DMP du projet "From halogenation in marine fungi to biocatalysts"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Plan Details

Plan title	DMP du projet "From halogenation in marine fungi to biocatalysts"
Version	Version initiale
Fields of science and technology (from OECD classification)	
Language	eng
Creation date	2021-12-09
Last modification date	2022-09-16
Identifier	DMP "HALO-CAT" project_ANR-21-CE44-0003-01
Identifier type	doi
License	**Name**: Creative Commons Attribution Non Commercial 4.0 International **URL**: http://spdx.org/licenses/CC-BY-NC-4.0.json

Project Details

Project title	From halogenation in marine fungi to biocatalysts
Acronym	HALO-CAT
Abstract	Preliminary work in the laboratory has shown the potential of marine fungi in the production of natural halogenated compounds with the detection not only of several new molecules but also of enzymes catalyzing the halogenation of substrates. In this project, we would like to both describe and understand the mechanisms involved in halogenation in these organisms, but also to propose innovative strategies for obtaining halogenated molecules of interest for therapeutics. Thanks to the expertise acquired in the isolation of natural products, in metabolomics and bioinformatics, and thanks to our preliminary work on v-HPO enzymes as biocatalysts, we hope to provide solutions based on Nature, more respectful of the environment, to fight against antibiotic resistance in particular. This work will therefore be valuable both in the fields of Health, Green chemistry and Biotechnologies.
Funding	• Agence Nationale de la Recherche : ANR-21-CE44-0003-01 • French National Research Agency : ANR-21-CE44-0003
Start date	2021-12-01
End date	2025-11-30
Partners	• Nantes Université 03gnr7b55

Research outputs :

1. HPLC-MS/MS data on the different extracts (Jeu de données)
2. Extract collection from fungal cultures and derivatives (after enzymatic reactions) (Collection)
3. Pure compounds isolated throughout the project (Collection)
4. Enzymes produced and purified throughout the project (Collection)
5. Scripts (R and/or Python) and software developed throughout the project for data treatment (Logiciel)
6. List of identified molecules by dereplication from the different extracts (Texte)
7. NMR data on pure compounds obtained (Jeu de données)
8. Biological assays results (on extracts, fractions and compounds) (Jeu de données)

Contributors

Name	Affiliation	Roles
Roullier Catherine - https://orcid.org/0000-0001-6486-6821	UR2160 ISOMer_Institut des Substances et Organismes de la Mer - 199613759C	• DMP manager • Personne contact pour les données (Extract collection, HPLC-MS/MS data, Compound collection, Enzyme collection, Scripts and software, Dereplication, NMR data, Biological data)
ROULLIER Catherine	MER, MOLECULES ET SANTE - 199613759C	• Project coordinator

DMP du projet "From halogenation in marine fungi to biocatalysts"

1. Data description and collection or re-use of existing data

HPLC-MS/MS data on the different extracts

1a. How will new data be collected or produced and/or how will existing data be re-used?

HPLC-MS/MS data will be acquired after analyses on HPLC-MS/MS instruments (Shimadzu IT-TOF, Thermo LT-Q or other instruments). Existing data acquired in the lab (on the same type of instruments) prior to the project, which have not been published yet, can also re-used. Data provenance will be documented from the folders generated on both the instrument computer and the user computer.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Data are first obtained in a raw format (*.lcd, *.raw, ...) on the different instruments and further converted to *.mzXML and/or *.cdf files using either the instrument vendor softwares, the msconvert tool or any other suitable script and software. Existing HPLC-MS/MS data in the converted format *.mzXML and *.cdf files will also be used.

The *.mzXML and *.cdf file formats are open and standard formats for MS data, which facilitate sharing and long-term re-use of data.

Extract collection from fungal cultures and derivatives (after enzymatic reactions)

1a. How will new data be collected or produced and/or how will existing data be re-used?

All fungal extracts will be produced by means of extraction using appropriate solvents on fungal cultures performed in the lab. If fungal extracts previously obtained are to be used in the project, their internal reference code will be used.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The extracts correspond to chemical material (mixtures of compounds), which will be dried and weighed in appropriate containers (labeled hemolysis tubes for small-scale extracts <200 mg and larger flasks for large-scale extracts > 200 mg).

Pure compounds isolated throughout the project

1a. How will new data be collected or produced and/or how will existing data be re-used?

