Data Management Plan

Cobitis projects

Contact person: **Karel Janko** (*janko@iapg.cas.cz*)
[Czech Academy of Sciences, Institute of Animal Physiology and Genetics](#)

Based on: *Common DSW Knowledge Model, 2.4.4 (dsw:root:2.4.4)*

Project phase: *Before Submitting the Proposal*

Created by: **Patrik Horna** (*hornap@vscht.cz*)

Generated on: *04 Apr 2023*

Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»

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

Contributors

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

- Karel Janko
  janko@iapg.cas.cz
  Roles: Contact Person, Data Collector, Project Leader, Project Member, Researcher
  Affiliation: [Czech Academy of Sciences, Institute of Animal Physiology and Genetics](#)
- Marie Kaštánková
  Roles: Data Collector, Producer, Project Member, Researcher
  Affiliation: [Czech Academy of Sciences, Institute of Animal Physiology and Genetics](#)
- Dmitrii Dedukh
  Roles: Data Collector, Project Member, Researcher
  Affiliation: [Czech Academy of Sciences, Institute of Animal Physiology and Genetics](#)
- Daniel Kulik
  Roles: Data Collector, Project Member, Researcher
- R. Hobza
  Roles: Data Collector, Project Member, Researcher
- Jan Pačes
  hpaces@img.cas.cz
  Roles: Data Curator, Data Manager, Project Manager, Project Member, Researcher
  Affiliation: [Czech Academy of Sciences, Institute of Molecular Genetics](#)
- Patrik Horna
  hornap@vscht.cz
  Roles: Data Curator, Data Manager, Data Steward, Project Member, Researcher
  Affiliation: [Czech Academy of Sciences, Institute of Molecular Genetics](#)
- Martin Pšenička
  Roles: Data Collector, Project Manager, Project Member
  Affiliation: [University of South Bohemia in České Budějovice](#)
- Tomáš Tichopád
  Roles: Data Collector, Project Member, Researcher
  Affiliation: [University of South Bohemia in České Budějovice](#)

Projects

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

GAČR 2024

Start date: *N/A*

End date: *N/A*

Funding: [*Grantová Agentura České Republiky*](#): *grant number not yet given (planned)*

Abstrakt TBD

Section A: Data Collection

1. What data will you collect or create?

Data formats and types

We will be using the following data formats and types:

- **FASTA Sequence Format**

  It is a standardized format. This is a suitable format for long-term archiving. We expect to have 100 GB of data in this format.

- **FASTQ Sequence and Sequence Quality Format**

  It is a standardized format. This is a suitable format for long-term archiving. We expect to have 1000 GB of data in this format.

- **Binary Alignment Map Format**

  It is a standardized format. This is a suitable format for long-term archiving. We expect to have 2000 GB of data in this format.

- **Variant Call Format**

  It is a standardized format. This is a suitable format for long-term archiving. We expect to have 1000 GB of data in this format.

- **Joint Photographic Experts Group Format**

  It is a standardized format. This is a suitable format for long-term archiving. We expect to have 1000 GB of data in this format.

2. How will the data be collected or created?

Data storage and file conventions

We will use a filesystem with files and folders with the following folder conventions:

- There will be a **folder for each sample/subject**.
- There will be a **(sub)folder for each (repeated) analysis**.
- There will be a **(sub)folder for each step in the analysis workflow**.

Moreover, we have made appointments about naming the files.

We will not be storing data in an "object store" system.

We will not use a relational database system to store project data.

Section B: Documentation and Meta-data

3. What documentation and meta-data will accompany the data?

List of data to be published is given in Section E, Question 9. This also includes information about catalogs where the data can be found. Information about data types used is given in Section A, Question 1.

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

We will be documenting the data with Dublin Core, DataCite, and DDI (Data Documentation Initiative) metadata standards. The provenance will be captured using W3C PROV.

Section C: Ethics and Legal Compliance

4. How will you manage any ethical issues?

5. How will you manage copyright and Intellectual Property Rights (IPR) issues?

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

All of our data can become completely open immediately.

Limited embargo will not be used as all data will be opened immediately.

All data will be owned by the institute.

Section D: Storage and Backup

6. How will the data be stored and backed up during the research?

Storage needs are large at the beginning and will be reduced later.

All essential data is also stored elsewhere to prevent a total loss of data. All project data stored outside of the working area will be adequately backed up.

7. How will you manage access and security?

Project members can carry data with them on password-protected laptops. All data centers where project data is stored carry sufficient certifications. All project web services are addressed via secure HTTP (https://...). Project members have been instructed about both generic and specific risks to the project.

We will mitigate information loss risk for the project or organization. The possible impact to the project or organization if information is leaked is small. We will mitigate information vandalism risk for the project or organization.

We are not using any personal information.

Only project members will have read access; only selected project members will be able to write data.

Section E: Selection and Preservation

8. Which data are of long-term value and should be retained, shared, and/or preserved?

9. What is the longterm preservation plan for the dataset?

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.

Section F: Data Sharing

10. How will you share the data?

Information about used repositories (i.e. where will potential users find out about

the data) is provided in Section E, Question 9.

Embargo on the data is described in Section C, Question 5, and Section F, Question 11.

11. Are any restrictions on data sharing required?

Ethical and legal restrictions are documented under Section C. We have used the Data Stewardship Wizard, which made us aware of options to minimize the restrictions.

No data sharing agreement will be required.

Section G: Responsibilities and Resources

12. Who will be responsible for data management?

Patrik Horna is responsible for implementing the DMP, and ensuring it is reviewed and revised.

Jan Pačes and Patrik Horna are responsible for reviewing, enhancing, cleaning, or standardizing metadata and the associated data submitted for storage, use and maintenance within a data centre or repository.

Karel Janko, Marie Kaštánková, Dmitrii Dedukh, Daniel Kulik, R. Hobza, Martin Pšenička, and Tomáš Tichopád are responsible for finding, gathering, and collecting data.

Jan Pačes and Patrik Horna are responsible for maintaining the finished resource.

Patrik Horna is responsible for the management and proficiency of data including data processing, data policies, data guidelines, and data availability.

13. What resources will you require to deliver your plan?

To execute the DMP, no additional specialist expertise is required.

Charges applied by data repositories (if any) are mentioned already in Section E, Question 9.