



NFDI₄Chem

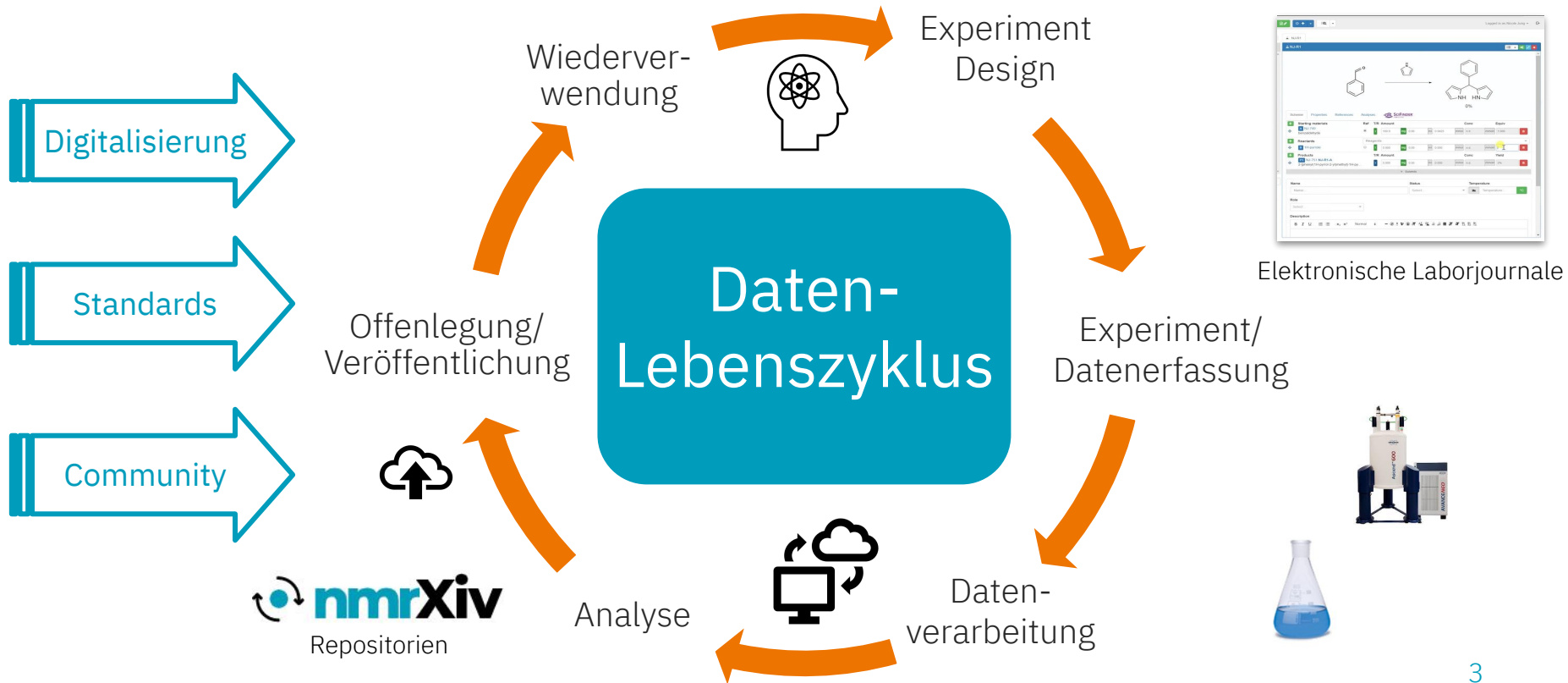
ENHANCE
YOUR
DATA.

NFDI4Chem: Auf dem Weg zu FAIRen NMR-Daten

Tillmann Fischer, Computergestützte Pflanzenbiochemie, Leibniz-Institute für Pflanzenbiochemie, Halle (Saale), PPNMR 2023, 21. März 2023, 14:30 s. t., Max Delbrück Center, Berlin/Buch







NFDI4Chem – NFDI Konsortium für Chemie



FAIR ist kein Modewort, es ist eine Einstellung!



