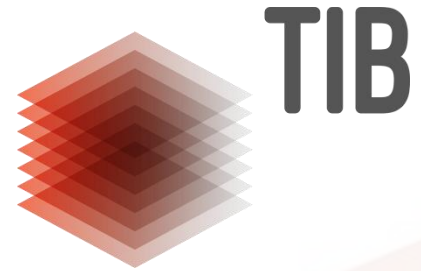


LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK



# Forschungsdatenmanagement und Nationale Forschungsdateninfrastruktur für die Chemie

Dr. Oliver Koepler, Dr. Janna Neumann  
Webinar  
Hannover, 13.12.2018

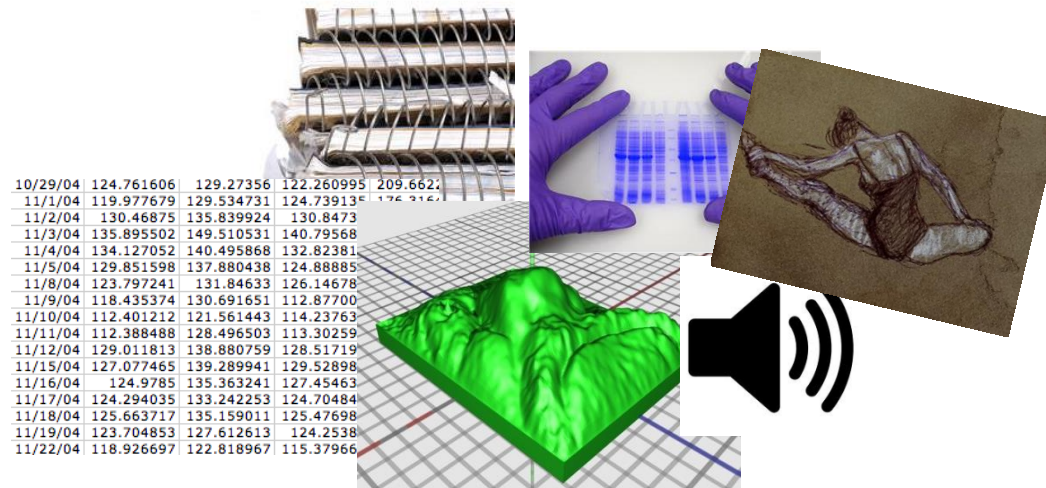


# Agenda

1. Forschungsdatenmanagement in der Chemie
2. Nationale Forschungsdateninfrastruktur
  - Allgemeines
  - Das Fachkonsortium Chemie (NFDI4Chem)
3. Internationale Initiativen im Kontext der NFDI4Chem
4. Labvolution/SmartLab

Forschungsdaten sind:

- Daten, die im Laufe der wissenschaftlichen Tätigkeit entstehen und als Grundlage für wissenschaftliche Forschungsergebnisse dienen.



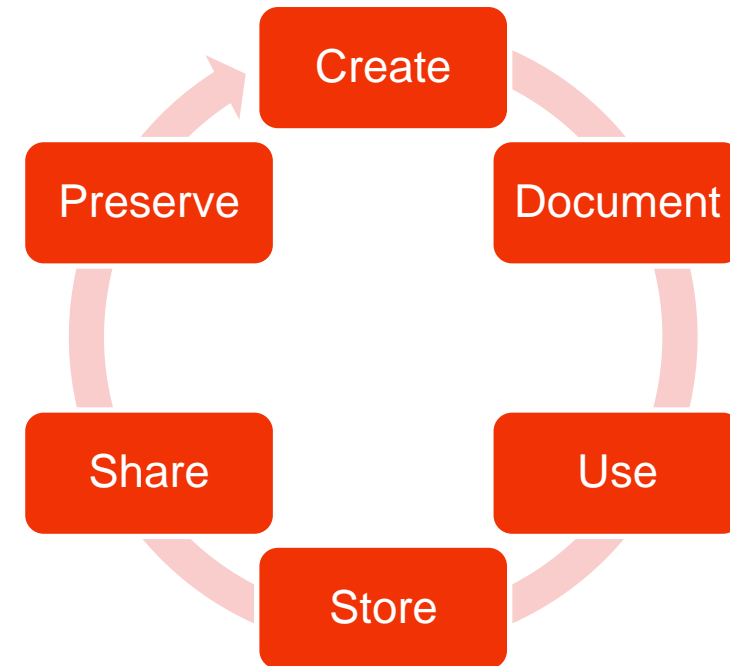
|          |            |            |            |          |
|----------|------------|------------|------------|----------|
| 10/29/04 | 124.761606 | 129.27356  | 122.260995 | 209.662  |
| 11/1/04  | 119.977679 | 129.534731 | 124.739135 | 176.3161 |
| 11/2/04  | 130.46875  | 135.839924 | 130.8473   |          |
| 11/3/04  | 135.895502 | 149.510531 | 140.79568  |          |
| 11/4/04  | 134.127052 | 140.495868 | 132.82381  |          |
| 11/5/04  | 129.851598 | 137.880438 | 124.88885  |          |
| 11/8/04  | 123.797241 | 131.84633  | 126.14678  |          |
| 11/9/04  | 118.435374 | 130.691651 | 112.87700  |          |
| 11/10/04 | 112.401212 | 121.561443 | 114.23763  |          |
| 11/11/04 | 112.388488 | 128.496503 | 113.30259  |          |
| 11/12/04 | 129.011813 | 138.880759 | 128.51719  |          |
| 11/15/04 | 127.077465 | 139.289941 | 129.52898  |          |
| 11/16/04 | 124.9785   | 135.363241 | 127.45463  |          |
| 11/17/04 | 124.294035 | 133.242253 | 124.70484  |          |
| 11/18/04 | 125.663717 | 135.159011 | 125.47698  |          |
| 11/19/04 | 123.704853 | 127.612613 | 124.2538   |          |
| 11/22/04 | 118.926697 | 122.818967 | 115.37966  |          |

Bilder aus: Davidson, J. et. al. Introduction to RDM: benefits, drivers and the role of the university, Digital Curation Centre, 2013, <http://www.dcc.ac.uk/resources/publications/2013>

Forschungsdatenmanagement:

- alle Aktivitäten, die mit der Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten verbunden sind

(Simukovic, Kindling, Schirnbacher,(2013))



