksichtigen, um chemiespezifische Suchmethoden, wie die Struktursuchen, zu ermöglichen. Aktuell fehlen für Datenveröffentlichungen in der Chemie solche Standards für Metadaten.

Bisher ist die breite Verfügbarkeit und Nutzung von Repositorien noch nicht gegeben. Daten werden eher in institutionseigenen Speicherlösungen oder disziplinübergreifenden Datenrepositorien der Universitäten abgelegt. Meist werden jedoch nur Teile von Daten oder abgeleitete Daten innerhalb von Aufsätzen oder als Supplement zur Verfügung gestellt.

Zwar existieren chemiespezifische, frei zugängliche Repositorien (z. B. NMRshiftdb¹⁶) als auch kostenpflichtige Datenbanken (z. B. SciFinder¹⁷), die für eine Datensuche genutzt werden können. Jedoch sind die bereits existierenden Repositorien häufig we-

¹⁴Deutsche Forschungsgemeinschaft. „Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex.“

¹⁵Deutsche Forschungsgemeinschaft. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html (letzter Zugriff: 2021-01-31).

¹⁶<https://nmrshiftdb.nmr.uni-koeln.de/> (letzter Zugriff: 2021-01-31).

¹⁷<https://www.cas.org/products/scifinder> (letzter Zugriff: 2021-01-31).

der ausreichend miteinander verknüpft, noch bekannt genug, um von der gesamten Chemie-Gemeinschaft genutzt zu werden. Ein Problem ist hierbei die Spezialisierung auf lediglich kleine Gebiete der Chemie. Kostenpflichtige Datenbanken entsprechen nicht dem FAIR-Gedanken. Eine Datensuche in diesen Datenbanken verweist meist auf in Fachzeitschriften veröffentlichte Artikel, in denen manuell nach den gewünschten Informationen gesucht werden muss. Oft sind die Informationen über die wichtigen Forschungsprozesse in der Supplemental Information der Artikel angefügt, jedoch lediglich als PDF verfügbar. Um zukünftig in umfassender Weise Forschungsdaten zugänglich machen zu können, sollen Repositorien, die sich die Speicherung von Daten entsprechend der FAIR-Prinzipien zum Ziel setzen, in der Breite in der Chemie-Gemeinschaft etabliert werden. Dies soll durch diverse Maßnahmen erreicht werden. Die Repositorien werden innerhalb der NFDI4Chem gefördert und übergreifend durchsuchbar sein, auch mit spezifischen Suchmethoden der Chemie. Ein Datenportal mit fachspezifischen Suchfunktionen wird hierfür der zentrale Einstiegspunkt sein. Damit bildet die NFDI4Chem eine wichtige, offene Infrastruktur für Datenpublikationen und deren Vernetzung mit anderen Publikationsformen, wie Zeitschriftenartikeln. Sie ist damit ein eminentes Gegengewicht zu kommerziellen Systemen zum Forschungsdatenmanagement, die im Umfeld von, ebenfalls kommerziellen, Publikationsplattformen und bibliographischen Datenbanken existieren.

Mit der bereits gut etablierten Referenzimplementierung von Chemotion als ELN, das den Datenaustausch mit dem komplementären Chemotion Repository ermöglicht, wird die unmittelbare Datenpublikation aus der Smart Lab-Umgebung heraus in Datenrepositorien als integrierte Umgebung aufgezeigt und weiterentwickelt.^{18,19,20} Dies ermöglicht eine durchgehende und benutzerfreundliche Digitalisierung und Vernetzung aller Schritte von der Datengenerierung über die Datenverarbeitung bis zur Datenpublikation mit minimalen Aufwänden für die Forschenden. Über standardisierte Schnittstellen und Metadaten (Minimalstandards) soll dieses Prinzip auch auf weitere Forschungsdatenrepositorien übertragen werden.