-  **F**_{indable}
 - **Findable**: Leicht auffindbar, unter Verwendung eines eindeutigen und **dauerhaften Identifikatoren** (PIDs). **Beschreibende Metadaten**, die das Auffinden ermöglichen.
-  **A**_{ccessible}
 - **Accessible**: Leichtes Abrufen durch Maschinen und Menschen in **Repositorien** unter Verwendung von **Standardprotokollen**, mit Metadaten und Archivierung.
-  **I**_{nteroperable}
 - **Interoperable**: Verfügbar in **Formaten**, die ausgetauscht und auch nach langer Zeit interpretiert werden können, einschließlich Metadaten.
-  **R**_{eusable}
 - **Reusable**: **Metadaten** gewährleisten die Wiederverwendung für zukünftige Forschung, mit klarer **Lizenz** und **Provenienz**, entsprechend den **Gemeinschaftsstandards**.



Lücken im Datenlebenszyklus von NMR-Daten:



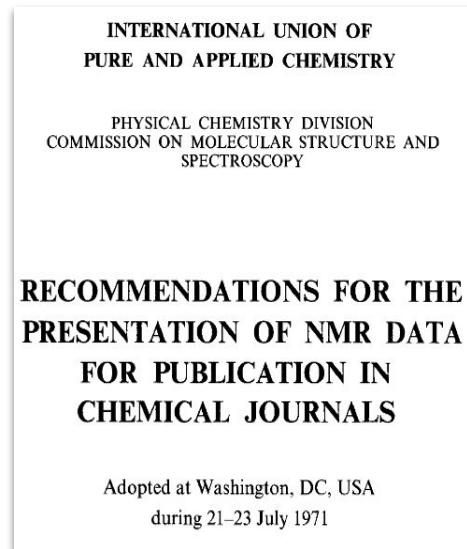
Publikation von NMR-Daten (NMR Text)

-> Basierend auf IUPAC Empfehlungen von 1971 und 1975. Messdaten werden mehrheitlich nicht publiziert.



Wiederverwendung von NMR-Daten

-> schwierig, da Daten oft nicht interoperabel.



Generische Repositorien – RADAR(4Chem)



RADAR
FIZ Karlsruhe – Leibniz Institute for Information Infrastructure

ABOUT RADAR FEATURES RADAR VARIANTS AGREEMENTS & PRICES SUPPORT

RADAR
FIZ Karlsruhe – Leibniz Institute for Information Infrastructure

You are here: [Linderazulen aus einer invas...](#)

Dataset: Linderazulen aus einer invasiven Pflanze

RADAR Metadata Content Statistics Technical Metadata

Related identifier: (References) [10.1002/ciuz.201900868](#) - DOI

Creator/Author: Keltsch, Nils [Leipzig University, Leipzig, Germany]
Munzert, Viola [University of Tübingen, Tübingen, Germany]
Zeller, Klaus-Peter [University of Tübingen, Tübingen, Germany]
Siehl, Hans-Ullrich [University of Ulm, Ulm, Germany]
Berger, Stefan [Leipzig University, Leipzig, Germany] [ORCID](https://orcid.org/0000-0002-8128-9613)
Sicker, Dieter [Leipzig University, Leipzig, Germany] [ORCID](https://orcid.org/0000-0003-3949-782X)

Contributors: (Data Curator) Fischer, Tillmann G. [Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany] [ORCID](https://orcid.org/0000-0003-4480-8661)

Title: Linderazulen aus einer invasiven Pflanze

Additional titles: (Subtitle) Delphi und sein violettes Wunder

Keywords: natural products

Language: English

DOI: [10.22000/786](#)
Publication date: 2022-10-05

Download Dataset
DOWNLOAD (407.2 MB)

Download Metadata
RADAR **DOWNLOAD**

Statistics
13 Views 1 Downloads

Rights statement for the dataset
This work is licensed under [CC BY 4.0 Attribution](#)

Filter by

- Creators
- Publishers
- Production year
- Language
- Subject areas
- Resource
- License
- Rights holders

Feld-spezifische Repositorien – Chemotion Repository



The screenshot displays the Chemotion Repository interface. On the left, a sidebar contains navigation options: Chemotion, Scheme-only reactions, My Published Elements, Pending Publications (highlighted), Embargoed Publications, My Collections, All, ELN Gate, and My Data. The main area shows a list of chemical entries with search filters (All, IUPAC, InChI, SMILES, RInI) and various icons for actions like adding, deleting, and printing. The entries include:

- (5R)-5-methyl-2-propan-2-ylidene-1-cyclohexanone** (C₁₀H₁₆O - abs: (S))
- TIFI-4 (R)-(+)-Pulegon** (C₁₀H₁₆O)
- 1,5,8-trimethylazuleno[6,5-b]furan** (C₁₅H₁₄O)
- TIFI-2 Linderazulene** (C₁₅H₁₄O) - selected
- 3,5,6,8-tetrahydroxy-1-methyl-9,10-dioxo-7-[(2S,3R,4R,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydropyran-2-yl]anthracene-2-carboxylic acid** (C₂₂H₂₀O₁₃)
- TIFI-6 Carmine acid** (C₁₁H₁₀O₇)

The detailed view for **TIFI-2 Linderazulene** is shown on the right. It includes the chemical structure, molecular formula (C₁₅H₁₄O), and physical properties (210.271060 g/mol, Exact mass: 210.104465 g/mol). The **1H NMR** and **13C NMR** data are highlighted with an orange circle:

1H NMR
Type: 1H nuclear magnetic resonance spectroscopy (1H NMR)
Status: Confirmed Instrument: 1/1
Content: 1H NMR (400 MHz, Benzene [7.16 ppm], ppm) δ = 8.22 (s, 1H), 7.51 (d, J = 3.0 Hz, 1H), 7.30–7.29 (m, 2H), 6.95 (d, J = 1.1 Hz, 1H), 2.66 (s, 3H), 2.56 (s, 3H), 1.88 (d, J = 1.3 Hz, 3H). The data was retrieved from automated ...

13C NMR
Type: 13C nuclear magnetic resonance spectroscopy (13C NMR)
Status: Confirmed Instrument: 1/1
Content: 13C NMR (100 MHz, Benzene [128.4 ppm], ppm) δ = 159.3, 139.9, 139.5, 137.0, 134.2, 132.6, 127.5, 125.3, 121.7, 120.1, 117.3, 111.9, 24.8, 13.6, 8.0. The data was retrieved from automated analysis of the uploaded spectra ...

Feld-spezifische Repositorien – nmrXiv.org



Sample chemical composition

100% 1S/C15H14O/c1-9-4-5-12-10(2)6-15-14(7-13(9)12)11(3)8-16-15/h4-8H.1-3H3

SMILES ¹

Or

Need help?

Sample chemical composition



Generische Metadaten

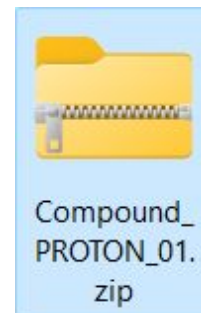
- Dublin Core Metadaten (aus OCLC/NCSA Metadata Workshop 1995 in Dublin/Ohio):
 - ...Typ, Sprache, Autor, Titel, Verleger, Rechte...
-> Bibliotheken und Museen
- DataCite Metadaten (seit 2009. Hauptsitz an der Technischen Informationsbibliothek (TIB) Hannover):
 - DOI-Registrierungsagentur (CrossRef für wissenschaftliche Publikationen)
 - Datensätze, Software, Preprint; auch physische Objekte (PIDINST) -> siehe Dokumentation: resourceTypeGeneral



(Meta-)Daten



Fachspezifische Metadaten
...Struktur, Temperatur, Lösungsmittel...



Mindestinformationstandards (MI-Standards)



“Von der wissenschaftlichen Gemeinschaft **akzeptierter Konsens über die Mindestinformationen**, die über eine chemische Untersuchung berichtet werden sollen, sowie über die Daten und Metadaten, die in einem Repository hinterlegt werden sollen.”

-> **MICHI** = **M**inimum **I**nformation about a **C**hemical **I**nvestigation

FAIR NMR Research Data Management Workshop June/2022 in Karlsruhe,
[Link to Report](#), Chemotion ¹H NMR MI Template ([GitHub](#), [Visualisierung](#))

NMR MICHI? Bereits in den Messdaten enthaltene (Meta-)Daten?