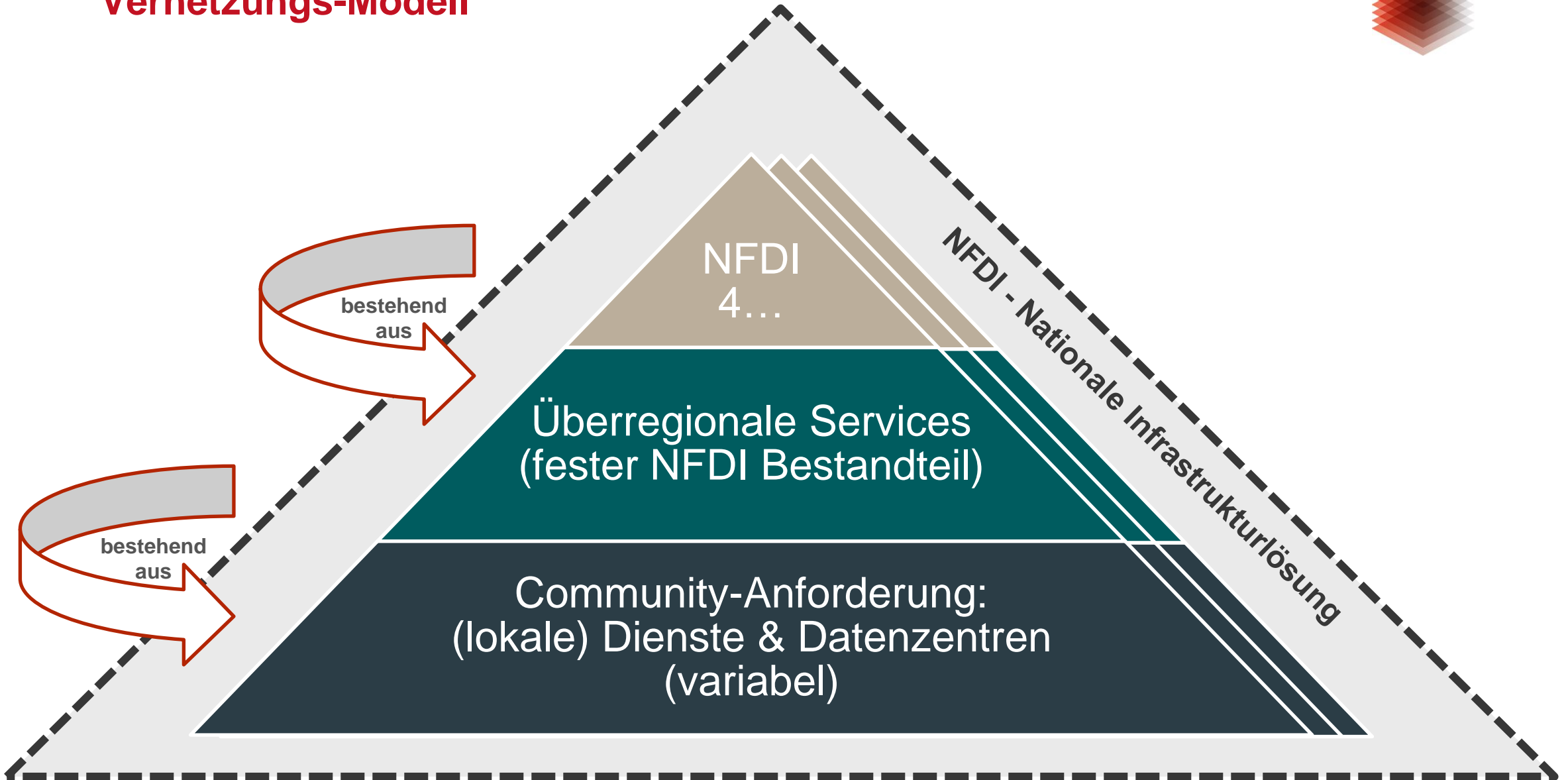
- Erzeugung großer Mengen von experimentellen Daten
- Heterogene Datenformate
  - Vielzahl analytischer Methoden
  - Proprietäre Formate
- Heterogener Umgang mit Daten
  - Unzureichende digitale Vernetzung
  - Keine nachhaltige Beschreibung erhobener Daten
  - Keine Verwendung von offenen, standardisierten Austauschformaten
  - Individuelle Regeln in jeweiligen Instituten
- Wenig etablierte Strukturen, die sich mit Forschungsdatenmanagement beschäftigen
- wenig Repositorien /Initiativen mit Nachnutzungsmodelle für Forschungsdaten aus der Chemie
- Publikation wissenschaftlicher Ergebnisse in Form von Zeitschriftenaufsätzen mit reduzierter Darstellung der Forschungsdaten zur Beweisführung
  - Handhabung von Kristallstrukturdaten in Repository wie der Cambridge Structural Database

## Forschungsdatenmanagement in Deutschland

- Dramatisches Anwachsen der Menge/Heterogenität an Forschungsdaten in Deutschland
  - z.B. durch Beobachtungen, Experimente, Simulationsrechnungen, Erhebungen, Befragungen, Quellenforschungen, Aufzeichnungen, Digitalisierung, Auswertungen.
- Viele informative Stellungnahmen zu FDM, aber „Vollzugsdefizit“
- Erheblicher Finanzierungs- und Personalbedarf
- Zahlreiche Hürden bremsen den Wandel
  - Heterogene Förderlandschaft, fehlende Koordination, Technik- statt Prozessorientierung, fachkulturelle Diversität und unklare Qualitätssicherungs- sowie Reputationsmechanismen, Unsicherheiten der Akteure in Bezug auf strategische Investitionen

# Nationale Forschungsdateninfrastruktur

## Vernetzungs-Modell



# Verwaltungsvereinbarung zu Aufbau und Förderung einer NFDI

## Bund-Länder-Vereinbarung zur NFDI (November 2018)

- Gemeinsame Förderung der NFDI
- Zusammenarbeit von Konsortien, Konsortialversammlung, wiss. Senat, Direktorat
- Drei Ausschreibungsrunden mit bis zu 30 Konsortien
- Begutachtungsverfahren durch die DFG (Expertengremium)
- Entscheidung zur Förderung durch die GWK auf Grundlage der DFG-Förderempfehlung

## Mittelbereitstellung

- Bis zu 90 Millionen Euro p. a. im Endausbau von 2019-2028
- Konsortien mit 85 Millionen Euro p.a.; pro Konsortium ca. 2 bis 5 Millionen Euro p.a. (Zuwendung DFG)
- Konsortien schließen Kooperationsvereinbarung (Ziele, Mittelfluss, Verwendung der Mittel)
- Direktorat jährlich mit bis zu 2,5 Millionen Euro ab 2019 (Zuwendung BMBF)

## Evaluation

- Wissenschaftsrat führt Strukturevaluation der gesamten NFDI durch (bis 2025)
- DFG evaluiert einzelne Konsortien und GWK entscheidet über weitere Förderung
- GWK entscheidet 2026 auf Basis der Strukturevaluation über weitere Ausgestaltung der NFDI und Förderung ab 2029



# NFDI Konsortien

## NFDI-Konsortien

- Struktur zur fortlaufenden Gestaltung von Dienste-Portfolio im Hinblick auf wissenschaftlichen Bedarf
- Sprech- und Handlungsfähigkeit in fachwissenschaftlicher Breite
- Forschende als Nutzer/innen und Mitverantwortliche in der Ausgestaltung (kontinuierlich und strukturell)
- Abdeckung fachlicher Breite und Vielfalt, bemessen an Methode nicht nach Fachdisziplin
- Dynamische Konsortien durch Vergrößerung des Fokus und Einbindung weiterer Akteure (auch international)
- Entwicklung von Beratungsdienstleistungen und Kompetenzen für digitale Wissenschaft

## Dienste-Portfolio

- Essentielle bedarfsorientierte Dienste (weiter-)entwickeln
- Gemeinsame (generische) und zielgruppenspezifische Dienste

