

# Chemisches Forschungsdatenmanagement – Herausforderungen und Lösungsansätze bei der digitalen Ablage von (Meta)Daten von Molekülen und Reaktionen



**Christoph Steinbeck\***

Friedrich-Schiller-University Jena

**Oliver Koepler\***

TIB Leibniz Information Centre of Science and Technology

**Nicole Jung**

Karlsruhe Institute of Technology (KIT)

**Sonja Herres-Pawlis & Alexander Hoffmann**

RWTH Aachen University

**Matthias Razum**

FIZ Karlsruhe - Leibniz Institute for Information  
Infrastructure

**Felix Bach**

Karlsruhe Institute of Technology (KIT)

**Johannes Liermann**

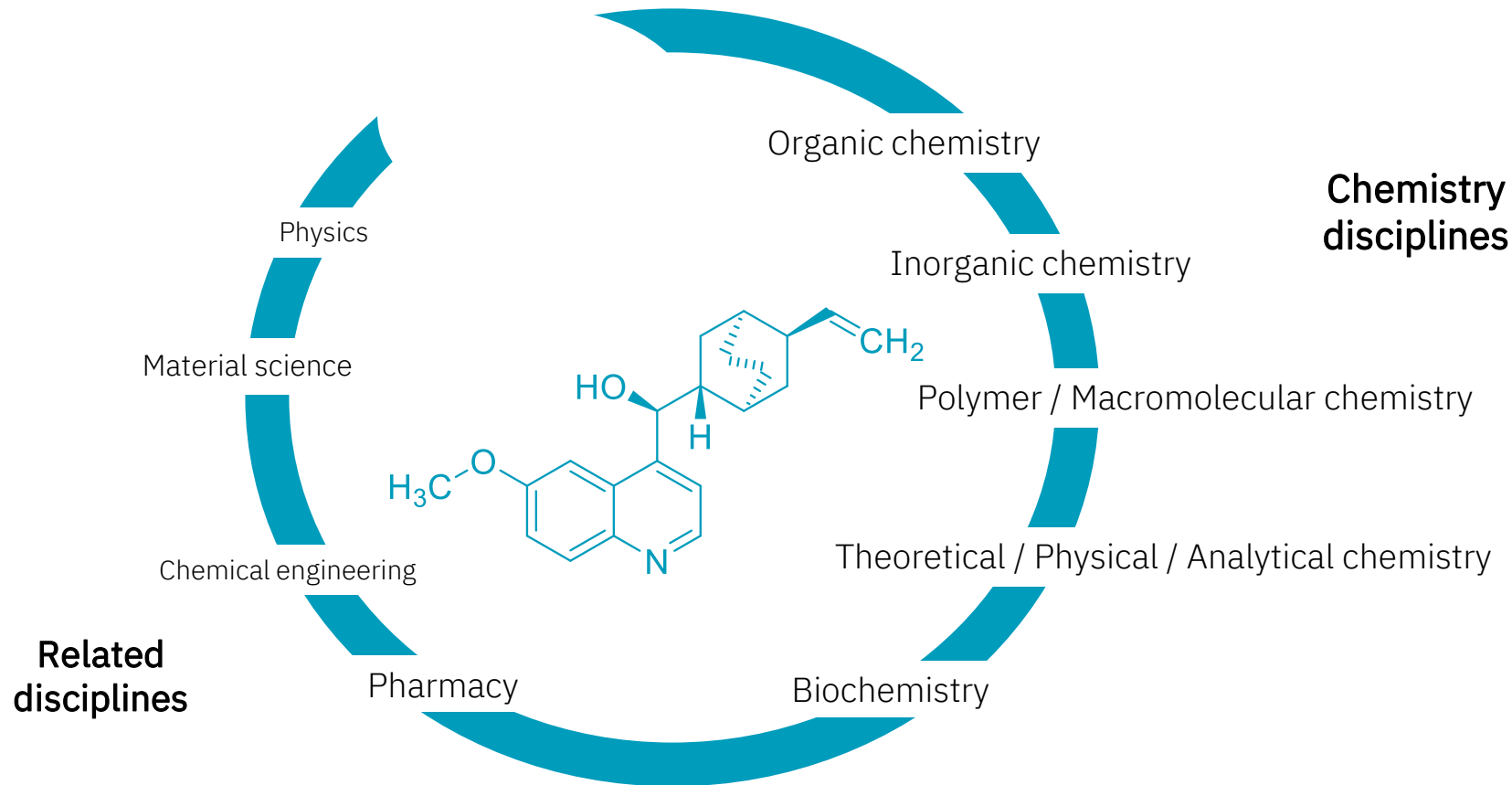
Johannes Gutenberg University Mainz

**Steffen Neumann**

Institut für Pflanzenbiochemie, Halle

***Tag der Forschungsdaten NRW 2021***

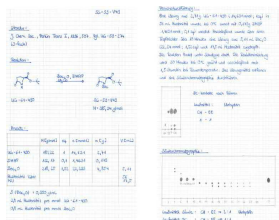
# Our scientific community



# Molecules and related (meta)data

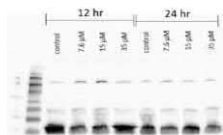


Reactions



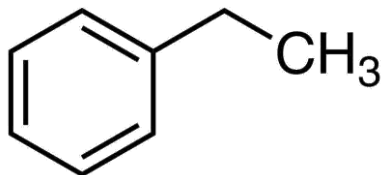
Experimental Description  
Starting compounds / Reagents  
Conditions ...

