Pillarimines and pillaramines: Essential pillars within supramolecular chemistry

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
OBSERVATIONAL DATA AND PROCEDURES ON CHEMICAL SYNTHESIS	Detailed notes on the used amount of reagents, procedures, observations and the obtained amount of product, created by chemical synthesis, written down in analog lab notebooks or the electronic lab notebook (ELN).	***N***	**D** and **P**	**I**mages **T**extual	The digitalized data of the lab notebooks and ELN are available in PDF format. The processed data, involving the procedure, purification, yield and characterization ready for implementation in papers, have been stored as .docx files.	<1GB	1-2 lab notebooks of 100 pages each
CHEMICAL COMPOUNDS	Vials containing mg to g quantities of synthesized compounds (mainly for final compounds).	***N***	**D** and **P**	**N** **T**	An inventory is kept in .xls files	<1GB	50-60 samples

NMR	Spectra are recorded on the NMR spectrometers (https://nmrfacilities.chem.kuleuven.be/instrument_specifications.php) Data folders generated by the spectrometer.	***N***	**D**	**I N T**	Data folders generated by the spectrometer containing raw data (.txt and .exec) and all relevant acquisition and processing parameters. Processed data are in the form of PDF files, images or textual interpretation (.docx) of the spectra ready for implementation in a paper (e.g. 1H NMR (D2O, 300 MHz): δ 1.61 (s, 6H), 1.68 (s, 3H), 1.72 (s, 3H), 2.17-1.99 (m, 8H), 4.45 (dd, 2H, J(H,H) = 6 Hz, J(P,H) = 6 Hz), 5.23- 5.15 (m, 2H), 5.46 (t, 1H, J = 6 Hz).)	<100GB	
UV-VIS-NIR	Spectra are recorded on the UV-VIS-NIR spectrometer	***N***	**D**	I N T	Data folders generated by the spectrometer containing raw data and all relevant acquisition parameters in .BSW format. Processed data are available in .xls and image format.	<1GB	

FLUORESCENCE SPECTROSCOPY	Spectra are recorded on a fluorescence spectrometer	***N***	**D**	I N T	Data folders generated by the spectrometer containing raw data and all relevant acquisition parameters, in .fs format, are kept by the colleague recording the data. Processed data are available in .csv, .xls and image format	<1GB	
CD SPECTROSCOPY	Spectra are recorded on a CD spectrometer	***N***	**D**	I N T	Data folders generated by the spectrometer containing raw data and all relevant acquisition parameters, stored as .jws files which can be converted to .txt format. Processed data are available in .xls and image format.	<1GB	
HRMS	Spectra are recorded on an HRMS spectrometer	***N***	**D**	I	Data folders generated by the spectrometer containing raw data and all relevant acquisition and processing parameters are kept by the colleague recording the data in .mzXML format. Processed data are available as PDF files.	<1GB	

MPLC	Chromatograms are recorded on a medium pressure liquid chromatography apparatus.	***N***	D		Data folders generated by the chromatography apparatus containing raw data and all relevant acquisition and processing parameters in .txt format. Processed chromatograms are stored as PDF files.	<1GB	
SINGLE CRYSTAL X-RAY DATA	The raw datasets are recorded on the X-ray diffractometer. After integrating the reflections and processing, the cif files are provided to the researchers and uploaded to https://www.ccdc.cam.ac.uk in the framework of publications.	***N***	D	I T	Data folders generated by the X-ray diffractometer containing raw data and all relevant acquisition and processing parameters in .cif format. The cif format is a standard format drafted by the international union of pure and applied crystallography. It contains the reflections (indices h,k,l), observed structure factors and standard deviation. Processed data are stored as image files.	<1GB	

THEORETIC CALCULATIONS	The raw datasets are generated and processed by calculations using Gaussian software in collaboration with the Quevedo group (National University of Cordoba, Argentina).	***N***	**D**	I N T	Data folders generated by calculations using Gaussian software containing the raw data, the coding and the processing parameters are kept by the colleague recording the data, obtained in .log format and converted to .mol2 format. Processed data are stored as image files.	<1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