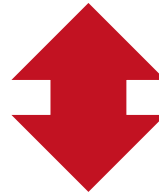
## Einstieg in die NFDI

- Erhebung von digitalen Bedarf
- Beschreibung von Regeln und Diensten
- Bestandsaufnahme vorhandener Dienste und Rolle der Akteure
- Bedarf für Archivierung
- Auswahl und Erhalt von Diensten

## Zusammenfassung NFDI

- Die Fachgemeinschaften definieren Ihre Bedarfe
- Fachgemeinschaften und Infrastruktureinrichtungen erzeugen/verknüpfen Lösungen für FDM

**Häufige Anforderung aus den Fachgemeinschaften:** Die Wissenschaft muss die Kontrolle über ihre Daten behalten



**Anforderung Rfll:**

Ein stimmiges Nutzungs-/Datenzugangskonzept & ein plausibles Betriebsmodell, sowie durchdachte strukturelle Effekte für das wissenschaftliche Gesamtsystem

# Konsortium Chemie (NFDI4Chem)

# Fachgespräch NFDI4Chem, 23. April 2018

23. April 2018

Fachgespräch Nationale Forschungsdateninfrastruktur für die Chemie, Hannover



August 2018

Thesenpapier NFDI4Chem: <http://doi.org/10.5281/zenodo.1404201>

# NFDI4Chem Workshop, 30. Oktober 2018

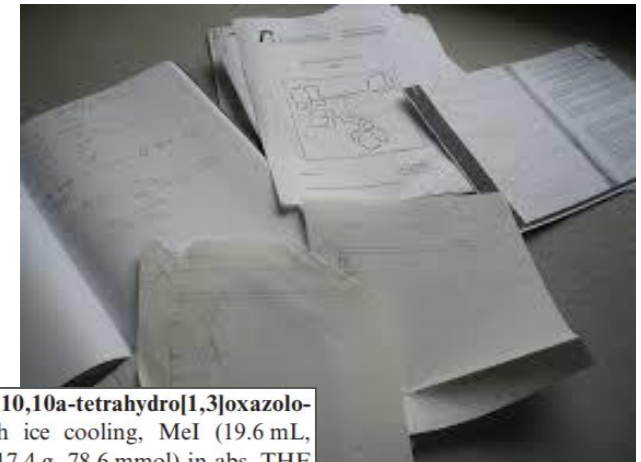
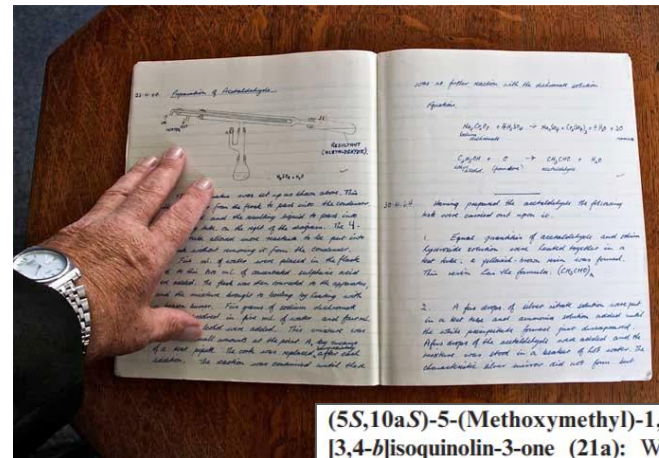


# **„Infrastruktur wissenschaftsgeleitet aufbauen“**

# Forschungsdatenmanagement in der Chemie

Idee → Experiment → Datenerhebung → Analyse → Interpretation → Ergebnispräsentation

- Digital und Analog
- Daten zerstreut
- Heterogene Systeme und Datenformate
- Aufbereitung mit Ziel einer Journal-Publikation



**(5*S*,10*aS*)-5-(Methoxymethyl)-1,5,10,10*a*-tetrahydro[1,3]oxazolo[3,4-*b*]isoquinolin-3-one (**21a**):** With ice cooling, MeI (19.6 mL, 314 mmol) and a solution of **20a** (17.4 g, 78.6 mmol) in abs. THF (45 mL) was added dropwise to a suspension of NaH (4.77 g, 121 mmol) in abs. THF (230 mL). The reaction mixture was stirred at room temp. for 16 h prior to hydrolysis with satd. NH<sub>4</sub>Cl solution (30 mL). The organic solvent was removed, and the aqueous layer extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 100 mL). The combined extracts were dried (MgSO<sub>4</sub>) and concentrated. The residue was purified by flash chromatography (SiO<sub>2</sub>; pentane/EtOAc = 5:1) to give **21a** as a colorless solid (16.7 g, 71.6 mmol, 91%), m.p. 72.5 °C, [α]<sub>D</sub><sup>25</sup> = -188.5 (c = 1.24, CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 7.26–7.21 (m, 3 H, 7-H, 8-H, 9-H), 7.14 (d, *J* = 7.1 Hz, 1 H, 6-H), 5.02 (t, *J* = 4.3 Hz, 1 H, 5-H), 4.60 (t, *J* = 8.6 Hz, 1 H, 1-H<sub>a</sub>), 4.25 (dddd, *J* = 10.8, *J* = 8.6, *J* = 4.5, *J* = 4.2 Hz, 1 H, 10a-H), 4.13 (dd, *J* = 8.6, *J* = 4.2 Hz, 1 H, 1-H<sub>b</sub>), 3.81 (dd, *J* = 10.1, *J* = 4.3 Hz, 1 H, 1'-H<sub>a</sub>), 3.78 (dd, *J* = 10.1, *J* = 4.3 Hz, 1 H, 1'-H<sub>b</sub>),