Biological activity



Organism  
Conditions  
Activity  
...

Properties

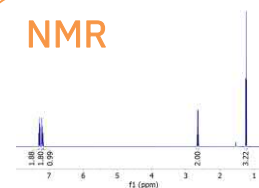


Name  
Formula  
CAS  
InChI Key  
...

Ethylbenzene  
 $C_8H_{10}$   
100-41-4  
YNQLUTRBYVCPMQ-  
UHFFFAOYSA-N

Assignment

NMR



Nucleus  
Pulse sequence  
Solvent ...

MS



Ionisation method  
Voltage ...

Spectra



## 1927

1985



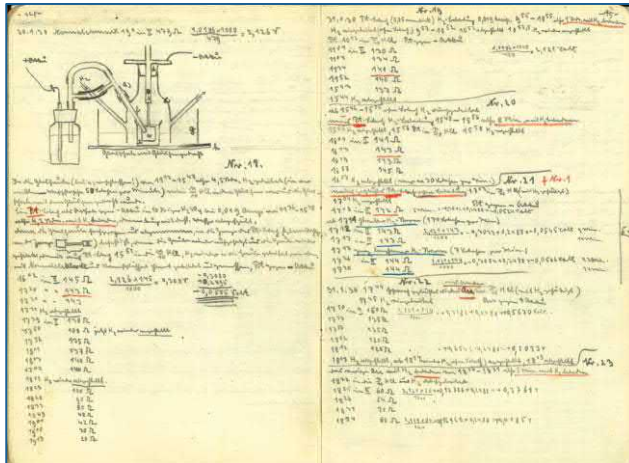
Cray-2, 1985  
1.9 GFlops  
244 MHz CPU Speed

2017



iPhone 8  
325.0 GFlops  
2.39 GHz CPU Speed

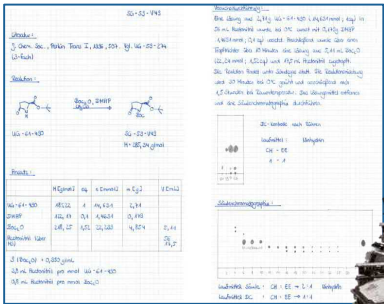
2019

[illegible]

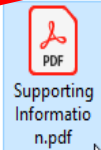
# Status Quo



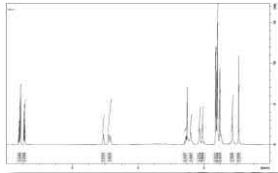
Analogue / Digital



**UnFAIR**



Supporting Information n.pdf



Digital



Not machine-readable



**Leitlinien zur Sicherung  
guter wissenschaftlicher Praxis**

Kodex

**DFG**

*Practical questions in the organisation of a working group:*



Where are the NMR data from 10 years ago?



Where are the synthesis details from 15 years ago?



Has anyone in the group already made compound xy?



