Research Data Management Policy of the J. Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences

Scope and Background

The J.Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences, v.v.i., "The Heyrovsky Institute", is committed to ensuring that the outputs of its research are handled in accordance with the European[1,2,3] and national[4] law on open science and research data management (RDM). This includes the incorporation of an open access (OA) policy which enables unrestricted access to published research articles (defined in the *'Open Access a základy managementu dat'* document). It also covers the whole life cycle of research data produced at The Heyrovsky Institute, incorporating the planning (*Data Management Plan Instructions for the J. Heyrovsky Institute of Physical Chemistry of the CAS*), re-use, acquisition, processing, storage, long-term preservation and sharing of data.

The Heyrovsky Institute is committed to the "Science without Boundaries" principle which highlights the EOSC definition for open data (which defines data as being "as open as possible; as closed as necessary"). It seeks to ensure that research data are both trustworthy and available without restrictions (where legally, ethically, and commercially appropriate). This approach will maximize the value of data produced by the Heyrovsky Institute for research, economics and for the overall benefit to society.

Researchers of The Heyrovsky Institute, as well as their Scientific Collaborators (Non-Profit and For-Profit Partners, Funders and Publishers etc.), should all assist and contribute to the development of this collaborative environment (including working conditions and culture). The institute will however provide personal support and the tools required to allow all researchers and collaborators to comply with the institute's regulations.

This document outlines the institute's policies on research data management which cover the expected practice to be followed by all scientific staff within the institute. These fundamental regulations will be followed by more detailed and further supporting procedures concerning: management of different data types; long term preservation; handling of sensitive data etc. The principles discussed in this document are consistent with the Heyrovsky Institute's policies on open access to published research, as well as intellectual property and patenting.

Positions and Responsibilities

Role	Responsible For
Data owner[a] – *a person with the rights to modify and delete a specific set of data. This is usually the person who produced the data in the first place. In line with institutional rules, data ownership may be delegated to another person (e.g., the head of the department, a data steward, etc.)*	Management and curation of data in accordance with this regulation. This includes the full annotation and long-term preservation of data.
Researcher – *employee of The Heyrovsky Institute involved in research*	Ensuring that the data are collected, processed and stored in line with the rules defined in this document.

---

[1] A new ERA for Research and Innovation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN

[2] European Commission, Directorate-General for Research and Innovation, *Realising the European open science cloud : first report and recommendations of the Commission high level expert group on the European open science cloud*, Publications Office, 2016, https://data.europa.eu/doi/10.2777/940154

[3] A European Strategy for Data. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066

[4] Novela publikována ve sbírce zákonů dne 31. 8. 2022 pod číslem 241/2022 Sb. a většina jejích ustanovení nabyla účinnosti hned 1. 9. 2022 (dále „implementační novela"). https://data.gov.cz/%C4%8Dl%C3%A1nky/implementace-sm%C4%9Brnice-o-otev%C5%99en%C3%BDch-datech – Czech only.

Principal Investigator[b]	Devising and implementing a Data Management Plan as part of the research proposal, in accordance with the data management and sharing requirements set out by the relevant funder (in addition to the rules described in this document).
Funding Body – any body providing financial support to researchers of The Heyrovsky Institute	Defining data governance principles and regulations[c] for supported projects.
Open Science Officer – defined in The Heyrovsky Institute's Organizational rules	Ensuring that the policies are reviewed and kept in line with national and European laws, as well as making sure that the funding requirements are also met.
Data Steward – oversees and governs the data produced by researchers at the institute.	Providing support for improving data management within a department.

[a]   *Owner with respect to the manipulation of data, not as defined in the Organizational rules*

[b]   *Applicant, co-applicant and subsequently the holder of the grant*

[c]   *All rules (institutional and from the funding body) apply.*

[1] A new ERA for Research and Innovation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN

[2] European Commission, Directorate-General for Research and Innovation, *Realising the European open science cloud : first report and recommendations of the Commission high level expert group on the European open science cloud*, Publications Office, 2016, https://data.europa.eu/doi/10.2777/940154

[3] A European Strategy for Data. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066

[4] Novela publikována ve sbírce zákonů dne 31. 8. 2022 pod číslem 241/2022 Sb. a většina jejích ustanovení nabyla účinnosti hned 1. 9. 2022 (dále „implementační novela"). https://data.gov.cz/%C4%8Dl%C3%A1nky/implementace-sm%C4%9Brnice-o-otev%C5%99en%C3%BDch-datech – Czech only.