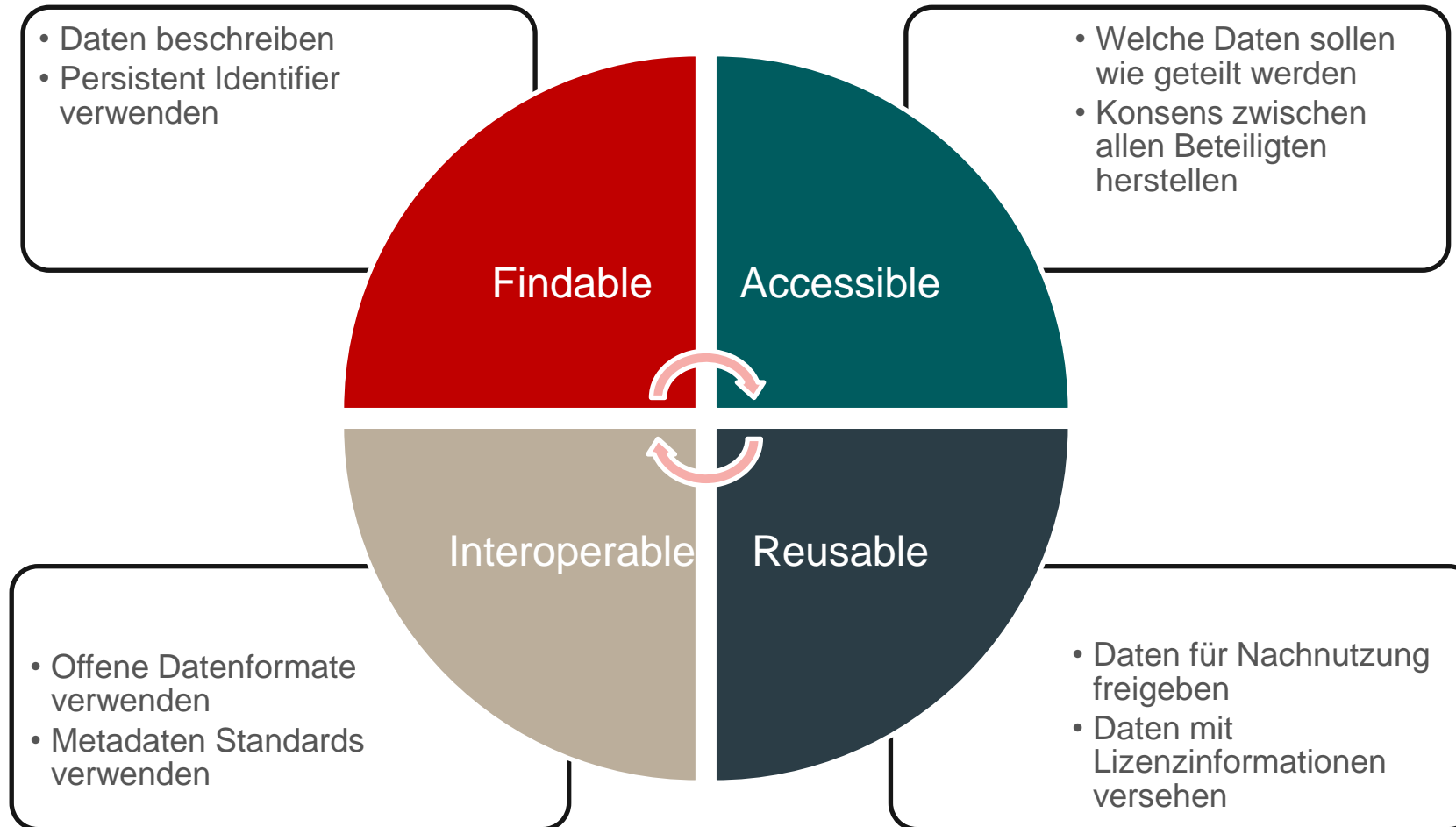
## Wo bleiben Ihre Forschungsdaten?

- Praktizieren Sie Forschungsdatenmanagement?
- Wissen Sie, wo die Daten der letzten Jahre sind?
- Können Sie auf Ihre Daten zugreifen?
- Können Sie ihre Daten noch lesen?
- Können Daten nachgenutzt werden?



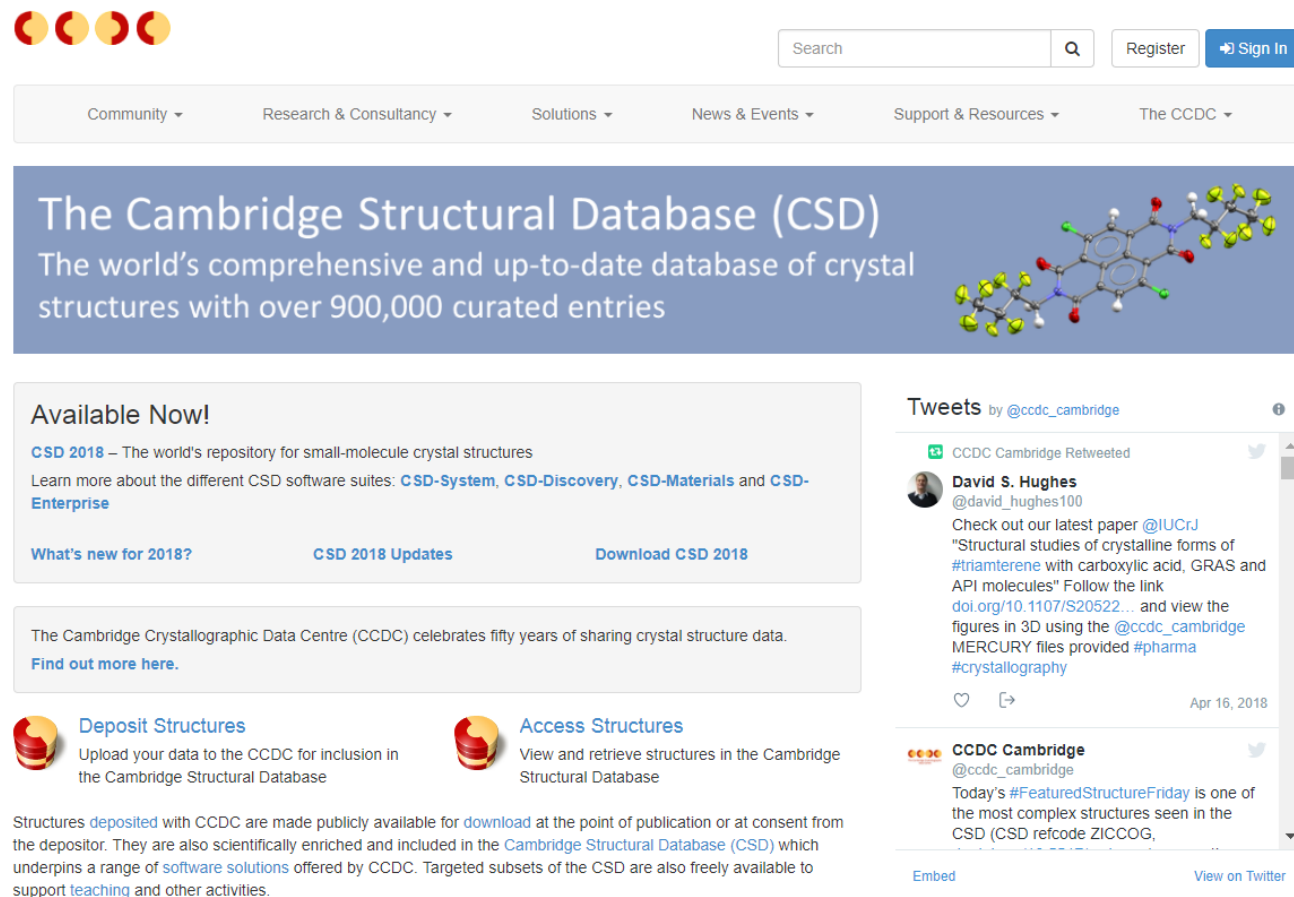


## FAIR Data Prinzipien



.... FAIR ist nicht zwingend Open oder Free Data

# Wo es funktioniert hat – Cambridge Structural Database



The screenshot shows the homepage of the Cambridge Structural Database (CSD). At the top, there are navigation links for Community, Research & Consultancy, Solutions, News & Events, Support & Resources, and The CCDC. A search bar and buttons for Register and Sign In are also present. The main banner features the text "The Cambridge Structural Database (CSD) The world's comprehensive and up-to-date database of crystal structures with over 900,000 curated entries" alongside a 3D molecular model. Below the banner, there is a section titled "Available Now!" with links for "What's new for 2018?", "CSD 2018 Updates", and "Download CSD 2018". A central message states that the CCDC celebrates fifty years of sharing crystal structure data, with a link to "Find out more here." Two main actions are highlighted: "Deposit Structures" (uploading data to the CCDC) and "Access Structures" (viewing and retrieving structures). A paragraph explains that structures deposited with CCDC are made publicly available for download at the point of publication or at consent from the depositor. On the right side, there is a "Tweets" section by @ccdc\_cambridge, featuring a tweet from David S. Hughes about a paper on triamterene and a tweet from CCDC Cambridge about a featured structure.