Frage an das Publikum



- *Welche NMR-Datenformaten sind bekannt?*



Rezept für erfolgreiche Datenformate



Zutatenliste:

- **Spezifikation** d.h. vollständige und klare Beschreibung des Formats incl. Datenmodells
 - Für XML und JSON auch im Format des Schemas (XSD, json.schema)
- Referenzimplementierung
 - **reader/writer, converter, validator**
 - als Programmbibliothek zur Implementierung in andere Software
 - weitere Implementierungen in anderen Programmiersprachen
 - **Testdaten**
- **Nutzer** und **Anbieter** von Formaten -> kann zu Henne-Ei-Problem führen



Datenformate für NMR: Herstellerformate



Beispiel: Bruker XWIN-NMR Format

- XWIN-NMR und TopSpin
- Rohdaten (FID/SER): **Binärformat**
- Parameterdaten (acqu*/proc*): **JCAMP-DX** (Text)

- Häufige Eigenschaften von Herstellerformaten:
 - **keine Spezifikation verfügbar**
- Referenzimplementierungen (insb. Validator) sind **keine frei nutzbare Software**

Binary

raw data files
(FID/SER)

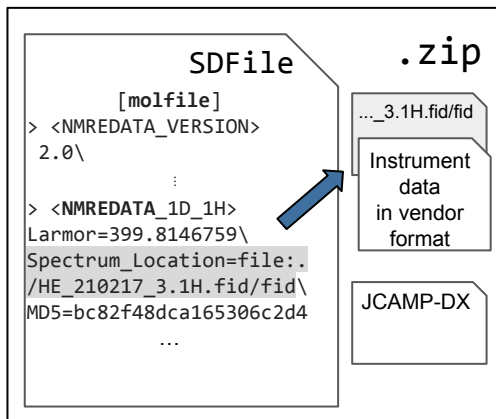
```
0000000000000000040000000000040000000000
00200000000002000000002000100000000000
100001600000000004440000000004000000
040000000444000004000000000000001000
000000000000000006000000000000142101
000240005702000250112510026737 0000100
03651015200612211007637502011001434412
75100312670000120142507041626130510157
32411310706247512351007000600001401415
06114670122110070166133706065305115510
```

...

Datenformate für NMR: Offene Formate



NMReDATA record Container



nmrML XML

```
<?xml version="1.0" encoding="UTF-8"?>
<nmrML xmlns="http://nmrml.org/schema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema
hema-instance" version="1.0.rc1"
xsi:schemaLocation="...nmrml.org/schema
http://nmrml.org/schema/...nmrML.xsd">
<cvList>
<cv id="NMRCV" fullName="Nuclear
Magnetic Resonance CV"
version="1.1.0"...>
</cvList>
<sampleList>
<sample >
<fieldFrequencyLock
fieldFrequencyLockName="cdc13"/> ...
```

JCAMP-DX Text/Tags

```
##TITLE= testntup
##JCAMP-DX= 6.0 $$$...
##DATA TYPE= NMR SPECTRUM
##DATA CLASS= NTUPLES
:
##PAGE= N=1
##DATA TABLE=(X++(R..R)), XYDATA
16383
B254931p506547j928560N577988J225
16375
H070280o319742J809880k462623o675
16367
d539532K5516321441737j691199N043
...
```