Two types of data will be gathered during this project. Chemical data from experiments (reaction conditions, experimental sequence, observations, amount of reagents used) will be stored by the researcher in Microsoft OneDrive (2 TB), MBook ELN(Electronic Lab Notebooks) and analog lab notebooks. Corresponding structural identification data for the obtained compounds (raw and processed) will be uploaded in the ELN and linked to the correct experiment making data traceable. All project data will be shared with the PI and other collaborators via Microsoft OneDrive and via the ELN.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

An Excel file will be provided to allow for fast and efficient searching through the data. The file gives an overview of the experiment number, the structure of the compound, the location in the lab book and the physical location of the sample. All digital data will be stored in the group's shared drive with a clear mention of the researcher's and project names. The data file will be subdivided per data type to increase the accessibility of the data. Additionally, the file directories of the ELN will be provided, which 6 groups all relevant synthetic and characterization data. For each binding study, an additional Excel file will be generated which allows for interpreting and reproducing the calculations.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- OneDrive (KU Leuven)
- Other (specify below)

Compounds will be stored at the laboratory in the KU Leuven Chem&Tech or in its stock room 01.186.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Other (specify below)

ELN cloud service

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

Microsoft OneDrive provides 2 TB of storage per user. Archival storage is rented at the KU Leuven ICTS data center and is expanded if needed.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

OneDrive and the ELN are not publicly accessible and password protected. Access needs to be granted by the researcher and will be limited to the PI. For collaborations, accessibility to the relevant files will be granted if described in a non-disclosure agreement. Archival storage at KU Leuven is not publicly accessible and only people with permission can access the data.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The use of OneDrive is free of charge if the capacity of 2 TB is not exceeded. Archival data storage is centrally offered via KU Leuven at 270 Euro/TB/Year. MBook ELN and the cloud service are offered by Mestrelab at 10 Euro/month for the subscription and a one-time license fee of 120 Euro. These costs are being covered by the general operating budget of the research group, or by the individual bench fee of researchers.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy
- Certain data cannot be kept for 10 years (explain below)

Samples will be stored for 10 years in the depository as long as the long-term stability and quantity of the compound allow it.

Where will these data be archived (stored and curated for the long-term)?

- Shared network drive (J-drive)
- Other (specify below)

Physical samples are stored at the laboratory in the KU Leuven Chem&Tech or in its stock room 01.186. All data will also be backed up in the ELN.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Data archival at KU Leuven is currently offered at 270 Euro/TB/year. Since no large datasets are generated in this project, one TB of storage will be sufficient and the price for storage during 10 years is 2700 Euro. These costs are being covered by the general operating budget of the research group or by the individual bench fee of researchers.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as embargoed data (temporary restriction)
- Yes, as open data
- No (closed access)

Relevant synthetic procedures, the corresponding characterization data and interaction studies of the synthesized macrocycles with guest molecules are being bundled for publication in peer-reviewed journals and will be made accesible via open access or via the KU Leuven digital repository after a one year embargo (LIRIAS; https://limo.libis.be/). Incomplete datasets will not be made public, allowing for the set-up of new projects and the continuation of the group's research.

If access is restricted, please specify who will be able to access the data and under what conditions.

/

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other (specify below)

Complete datasets will be published in dedicated journals. Incomplete datasets will not be made public, allowing for the set-up of new projects and the continuation of the group's research.

When will the data be made available?

- Upon publication of research results

Complete datasets will be published in dedicated journals. Incomplete datasets will not be made public, allowing for the set-up of new projects and the continuation of the group's research.

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

Data from the project that can be shared will be made available under a creative commons attribution license (CC-BY 4.0).

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

What are the expected costs for data sharing? How will these costs be covered?

No costs are related to depositing data in the LIRIAS repository of KU Leuven. No charges apply to publishing supporting information related to publications. Exceptionally, data will be published at a publisher charging a publication fee (around 1500 Euro).

Responsibilities

Who will manage data documentation and metadata during the research project?

The researcher is responsible for collecting all relevant data files and entering the observational data in the ELN and analog lab notebooks.

Who will manage data storage and backup during the research project?

The researcher is responsible for storing all relevant data in Microsoft OneDrive, MBook ELN and analog lab notebooks.

Who will manage data preservation and sharing?

The PI is the end responsible for ensuring data preservation and reuse.

Who will update and implement this DMP?

The DMP is updated in agreement with the PI.