

Wytyczne dotyczące przygotowania planu zarządzania danymi wytworzonymi w pracy naukowej

Prawidłowe zarządzanie danymi wytworzonymi w pracy naukowej jest kluczowym działaniem zapewniającym rzetelność prowadzonych badań. Zgodnie z kodeksem Narodowego Centrum Nauki dotyczącego rzetelności badań naukowych i starania o fundusze na badania, każdy naukowiec powinien opracować plan zarządzania danymi w projekcie i przedstawić go na etapie składania wniosku o finansowanie. Wymóg taki istnieje również dla projektów finansowanych w ramach funduszy unijnych (np. Horizon Europe). Nie jest możliwe opracowanie planu uniwersalnego ze względu na indywidualny charakter każdego projektu i rodzaj generowanych danych. Niniejszy dokument przedstawia wyłącznie przykładowe odpowiedzi na pytania zawarte we wnioskach o finansowanie projektów z funduszy NCN. Są one zgodne z wytycznymi NCN oraz Komisji Europejskiej. Odpowiedzi podzielono na dwa typy w zależności od stopnia, w jakim sugeruje się edycję wzorcowej odpowiedzi:

(1) *kierownik projektu musi odpowiednio przereklamować odpowiedź uwzględniając planowane badania,*

(2) sugerowana odpowiedź jest zgodna z regulacjami wewnętrznymi Instytutu oraz obowiązującym prawem, zalecana ostrożność w zmianach.

Odpowiednia czcionka wskazuje sugerowane działanie.

Kolorem żółtym wyróżniono najważniejsze zmiany poczynione w stosunku do wcześniejszej wersji wytycznych – głównie związane z rekomendowanymi repozytoriami RepOD (<https://repod.icm.edu.pl/>) - dane naukowe) i OPEN (dawniej CeON) (<https://open.icm.edu.pl/>) - repozytorium publikacji).

Na czerwono zaznaczono pola, które kierownik projektu powinien edytować w kontekście projektu.

Guidelines for preparing research data management plan

Appropriate management of scientific data is key to research integrity. According to the National Science Centre, Poland, each person applying for research funds should set up a plan for research data management and include it in the submitted project. Such obligation exists also for EU-funded projects. While there is no way to prepare a universal plan that would fit any project, we prepared a template of answers to questions that you will find in most of NCN's project submission systems. These answers comply with NCN and EU Commission guidelines. The answers are divided into two types reflecting the suggested degree of edition:

(1) *the project leader must edit the answer, taking into account project details.*

(2) it is suggested that changes to these sections be limited as they already comply with the Institute's internal regulations and the law in force.

Appropriate font suggests the action that needs to be taken.

Critical changes compared to the previous version are highlighted in yellow –these concern mainly the use of the RepOD repository (<https://repod.icm.edu.pl/>) - open data repository) and OPEN repository (former CeON) (<https://open.icm.edu.pl/>) - publications repository).

Fields that the project manager should edit within the context of the project are marked in red.

Plan zarządzania danymi - propozycja wypełnienia *Data Management Plan – template*

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych **Data description and collection or re-use of existing data**

Sposób pozyskiwania i opracowywania nowych danych i/lub ponownego wykorzystania dostępnych danych.

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

Data will be collected from instruments dedicated to specific measurements (X, Y, etc.) and saved on their discs or on discs of the dedicated computers in instrument-specific formats such as formatted text files or images. File names and their metadata will be constructed in a way that eases the identification of measurement type, date, and collector. Raw output files will be moved to computers for further processing. Each instrument will be calibrated according to the manufacturer's guidelines.

Quantum-chemical calculations will be carried out using Z and R programs. The calculations will be run both locally as well as remotely in supercomputing centres. In the latter case, data transfer will be performed using SSH protocol. Data processing will be carried out locally.

Raw experimental data, as well as programs' inputs, outputs, and other files necessary to repeat calculations, will be deposited both locally and in cloud service. Data being substantial to publications should also be stored in the RepOD open repository.

Pozyskiwanie lub opracowywane dane (np. rodzaj, format, ilość)

What data (for example, the kinds, formats, and volumes) will be collected or produced?

The data will be acquired from instruments or generated by quantum-chemical codes as digital data in a format that is instrument-and software-dependent, e.g., plain text files (.asc, .csv, .txt, .inp, .out, etc.), image files (.tiff, .jpg, .png, etc.) or integral files (FCIDUMP, .gbw, etc.). Processed data will be saved in program-specific formats, e.g., formatted text (.odt, .doc, .docx, etc.), tabulated data (.ods, .xls, .xlsx, .opj, etc.), editable figures (.svg, .cdr, .xcf, .ps, etc.), images (.gif, .jpg, .png, etc.), animations (.mov, .avi, .mp4, etc.) or presentations (.odp, .ppt, .pptx, etc.). The output of the project includes various computer scripts that will automate data processing. These will be saved as plain text files.

We expect to process and store a peak maximum of 10 TB of data distributed over computers belonging to project team members.

Up to 1TB of secure backup will be dedicated to key project data selected by the PI to be stored locally. Selected data will be mirrored in the RepOD open data repository.

2. Dokumentacja i jakość danych **Documentation and data quality**

Metadane i dokumenty (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych) towarzyszące danym.

What metadata and documentation (for example, methodology or data collection and way of organising data) will accompany data?

