

Guidance Document

NMR Data Sharing on DATAVERSE

How to Prepare and Share Raw NMR Data along with Your Publication

Enhancing the Transparency and Reproducibility of Structure Elucidation while Improving the Dereplication Process

Contributors:

Seon Beom Kim (UIC)

Charlotte Simmler (UIC)

Jonathan Bisson (UIC)

Julian Gautier (Dataverse)

Mercè Crosas (Dataverse)

Philip Durbin (Dataverse)

Satishchandra Prabhakar Achanta (UIC)

Guido Pauli (UIC)



University of Illinois at Chicago



Overview

Consider in advance that you will need

- The raw NMR data
- The annotated, fully assigned ^1H and ^{13}C NMR spectra in some electronic format.

The file names should be chosen such that users will understand intuitively the content of each file.

The steps are as follows:

- Step 1** Create a new Dataset
- Step 2** Describe the detailed Metadata
- Step 3** Upload NMR data files on Dataset
- Step 4** Describe the uploaded NMR data

Stepwise Process

A. Steps 1-2: Create A New Dataset and Fill in the Metadata

- Testing the creation of a dataset only try on the **Demo Version Dataverse** (<https://demo.dataverse.org>)
- you can create your own dataverse (for your laboratory publications for example) and create a **new dataset** (<https://dataverse.harvard.edu/>) within this dataverse.
- **Your dataset will now have an attributed DOI**

HARVARD Dataverse

Search About User Guide Support Seon Beom Kim

Deposit and share your data. Get academic credit.
Harvard Dataverse is a repository for research data. Deposit data and code here.
82,260 datasets 6,436,086 downloads

Organize datasets and gather metrics in your own repository.
A dataverse is a container for all your datasets, files, and metadata.
3,184 dataverses

Add a dataset + **Click!** Add a dataverse +

Test_title Draft Unpublished

Kim, Seon Beom, 2019, "Test_title", <https://doi.org/10.7910/DVN/KUZTX0>, Harvard Dataverse, DRAFT VERSION

Cite Dataset

Learn about Data Citation Standards.

Description Test_Description (2019-04-02)

Subject Chemistry

Keyword Test_term

Related Publication Test_Citation issn: test

Notes Test_Note

Tap here! Specify contents in the metadata

Files Metadata Terms Versions Add + Edit Metadata

Citation Metadata

Dataset Persistent ID doi:10.7910/DVN/KUZTX0

Title Test_title

METADATA

*Asterisks indicate required fields

Citation Metadata ^

Title *
Full title by which the Dataset is known.

Enter title...
Add "Replication Data for" to Title

Author *
The person(s), corporate body(ies), or agency(ies) responsible for creating the work.

Name * Kim, Seon Beom
Affiliation University of Illinois at Chicago
Identifier Scheme Select...
Identifier

Contact *
The contact(s) for this Dataset

Name Kim, Seon Beom
Affiliation University of Illinois at Chicago
E-mail * skim592@uic.edu

Description *
A summary describing the purpose, nature, and scope of the Dataset.

This field supports only certain HTML tags.

Text *

Date YYYY-MM-DD

- **Title:** The title can be related to the type of compounds described in the NMR data or to your research article.
- **Author:** Add authors with name, affiliation, identifier (e.g ORCID #). Use your ORCID only, avoid the other author identifier schemes such as ISNI. If you do not have a ORCID, create one!
- **Contact:** Indicate who among the authors will be the contact for this dataset.
- **Description:** Indicate what this dataset contains
 - Purpose of the research (see also Related publication below)
 - Describe the NMR dataset (1D; ^1H and ^{13}C , 2D; COSY, HSQC, HMBC) with the name of your structure, its Pubchem ID if known (with <http://link>)
 - Indicate whether you also performed a purity determination of the corresponding compounds
- **Subject:** Choose the specific subject categories : eg. Chemistry
- **Keyword:** will help scientists to find your datasets in the repository (structure elucidation, quantitative NMR, NMR raw data). Just fill Term, unless you are using ontologies.
- **Related publication: IMPORTANT**

HERE you can actually link your dataset to your (just published) research article by adding its DOI. Also if your article is not yet published when you release your data, you can still add this information later. Everything stays together and related to each other.

Subject *
Domain-specific subject Categories that are topically relevant to the Dataset.

Agricultural Sciences
 Arts and Humanities
 Astronomy and Astrophysics
 Business and Management
 Chemistry

Keyword
Key terms that describe important aspect of the Dataset

Term

Vocabulary

Vocabulary URL
 Enter full URL, starting with http://

Related Publication
Publications that use the data from this Dataset

Citation

ID Type
 Select...

