DNA Display of Ligands or Reactive Groups for Multivalent Spatial Recognition, Recruitment and Arrangement of Proteins

Data Management Plan

08 January 2026

Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»
using Common DSW Knowledge Model v2.6.13 (dsw:root:2.6.13).

HISTORY OF CHANGES		
Version	**Publication date**	**Changes**
There are no named versions		

Contributors

The following contributors are related to the project of this DMP:

- **Prof. Michal Hocek, Ph.D., DSc.**
  hocek@uochb.cas.cz, ORCID: 0000-0002-1113-2047
  Roles: *Contact Person*, *Project Leader*, *Supervisor*
  Affiliation:
  Czech Academy of Sciences, Institute of Organic Chemistry and Biochemistry (ÚOCHB AV ČR)
- **RNDr. Marek Ondruš, Ph.D.**
  marek.ondrus@uochb.cas.cz, ORCID: 0000-0002-7653-1866
  Roles: *Creator of DMP*, *Data Collector*, *Data Manager*, *Data Steward*, *Project Manager*, *Researcher*
  Affiliation:
  Czech Academy of Sciences, Institute of Organic Chemistry and Biochemistry (ÚOCHB AV ČR)

Projects

We will be working on the following project and for those are the data and work described in this DMP.

DNA Display of Ligands or Reactive Groups for Multivalent Spatial Recognition, Recruitment and Arrangement of Proteins

Acronym:

N/A

Project Number:

N/A

Start date:

2025-01-01

End date:

2027-12-31

Funding:

Grantová Agentura České Republiky: 25-15592S (granted)

1. Data Summary

Instrument datasets

The following instrument datasets will be acquired in the project:

- **(no name given)**
  This dataset will be collected by experts in the project, with our own equipment.

  The equipment is very well described and known.

  Other researchers working in the same field of research could be interested in using this data.

Re-used datasets

We identified the following reference datasets considered for potential re-use:

- Worldwide Protein Data Bank
  The dataset can be used in the provided format without any conversion needed.

  We will keep a copy of the dataset and make it available with our results for the reproducibility.

There is no need to harmonize different sources of existing data in our case.

We will be using data that needs to be (re-)made computer readable first. We will provide machine readable, standardised metadata to others. The data itself will be made available in computer readable form to others through a standard repository.

Data formats and types

We will be using the following data formats and types:

- Nuclear Magnetic Resonance Extracted Data Format
  It is a standardized format. This is a suitable format for long-term archiving. We will have only a small amount of data stored in this format.

2. FAIR Data

2.1. Making data findable, including provisions for metadata

- **raw data** (published)
  There won't be different versions of this data over time.
  We will not be adding a reference to any data catalogue because the data will be stored in a repository that is the prime source of data for re-use in the field.

There are no 'Minimal Metadata About ...' (MIA...) standards for our experiments. However, we have a good idea of what metadata is needed to make it possible for others to read and interpret our data in the future.

We will use an electronic lab notebook to make sure that there is good provenance of the data analysis.

We made a SOP (Standard Operating Procedure) for file naming. We will be keeping the relationships between data clear in the file names. All the metadata in the file names also will be available in the proper metadata.

2.2. Making data accessible

We will be working with the philosophy *as open as possible* for our data.

All of our data can become completely open over time.

Data that is not legally restrained will be released after a fixed time period (immediately after legal restrictions), unconditionally.

Metadata will be openly available including instructions how to get access to the data. Metadata will be available in a form that can be harvested and indexed.

All data will be owned by the institute.

For the reference and non-reference data sets that we reuse, conditions are as follows:

- Worldwide Protein Data Bank
  It is freely available for any use (public domain or CC0).

For our produced data, conditions are as follows:

- **raw data** (published)
  The dataset will published after an embargo.

2.3. Making data interoperable

We will be using the following data formats and types:

- Nuclear Magnetic Resonance Extracted Data Format
  It is a standardized format.

We will be using the following standards (encodings, terminologies, vocabularies, ontologies):

-

2.4. Increase data re-use

The metadata for our produced data will be kept as follows:

- **raw data** (published) – This data set will be kept available as long as technically possible. – The metadata will be available even when the data no longer exists.

