A cross-platform format to associate NMR-extracted data (NMReDATA) to chemical structures

Academic partners: Julien Wist[1,§], Jean-Marc Nuzillard[2], Nils E. Schloerer[3], Stefan Kuhn[3], Marion Pupier[4], Mate Erdelyi[5], Christoph Steinbeck[6], Antony J. Williams[9], Craig Butts[10], Bozhana Mikhova[11,*], Damien Jeannerat[4,§]

Industrial partners: Kessler Pavel[#], Fabrice Moriaud,[#] Mikhail Elyashberg[%], Dimitris Argyropoulos[%], Manuel Pérez[&,§]

Publishing partner: Patrick Giraudeau[7,§], Roberto R. Gil[8,@], Gary Martin[‡,@], Paul Trevorrow[£]

[1] **University of Valle**, Cali, Colombia
[2] CNRS and **University of Reims-Champagne-Ardenne**, Reims, France
[3] IDNMR, **University of Cologne**, Cologne, Germany
[4] Department of organic chemistry, **University of Geneva**, Geneva, Switzerland
[5] Department of Chemistry - BMC, **University of Uppsala**, Sweden
[6] Cheminformatics and Chemometrics, **Friedrich-Schiller-University Jena**, Germany
[7] CEISAM, **University of Nantes**, Nantes, France
[8] Department of Chemistry, **Carnegie Mellon University**, Pittsburgh, USA
[9] National Center for Computational Toxicology, **EPA**, USA
[10] **University of Bristol**, UK
[11] Institute of organic chemistry, **Bulgarian academy of sciences**, Bulgaria

[#] **Bruker BioSpin GmbH**, Rheinstetten, Germany
[%] **ACD/Labs**, Toronto, Canada
[&] **Mestrelab Research**, Santiago de Compostela, Spain
[‡] **Merck and co**. New-Jersey, USA

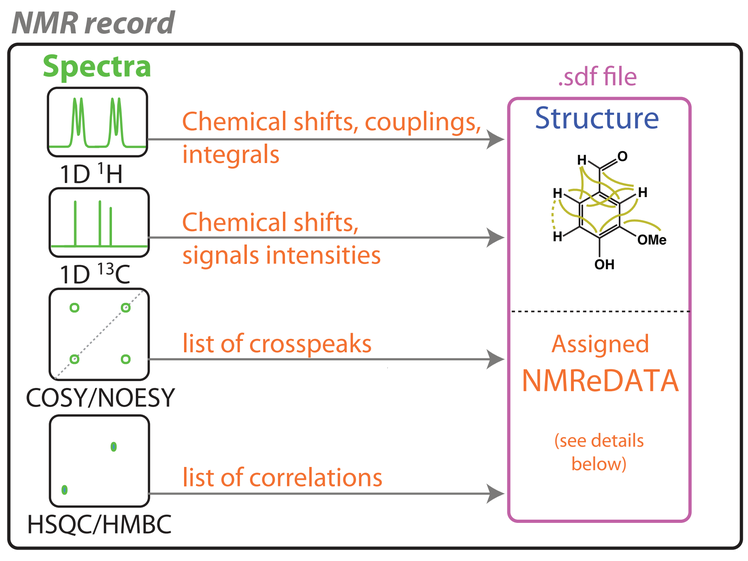
Magnetic Resonance in Chemistry, Wiley, Chichester, UK
[@] Coeditors in Chief
[*] Member of the *Editorial board*
[§] Member of the *Associate board*
[£] Executive journal editor