ID Number

URL
 Enter full URL, starting with http://

Notes
Additional important information about the Dataset

Depositor

Deposit Date

Contributor
The organization or person responsible for either collecting, managing, or otherwise contributing in some form to the development of the resource.

Type
 Select...

Name

Grant Information
Add grant here

Grant Agency

Grant Number

- **Subtitle:** you can possibly add a secondary title or certain limitations on the main title.
- **Alternative Title:** A title which commonly referred or abbreviation of the title.
- **Producer:** Person or organization with the financial or administrative responsibility over this Dataset
- **Producer Date:** Date when the data collection or other materials were produced (not distributed, published or archived)
- **Production Place:** The location where the data collection and any other related materials were produced.
- **Contributor: DON'T FORGET TO ACKNOWLEDGE ANYBODY WHO PARTICIPATED IN DATA ACQUISITION!**

B. Steps 2-3: Prepare and upload your NMR data

Files

For more information about supported file formats, please refer to the User Guide.

Upload with HTTP via your browser [^](#)

File upload limit is 2.5 GB per file. Select files or drag and drop into the upload widget.

+ Select Files to Add

Upload your NMR data Drag and drop here. Drag and drop files here.
Also, possible via the dropbox.

Select files from Dropbox.

Upload from Dropbox

Metadata Tip: After adding the dataset, click the Edit Dataset button to add more metadata.

Save Dataset Cancel

- Prepare your NMR data **on your personal computer:**

Create your data folder containing the following information

1. Folder with all the raw 1D and 2D NMR data

- a. Name your folder with Compound_Magneticstrength_Solvent_NMRdataformat
Example Naming: NameOfCpd_RawNMRdata_TypeOfInstr_Solv_MagFld
Basically the name of your folder should contain essential information describing your NMR data
- b. Each spectrum/subfolder within this folder should have its title:
Example: ^1H , gCOSY, HSQC, ^{13}C _DEPT135

2. Folder containing the “jdx” formats of all your 1D and 2D NMR data (optional)

- a. With the same name scheme and title as described above
Example Naming: NameOfCpd_NMRdata_TypeOfInstr_Solv_MagFld.jdx
Offering you NMR data as jdx. format should allow a more “universal” use of your data with different NMR software or platforms.

3. PDF/ png files with fully assigned/annotated ^1H and ^{13}C NMR spectra

- Although it is not absolutely necessary for sharing your NMR data, this type of document will facilitate the interpretation of your data. The annotated spectra go hand in gloves with any of your publication related to the description of new structures. Such document will promote transparency, and is educational at the same time!
- a. If the structure is complex, create a document with expansion of the region of interest for more clarity, and then save/export your compiled documents in PDF (or png) formats.

- b. Name your document with essential information:
Example Naming: NameOfCpd_annotated_NMRspct_Solv_MagFld.pdf

4. The structure of your compound as a “.mol” file

- a. Name of compound (e.g. TestCpd01)
Example Naming: NameOfCpd.mol

5. Folder with quantitative NMR analysis

If you use quantitative NMR (qNMR) for purity determination or quantitation of compounds. This folder should contain all the elements necessary to reproduce or calculate the purity of your compound.

- a. ^1H NMR spectrum acquired under quantitative conditions (Bruker format or *.jdf, and *.jdx)
b. The calculation spreadsheet used to determine the purity, using a calibrant or the 100% method
See also <https://gfp.people.uic.edu/qnmr/content/qnmrcalculations/index.html>
c. The annotated ^1H NMR spectrum with the integrals for each considered signals (*.pdf or *.png)

Example of your dataset on your personal computer:

Name	Date modified	Type	Size
TestCpd01_HiFSA_Solv_MagFld	4/3/2019 11:12 AM	File folder	
TestCpd01_jdx_Solv_MagFld	4/3/2019 1:03 PM	File folder	
TestCpd01_TestInstr_Solv_MagFld	4/3/2019 11:11 AM	File folder	
TestCpd01.mol	11/13/2017 3:52 PM	Molecule Definition	3 KB
TestCpd01_Annot_1HNMRspec_Solv_Magfld.pdf	2/23/2018 3:29 PM	Adobe Acrobat D...	48 KB
TestCpd01_Annot_1HNMRspec_Solv_Magfld.png	12/7/2017 5:35 PM	PNG File	59 KB
TestCpd01_Annot_13CNMRspec_Solv_Magfld.pdf	2/23/2018 3:35 PM	Adobe Acrobat D...	71 KB
TestCpd01_Annot_13CNMRspec_Solv_Magfld.png	2/23/2018 3:25 PM	PNG File	46 KB

Note: [HiFSA](#) refers to ^1H iterative full spin analysis, i.e the complete description of your NMR spin parameters detailing altogether the chemical shifts, coupling constants and line width of your ^1H NMR spectrum.