**Available Now!**

**CSD 2018** – The world's repository for small-molecule crystal structures

Learn more about the different CSD software suites: [CSD-System](#), [CSD-Discovery](#), [CSD-Materials](#) and [CSD-Enterprise](#)

[What's new for 2018?](#)      [CSD 2018 Updates](#)      [Download CSD 2018](#)

The Cambridge Crystallographic Data Centre (CCDC) celebrates fifty years of sharing crystal structure data.

[Find out more here.](#)

**Deposit Structures**  
Upload your data to the CCDC for inclusion in the Cambridge Structural Database

**Access Structures**  
View and retrieve structures in the Cambridge Structural Database

Structures [deposited](#) with CCDC are made publicly available for [download](#) at the point of publication or at consent from the depositor. They are also scientifically enriched and included in the [Cambridge Structural Database \(CSD\)](#) which underpins a range of [software solutions](#) offered by CCDC. Targeted subsets of the CSD are also freely available to support [teaching](#) and other activities.

**Tweets** by @ccdc\_cambridge

CCDC Cambridge Retweeted

**David S. Hughes**  
@david\_hughes100  
Check out our latest paper @IUCrJ "Structural studies of crystalline forms of #triamterene with carboxylic acid, GRAS and API molecules" Follow the link [doi.org/10.1107/S20522...](https://doi.org/10.1107/S20522...) and view the figures in 3D using the @ccdc\_cambridge MERCURY files provided #pharma #crystallography

Apr 16, 2018

**CCDC Cambridge**  
@ccdc\_cambridge  
Today's #FeaturedStructureFriday is one of the most complex structures seen in the CSD (CSD refcode ZICCOG,

[Embed](#)      [View on Twitter](#)

## Verankerung im Publikationsprozess

“.. Note that CIFs, structure factor tables, and CheckCIF reports must be submitted to the Cambridge Crystallographic Data Centre (CCDC) **prior to manuscript submission**. The CCDC deposition number(s) should be entered into the ACS Paragon Plus Environment during submission, and any explanations for A and/or B level alerts should be described in a separate document and uploaded as Supporting Information for Review Only...”

“Peer Review and Acceptance While the manuscript is undergoing peer review, the CIFs and structure factor tables will be held confidentially in the CCDC’s archive where they will be accessible only by ACS and reviewers during the peer review process. If the manuscript is accepted and published, ACS will notify the CCDC and the data will be released from embargo where it will then be freely accessible through the CCDC, via direct link(s) in the article when it is published online.”

Aus den Autoren-Guidelines, ACS Publications, 1. November 2017

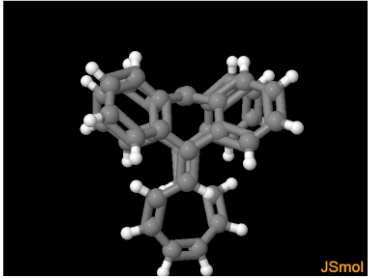
# Erfolgsfaktor Nachnutzung von publizierten Daten

Daten der CSD waren von Beginn an Gegenstand von wissenschaftlichen Arbeiten.

Mehr als 3000 Publikationen nutzen Erkenntnisse aus den Daten der CSD.

BEMLIP : anti-1ZcZcZl-10,10'-(Octa-2,4,6-triene-1,8-diyldiene)-9,9'10,10'-tetrahydro-9,9'-bianthracene  
**Space Group:** C 2/c (15), **Cell:** a 16.280(2)Å b 10.9590(10)Å c 14.373(2)Å,  $\alpha$  90°  $\beta$  110.41(2)°  $\gamma$  90°

3D viewer



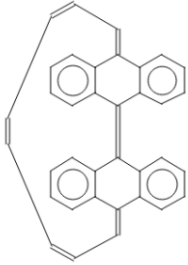
JSmol

H Disorder  $\phi$  Menu Open

Style Labels Packing Measure



Ball and Stick No Labels None None

Chemical diagram

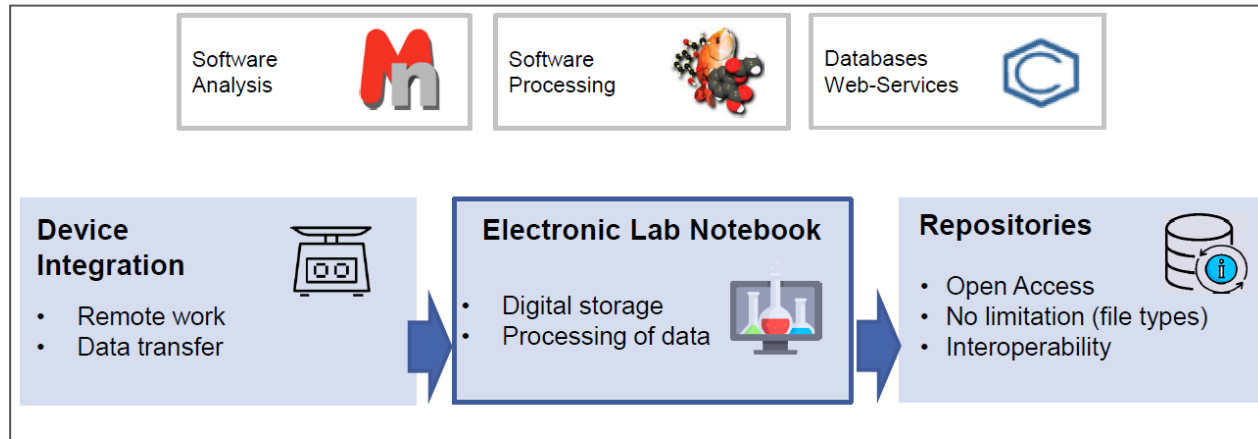


View group symbols key

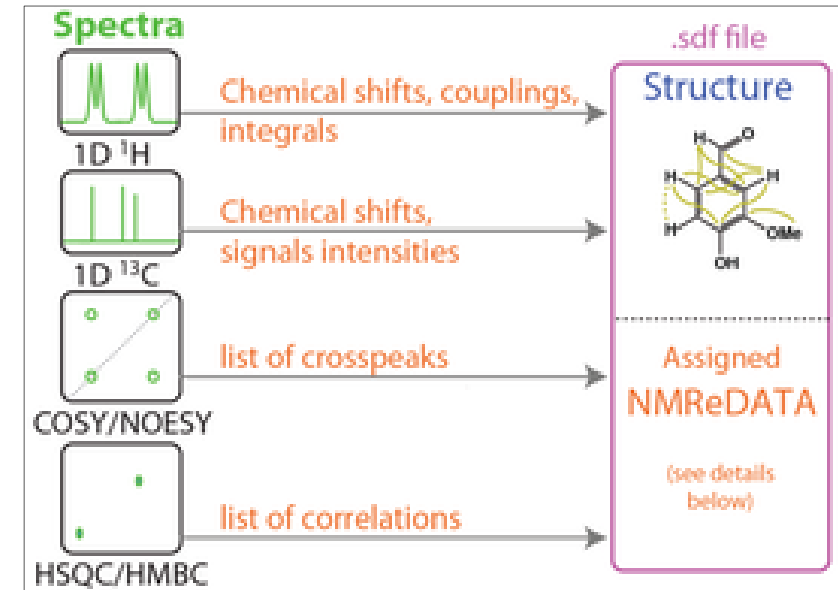
| Additional details                     |   |
|--|---|
| <b>Deposition Number</b>               | 223927  |
| <b>Data Citation</b>                   | D.Ajami, O.Oeckler, A.Simon, R.Herges CCDC 223927: Experimental Crystal Structure Determination, 2014, DOI: <a href="https://doi.org/10.5517/cc7j0g6">10.5517/cc7j0g6</a> |
| <b>Additional Database Identifiers</b> | BEMLIP01  |
| <b>Synonyms</b>                        | 10,10'-(Octa-2,4,6-triene-1,8-diyldiene)-9,9'10,10'-tetrahydro-9,9'-bianthracene  |
| <b>Deposited on</b>                    | 08/11/2003  |