Über Chemotion hinaus wurden sechs weitere spezifische Repositorien ausgewählt (Abbildung 2), in denen molekülbezogene Daten abgelegt werden können und die daher für NFDI4Chem bedeutsam sind. Sie sollen in der ersten Förderperiode weiterentwickelt und in der Chemie-Gemeinschaft etabliert werden. Dabei bilden sie den Nukleus eines förderierten Verbunds, der perspektivisch um weitere Repositorien ergänzt werden kann. Die Inhalte der Repositorien werden durch geeignete Schnittstellen und Metadatenstandards einfach auffindbar und verknüpfbar gemacht. Für allgemeinere, weniger chemiespezifische Datentypen werden ferner zwei generische Repositorien eingebunden.

¹⁸Tremouilhac, Pierre et al. „Das Repository Chemotion“.

¹⁹Tremouilhac, Pierre et al. „Chemotion Repository“.

²⁰<https://www.chemotion-repository.net/home/welcome> (letzter Zugriff: 2021-01-31).

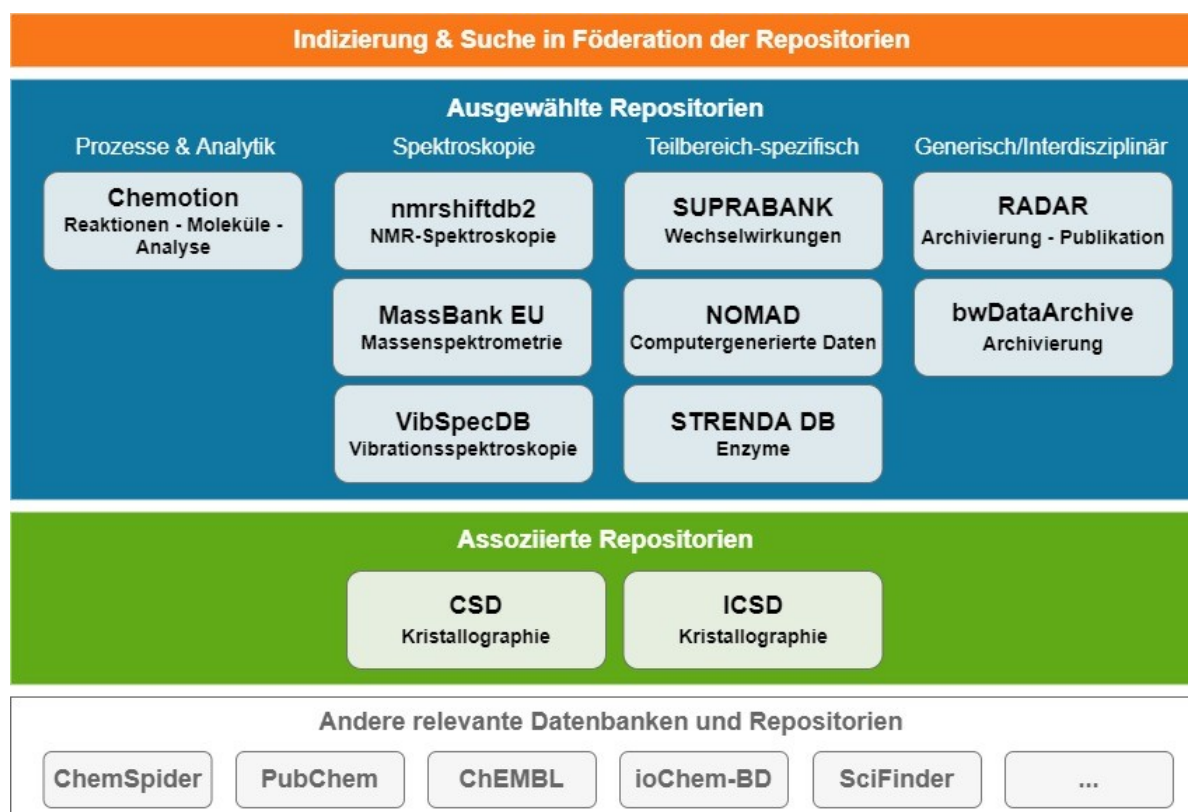


Abbildung 2: Ausgewählte Repositorien der NFDI4Chem.