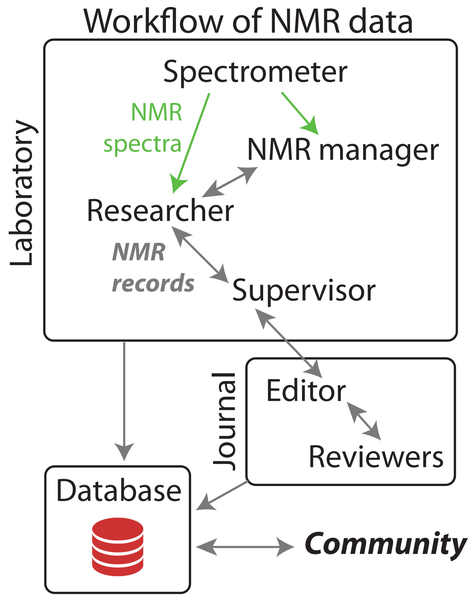
The NMReDATA initiative

The goal of the NMReDATA initiative is to introduce a manner to associate the data extracted from a "full" NMR analysis (the NMReDATA) to a chemical structure.

We introduced a file format based on the commonly used "Structure Data Format" (.sdf) to combine the chemical shifts, couplings, lists of 2D correlations and assignment (NMReDATA) with a chemical structure in the .mol format.

NMR records (NMR spectra + NMReDATA) including the .sdf file will be generated by computer-assisted structure elucidation software or web-based tools under development.





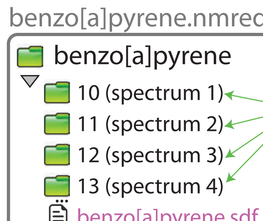
Important benefits of the new format

- **Improved quality** of the NMR data for researchers and the community
- Straightforward **inclusion** of NMR data in **reports and journal articles**
- **Simplified referee work**
- **Compatibility** with electronic storage in **database**
- Easier **comparison** of dataset
- Improved **searchability** of NMR data

D. Jeannerat, Magn. Res. in Chem., 2017, 55, 7-14.
J. Bisson, C. Simmler, S.-N. Chen, J. B. Friesen, D. C. Lankin, J. B. McAlpine, G. F. Pauli, Nat. Prod. Rep., 2016, 33, 1028

Detailed structure of *NMR records*

benzo[a]pyrene.nmredata.zip



A ***record*** is a folder or database entry containing:
1) All the NMR spectra of a compound (including FID's, acquisition and processing parameters).
2) The .sdf file containing the NMReDATA

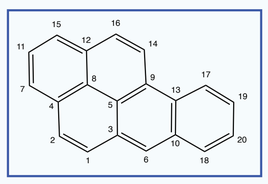
"structure" part of the file

```
benzo(a)pyrene
demo of .sdf file containing NMReDATA
test_generationV1 1                    0999 V2000
   -1.6583    2.2334    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.6633    3.1556    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.3753    0.8191    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7234    2.7603    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.3848    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
 ....
    2.6960   -0.4398    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.5264   -3.3746    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.1687   -2.4653    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.5724   -4.2723    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.9104   -3.8155    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0
  1  3  1  0  0  0
  2  4  1  0  0  0
  3  5  1  0  0  0
  3  6  2  0  0  0
  4  7  1  0  0  0
  4  8  2  0  0  0
  5  8  1  0  0  0
  5  9  2  0  0  0
  6 10  1  0  0  0
  7 11  2  0  0  0
  8 12  1  0  0  0
  9 13  1  0  0  0
  9 14  1  0  0  0
 10 13  2  0  0  0
 10 18  1  0  0  0
 11 15  1  0  0  0
 12 15  2  0  0  0
 12 16  1  0  0  0
 13 17  1  0  0  0
 14 16  2  0  0  0
 17 19  2  0  0  0
 18 20  2  0  0  0
 19 20  1  0  0  0
M  END
```

The format is the same as .mol files (compatible with most chemical structure editor)

list of atoms with coordinates (either 2D as in "flat" structures used for drawing or 3D)

list of bonds



Explicit structures (including all hydrogen atoms) should be favored to facilitate the assignment of non-equivalent hydrogens. For aromatic compounds implicit hydrogen is fine.