## What do the colleagues say...

---

*Critical questions from colleagues:*



How much time does it cost my students?



How much money does it cost me?



Do I really have to do this?

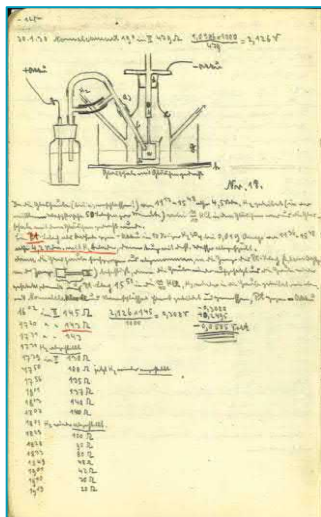
*More critical statements from colleagues:*

- The supporting information is sufficient.
- No one else does this...
- It is a better time invest when my students are not sitting too much at the computer.

# Our Vision



1927



Digitalisation

Standards

Community



Data re-use

Data publication

Experiment  
& Data  
collection



Data  
processing



Data  
Analysis

Data  
preserved



FAIR  
DATA



# Key objectives

---

**Objective 1:** Connect existing **data repositories**, fill in missing research data repositories, and link them to international repositories.

**Objective 2:** **Minimum information (MI) standards for data and machine-readable metadata**, open data standards, in order to support the FAIR principles for research data.

**Objective 3:** Foster **Electronic Laboratory Notebooks (ELN)**, tools and APIs between between instrumentation and software towards a embedded, digital information architecture. Capture research data in well-annotated electronic form at the **earliest possible point in time in the research process**.



# Key objectives

---



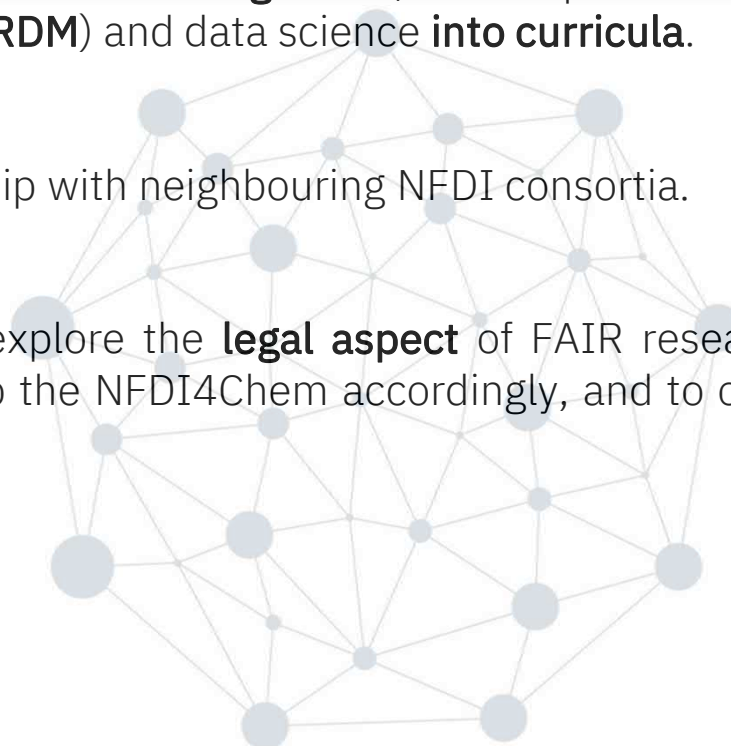
**Objective 4:** Create **awareness** for **FAIR data management**, initiate processes to **integrate** research data management (RDM) and data science **into curricula**.



**Objective 5:** Maintain a close relationship with neighbouring NFDI consortia.



**Objective 6:** Engage with experts to explore the **legal aspect** of FAIR research data management, design and develop the NFDI4Chem accordingly, and to offer advice for the research community.