3.3 Nachnutzung von Forschungsdaten durch offene Daten- und Metadatenstandards sowie Ontologien ermöglichen

Wichtige Kriterien für ein FAIRes Forschungsdatenmanagement sind unter anderem die **F**indbarkeit, Zugänglichkeit (**A**ccessibility), die **I**nteroperabilität und die **R**e-Usability, also die Nachnutzbarkeit von Daten. Repositorien, persistente Identifikatoren (PID) und chemiespezifische Suchfunktionen stellen die **F**indbarkeit sicher. Dies wird ergänzt durch umfangreiche Metadaten, die aus Kernelementen (Autorinnen oder Autor, Datum oder Lizenzangaben), aber auch technischen (aus dem erzeugenden Messgerät) sowie weiteren kontextuellen (experimentelles Umfeld der Datenerzeugung) und chemiespezifischen Metadaten (chemische Strukturinformationen) bestehen. Für diese Elemente bestehen in der Chemie nur teilweise Standards, auch sind unzureichend Minimalinformationen definiert. Dies hat, genauso wie fehlende Standards für offene Dateiformate, Auswirkungen auf die **I**nteroperabilität von Metadaten und Daten. Mit der Erfüllung dieser Kriterien ist eine langfristige Nachnutzbarkeit (**R**e-Usability) über die Repositorien sowie eine Reproduzierbarkeit der Daten und daraus abgeleiteter Erkenntnisse möglich.

Das Ziel der NFDI4Chem ist es, gemeinsam mit der nationalen und internationalen Chemie-Gemeinschaft, aber auch mit Messgeräte- und Softwareherstellern, fehlende

Standards für Datenformate und Metadaten sowie Ontologien zur semantischen Annotation zu entwickeln. Dabei wird auf den Erkenntnissen und Erfahrungen aus den Minimum Information for Biological and Biomedical Investigations (MIBBI)²¹ und der Entwicklung offener Datenformate wie mzML, nmrML und NMReData²² aufgebaut. Die Digitalisierung der Arbeitsschritte im Smart Lab und darüber hinaus ermöglicht die durchgehende digitale Erfassung von Metadaten in kleinen, zeitlich überschaubaren Schritten entlang des Forschungsdatenlebenszyklus. Alle erfassten Daten und Metadaten stehen somit unmittelbar am Ende eines Projektes und zum Zeitpunkt der Publikation zur Verfügung. Die Quantität und Qualität der Metadaten steigt dadurch beträchtlich. Offene Dateiformate ermöglichen die langfristige Interoperabilität und Nachnutzbarkeit, ein Defizit von software- und versionsgebundenen proprietären Dateiformaten. Durch die niedrighschwellige Bereitstellung von kontrollierten Vokabularen bzw. Ontologien innerhalb der digitalen Werkzeuge können Daten auch durch Forschende semantisch annotiert und eindeutig beschrieben werden. Mit chemiespezifischen Ontologien können so Forschungsdaten, aber auch experimentelle Methoden präzise, maschinenles- und interpretierbar beschrieben werden. Das ist in den experimentellen Beschreibungen von wissenschaftlichen Publikationen in dieser Form bisher nicht möglich. So können heterogene Datenquellen leicht zu Linked Data zusammengeführt und Daten für neue Forschungsfelder, wie Big Data, Machine-Learning-Ansätze und Künstliche Intelligenz in der Vorhersage von Reaktivitäten oder Eigenschaften von Molekülen, verfügbar gemacht werden.^{23,24,25}

Bei der Entwicklung der genannten Standards und Ontologien setzt NFDI4Chem auf den kontinuierlichen Austausch mit der Chemie-Gemeinschaft und wird dazu regelmäßige Workshops und (Web)-Seminare durchführen.²⁶ Die internationale Verankerung der Standardentwicklung erfolgt durch eine enge Kooperation mit den oben genannten internationalen Gremien.

²¹Taylor, Chris F., Dawn Field, Susanna-Assunta Sansone et al. „Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project“. *Nature Biotechnology* 26, Nr. 8 (2008): 889-96, <https://doi.org/10.1038/nbt.1411>.