D. Rauh, C. Blankenburg, T. G. Fischer, N. Jung, S. Kuhn, U. Schatzschneider, T. Schulze, S. Neumann, *Pure Appl. Chem.* **2022**, *94*, 725. DOI: [10.1515/pac-2021-3101](https://doi.org/10.1515/pac-2021-3101). M. Pupier, J.-M. Nuzillard, J. Wist, N. E. Schlörer, S. Kuhn *et al.*, NMReDATA, a standard to report the NMR assignment and parameters of organic compounds, *Magn. Reson. Chem.* **2018**, *56*, 703. DOI: [10.1002/mrc.4737](https://doi.org/10.1002/mrc.4737). S. Kuhn, L. H. E. Wieske, P. Trevorror, D. Schober, N. E. Schlörer *et al.*, NMReDATA: Tools and applications, *Magn. Reson. Chem.* **2018**, *59*, 792. DOI: [10.1002/mrc.5146](https://doi.org/10.1002/mrc.5146). NMReDATA initiative, URL: <http://nmredata.org/>. NMReData initiative format specification, URL: http://nmredata.org/wiki/NMReDATA_tag_format_2.0. NMReData Initiative, GitHub, URL: <https://github.com/NMReDATAinitiative>. Reference implementation, converter and validator in java available via GitHub. D. Schober, D. Jacob, M. Milson *et al.*, nmrML: A Community Supported Open Data Standard for the Description, Storage and Exchange of NMR Data, *Anal. Chem.* **2018**, *90*, 649. DOI: [10.1021/acs.analchem.7b02795](https://doi.org/10.1021/acs.analchem.7b02795). nmrML, <https://www.nmrml.org/>. nmrML/nmrML GitHub, <https://github.com/nmrML/nmrML>. J. Gasteiger, T. Engel, Uni Erlangen, **2004**, URL: <https://www2.chemie.uni-erlangen.de/projects/vsc/chemoinformatik/erlangen/daten/jcamp.html>. IUPAC, JACMP-DX, Committee on Printed and Electronic Publications (now Committee on Publications and Cheminformatics Data Standards), Subcommittee on Electronic Data Standards (now Subcommittee on Cheminformatics Data Standards), **2005**, URL: <http://www.jcamp-dx.org/>. A. N. Davies, R. M. Hanson, P. Lampen, R. J. Lancashire, An overview of the JCAMP-DX format, *Pure Appl. Chem.* **2022**, *94*, 705. DOI: [10.1515/pac-2021-2010](https://doi.org/10.1515/pac-2021-2010). L. R. McEwen, D. Martinsen, R. Lancashire, P. Lampen, A. N. Davies, Are your spectroscopic data FAIR? *Spectrosc. Eur.* **2018**, *30*, 21. DOI: [10.1255/sew2018.a2](https://doi.org/10.1255/sew2018.a2).



Follow us | subscribe | stay up to date

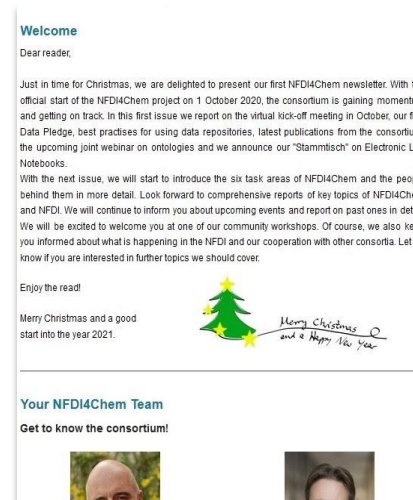
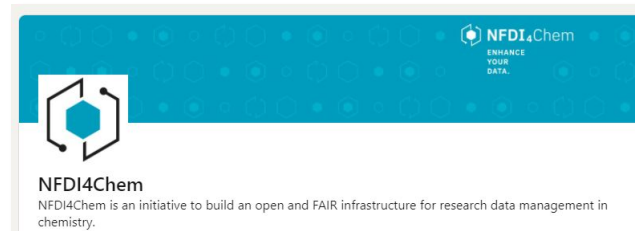


twitter.com/Nfdi4Chem



linkedin.com/company/nfdi4chem

Homepage: nfdi4chem.de



Subscribe to our newsletter which is released every 3 months via our website!

NFDI4Chem Knowledge Base



- Access to Research Data Management (RDM) knowledge via various points of entry

NFDI4Chem Knowledge Base

A place for all knowledge regarding Research Data Management (RDM) in Chemistry

Get started

Roles Domains Handling Data Topics & Concepts Lead by Example

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

NFDI4Chem is funded by DFG
Project Number 441958208

Community
[LinkedIn](#)
[Twitter](#)

Resources
[NFDI4Chem Website](#)
[NFDI4Chem FAQ](#)
[NFDI4Chem Helpdesk](#)
[NFDI4Chem Terminology Service](#)
[GitHub Repository](#)

Legal information
[About](#)
[Legal Notice](#)
[Privacy](#)

Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License, if not stated otherwise.
Copyright © 2021 NFDI4Chem. Built with Docusaurus.