Pure compounds will be produced by means of purification using appropriate methods such as chromatography, or by

buying them to suppliers. If compounds previously obtained are to be used in the project, their internal reference code will be used.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The pure compounds correspond to chemical material, which will be dried and weighed in appropriate containers (labeled hemolysis tubes for small-scale extracts <200 mg and larger flasks for large-scale extracts > 200 mg).

Enzymes produced and purified throughout the project

1a. How will new data be collected or produced and/or how will existing data be re-used?

Pure enzymes will be produced by means of heterologous expression in appropriate hosts or native production and further purification using appropriate methods such as affinity chromatography, or by getting them through collaboration. If enzymes previously obtained are to be used in the project, their internal reference code will be used.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The enzymes correspond to biological material, which will be preserved in buffered aqueous solution in appropriate containers (labeled eppendorf tubes).

Scripts (R and/or Python) and software developed throughout the project for data treatment

1a. How will new data be collected or produced and/or how will existing data be re-used?

Scripts and softwares will be produced by means of different tools such as R, Python or HTML encoding language. If pre-existing scripts or softwares will be used (such as MeHaloCoA or FiBiCo previously developed), they will be referred as with their code. Modifications of these latter might be considered to better apply our needs of data treatment.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Scripts and softwares produced will correspond to informatic language programming on individual files or complete packages including different files to be installed on a computer.

List of identified molecules by dereplication from the different extracts

1a. How will new data be collected or produced and/or how will existing data be re-used?

Dereplication will be performed on LC-MS data by means of peak picking, adduct searching, accurate mass calculating, molecular formula predicting, and search in appropriate natural product databses such as DNP, NPAtlas, Lotus, SciFinder. MS/MS fragmentation pattern may also be used to complete the dereplication process by using either the GNPS (https://gnps.ucsd.edu/) online molecular networking approach and associated tools (MetGem, MS2LDA, ...) or by using in-house developed solutions. All this work may be done manually or with more automated tools.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The data (dereplication results) will mainly correspond to files (such as .xls, or .cy for molecular networks) recording all data annotated with at least the following information : m/z of the parent ion, retention time, accurate mass, adduct information, molecular formula predicted, and names of hit compounds.

NMR data on pure compounds obtained

1a. How will new data be collected or produced and/or how will existing data be re-used?

NMR data will be acquired after analyses on NMR instruments (Bruker or other, depending on external outsourcing). Existing data acquired in the lab (on the same type of instruments) prior to the project, which have not been published yet, can also re-used. Data provenance will be documented from the folders generated on both the instrument computer and the user computer.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Data are first obtained in a raw format on the different instruments and further analysed by means of appropriate softwares such as MestReNova or Topspin. The NMR file formats are standard formats, which facilitate sharing and long-term re-use of data.

Biological assays results (on extracts, fractions and compounds)

1a. How will new data be collected or produced and/or how will existing data be re-used?

Biological data will be produced after testing extracts, fractions or pure compounds on different targets though different outsourcing contractors.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Data collected will correspond to growth inhibition percentages at different concentration, minimal 50% inhibiting concentration (IC50), inhibition diameter on microbiological cultures or other kind of biological activities. These will be provided in electronic files (.xls or .doc).

2. Documentation and data quality

HPLC-MS/MS data on the different extracts

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Raw data files contain the following information : Name of instrument, Date of acquisition, Mode of acquisition, Method used and post-acquisition treatments performed such as calibration. Their names contain the date of the acquisition (first day of the sequence launched).

For each sequence of analyses (at a given time), generated converted files will be accompanied by a sequence file (text file) describing the date of acquisition, the name of the instrument, the mode of acquisition, the methods used, post-acquisition treatment and the method used for conversion.

2b. What data quality control measures will be used?

All data will be calibrated using a calibration solution according to the specifications of the instruments. Moreover, mixtures of samples (named QC) will be acquired for each sequence of analyses several times during the sequence to insure reproducibility and quality of the analyses. Blank samples will also be systematically added to any analysis.

Extract collection from fungal cultures and derivatives (after enzymatic reactions)

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata associated to fungal extracts correspond to all the information around their obtention (date, operator, methodology,strain code, type of culture, duration, mass obtained, mass available, location). All these information will be gathered in a file.

2b. What data quality control measures will be used?

Quality of extracts will be controlled by standardized procedures of extraction, involving solvents and appropriate filtration.