PRINCIPLES

1. **The Heyrovsky Institute subscribes to the Czech national and European principles on Data Policy[4,5,6]**, the main principle being that:

Publicly funded research and the resulting data produced are a public good, produced in the public interest. These data should be made openly available with as few restrictions as possible in a timely and responsible manner, for the easiest possible re-use.

2. **The Heyrovsky Institute is committed to making research data available** to the research community in a timely manner and without any unreasonable restrictions. The availability of all data should be clearly defined in a license associated with the data (metadata) and any restrictions should be proportional to the funding terms and conditions, previously agreed. The institute must not restrict or delay access. The access is managed via the National Data Infrastructure and it should be given to all relevant users on equal terms as long as there are no legal, ethical or security-related reasons to preclude this. Sensitive data are treated by special access rules defined by the Czech National Data Infrastructure – *Working Group on Sensitive Data[6]*. Wherever possible, data will be stored in established public repositories e.g., PDB[7], Cambridge Energy Landscape Database[8], NIST[9], etc.

3. **The Heyrovsky Institute is committed in enabling public access to research data** in all justified/legitimate cases (and as far as possible, free of charge). Public access is enabled within the framework of existing legislation on sensitive data protection and on intellectual property, as well as in compliance with the provisions included in The Heyrovsky Institute Organizational rules. Public access is managed via the Czech National Data Infrastructure.

4. **The Heyrovsky Institute will ensure research data is collected, managed, stored and curated appropriately.**
- All data must be accompanied with the minimal metadata required (defined by the Czech National Data Infrastructure and National Metadata Archive)[10]. Producing detailed accompanying metadata with consistent terminology (which also uses the international standards as defined by the FAIR data principles[11]), is strongly recommended.
- All data must have defined persons responsible for the content – this will either be the data owner or a data steward.
- The assignment of unique identifiers e.g., Digital Object Identifiers (DOIs) to 'published' datasets is required for all archived datasets.
- Both data and metadata must be machine-readable.
- The use of electronic laboratory notebooks (ELNs) for data management is strongly recommended.
- The use of standard formats for research data is strongly recommended.
- In general, all data[d] that are relevant to the research field of the researcher; or which are of public benefit must be stored and archived on adequate repositories (see below). This is even if the data are not previously declared in the specific Data Management Plan for a project. This applies especially to data supporting publications.

[4] Novela publikována ve sbírce zákonů dne 31. 8. 2022 pod číslem 241/2022 Sb. a většina jejích ustanovení nabyla účinnosti hned 1. 9. 2022 (dále „implementační novela"). https://data.gov.cz/%C4%8Dl%C3%A1nky/implementace-sm%C4%9Brnice-o-otev%C5%99en%C3%BDch-datech – Czech only.

[5] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024

[6] Architektura implementace iniciativy EOSC ČR – Czech only.

[7] https://www.rcsb.org/

[8] https://www-wales.ch.cam.ac.uk/CCD.html

[9] https://materialscommons.org/

[10] Currently NDK - https://old.ndk.cz/standardy-digitalizace/metadata - Czech only. Techlib develops NMA (should be available in 2023/2024).

[11] https://www.go-fair.org/fair-principles/

- In line with funder policies, researchers must ensure that any intellectual property in the outputs of their research are protected and managed in accordance with the Institutional directive on intellectual property. Delays or restrictions on data sharing may be necessary to gain intellectual property protection, or to further develop a technology for public benefit.

5. **The Heyrovsky Institute provides an infrastructure for data management** and sharing, both for use within the institute and externally. The use of the HeyRACK institutional repository is recommended for 'hot' and 'warm' data; generic and domain-specific repositories of the NDI (as well as recognised global public repositories) are preferred for 'cold' data. The heads of department are responsible for the timely upload of data to the appropriate repository. They may delegate this responsibility to the principal investigators.

6. **The Heyrovsky Institute provides tools for data management** and sharing both for use within the institute and externally. The two available systems of electronic laboratory notebook (ELN) as well as one basic website which allows access to the repository, are managed by delegated persons at The Heyrovsky Institute.

7. **The Heyrovsky Institute supports the need to preserve and share research datasets** in a manner that maximises their long-term value. The National Data Infrastructure (both its generic and domain-specific repositories) as well as other public global repositories are recommended for long-term data preservation. Research data must be stored and accessible in a safe and secure manner for a minimum of 10 years after the end of a project (or their collection). Data deposited for long-term preservation must be adapted accordingly and use formats which ensure that they remain reusable in future. (The regulations surrounding this will be defined in the '*Data preservation strategy and standards*' document – to be released in 2024).