NFDI₄Chem

ENHANCE
YOUR
DATA.

tfischer@ipb-halle.de
Slides: <https://t1p.de/r1b6w>



Datenformate für NMR: Bruker XWIN-NMR Format



Format in XWIN-NMR und TopSpin

- Rohdaten (FID/SER): **Binärformat**
- Parameterdaten (acqu*/proc*): **JCAMP-DX** (Text)
- *Data Formats User Manual* verfügbar (Beschreibung des Datenmodells), jedoch **keine Spezifikation**
- Referenzimplementierung = TopSpin, jedoch **keine frei nutzbare Software**
- **Validator nicht verfügbar**

Binary

raw data files
(FID/SER)

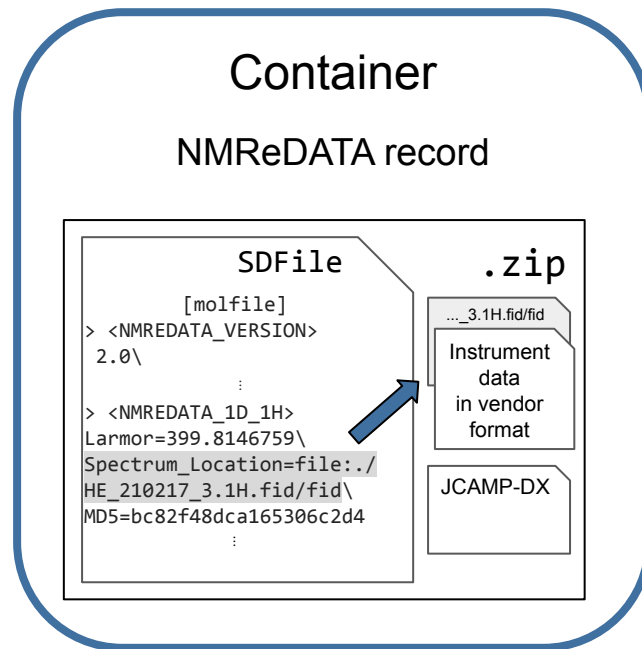
```
00000000000000004000000000004000000000
00200000000002000000002000100000000000
10000160000000000444000000000040000000
0400000000444000004000000000000001000
00000000000000000060000000000000142101
000240005702000250112510026737 0000100
03651015200612211007637502011001434412
75100312670000120142507041626130510157
32411310706247512351007000600001401415
06114670122110070166133706065305115510
```

...



NMReDATA record **Ordner/ZIP**

- Daten in **Herstellerformat**, optional zusätzlich als **JCAMP-DX**
- **SDFfile** (ein CT format)
 - Struktur (molfile)
 - Chem. Verschiebungen
 - Kopplungskonstanten
 - Integrale
 - Zuordnungen,
 - Kreuzpeaks und Korrelatione
- reader/writer, converter verfügbar
- Kein Validator verfügbar





XML-basiertes Format

- Entwickelt für Metabolomics Daten
 - Format für NMR daten im Repositorium [Metabolights](#)
- **Spezifikation, XSD Datei, XSD Dokumentation und nmrCV Ontologie verfügbar**
- **reader/writer, converter und Validator verfügbar**

XML extensible markup language nmrML

```
<?xml version="1.0" encoding="UTF-8"?>
<nmrML xmlns="http://nmrml.org/schema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="1.0.rc1"
xsi:schemaLocation="...nmrml.org/schema
http://nmrml.org/schema/.../nmrML.xsd">
<cvList>
<cv id="NMRCV" fullName="Nuclear Magnetic
Resonance CV" version="1.1.0"...>
</cvList>
  <sampleList>
    <sample >
      <fieldFrequencyLock
fieldFrequencyLockName="cdcl3"/>
      :
    </sample >
  </sampleList>
</nmrML>
```

Datenformate für NMR: JCAMP-DX



```
##TITLE= ACET.SPOROPO.TYPHA/CH-CORR
##JCAMP-DX= 6.00 $$ ISAS JCAMP-DX program (Version 2.0 beta)
##DATA TYPE= NMR SPECTRUM
##DATA CLASS= NTUPLES
##NUM DIM= 2
##ORIGIN= Dr. J. Lambert, ISAS Dortmund, Germany
##OWNER= COPYRIGHT (C) 1999 by ISAS Dortmund, Germany
##.SOLVENT NAME= CDCL3
##SPECTROMETER/DATA SYSTEM= JEOL GX 400
##.FIELD= 9.1
##.OBSERVE FREQUENCY= 100.40
##.OBSERVE NUCLEUS= ^13C
##.PULSE SEQUENCE= HETCOR
s1= 0.0
s3= 0.0
s4= 14
p1= 10.500
p2= 21.000
p3= 28.700
rd= 1000.000
id1= 0.114
d1= 0.114
d2= 10.000
NS= 2720
##.ACQUISITION SCHEME= NOT PHASE SENSITIVE
##.DIGITISER RES= 16
##.ZERO FILL= 0
##RESOLUTION= 5.87
```