# Involvement of the Community

---



Community requirements

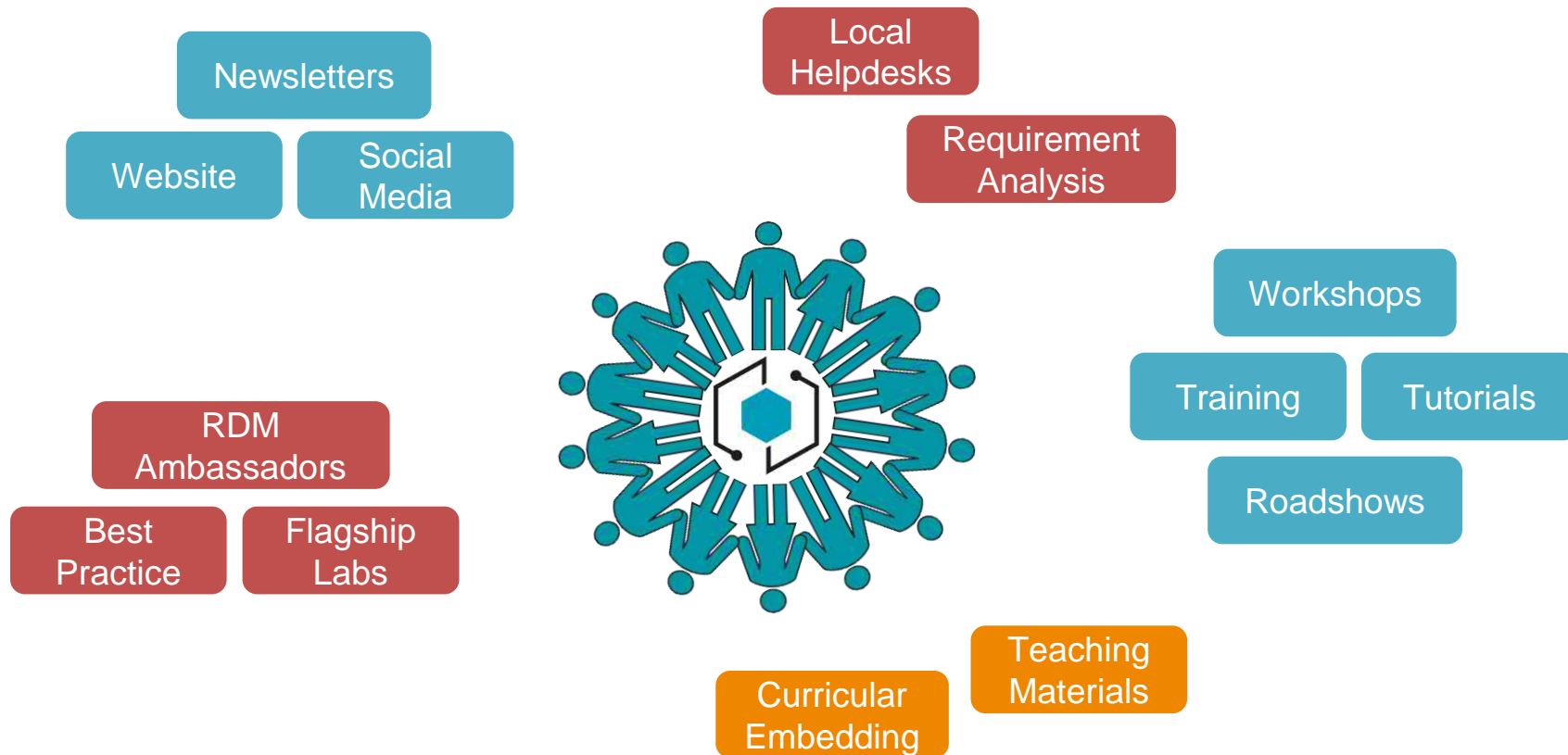


Fostering cultural change

Raising RDM awareness

RDM  
Infrastructure

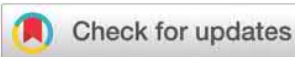
# Involvement of the Community





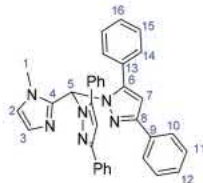
## Dalton Transactions

PAPER



Cite this: DOI: 10.1039/d1dt00832c

<https://www.chemotion-repository.net/welcome>



$^1\text{H}$  NMR ( $\text{CD}_2\text{Cl}_2$ , 400 MHz):  $\delta$  = 7.80 (d,  $^3J$  = 7.8 Hz, 4H, H-10), 7.67 (s, 1H, H-5), 7.39 (t,  $^3J$  = 7.4 Hz, 4H, H-11), 7.32 (d,  $^3J$  = 7.3 Hz, 4H, H-12 + H-16), 7.27 (t,  $^3J$  = 7.3 Hz, 4H, H-15), 7.12 (d,  $^3J$  = 7.1 Hz, 4H, H-14), 6.94 (d,  $^3J$  = 3.1 Hz, 2H, H-2 + H-3), 6.65 (s, 2H, H-7), 3.70 (s, 3H, H-1) ppm.  $^{13}\text{C}$   $\{^1\text{H}\}$  NMR ( $\text{CD}_2\text{Cl}_2$ , 100 MHz):  $\delta$  = 151.8 (C-8), 146.4 (C-6), 142.5 (C-4), 133.7 (C-9), 129.5 (C-13), 129.4 (C-14), 129.1 (C-15), 129.1 (C-12 + C-16), 128.5 (C-11), 128.0 (C-3), 126.3 (C-10), 124.3 (C-2), 106.3 (C-7), 70.0 (C-5), 34.9 (C-1) ppm. HRMS (ESI $^+$ , MeOH):  $m/z$  (found) = 533.24353 (100%), 534.24664 (37%), 535.25024 (7%), 536.25299 (1%);  $m/z$  (calc.) = 533.24537 (100%,  $^{12}\text{C}_{35}^{1}\text{H}_{29}^{14}\text{N}_6^+$ ), 534.24872 (38%,  $^{12}\text{C}_{34}^{13}\text{C}^1\text{H}_{29}^{14}\text{N}_6^+$ ), 535.25207 (7%,  $^{12}\text{C}_{33}^{13}\text{C}_2^1\text{H}_{29}^{14}\text{N}_6^+$ ), 536.25543 (1%,  $^{12}\text{C}_{32}^{13}\text{C}_3^1\text{H}_{29}^{14}\text{N}_6^+$ ). IR (ATR, neat),  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ]: 1605 (vw), 1551 (w), 1459 (w), 1437 (w), 1410 (vw), 1298 (vw), 1281 (vw), 1259 (w), 1202 (w), 1138 (vw), 1076 (w), 1027 (w), 1007 (w), 957 (w), 916 (w), 868 (w), 841 (w), 834 (w), 816 (m), 803 (w), 767 (m), 759 (vs), 752 (s), 704 (m), 691 (vs), 678 (m), 666 (m), 573 (m), 524 (w), 432 (vw).