8. **The Heyrovsky Institute expects all users of research data to acknowledge the sources** of data as well as other digital(ised) objects re-used for the generation of datasets. If no third-party rights, legal requirements or property laws prohibit it, research data may receive a free license for its open use (granting traceability and credit to the original data and authors).

9. **The Heyrovsky Institute will recognise the contributions of researchers** who generate, preserve and share key research datasets (see '*Agreement on the Reforming Research Assessment*'[12]).

10. **The Heyrovsky Institute is committed to the open-source sharing of research software/code** that supports the analysis or processing of research data. There is an expected obligation for employees to share their software/code along with the associated project/dataset, where this is practical.

11. **The Heyrovsky Institute will provide its researchers with the appropriate support, training, guidance and advice on Research Data Management.**

d) *'All data' means positive and negative results (including those not associated with publications).*

---

[12] https://coara.eu/agreement/the-agreement-full-text/

Clarification of Terms

'Cold', 'hot' and 'warm' data – The data freshly produced (e.g., from an instrument), which undergo intensive processing are called '*hot*'. On the contrary, finalised data ready for publishing and long-term preservation are referred to as '*cold*'. No modification of 'cold' data is expected. To modify published data, versioning should be applied. The term '*warm*' data may sometimes be used. Such data do not undergo intensive processing but are still expected to be modified (e.g., missing metadata, verification by a collaborator, etc …).

Data – the term 'data' in this document covers a broader range of scientific outputs including scientific data, protocols and workflows, software and algorithms, audio and video recordings, digitalised objects (artefacts) at all stages or their production. This comprises also unpublished and intermediate outputs.

Data Management Plan – A document describing how research data and digital(ised) objects from projects or other research activities are managed throughout the whole lifecycle (starting before the activity and finishing as long-term preservation).

Data Storage versus Repository – Data storage is a general space in which data may be stored in a less formal way. Organisation of data in data storage is not pre-defined. On the contrary, a repository is a data storage location for the deposition and sharing of annotated and structured data (usually FAIR data). A data repository has a predefined structure as well as rules for data storage and sharing. It is also used for the long term and secure perseveration of data. If there is a domain specific repository then such repository is advisable to use.

Metadata – Valuable information/data describing the principle dataset in terms of provenance, technical specifications, specific experimental protocols, software/algorithms/code used for data processing, as well as other parameters which are essential for future users to know (needed for the data to be used correctly and reliably). Metadata may be thought of as 'Data about a dataset'.

The metadata:

- Must enable others to search for and re-use the data efficiently and effectively.
- Should follow international standards.
- Should describe the quality of the data.

Persistent Identifier (PID) – A persistent and unique identifier for research data, digital objects and publications. Examples include a DOI (Digital Object Identifier – provided by Datacite[13]) and Handle[14] (provided by CNRI).

Public Repository – A data storage location which provides curatorial services for scientific data and allows for the permanent storage of data. It may be easily accessed by the general public. Public repositories are usually domain- or topic-specific (e.g., PDB – database of protein structures).

Published Datasets – Datasets which are released in the published form for use by other users of the repository. Published datasets are accompanied by a persistent identifier (e.g., DOI) and should be partnered with a set of rich/comprehensive metadata.

Repository – see *Data storage*.

---

[13] https://www.doi.org/
[14] https://www.handle.net/

Research data - all data that are generated, gathered, simulated, or derived to describe the facts and processes discovered or observed during research activities and projects. Research data may also be processed and reported.

Acknowledged resources:

Research Data Management Policy of Universita degli Studi di Milano: https://www.unimi.it/en/research/research-data-and-outputs/open-science/research-data-management

Scientific Research Data Management of The Francis Crick Institute (Issued 18.1.2016): https://www.crick.ac.uk/news/2016-10-28-francis-crick-institute-publishes-open-access-and-data-management-policies

Research Data Management Policy (MPF1242) of the University of Melbourne (Issued 9.6.2022): https://policy.unimelb.edu.au/MPF1242/

The Research Council of Norway's Policy for Open Access to Research Data (December 2017): https://www.forskningsradet.no/en/research-policy-strategy/open-science/research-data/

[13] https://www.doi.org/
[14] https://www.handle.net/