"NMReDATA" part of the file

```
>  <NMREDATA_VERSION>
0.96

>  <NMREDATA_ID>
no ide yet

>  <NMREDATA_SOLVENT>
cdcl3

>  <NMREDATA_TEMPERATURE>
298.15

>  <NMREDATA_ASSIGNMENT>
(1), 128.1712, 1
(2), 127.9300, 2
(3), 129.9, 3
(4), 131.6, 4
(5), 123.8, 5
(6), 124.8, 6
(7), 124.9, 7
(8), 128.3, 8
(9), 127.4, 9
(10), 131.35, 10
(11), 126.09, 11
(12), 131.4, 12
(13), 128.23, 13
(14), 122.2, 14
(15), 125.6, 15
(16), 127.5, 16
(17), 123.0, 17
(18), 128.9, 18
(19), 125.97, 19
(20), 126.06, 20
H(1), 8.02, H1
H(2), 7.94, H2
H(6), 8.53, H6
H(7), 8.10, H7
H(11), 7.98, H11
H(14), 9.07, H14
H(15), 8.25, H15
H(16), 8.34, H16
H(19), 7.85, H19
H(20), 7.79, H20
H(17), 9.06, H17
H(18), 8.30, H18
```

The .sdf file can include "tags". Each tag has a name between "> <" and ">" and data (one or more lines terminating with an empty line.

NMReDATA tags start with "NMREDATA_".

This tag associates the labels used in the assignment and the atom(s) of the molecule.

Labels used for the assignment (see right column). In this example, the proton bound to the atom 1 in the structure part (in blue above) was called "H(1)" and the carbon "(1)", but the authors can chose the labels according to IUPAC rules or the requirement of the journal where the data are published.

Chemical shifts in ppm

Atom number in the .mol part (see the blue frame above). When using implicit hydrogen, the atom number to which the hydrogen atom is bound is given following "H".

1D spectra

```
>  <NMREDATA_1D_1H>
;note that integrals (E) were not measured in the spectrum but set
to 1
Spectrum_Location=file:./nmr/10/pdata/1/
Larmor=500.13
9.0600, L=H(17), S=d, J=8.44(H(19)), E=1
9.0700, L=H(14), S=d, J=9.13(H(16)), E=1
8.5300, L=H(6), S=s, E=1
8.3400, L=H(16), S=d, J=9.13(H(14)), E=1
8.3000, L=H(18), S=d, J=8.06(H(20)), E=1
8.2500, L=H(15), S=d, J=7.65(H(11)), E=1
8.1000, L=H(7), S=d, J=7.31(H(11)), E=1
8.0200, L=H(1), S=d, J=9.07(H(2)), E=1
7.9800, L=H(11), S=dd, J=7.31(H(7)), 7.65(H(15)), E=1
7.9400, L=H(2), S=d, J=9.07(H(1)), E=1
7.8500, L=H(19), S=dd, J=6.77(H(20)), 8.44(H(17)), E=1
7.7900, L=H(20), S=dd, J=6.77(H(19)), 8.06(H(18)), E=1
Spectrum_Location =file:./10/1/pdata/1


>  <NMREDATA_1D_13C>
;note that intensisites were not measured in the spectrum 110 for CH
100 for Cq
Spectrum_Location=file:./nmr/13/pdata/1/
Larmor=125.0
Decoupled=1H
131.6000, L=(4), I=100.00
131.4000, L=(12), I=100.00
131.35, L=(10), I=100.00
129.9, L=(3), I=100.00
128.9, L=(18), I=110.00
128.17, L=(1), I=110.00
128.23, L=(13), I=100.00
127.9, L=(2), I=110.00
127.4, L=(9), I=100.00
127.5, L=(16), I=110.00
126.09, L=(11), I=110.00
126.06, L=(20), I=110.00
128.3, L=(8), I=100.00
125.97, L=(19), I=110.00
125.6, L=(15), I=110.00
124.8, L=(6), I=110.00
124.9, L=(7), I=110.00
123.8, L=(5), I=100.00
123.0, L=(17), I=110.00
122.2, L=(14), I=110.00
Spectrum_Location =file:./11/1/pdata/1
```

Chemical shift (or range in case of overlap or complex multiplet)

Label (see left)

Multiplicity ("d" for doublet, etc.)