Additional information on the NMR of the target compound including original data files is available via Chemotion Repository: <https://dx.doi.org/10.14272/LSGGPBYVWWQPOY-UHFFFAOYSA-N.1>



View Article Online

View Journal

## Copper nitrenes and amination

ai C. Göbgen,  
mann  and

as intermediates in the copper catalyzed amination of aryl amines. The copper complexes were synthesized at low temperatures from (pyrazolyl)methane family. The copper complex with  $^5\text{PhINTs}$  in dichloromethane.

# Chemotion Repo



Chemotion Repository

www.chemotion-repository.net/mydb/collection/2171/sample/49359

Repository

All IUPAC, InChI, SMILES, RIn

Chemotion

Scheme-only reactions

My Published Elements

Pending Publications

Embargoed Publications

My Collections

All

ELN Gate

My Data

FT-17

C28H38N5

FT-22

C28H37Cu2F12N10P2

FT-15

C12H11N5  
2-[di(pyrazol-1-yl)methyl]pyridine

FT-7

C38H38N5  
1-[(3,5-diphenylpyrazol-1-yl)-(1-methylimidazol-2-yl)methyl]-3,5-diphenylpyrazole

FT-6

C38H38ClN5  
2-[bis(3,5-diphenylpyrazol-1-yl)methyl]-4-chloropyridine

FT-5

C28H22CuF9N9P  
acetonitrile,2-[bis(3-phenylpyrazol-1-yl)methyl]pyridine,copper(1+),hexafluorophosphate  
627.003260 g/mol  
Exact mass: 626.084373 g/mol

Properties

Analyses

QC & curation

Literature

Results

Molecule

Stereo Abs

Stereo Rel

C28H22CuF9N9P

any

any

Top secret

Name

External label

Solvent

Amount

Boiling point

0.000

0.0000

0.000

0.000

°C

Density

Molarity

Purity

Melting point

0.0000

g/ml

1.0000

°C

Solvents

Description

Location

Elemental composition

Chemical identifiers

Close

Save

Download Analysis



# Integration into Curricular Teaching



Empfehlungen  
der GDCh-Studienkommission

zum  
Bachelorstudium Chemie  
an Universitäten

2021

Curricular recommendations of GDCh just came out!

