Data Management Plan of the Institute of Physics of Materials of the CAS, v. v. i.

1. Definition of "data"

We consider as "data" relevant to this plan only those datasets that (1) must be made available according to the rules and principles of open access to scientific data ("open data"); the data that may become mandatory disclosures under the current rules in the future; and data generated for the benefit of non-institutional bodies in the context of contract research or infrastructure activities. All other data collected or processed in the course of research activities of the Institute of Physics of Materials of the CAS, v. v. i., (hereinafter "IPM") we consider to be internal data, which are either of a non-important nature (i.e., they do not need to be made available to support the conclusions of the research), or they are data for which there is a reason preventing disclosure, i.e., a valid exemption.

Due to the technological and instrumental focus of the IPM, data on the physical description of processes in materials are an essential part of the research, which clarifies the relationship between the behaviour and properties of materials and their structural and microstructural characteristics. The research data are generated both by experimental study of the physical nature of the processes occurring in materials and by theoretical multilevel modelling of the properties and behaviour of materials under different loading conditions. An integral part of the research data is the characterisation of the structure of materials using the latest electron and light microscopy techniques and other analytical techniques. Data (information) in the field of mechanical engineering in the form of design solutions can also be output. This may be engineering drawings, electrical schematics, software, etc. The actual data in the usual sense of data can then be data measured by experimental instrumentation, image data and data from theoretical models. However, the relevance of the data for external stakeholders is minimal, mainly due to its origin in specific instrumentation platforms, including instruments of their own design. Their comparability and reproducibility at sites outside their origin may therefore be very limited.

The data formats worked with at the IPM can be very diverse, reflecting the wide technological range of the department. The data can be categorized as follows:

1. Data from traditional formats can be processed by common, widely used software.
This category includes numeric and text files, image data in common formats, etc.

2. Data obtained by proprietary software that is part of special but commercial instruments operated at the IPM.

3. Data is recorded by instruments constructed at the IPM and stored in a format specific to the particular apparatus.

4. Design data.

The use of data in the internal environment and its management for the research needs of the research teams of individual projects is the responsibility of the departments, groups, and researchers.

2. FAIR data

Ensuring data findability, including provisions for metadata

Metadata is an essential part of the stored data files and carries information about the time of creation of the data file, the circumstances of its recording, the type of experiment, and the project within which it was taken. Metadata related to their interpretation may, in many cases, include intellectual property elements and their disclosure or transfer to external parties is subject to the rules for disclosure of data with application potential (see below).

Due to the broad disciplinary scope of the IPM, further details of the metadata structure are specifically linked to the practices of the respective disciplines. In this data management plan, they are not, and cannot be, nominally modified just because of the variability of individual projects and their specific requirements.

With few exceptions, IPM data do not have the character of data collected on sets of objects or samples that would be relevant to external machine processing (data harvesting). Modifying them for machine reading would not be meaningful and could lead to undesirable misinterpretation.

Data accessibility (Accessibility)

The forms of access to the data of the IPM can be categorized according to the nature of the subject to which they are to be made available:

1. Access for the IPM research team

Access to data for the internal needs of the research teams for projects solved at the IPM is full and unrestricted. Access to data repositories is protected by standard data security procedures. Metadata reflects the clarity needed internally within the research team.

2. Access for IPM staff outside the research team of a specific project

Data disclosure to persons outside the specific research team within the IPM is full, as in point 1, unless the research team is contractually bound by confidentiality (NDA).

3. Access for members of the consortium of projects with which the IPM is bound by a cooperation agreement

In the case of data generated within projects involving external co-principal investigators, the forms of data storage, the nature of metadata, and the forms of data sharing are determined by the contracts with the partners or by the procurement documents of the individual consortium projects. These project and contractual requirements are overridden in specific projects by this data management plan.

4. Access to external entities

Data that are identified as data for which disclosure to external parties is permitted are disclosed on a "upon reasonable demand" basis - upon a meaningful request from an external party. Forms can be used for data requests on the IPM website. The request is then assessed by the Data Steward (see below), who will pass it on to the relevant Institute staff member. The decision as to whether the request is 'reasonable' rests with the project investigator of the project in which the data was generated. If necessary, the investigator will consult with the Group Leader, Department Head, or Data Steward.

The disclosure of data to external entities is not anonymous, the request for data transfer is linked to the completion of a questionnaire which includes the name, affiliation and email address of the requestor. The actual request must be made

in a way that allows the identity of the applicant to be verified, i.e., by data box or registered letter. Applications submitted by email will not be taken into account.

The forms of access to data of the IPM can also be categorised according to the nature of the project within which it was created and its outputs:

1. Access to publication-related data

Data related to publications are subject to the rules of the publishers and are made available in the appropriate form requested by the publishers.

2. Access to data generated by projects of different providers

The tender documents and rules of the providers of earmarked project support usually lay down their own rules for working with data. Where these rules are not in line with the IPM Data Management Plan, but the Director of IPM has approved the submission of a project application for specific individual projects, the rules of the Earmarked Providers override this plan.

The resulting data are made available when the logical units of research are completed and it is clear what the result of the unit is, how the result will be handled, whether the result is properly processed and prepared in that spirit. Typically, this involves the disclosure of data after/submission of the publication manuscript, completion of the project, etc., so that the data disclosed are relevant to the research conducted and it is clear that disclosure will not jeopardise the legitimate interests of the IPM or other IP protection entities or violate obligations arising, e.g., from contracts with application partners.