Pure compounds isolated throughout the project

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata associated to pure compounds correspond to all the information around their obtention (date, operator, purification methodology, extract code, fraction involved, mass obtained, mass available, location, analyses and assays performed on it). All these information will be gathered in a file.

2b. What data quality control measures will be used?

Quality of pure compounds will be controlled by means of different types of analyses such as LC-MS, LC-UV and NMR.

Enzymes produced and purified throughout the project

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata associated to enzymes correspond to all the information around their obtention (date, operator, purification methodology, mass obtained, mass available, location, analyses and assays performed on it). All these information will be gathered in a file.

2b. What data quality control measures will be used?

Quality of enzymes will be controlled by using standardized procedures and by appropriate methods of analysis (of their size and activity).

Scripts (R and/or Python) and software developed throughout the project for data treatment

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata and documentation accompanying the data (scripts and softwares) will be both in the form of explicative sentences throughout the lines of program, and as a separate text file explaining how to use the data.

2b. What data quality control measures will be used?

Data (scripts and softwares) quality will be controlled by testing the scripts and softwares on different computers (PC) by different users (beta testers).

List of identified molecules by dereplication from the different extracts

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata accompanying dereplication results will correspond to the operator, the methodology used, the different parameters and thresholds defined at each step of the process, the reference of the data and samples treated, the total number of peaks and the number of peaks annotated, and the location of the results. All these information will be gathered in a file.

2b. What data quality control measures will be used?

Data quality (dereplication) will be insured by crossing the results from different approaches and methods (different databases, different parameters applied for peak picking, different softwares and scripts).

NMR data on pure compounds obtained

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Raw data files contain the following information : Name of instrument, Date of acquisition, Mode of acquisition, Method used and post-acquisition treatments performed such as calibration. Their names contain the date of the acquisition and the type of experiment performed.

For each set of analyses, a folder gathering all analyses performed on one sample is provided.

2b. What data quality control measures will be used?

All data will be obtained from appropriately maintained and calibrated instruments, according to their specifications. Moreover, to insure quality of data and stability of samples analysed, the same analysis will be repeated several times thoughout the sequence to insure reproducibility and quality of the analyses.

Biological assays results (on extracts, fractions and compounds)

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Metadata will consist in separate files (.doc) gathering all the information about methodology, date of testing and comments on the results provided.

2b. What data quality control measures will be used?

All biological assays will be performed with positive and negative controls to insure the validity of the results.

3. Storage and backup during the research process

HPLC-MS/MS data on the different extracts

3a. How will data and metadata be stored and backed up during the research?

All data and metadata will be first stored on the instrument computers in a folder with the following data path : DATA\Catherine\name of student\20XX-XX-XX\ . These data are regurlarly backed up on the University servers. Moreover, all converted data will also be stored both on individual computers (from students and coordinator) and on a shared online folder on UNCLOUD developed by Nantes Université.

3b. How will data security and protection of sensitive data be taken care during the research

The protection of sensitive data is insured by the Nantes University Information Technology department.

Extract collection from fungal cultures and derivatives (after enzymatic reactions)

3a. How will data and metadata be stored and backed up during the research?

Data (extracts) will be preserved dried at -20°C. Backing up will be provided by preparing replicates of each extract when possible and preserve them in different freezers.

All metadata associated to extracts obtained will be stored both on individual lab books from the students (electronic or paper version) and on a shared online folder on UNCLOUD developed by Nantes Université. Moreover, the extracts generated throughout the project will implement the in-house collection of compounds at UR2160 (ISOMer lab) and associated metadata will be recorded in the in-house database of extract collection.

3b. How will data security and protection of sensitive data be taken care during the research

Security and protection of data (extracts) will be provided by restricted access (with magnetic card) to the building.

Security and protection of sensitive metadata will be insured by the retention of lab books inside the lab building and by informatic security provided by Nantes University Information Technology department.

Pure compounds isolated throughout the project

3a. How will data and metadata be stored and backed up during the research?

Data (pure compounds) will be preserved dried at -20°C. Backing up will be provided by preparing replicates of each compound when possible and preserve them in different freezers.

All metadata associated to pure compounds obtained will be stored both on individual lab books from the students (electronic or paper version) and on a shared online folder on UNCLOUD developed by Nantes Université. Moreover, the compounds generated throughout the project will implement the in-house collection of compounds at UR2160 (ISOMer lab) and associated metadata will be recorded in the in-house database of compound collection.