| Associated publications   |   |
|---|---|
|  | D.Ajami, O.Oeckler, A.Simon, R.Herges <i>Nature (London)</i> , 2003, 426, 819, DOI: <a href="https://doi.org/10.1038/nature02224">10.1038/nature02224</a>   |
|  | D.Ajami, K.Hess, F.Kohler, C.Nather, O.Oeckler, A.Simon, C.Yamamoto, Y.Okamoto, R.Herges <i>Chemistry-A European Journal</i> , 2006, 12, 5434, DOI: <a href="https://doi.org/10.1002/chem.200600215">10.1002/chem.200600215</a> |

# Initiativen, Visionen



## Digitalisierung von Workflows mit ELNs



## Verbesserte Datenformate: NMRReData Initiative

## Vision NFDI4Chem

### Ziel:

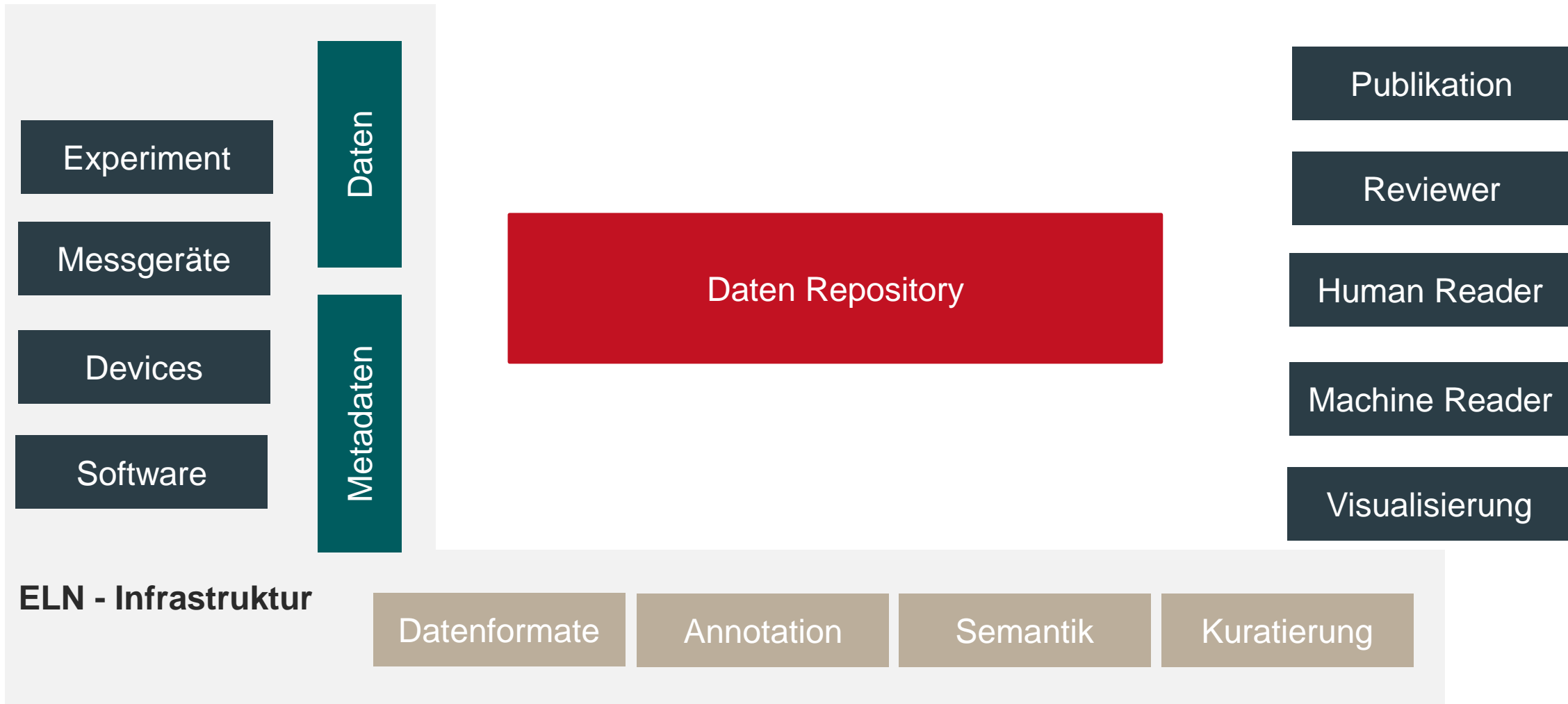
Es soll ein **digitaler Wandel im Umgang mit Forschungsdaten in der Chemie** angestoßen werden. Für die Weiterentwicklung des Forschungsdatenmanagements sollen notwendige **Veränderungsprozesse** forciert und Infrastrukturen geschaffen sowie vernetzt werden

### Anforderungen NFDI4Chem:

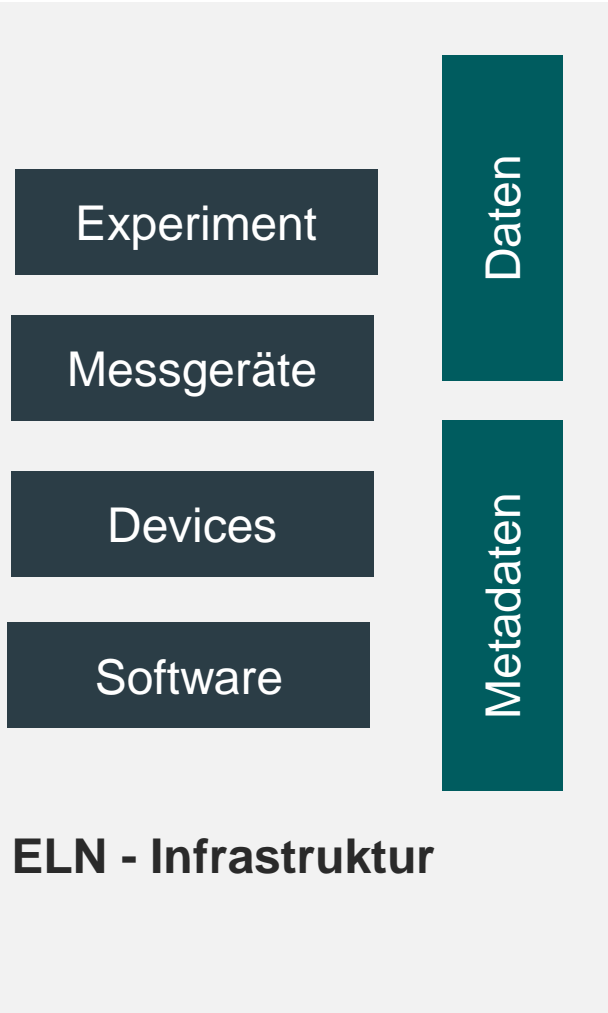
- Diskussion & Gestaltung von Policies zum FDM
- Hohe Datensicherheit, differenzierte Betrachtung von Rollen & Rechten
- Niedrigschwellige Nutzung der Infrastruktur und von Tools & Plattformen
- Gesicherte, langfristige Finanzierung des Forschungsdatenmanagements (FDM)