Coupling constants. They can be assigned as here: $J_{H(11),H(7)}$ = 7.3 Hz

Integral (crude values - here they were normalized)

Chemical shift in ppm

Label (see the NMREDATA_ASSIGNMENT tag)

Signal intensity (usefull to distinguish signals with NOE and, in DEPT spectra, to determine the sign of the peak. Here they were set to 100 and 110 but the experimental values should be given here.)

2D spectra

```
>  <NMREDATA_2D_1H_NJ_1H>
Larmor=500.13
CorType=COSY
H(1)/H(2)
H(7)/H(11)
H(11)/H(15)
H(14)/H(16)
H(19)/H(20)
H(19)/H(17)
H(20)/H(18)
Spectrum_Location =file:./12/1/pdata/1/


>  <NMREDATA_2D_13C_1J_1H>
;one-bound correlations test of comment line
Larmor=500.13/comment 2aa
CorType=HSQC
(1)/H(1)
(2)/H(2)
(6)/H(6)
(7)/H(7)
(11)/H(11)
(14)/H(14)
(15)/H(15)
(16)/H(16)
(17)/H(17)
(18)/H(18)
(19)/H(19)
(20)/H(20)
Spectrum_Location =file:./13/pdata/1/
...


>  <NMREDATA_LITERATURE>
Source=Journal
DOI=....


>  <...>


$$$$
```

Isotope in F1 / type of mixing / isotope in F2

"N": multiple-bond, "J": scalar coupling mixing

Correlation of "H(1)" with "H(2)".

Correlation of "(1)" with "H(1)".

Link to the spectra (the data must include the fid, and the processed data). This points either to a folder where the .sdf file is located or to a permanent record in an open-access database.

HMBC, NOESY data would follow with the same format

Reference to the main publication associated to the record.

Other data such as the origin of the sample, the sample preparation and other analytics data could (and should) be added to .sdf file ***outside*** the NMReDATA initiative.

End of file code

Timeline of the initiative

Mid-2016: Proposition by the members of the Associate editorial board of Magnetic resonance in Chemistry to request authors to submit NMR spectra and the extracted data in a manner allowing serious reviewing and to become a source or reliable peer-reviewed NMR data.

September 2016: **Decision** of the Editorial board of Magnetic Resonance in Chemistry to **request NMR data for structure papers**.

Until March 2017: Elaboration of a beta version of the format to include NMR data in .sdf files.

July 2017: **Official announcement of the *Initiative*** at the Euromar 2017, Warsaw, Poland.

By August 2017: **Elaboration and tests of the cheminfo.org web-based platform** to host the embargoed NMR records during the peer-review process and open them to the public upon acceptance for publication.

By September 2017: **Implementation of import/export** features by the providers of computer-assisted structure elucidation software.

September 2017: **Round-table discussion at the SMASH** conference (Baveno, Italy) and decision on the version 1.0 of the file format.

From January 2018: NMReDATA and spectra will be requested for all structure papers submitted to Magnetic Resonance in Chemistry.

Mid-2018: Evaluation of the initiative. If satisfactory, invitation to other journals to join the initiative.

During 2018: Contact with IUPAC to consider changes in the recommendation to report small-molecule NMR data.

Follow the progress of the initiative on
www.nmredata.org

Acknowledgments

Damien Jeannerat thanks the State of Geneva and the Swiss NSF funds 200021-147069, 200020-126650, 206021-1288746 for financial support.This work does not necessarily reflect U.S. EPA policy.