3b. How will data security and protection of sensitive data be taken care during the research

Security and protection of data (pure compounds) will be provided by restricted access (with magnetic card) to the building.

Security and protection of sensitive metadata will be insured by the retention of lab books inside the lab building and by

informatic security provided by Nantes University Information Technology department.

Enzymes produced and purified throughout the project

3a. How will data and metadata be stored and backed up during the research?

Data (enzymes) will be preserved in buffered solutions at -20°C. Backing up will be provided by preparing replicates (different aliquots of each enzyme) and preserve them in different freezers.
All metadata associated to enzymes obtained will be stored both on individual lab books from the students (electronic or paper version) and on a shared online folder on UNCLOUD developed by Nantes Université.

3b. How will data security and protection of sensitive data be taken care during the research

Security and protection of data (enzymes) will be provided by restricted access (with magnetic card) to the building.
Security and protection of sensitive metadata will be insured by the retention of lab books inside the lab building and by informatic security provided by Nantes University Information Technology department.

Scripts (R and/or Python) and software developed throughout the project for data treatment

3a. How will data and metadata be stored and backed up during the research?

Data (scripts and softwares) and metadata will be stored both on individual computers (of the students and the coordinator), on a shared online folder on UNCLOUD developed by Nantes Université, and on GitHub after publication. This will insure backing up of the research.

3b. How will data security and protection of sensitive data be taken care during the research

Security and protection of sensitive metadata will be insured by the informatic security provided by Nantes University Information Technology department.

List of identified molecules by dereplication from the different extracts

3a. How will data and metadata be stored and backed up during the research?

Data (dereplication results) and metadata will be stored both on individual computers (of the students and the coordinator), on a shared online folder on UNCLOUD developed by Nantes Université, and online as supplementary information after publication. This will insure backing up of the research.

3b. How will data security and protection of sensitive data be taken care during the research

Security and protection of sensitive metadata will be insured by the informatic security provided by Nantes University Information Technology department.

NMR data on pure compounds obtained

3a. How will data and metadata be stored and backed up during the research?

All data and metadata will be first stored on the instrument computers from outsourcing contractor. Moreover, all data provided will also be stored both on individual computers (from students and coordinator) and on a shared online folder on UNCLOUD developed by Nantes Université.

3b. How will data security and protection of sensitive data be taken care during the research

The protection of sensitive data is insured by the Nantes University Information Technology department.

Biological assays results (on extracts, fractions and compounds)

3a. How will data and metadata be stored and backed up during the research?

All data and metadata will be first stored on the computers from outsourcing contractor. Moreover, all data provided will also be stored both on individual computers (from students and coordinator) and on a shared online folder on UNCLOUD developed by Nantes Université.

3b. How will data security and protection of sensitive data be taken care during the research

The protection of sensitive data is insured by the Nantes University Information Technology department.

4. Legal and ethical requirements, code of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

No personal data will be processed in this project.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

The data will remain accessible to the people involved in the project among the UR2160 research team. They might be also re-used for further projects. They will only be made publically available after publication of corresponding research articles.

Intellectual property rights will be questioned and evaluated at each step of the research project by the coordinator. When needed, the coordinator will get in touch with the Nantes Université service for innovation and valorisation.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

No ethical issues are present in this project.

---

5. Data sharing and long-term preservation

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Data will be first retained in the laboratory where the research is conducted (UR 2160). It will be stored for 5 years and re-evaluation of their storage will be performed every 5 years.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Data will be made publically available only after publication of research articles on suitable repositories such as GNPS.ucsd.edu for MS/MS data or as supplementary information in related articles.

5c. What methods or software tools are needed to access and use data?

Data will be shared in open and re-usable formats. No specific tool will be needed.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

As data will only be shared through article publication, a persistent identifier (doi) will be provided for each set of data.

---

6. Data management responsibilities and resources

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The coordinator of the project (Catherine Roullier) as well as Aurore Michaud (Assistant Engineer) will be responsible for data management throughout the project.

A common document will be shared with all people involved in the production of new data thoughout the project to gather all information about the data collected and their storage. This document will be reviewed by the coordinator and will allow the safe retrieval of data.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Data storage and sharing has been anticipated in the project by allocating funds to the acquisition of preservation material such as freezer and freeze-dryer, as well as for the fees for open-source publications.