→ Anchoring data literacy and research data management modules already in the B.Sc. studies

BUT: long march through the institutions...





# Integration into Curricular Teaching

---

**More direct and faster way by hidden integration:**

***Example 1: in a Master lecture***

Master lecture (2h per week) in the free area of the Master studies @RWTH

- Sustainable coordinative polymerisation catalysis
- 80-100 students, 50 take the exam
- Explaining chemistry with case studies
- Dissecting the RDM of the case studies (good and bad examples)
- Integrating videos on the basics of RDM from RWTH library/NFDI4Chem



**More direct and faster way by hidden integration:**

***Example 2: lab stage in advanced inorganic chemistry***

- 5th semester in bachelor studies @RWTH
- ~ 130 students
- Introduction of the Electronic Laboratory Notebook Chemotion
- Integrating videos on the basics of research data management, FAIR principles, data management plan, metadata and InChI and SMILES

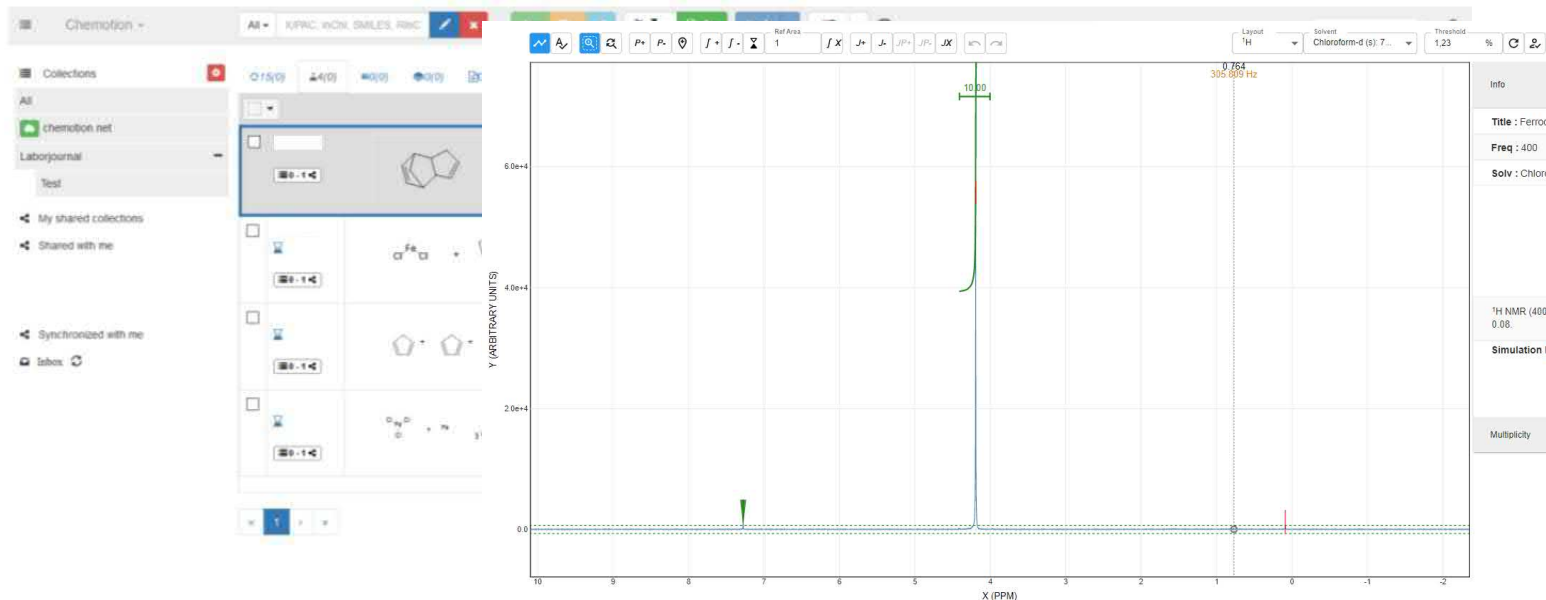


# Integration into Curricular Teaching



More direct and faster way by hidden integration:

*Example 2: lab stage in advanced inorganic chemistry*





- The teaching project "FAIRThesis" is supported by the Fonds der Chemischen Industrie (FCI).
- Focus on the digitization of research data obtained in research internships, Bachelor and Master theses and the processing of FAIR research data (Findable, Accessible, Interoperable and Reusable) in the electronic laboratory notebook Chemotion.
- Installation of Chemotion on a server available via VPN of the RWTH Aachen University
- In addition, students will be provided with in-depth knowledge of research data management and electronic laboratory notebook. We are supported by Dr. Nicole Jung of the Karlsruhe Institute of Technology in the implementation of the project.
- Available starting in January 2022, Contact @RWTH: Dr. Alexander Hoffmann



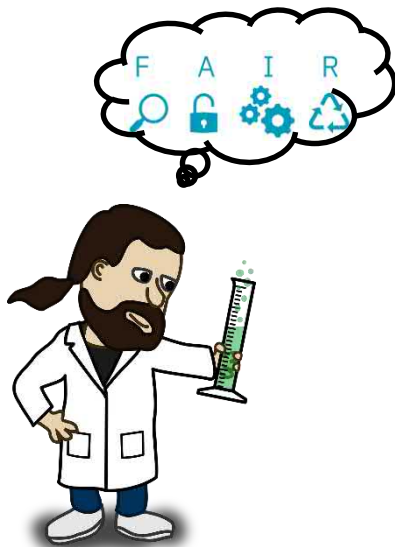


## What's it about?

Award is given for published chemistry research datasets that best meet the [FAIR principles](#).

## How are datasets evaluated?

Available FAIR assessment tools and a jury will evaluate the best dataset.



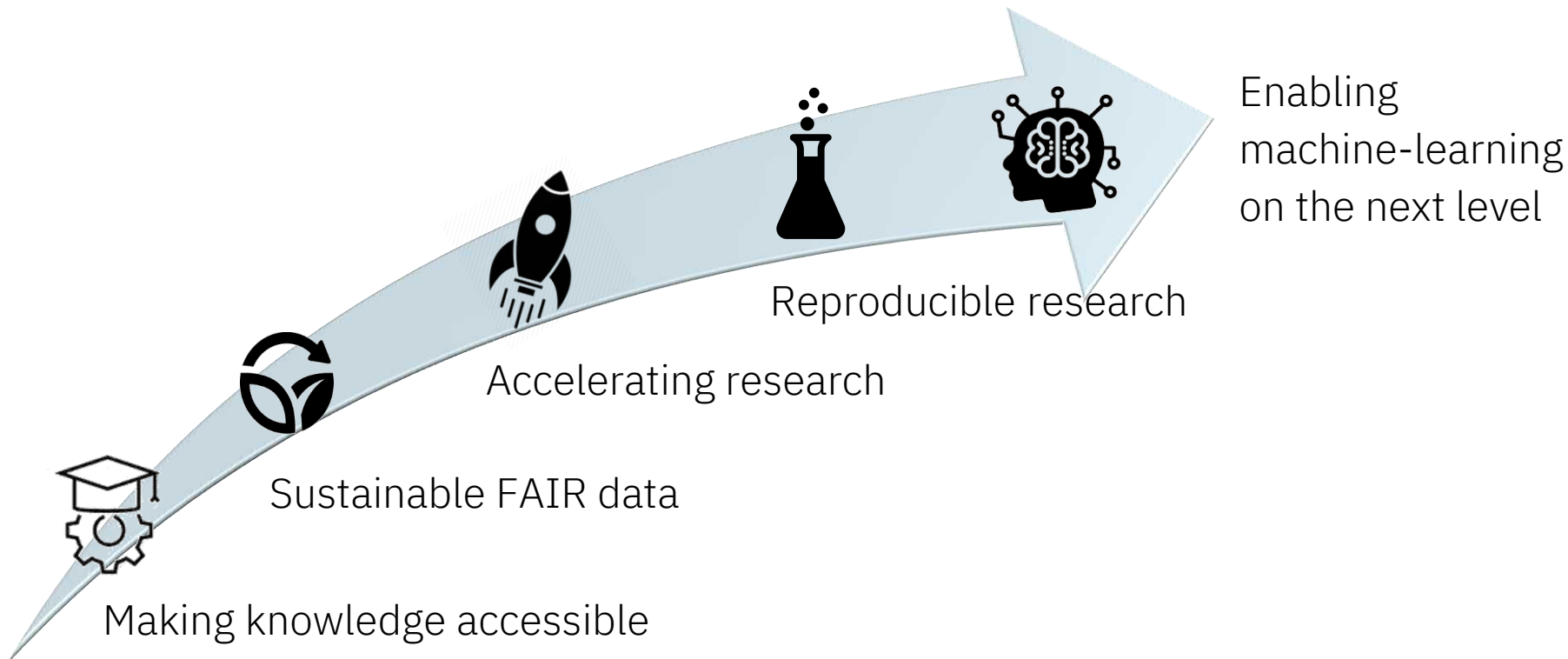
## What's the prize?