● **Upload your NMR data on your created Harvard Dataset**

1 File

Click Here!!

+ Upload Files

Edit Files -

Download

test-1.zip
ZIP Archive - 1.5 MB - Apr 3, 2019 - 0 Downloads
MD5: 056fcac763dd71fe8202fc87538203a9
Test_Description



To keep your Bruker raw NMR data together in one single folder you may need to **double zip*** your folder before uploading it. Otherwise a single zip is preferred for uploading your folder.



C. Step 4: Describe your data

.....As important as your Overarching Metadata section.....

1) Describe the NMR information for your uploaded raw NMR data

- Type of NMR spectroscopy (e.g. 1D and 2D)
- Type of NMR file (e.g. *.jdx, Bruker fid and *.jdf)
- NMR tube and Solvent (provider, batch, lot)
- Qty of compound analysed in the NMR tube
- Precise whether your compound has been isolated from natural source, synthesized, or whether it is a commercial compound
- Pubchem ID of structure (If already known compound)

2) Describe the qNMR information for your uploaded quantitative analyses

- Method used for purity determination : EC (External Calibration), IC (Internal Calibration) or 100% Methods. See related article [here](#).
- Identity of the NMR calibrant, qty in the NMR tube
- Qty of compound in the NMR tube
- Volume of solvent in the tube
- Purity certificate of your chosen NMR calibrant (e.g. CRM or 100% method qHNMR spectrum)
- [Calculation spreadsheet](#) utilized for the determination of purity
- Annotated NMR spectrum with integrals used for purity determination

→ You can also check a few examples of uploaded NMR datasets at :

<https://dataverse.harvard.edu/dataverse/cenaptnmr>

D. Final step: Publish your dataset!

You can save a "draft" of your dataset before publishing it, and add gradually more information to metadata and data at a later time when they are ready.

Once everything has been uploaded (and your research article is accepted for publication) and you are happy with everything, you can push on the "publish" button.

Your NMR data can now be used and **cited** by other scientists thanks to the attributed DOI: you just have given a second life to your data, while benefiting the scientific community and enhancing your impact!

Congratulations !!

E. Some [References](#)

1. Choules, M. P., Bisson, J., Gao, W., Lankin, D. C., McAlpine, J. B., Niemitz, M., ... Pauli, G. F. (2019). Quality Control of Therapeutic Peptides by ¹H NMR HiFSA Sequencing. *The Journal of Organic Chemistry*, 84(6), 3055–3073. <https://doi.org/10.1021/acs.joc.8b02704>
2. McAlpine, J. B., Chen, S.-N., Kutateladze, A., MacMillan, J. B., Appendino, G., Barison, A., ... Pauli, G. F. (2019). The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Natural Product Reports*, 36(1), 35–107. <https://doi.org/10.1039/C7NP00064B>
3. Bisson, J., Simmler, C., Chen, S.-N., Friesen, J. B., Lankin, D. C., McAlpine, J. B., & Pauli, G. F. (2016). Dissemination of original NMR data enhances reproducibility and integrity in chemical research. *Natural Product Reports*, 33(9). <https://doi.org/10.1039/c6np00022c>
4. Pauli, G. F., Chen, S., Simmler, C., Lankin, D. C., Go, T., Jaki, B. U., ... Napolitano, J. G. (2014). Importance of Purity Evaluation and the Potential of Quantitative ¹H NMR as a Purity Assay. *Journal of Medicinal Chemistry*, 57, 9220–9231. <https://doi.org/10.1021/jm500734a>
5. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. Retrieved from <https://doi.org/10.1038/sdata.2016.18>
6. Wilkinson, M. D., Dumontier, M., Jan Aalbersberg, I., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2019). Addendum: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 6(1), 6. <https://doi.org/10.1038/s41597-019-0009-6>

***Zip or double Zip**, note from Mercè Crosas: the latest version of Dataverse maintains the folder structure when you upload the data in a zip file. If you want to take advantage of this feature, you don't want to double zip. Only if you have a very large number of files in the zip, double zip could have an advantage since extracting the folders and files from a zip could be a performance issue in some cases. If performance is not an issue, extracting the zip file into the folder structure is recommended for transparency and preservation.