Text/Tags with
encoded data

JCAMP-DX

```
##TITLE= testntup
##JCAMP-DX= 6.0 $$ Bruker...
##DATA TYPE= NMR SPECTRUM
##DATA CLASS= NTUPLES
:
##PAGE= N=1
##DATA TABLE=(X++(R..R)), XYDATA
16383 B254931p506547j928560N577988J225
16375 H070280o319742J809880k462623o675
16367 d539532K5516321441737j691199N043
:
```

Datenformate für NMR: JCAMP-DX



Table 2: Example of an IR data file.

```
##TITLE= o-eugenol
##JCAMP-DX= 5.01 $$export from JSpecView
##DATA TYPE= INFRARED SPECTRUM
##DATA CLASS= XYDATA
##ORIGIN= Dept of Chem, UWI, JAMAICA
##OWNER= public domain
##LONGDATE= 1997/07/01 09:24:00
##SPECTROMETER DATA SYSTEM= PERKIN-ELMER 1000 FT-IR
##INSTRUMENTAL PARAMETERS= 4400.00,450.00 cm-1; 16 scans; mode ratio; apod strong
##STATE= liquid
##IUPAC NAME= 2-allyl-6-methoxyphenol
##MOLFORM= C10 H12 O2
##BP= 271
##RESOLUTION= 2
##XUNITS= 1/CM
##YUNITS= TRANSMITTANCE
##XFACTOR= 3.93E-03
##YFACTOR= 9.40262357E-07
##FIRSTX= 470
##FIRSTY= 0.00520004
##LASTX= 4400
##NPOINTS= 3931
##XYDATA= (X+(Y...Y)) 119593E530R12K042n65I78n02q8OJ885K545o2j634K702r4r3j760J759O29K859A7942
124427B0990L048J89k891K263L362J288K325L237N97I14K891K828091K043L865D8454
128753E1031K733M305J666K83L802M902J319L048N311J603n35K703K891N122M022I5934
133079I8416M431L991M336L174L362M493K766N813N844N374N122O630Q7980945A74114
137150A85238J0401q673m682m148m745I519I268K514k828I017I173m368m714o096A32353
.....
1119593H75220 $$checkpoint
##END=
```

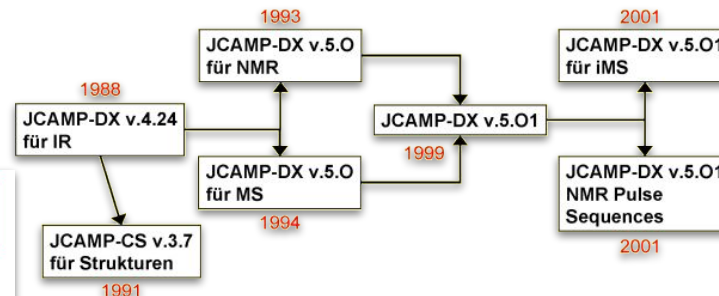
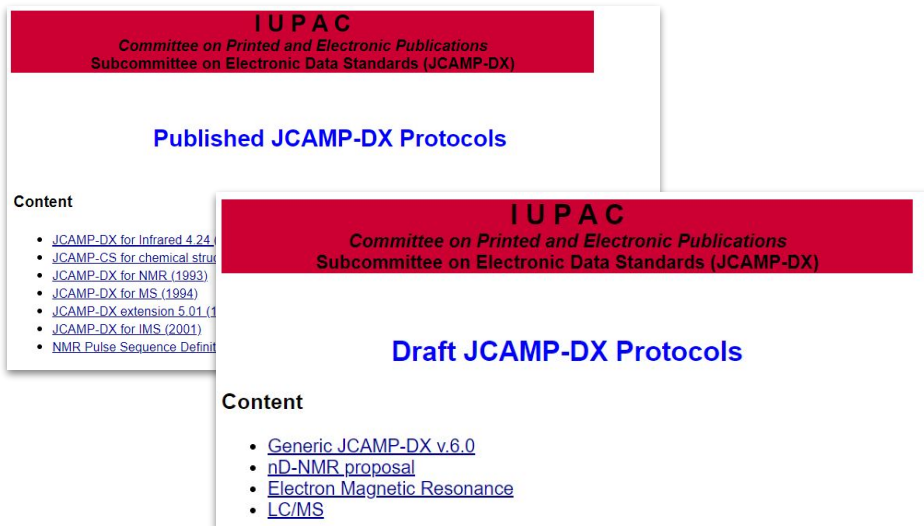
Table 6: Example Bruker NMR JCAMP-DX file showing `private` labels and comments.

```
$$ Bruker specific parameters
$$ -----
##$DATPATH= <C:\Bruker\data\cga\nmr>
##$EXPNO= 28
##$NAME= <cyclosporine-0813>
##$PROCNO= 1
##$ACQTO= -6.36618283677107
##$AMP= (0...31)
100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
##$AMPCOIL= (0...19)
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##$ANAVPT= -1
##$AQSEQ= 0
##$AQ_mod= 3
##$AUNM= <au_sel180zg>
##$AUTOPOS= <>
.....
$$ End of Bruker specific parameters
$$ -----
```