Repositories

The trustworthiness of data repositories and repositories is a subject of great attention and care. The internal data repositories, which are part of the Institute's computer network infrastructure, are essential for the management of IPM data. It is the responsibility of the Economics and Operations Department to ensure proper backup, security, and operation reliability. Individual scientific departments have dedicated data storage space and use it in a mode appropriate to their own needs. Further development and strengthening of the internal data storage facilities, both in terms of increasing the volume of data stored and in terms of strengthening the cyber security of the stored data, is one of the priorities of the IPM.

In addition to the internal data repositories, the IPM uses a number of other repositories and storage facilities. They can be divided as follows:

1. Subject-specific data repositories are widely used by the international community, with credibility determined by the breadth of their use in science. For example, arXiv or ZENODO.
2. Trusted national repositories offered by national infrastructures for science and research, namely EOSC-CZ, KNAV-ASEP, and CESNET.
3. Repositories of publishers of scientific publications whose use for publication-related data is an obligation for authors.
4. Domain-specific repositories are dedicated to a narrower international scientific community, the use of which is usually necessary in the context of disciplinary consortium projects.

The use of other data repositories and repositories is possible after due verification of their credibility and with the approval of the head of the relevant scientific department and the Data Steward.

Data

With regard to the technological focus of the institute, many of the results have application potential and/or are created directly in collaboration with partners from the application sphere. Therefore, there is a possibility of future commercialisation or transfer into practice in a broader sense. Access to the data is then restricted or embargoed to ensure the protection of intellectual property. In particular, research data that can be used by a competitor to undermine IP protection or to prevent the implementation of protection (e.g., to prevent patenting or licensing) will not be provided. This, therefore, means that all research data relating to potentially commercially useful and applicable outputs are considered to fall outside the category of research data covered by the open access principles.

Non-disclosure of data - data is not disclosed in the following cases:

1. Applied research project data

This restriction does not apply to data related to publications produced in the framework of these projects.

2. Data from the non-publicly funded contract and collaborative research projects and data generated by "Other Activities."
3. Data with application potential and data containing intellectual property elements

The decision of whether a dataset has application potential or contains an element of intellectual property rests with the project investigator within whose project the data was created, and consultation with the group or department leader is recommended.

4. Data bound by non-disclosure agreements (NDAs)
5. Data of constructional nature - engineering drawings, electrical schematics, software (except software distributed in the form of Open Source)
6. Data that has been created by proprietary software that is bound by the constraints of the licensing arrangements for that software and associated metadata
7. Data with ethical constraints
8. Data that violates the principles of data protection (GDPR)

Data interoperability

Data interoperability is essential for sharing, exchanging, and reusing data. The primary imperative is to ensure interoperability within solution teams for internal use. Therefore, the data are provided with adequate metadata and an understandable description. The formats and standards are consistent with industry practices and the needs of the research teams. Project investigators are encouraged to use the standard metadata suggested by The Research Data Alliance unless otherwise specified in the project specifications or project collaboration agreements. Standard vocabulary will be used for all types of research data generated by the project.

The interoperability of data relevant to scientific publications will be ensured from a technical point of view by depositing a machine-readable electronic version of the published publication or the final version of a peer-reviewed manuscript accepted for publication in the relevant repositories. In legal terms, open access to research data deposited in these repositories under the terms of the latest available version of the CC BY public licence or its equivalent.

Ensuring data reuse (Reusability)

The internal usability of the data is ensured by storing it in internal data repositories, storing it in formats used by standard research teams, operating laboratory apparatus, and related software, and providing adequate metadata to ensure readability.

Data open for direct external use are data linked to specific publication outputs in the respective repositories. These include information on partial research results or the tools and instruments needed to reuse or validate research data. Open access to the data will be provided in these repositories, subject to compliance with the terms of the latest available version of the CC BY public licence or its equivalent.

Data to be shared within consortium projects is available to consortium members by project partners to use. The quality of the research data will be assured by the researcher's codes of ethics applicable to the consortium institutions.

3. Data management resource allocation

The resources allocated to data care at the IPM are mainly directed to investments in computer network infrastructure and data storage. An adequate annual budget is allocated for regular upgrades of network, computer, and data resources. Purchases of a capital nature are included in the IPM instrument investment plan as appropriate.

Human resources dedicated to data management

1. Data management at the project level is the responsibility of the project investigators.
2. Data interoperability at the level of scientific groups is entrusted to group leaders and at the level of departments to heads of departments.
3. Care of the computer network, the Institute's data storage, data security, and data backup is the task of the Economic and Operational Department.
4. The role of the Data Steward, whose tasks are defined in the chapter "Making Data Available," is assigned to the Institute's library staff.

4. Data security

The security of data stored on internal data repositories is ensured by the Economic and Operational Department of the IPM. This includes security against hostile attacks, technical failure, and other risks of data loss.

Data storage on external data repositories and repositories is secured by choosing repositories and repositories that are trusted and properly certified.

5. Ethics

The data produced at the IPM are data about inanimate nature and are therefore not subject to any ethical constraints. Research towards medical and other applications, where ethical considerations in the handling of data must also be taken into account, is always carried out in a collaborative manner and organized so that any experiments on animal models or research on human subjects are carried out at a collaborating facility that has the appropriate permissions and procedures, rather than at the IPM.

This Data Management Plan shall take effect on 1 January 2024.