²²Pupier, Marion, Jean-Marc Nuzillard, Julien Wist et al. „NMReDATA, a standard to report the NMR assignment and parameters of organic compounds“. *Magnetic Resonance in Chemistry* 56, Nr. 8 (2018): 703-15, <https://doi.org/10.1002/mrc.4737>.

²³Taylor, Michael G., Tzuhsiung Yang, Sean Lin et al. „Seeing is Believing: Experimental Spin States from Machine Learning Model Structure Predictions“. *The Journal of Physical Chemistry A* 124, Nr. 16 (2020): 3286-99, <https://doi.org/10.1021/acs.jpca.0c01458>.

²⁴Coley, Connor W., Wengong Jin, Luke Rogers et al. „A graph-convolutional neural network model for the prediction of chemical reactivity“. *Chemical Science* 10, Nr. 2 (2019): 370-7, <https://doi.org/10.1039/C8SC04228D>.

²⁵Ahneman, Derek T., Jesús G. Estrada, Shishi Lin et al. „Predicting reaction performance in C-N cross-coupling using machine learning“. *Science* 360, Nr. 6385 (2018): 186-90, <https://doi.org/10.1126/science.aar5169>.

²⁶Weiterführende Informationen zur Arbeit des Konsortiums sowie zu Veranstaltungen: <https://nfdi4chem.de/> (letzter Zugriff: 2021-01-31).

4 Welche Herausforderungen stellen sich für die praktische Arbeit des Konsortiums?

Die Diversität ihrer Stakeholder stellt eine große Herausforderung für die Umsetzung der Vision der NFDI4Chem dar. Unterschiedlichste Interessen müssen in die Veränderungsprozesse, die den Datenlebenszyklus betreffen, eingebunden werden. Dies beginnt bei der Datenerzeugung mit zahlreichen Geräte- und Softwareherstellern, verschiedenen LIMS (Laboratory Information Management System) und ELNs für das Datenmanagement und Software zur Datenanalyse, erstreckt sich über unterschiedliche Repositorien- und Datenbankbetreiber bis zu Verlagen, die Publikationsprozesse anpassen und Vernetzungen mit Datenrepositorien hinzufügen sollten. Die Stakeholder müssen nicht nur miteinander in Einklang gebracht werden, sondern auch von der Nutzung von offenen Standards und Dateiformaten überzeugt werden. NFDI4Chem wird diese Veränderungsprozesse gemeinsam mit den genannten Stakeholdern anstoßen, gestalten, moderieren und fördern. Die Entwicklung von Standards und Ontologien wird NFDI4Chem national und international durch Thesenpapiere, Beiträge für Konferenzen, (Web-)Seminare und Workshops gestalten und vorantreiben.

Eine große Herausforderung bei der zukünftigen Weiterentwicklung von elektronischen Laborjournalen ist das Zusammenbringen der Bedarfe der Nutzenden unterschiedlicher Teildisziplinen der Chemie und die effiziente Entwicklung und Bereitstellung der benötigten Funktionen. Es müssen weiterhin Mechanismen erarbeitet werden, die eine Integration von Mess- und Analysegeräten in möglichst umfassender Weise ermöglichen.

Die Veränderungen durch Digitalisierung in der Chemie erfordern eine neue Wahrnehmung und Bewertung des Umgangs mit Forschungsdaten für die Qualität chemischer Forschung und die wissenschaftliche Reputation. Die Einleitung und Gestaltung eines kulturellen Wandels in der Chemie²⁷ ist von entscheidender Bedeutung für die praktische Arbeit des Konsortiums. Dabei steht vor allem die Anerkennung der Datenpublikation in einem Repository als zitierbare Publikation im Vordergrund. Dies führt zu grundlegenden Änderungen im Verständnis des Forschungs- und Publikationsprozesses. Damit erlangt die Datenpublikation direkten Einfluss auf die wissenschaftliche Reputation der Forschenden. Dieser Transitionsprozess bedarf einer erheblichen Kraftanstrengung und enger Einbindung der wissenschaftlichen Gemeinschaft, sowohl des Nachwuchses als auch der Forschenden mit langjähriger Erfahrung. Dazu sollen dem wissenschaftlichen Nachwuchs schon frühzeitig Kompetenzen im Bereich Daten und Datenmanagement vermittelt werden, eine entscheidende Qualifikation der Studierenden in der zunehmenden Digitalisierung. Gemeinsam mit den Fachgesellschaften sollen Empfehlungen und Umsetzungspläne für die Integration von Datenkompetenzen