Ausgewählte Erweiterungen:

- *Private Dateneinträge (private data labels) ##\$Descriptor = Wert*
- *Kommentare (comments) \$\$ (one line)*

Datenformate: JCAMP-DX



Weiterentwicklung von JCAMP-DX 2001 mit IUPAC gestoppt zur Vereinigung der Kräfte zur Entwicklung eines XML basierten Formats (AniML).

Entwurf für 2D NMR Daten -> implementiert in TopSpin.

J. Gasteiger, T. Engel, Uni Erlangen, **2004**, URL: <https://www2.chemie.uni-erlangen.de/projects/vsc/chemoinformatik/erlangen/daten/jcamp.html>. D. Rauh, C. Blankenburg, T. G. Fischer, N. Jung, S. Kuhn, U. Schatzschneider, T. Schulze, S. Neumann, Data format standards in analytical chemistry, *Pure Appl. Chem.* **2022**, 94, 725, DOI: [10.1515/pac-2021-3101](https://doi.org/10.1515/pac-2021-3101). IUPAC, JACMP-DX, Committee on Printed and Electronic Publications (now Committee on Publications and Cheminformatics Data Standards), Subcommittee on Electronic Data Standards (now Subcommittee on Cheminformatics Data Standards), **2005**, URL: <http://www.jcamp-dx.org/>. A. N. Davies, R. M. Hanson, P. Lampen, R. J. Lancashire, An overview of the JCAMP-DX format, *Pure Appl. Chem.* **2022**, 94, 705, DOI: [10.1515/pac-2021-2010](https://doi.org/10.1515/pac-2021-2010). L. R. McEwen, D. Martinsen, R. Lancashire, P. Lampen, A. N. Davies, Are your spectroscopic data FAIR? *Spectrosc. Eur.* **2018**, 30, 21, DOI: [10.1255/sew.2018.a2](https://doi.org/10.1255/sew.2018.a2).



Quick Start

Quick Start Guide (English)

Quick Start Guide (Deutsch)

General tools

1D spectra

2D spectra

Structure and Assignment

Teaching

For developers

Include NMRium react component

Loading files from URL

NMRium file format

GitHub

NMRium file format

The NMRium file format is a JSON that may (or may not) be zipped.

In order to determine if the format is zipped you can check the first 4 bytes of the buffer to see if it matches hexadecimal code 50 4b 03 04 (file signature for zip file, https://en.wikipedia.org/wiki/List_of_file_signatures).

Edit this page

Previous

Next

« Loading files from URL

Writing end-to-end tests »

- in JSON

JSON

JavaScript Object Notation

```
{
  "$schema": "http://.../"
  "title" : "...",
  "description" : "JSON-...",
  "type" : "object",
  "properties" : {
    "@id": { "type": "string",
    "filename": { "type" : "st
    "identifier": { "type" :
    "title" : { "type" : "stri
    "description" : { "type" :
    :
  }
```