As stated already in Section 2.2, all of our data can become completely open over time.

We will be archiving data in *cold storage* systems for long-term preservation already during the project. The data are expected to remain interpretable and reusable over time.

To validate the integrity of the results, the following will be done:

- We will run a subset of our jobs several times across the different compute infrastructures.
- We will be instrumenting the tools into pipelines and workflows using automated tools.
- We will use independently developed duplicate tools or workflows for critical steps to reduce or eliminate human errors.
- We will run part of the data set repeatedly to catch unexpected changes in results.

3. Other research outputs

We use Data Stewardship Wizard for planning our data management and creating this DMP. The management and planning of other research outputs is done separately and is included as appendix to this DMP. Still, we benefit from data stewardship guidance (e.g. FAIR principles, openness, or security) and it is reflected in our plans with respect to other research outputs.

4. Allocation of resources

FAIR is a central part of our data management; it is considered at every decision in our data management plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible.

- **DNA Display of Ligands or Reactive Groups for Multivalent Spatial Recognition, Recruitment and Arrangement of Proteins**
  Following resources will be dedicated to data management and ensuring that data will be FAIR:
    - **(no name given)** -
      This cost will be covered by funding grant.

We will be archiving data in cold storage systems for long-term preservation after the project but also already during the project. The costs for archiving data will be paid out of departmental budgets from one or more of the project participants. The minimum lifetime of the archive is 10 years. The archival period can be extended – one of the principle investigators involved in the project will decide. The decision whether or not to extend the renewal be based on the predicted use of the archived

data. Data formats in cold storage will be upgraded if they become obsolete. Archived data will be migrated regularly to more modern storage media (e.g. newer tapes).

None of the used repositories charge for their services.

We have a reserved budget for the time and effort it will take to prepare the data for publication. For making data or other research outputs FAIR, we budgeted: 50000 CZK.

RNDr. Marek Ondruš, Ph.D. is responsible for implementing the DMP, and ensuring it is reviewed and revised.

RNDr. Marek Ondruš, Ph.D. is responsible for finding, gathering, and collecting data.

RNDr. Marek Ondruš, Ph.D. is responsible for maintaining the finished resource.

RNDr. Marek Ondruš, Ph.D. is responsible for the management and proficiency of data including data processing, data policies, data guidelines, and data availability.

RNDr. Marek Ondruš, Ph.D. is responsible for implementing the DMP, and ensuring it is reviewed and revised.

To execute the DMP, additional specialist expertise is required and we have such trained support staff available.

We do not require any hardware or software in addition to what is usually available in the institute.

5. Data security

Project members will not store data or software on computers in the lab or external hard drives connected to those computers. They will not carry data with them (e.g. on laptops, USB sticks, or other external media). All data centers where project data is stored hold appropriate certifications. All project web services are accessible via secure HTTP (https://...). Project members have been instructed about both generic and specific risks to the project.

The possible impact of information loss is small. The risk of information leak, and vandalism is acceptably low.

We are not using any personal information.

The archive will be stored in a remote location to protect the data against disasters. The archive needs to be protected against loss or theft. It is clear who has physical access to the archives.

We are running the project in a collaboration between different groups and institutes. A collaboration agreement that describes who can have access to what data in the project is set.

6. Ethics

Data we produce

For the data we produce, the ethical aspects are as follows:

- **raw data**
  - It does not contain personal data.
  - It does not contain sensitive data.

Data we collect

We will not collect any data related to individuals, i.e. "personal data".

The data collection is not subject to ethical legislation.

7. Other issues

We use the Data Stewardship Wizard with its *Common DSW Knowledge Model* (ID: dsw:root:2.6.13) knowledge model to make our DMP. More specifically, we use the https://dsw.uochb.cas.cz/wizard DSW instance where the project has direct URL: https://dsw.uochb.cas.cz/wizard/projects/f39de16d-5249-4171-9c16-16783776331f.

We will be using the following policies and procedures for data management:

- **IOCB Research Data Management Policy**
  https://www.uochb.cz/en/research-data-management
- **Open Science Principles of the Czech Academy of Sciences**
  https://www.avcr.cz/en/about-us/legal-regulations/open-science-principles-of-the-czech-academy-of-sciences/