### Thesen NFDI4Chem ►

# Digitaler Wandel im Umgang mit Forschungsdaten



NFDI



Daten  
Repositories

Daten  
Repositories

Publikation

Reviewer

Daten  
Repositories

Daten  
Repositories

Human Reader

Daten  
Repositories

Daten  
Repositories

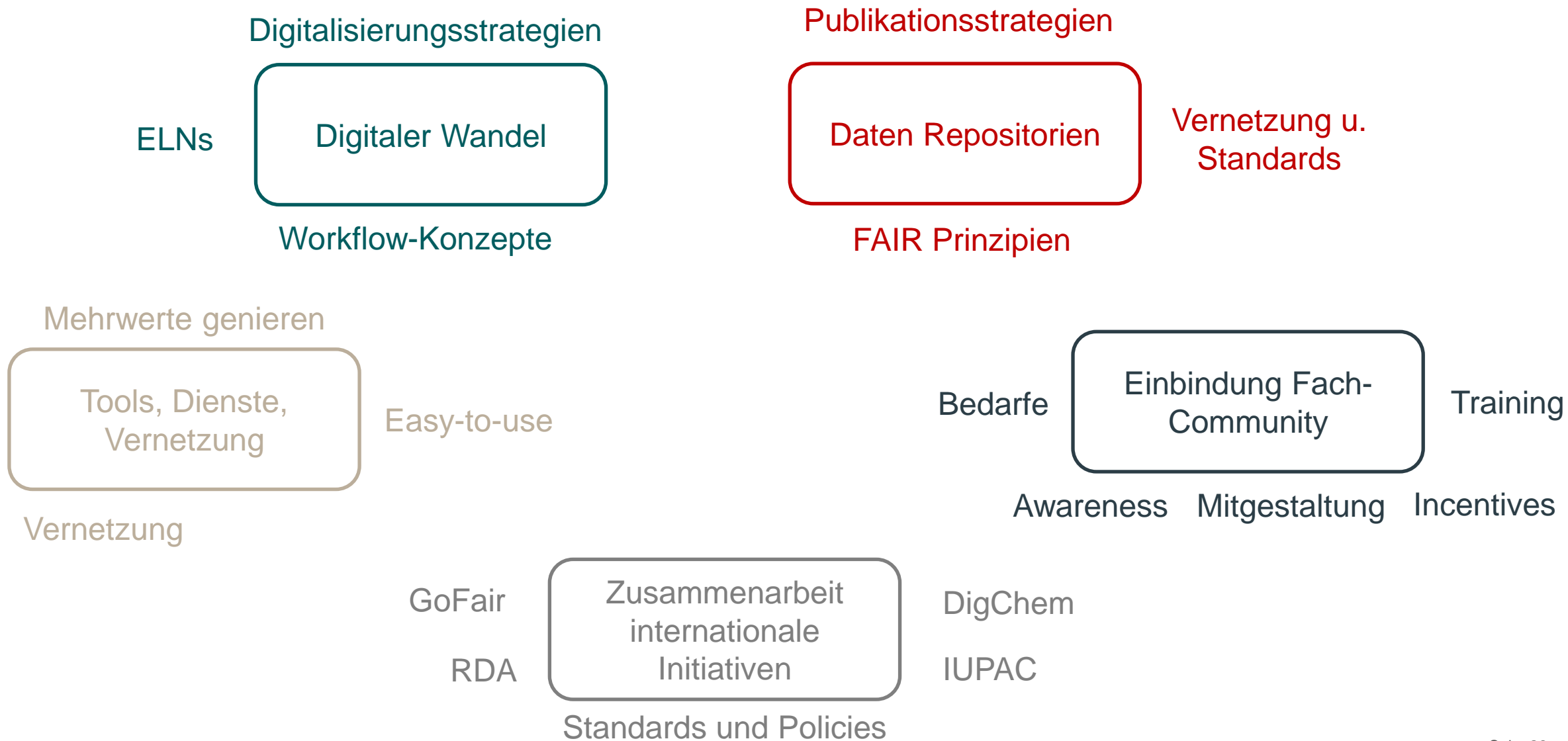
Machine Reader

Visualisierung

Gemeinsame Standards, APIs, Formate, Tools, Vokabulare



# NFDI4Chem Thesen / Themenschwerpunkte



# Internationale Initiativen

## Internationale Institutionen, Aktivitäten

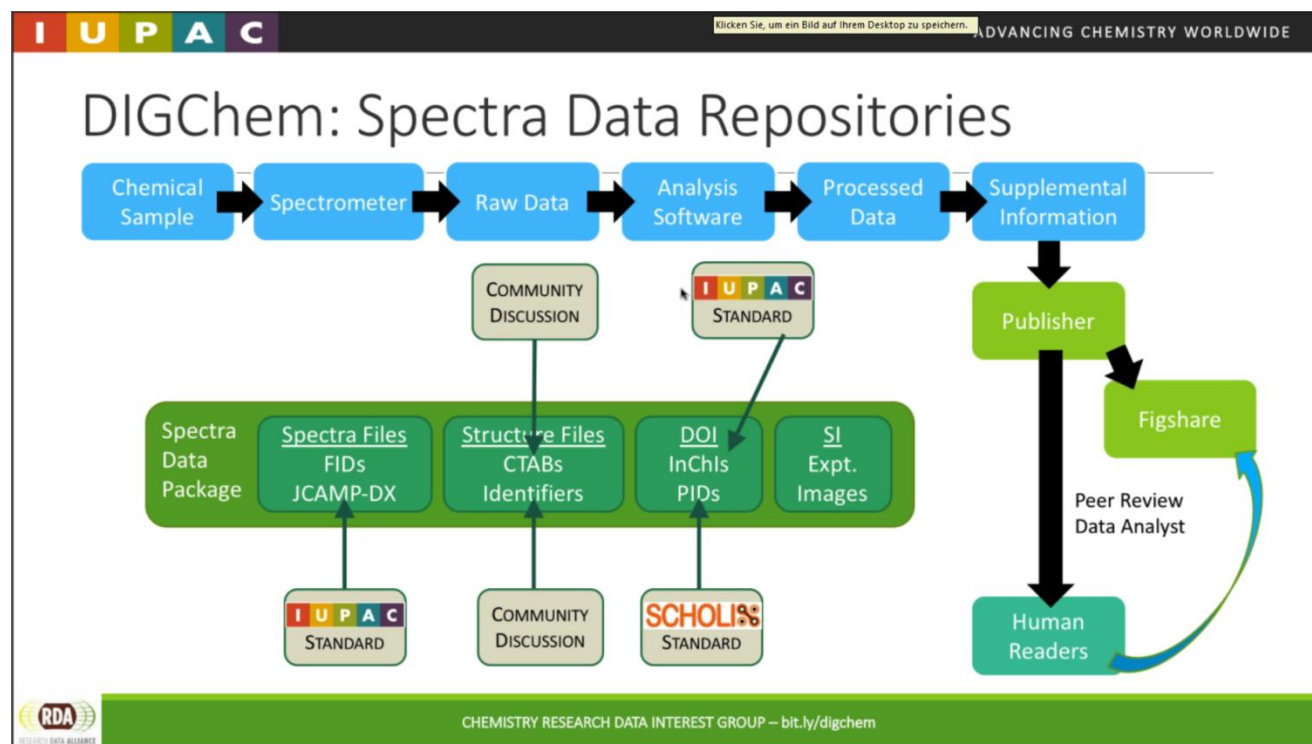
- CPCDS Committee on Publications and Chemoinformatics Data Standards, IUPAC
- SCDS Subcommittee on Chemoinformatics Data Standards, IUPAC
- Chemistry GO FAIR Implementation Network ChIN (IUPAC, CRDIG)
  