²⁷ Herres-Pawlis, Sonja, Oliver Koepler und Christoph Steinbeck. „NFDI4Chem: Digitalen und kulturellen Wandel in der Chemie gestalten“. *Angewandte Chemie* 131, Nr. 32 (2019): 10880-2, <https://doi.org/10.1002/ange.201907260>. „NFDI4Chem: Shaping a Digital and Cultural Change in Chemistry“. *Angewandte Chemie International Edition* 58, Nr. 32 (2020): 10766-8, <https://doi.org/10.1002/anie.201907260>.

in die universitären Lehrpläne entwickelt werden. Dieser Schritt ist essenziell für eine Zukunft, in der Forschende täglich routiniert ihre qualitativ hochwertigen Forschungsdaten nach den FAIR-Prinzipien managen und diese zur Reproduzierbarkeit einer Publikation einfach und schnell per Mausklick aus einer Smart Lab-Umgebung in offene und nachhaltige Repositorien übertragen. Dadurch wird sich langfristig die Nachnutzung der Daten deutlich verbessern. Auch der Einsatz von datenbasierten Methoden kann gefördert und es kann eine solide Datenbasis zur Anwendung analytischer Modelle (z. B. Machine-Learning) gebildet werden.

Die Einbindung, Zusammenarbeit und Unterstützung der Gemeinschaft wird eine wichtige Rolle in der praktischen Arbeit des Konsortiums spielen. Dabei gilt es zunächst, die Forschenden für die vielseitigen Facetten des FDMs zu sensibilisieren und darüber zu informieren. Die Aktivierung und Einbindung der Forschenden erfolgt durch vielseitige Maßnahmen, z. B. regelmäßige Newsletter, Webinare, Workshops, Schulungen und weitere Veranstaltungen zum Thema FDM,²⁸ die stetig erweitert werden sollen. Die Identifikation von besonders erfolgreichen Projekten und wiederkehrende Beschreibung in Best Practices²⁹ soll die Akzeptanz des digitalen und kulturellen Wandels unterstützen. Der Erfolg der Maßnahmen sowie die Erfassung von Anforderungen aus der Gemeinschaft werden durch regelmäßige Umfragen evaluiert und gesteuert.

5 Fazit

Die größte Herausforderung der NFDI4Chem, ihrer Dienste und Infrastrukturen ist eine dauerhafte und nachhaltige Implementierung und Akzeptanz in der Chemie weit über einen Projektstatus hinaus. Letztendlich werden jedoch funktionierende, nutzerfreundliche und vor allem für die Forschenden im Labor hilfreiche Werkzeuge und Dienste zur Bewältigung der täglich anfallenden Arbeiten mit Forschungsdaten über den Erfolg und eine breite Akzeptanz der anstehenden Veränderungen entscheiden. Diese Zukunft möchte das NFDI4Chem-Konsortium einleiten, gestalten und vorantreiben. Interessierte sind jederzeit herzlich eingeladen, Kontakt aufzunehmen, Vorschläge und Anregungen einzureichen, in einer der Gruppen mitzuarbeiten oder an einer der zahlreichen Veranstaltungen des Konsortiums teilzunehmen.³⁰

²⁸Weiterführende Informationen: <https://nfdi4chem.de/> (letzter Zugriff: 2021-01-31).

²⁹Weiterführende Informationen: <https://nfdi4chem.de/> (letzter Zugriff: 2021-01-31).

³⁰Weiterführende Informationen: <https://nfdi4chem.de/> (letzter Zugriff: 2021-01-31).