Prize money of 500 € (financed by FCI) and talk about winning dataset at JCF Frühjahrssymposium 2022.

## How to participate?

Submit the link to the dataset between October 15<sup>th</sup> and December 15<sup>th</sup> 2021 via [FAIR4Chem Award webpage](#).

# A Quantum Leap in Chemistry RDM



# Follow us | subscribe | stay up to date



[Homepage](#)



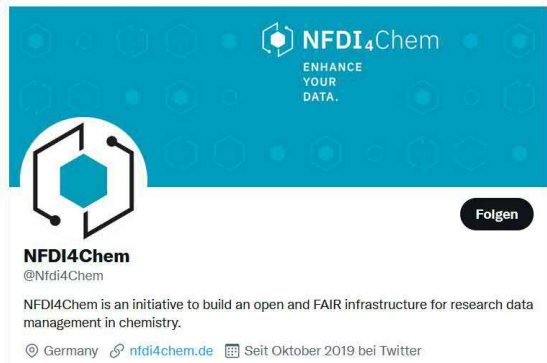
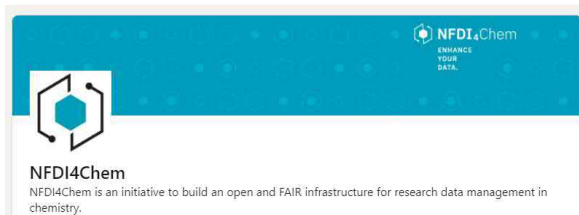
[Twitter](#)



[LinkedIn](#)



[YouTube](#)



NFDI4Chem @Nfdi4Chem · Oct 20

Last week we had our annual consortium meeting. Big thanks to all members of our advisory boards for attending & sharing their valuable insights & to all speakers for their contributions. It was amazing to see such passionate talks about solutions for [#RDM in #Chemistry](#). [@NFDI.de](#)



5 14

Welcome

Dear reader,

Just in time for Christmas, we are delighted to present our first NFDI4Chem newsletter. With the official start of the NFDI4Chem project on 1 October 2020, the consortium is gaining momentum and getting on track. In this first issue we report on the virtual kick-off meeting in October, our first Data Pledge, best practices for using data repositories, latest publications from the consortium, the upcoming joint webinar on ontologies and we announce our "Stammtisch" on Electronic Lab Notebooks.

With the next issue, we will start to introduce the six task areas of NFDI4Chem and the people behind them in more detail. Look forward to comprehensive reports of key topics of NFDI4Chem and NFDI. We will continue to inform you about upcoming events and report on past ones in detail. We will be excited to welcome you at one of our community workshops. Of course, we also keep you informed about what is happening in the NFDI and our cooperation with other consortia. Let us know if you are interested in further topics we should cover.

Enjoy the read!

Merry Christmas and a good start into the year 2021.



Your NFDI4Chem Team

Get to know the consortium!



**Subscribe to our newsletter which is released every quarter!**



**NFDI**<sub>4</sub>Chem

ENHANCE  
YOUR  
DATA.