- CRDIG Chemistry Research Data Interest Group, Research Data Alliance (RDA)
- **DIGChem: Projektgruppe von CRDIG**  
Community based working groups to survey the existing landscap in chemoinformatics

## Internationale Initiativen

Chemistry Research Data Interest Group, DIGChem Project

Arbeitsgruppen für Aktionsfelder

- Publishing Guidelines
- Survey
- Workflow-Tools
- JCAMP-DX
- NMR/Spectra Repositories
- Open-Structures
- DataCite Recommendations
- Education
- Professional Training
- Cheminformatics Color Book
- Gold Book Website



# Strukturierung Konsortium Chemie



# NFDI4Chem – Themen- und Arbeitsgruppen-Struktur

| Topics         | Topic 1<br>Digitalisierung                                   | Topic 2<br>Repositorien  | Topic 3<br>Software, Tools  | Topic 4<br>Einbindung Community   | Topic 5<br>Internat. Netzwerk   |
|----------------|--|--|---|---|---|
| Arbeitsgruppen | (1.2) Geräteanbindung<br>(1.3) ELNs<br>(1.4) Schnittstellen  | (2.2) Entwicklung Repos<br>Hosting/Speicherung, LZA<br>(2.3) Vernetzung von Repos,<br>Datenaustausch zwischen Repos<br>(2.4) Workflows für Publikation | (3.1) Tools die auf Repos<br>aufsetzen: Suche,<br>Visualisierung, Analyse,<br>Vorhersage..<br>(3.2) Tools für Einsatz<br>Digitalisierung: Konverter,<br>Uploader..<br>(3.3) Knowledge Graph | (4.1) Awareness<br>(4.2) Incentives/Benefits<br>(4.3) Training - Curriculare<br>Einbindung, Aufbau FDM<br>Kompetenzen)<br>(4.4) Support - Infrastrukturen<br>für den Aufbau und Betrieb<br>FDMs | (5.2) Policies<br>(Datenpublikation<br>internationale Journale)<br>(5.3) An- und Einbindung<br>internationaler Repositories an<br>NFDI4Chem |
|                | Daten- Metadatenstandards, Formate                           |  |   |   |   |
|                | Juristische Aspekte FDM                                      |  |   |   |   |
| Akteure        | (A) Bibliotheken<br>(B) Rechenzentren<br>(C) Wissenschaftler | (D) Entwickler   | (E) Fachgesellschaften<br>(F) Journale<br>(G) Fördergeldgeber   | (H) Gerätehersteller<br>(I) Softwarehersteller  | (J) Kollegium (international)<br>(K) Industrie  |

## NFDI4Chem – Arbeitsgruppen

- AG (Meta)Datenstandards und Formate
- AG ELN (Elektronische Laborjournale)
- AG API (Geräteanbindung und Schnittstellen)
- AG Software und Tools
- AG Juristische Betrachtung Forschungsdatenmanagement
- AG Repositories
- AG Community



## NFDI4Chem – Mögliche Abdeckung Fachbereiche

- Bereich 1: Organische Chemie, Synthesechemie
- Bereich 2: Anorganische Chemie, Katalyse
- Bereich 3: Physikalische Chemie, Theoretische Chemie
- Bereich 4: Polymerchemie, Oberflächenchemie
- Bereich 5: Pharmazeutische Chemie
- Bereich 6: Biochemie
- Querschnittsbereiche: Analytische Chemie, Materialwissenschaften

# Technical and operational readiness level in den Fachbereichen

Anstieg Reifegrad

## Level 1

Topic 1: analoges und digitales Datenmanagement

Topic 2: Keine Repositorien vorhanden

=> Aufbau Repositories mit Technologien / Erfahrungen aus Level 2 und 3

Topic 3: wenige bis keine Tools

=> Toolentwicklung mit Technologien / Erfahrungen aus Level 2 und 3, Nachnutzung

Topic 4: FDM unbekannt, keine Datenmanager vorhanden, keine Schulungsangebote

## Level 2

Topic 1: Digitales Datenmanagement, teilweise Vernetzung

Topic 2: Institutionelle Repositories geringer Reichweite vorhanden

=> Weiterentwicklung und Vernetzung mit Technologien aus Level 3

Topic 3: Tools vorhanden

=> Optimierung

Topic 4: FDM FAIR bekannt, teilweise implementiert, Datenmanager teilweise vorhanden, Schulungskonzepte in Entwicklung

## Level 3

Topic 1: Digitales Datenmanagement, hohe Vernetzung, FDM-Prozesse vorhanden

Topic 2: Repositories hoher Reichweite und Vernetzung vorhanden

=> Fachbereichsübergreifende und internationale Vernetzung

Topic 3: Tools vorhanden, etabliert, breite Anwendung

Topic 4: FDM, FAIR werden angewendet, Datenmanager, Stewards etabliert, Schulungen, Training vorhanden

# Die Zukunft?

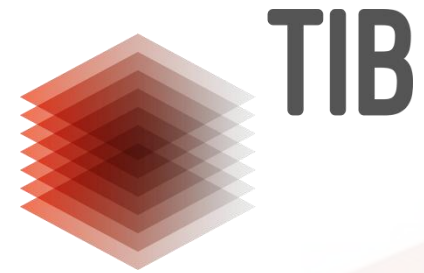
## Blick nach vorne: smartLab / Labvolution



LABVOLUTION 2019, 21.-23. Mai 2019 Hannover

---

LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK



<https://www.nfdi4chem.de>  
[contact@nfdi4chem.de](mailto:contact@nfdi4chem.de)

**Vielen Dank**

Dr. Janna Neumann, [janna.neumann@tib.eu](mailto:janna.neumann@tib.eu)  
Dr. Oliver Koepler, [oliver.koepler@tib.eu](mailto:oliver.koepler@tib.eu)