Data stored in RepOD will contain metadata, ensuring that data are findable. Metadata will include: (a) data description along with the date of deposition, venue of acquisition, authors involved, and (b) funding details. Data will be organised in folders named after measurement or calculation type. Sub-folders will be organised in a way to ease the identification of the data collector and the date of acquisition. If needed, data structure with additional comments will be stored in a README file. Additionally, project team members will record experimental conditions and other specific set-ups in lab notes. The form of lab notes can be digital (stored with the data) or hand-written (stored by the project team member in a locked cabinet). Meaningful and selected data from hand-written lab notebooks will be digitized when writing a paper or report and stored in the designated cloud service. Selected data will be attached as supplementary information, when possible, or stored in the RepOD.

Stosowane środki kontroli jakości danych

What data quality control measures will be used?

Data will be acquired from state-of-the-art research instrumentation and analysed using validated software. When possible, we will repeat experiments at least three times to decrease random errors. The quality of the data will be assessed by the project leader, and the assessment will be based on her/his experience. Experimental conditions, the instrument's specific set-up, or non-default program settings will be documented in lab notes and subsequently described in reports and scientific publications when possible.

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

Storage and backup during the research process

Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

How will data and metadata be stored and backed up during the research process?

During the project duration, the data will be stored at three levels:

(1) on local computers of the project team members (for data processing, a total of **about 10 TB**).

(2) on an external, group-enabled hard drive connected to a dedicated computer or stored securely after project completion (for secure data exchange, backup, and long-term data storage, up to **10TB**, minimum **10 years** storage).

(3) in a designated cloud service. **The Institute will provide up to 100 GB of cloud space. If the data exceeds this limit, the PI will provide funds for additional cloud space during the research process.**

(4) upon publication, selected data should be stored in the RepOD open repository <https://repod.icm.edu.pl/>.

The backup process to cloud service will be automated on each computer used for data processing. We plan to back up users' key data incrementally on a regular basis to external discs and RepOD.

Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

How will data security and protection of sensitive data be taken care of during the research?

The backup will be stored on two separate local devices (1) and (2) and a dedicated cloud service.

Adapted data storage structure and management allow data to be stored in multiple physical locations. The additional level of security is ensured by mounting long-term backup storage in the user's system only at the moment of backup creation.

Personal data will be collected only during the recruitment procedure and processed according to detailed internal regulations (*vide infra*).

4. Wymogi prawne, kodeks postępowania
Legal requirements, codes of conduct

Sposób zapewnienia zgodności z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych w przypadku przetwarzania danych osobowych

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

Regulations for the management of procedures for processing personal data at the Institute of Physical Chemistry of the Polish Academy of Sciences are defined by Annex 1 to the Resolution of Director of the Institute of Physical Chemistry No 55/18 (from 26.09.2018) “Handbook of Procedures for Processing Personal Data at the Institute of Physical Chemistry of the Polish Academy of Sciences”.

Sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własnością. Obowiązujące przepisy.

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

Regulations on the management of intellectual property rights and the principles of commercialization of the results of scientific research and development works at the Institute of Physical Chemistry of the Polish Academy of Sciences are defined by Appendix No. 1 to the Resolution of the Scientific Council No. 6/321/2024 (from 17.06.2024).

5. Udostępnianie i długotrwałe przechowywanie danych **Data sharing and long-term preservation**

Sposób i termin udostępnienia danych. Ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga

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

In principle, no embargo will be imposed. However, sharing and presenting unpublished results will be subject to the principal investigator's decision, who must consider possible conflicts of interest or patenting procedure. The outcomes of the project will be published in international scientific journals in an access mode that complies with the funder's rules. Selected source data, especially data necessary for replication of the published results, will always be accessible at the RepOD repository via a persistent DOI number or will be available on request free of charge. (These data may also be deposited along with the original manuscript and in external databases such as the Cambridge Structural Database or Crystallography Open Database).

Sposób wyboru danych przeznaczonych do przechowywania oraz miejsce długotrwałego przechowywania danych (np. repozytorium lub archiwum danych)

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

After the completion of the project, all data will be stored for a minimum of 10 years on two dedicated external hard drives, one of which will be securely deposited in the Institute and managed by the Library, and the other deposited with the PI.

Selected data, in the form of final peer-reviewed manuscripts along with supplementary material (SM), will be deposited in an online repository OPEN (<https://open.icm.edu.pl/>). Accompanying source data will be stored in an online data repository, RepOD (<https://repod.icm.edu.pl/>), linked with the original manuscript, which follows the rules of FAIR Data.

Data chosen for long-term storage in an online repository will be selected by the principal investigator, and the selection will be based on his/her expertise. The Institute's Data Officer will be in place to assist with the data selection.

Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie z danych

What methods or software tools will be needed to access and use the data?

Long-term and free access to the deposited data in the form of accepted manuscripts and accompanied research data will be possible via online repositories: OPEN (<https://open.icm.edu.pl/>) and RepOD (<https://repod.icm.edu.pl/>).

Sposób zapewniający stosowanie unikalnego i trwałego identyfikatora (np. cyfrowego identyfikatora obiektu (DOI)) dla każdego zestawu danych
How will the application of a unique and persistent identifier (such us a Digital Object Identifier (DOI)) to each data set be ensured?

OPEN (<https://open.icm.edu.pl/>) and RepOD (<https://repod.icm.edu.pl/>) repositories allow for the assignment of persistent digital identification to all stored items.

6. Zadania związane z zarządzaniem danymi oraz zasobami.
Data management responsibilities and resources.

Osoba (np. funkcja, stanowisko i instytucja) odpowiedzialna za zarządzanie danymi (np. data steward)

Who will be responsible for data management (i.e. data steward)?

The Data Protection Officer affiliated with the Institute of Physical Chemistry of the Polish Academy of Sciences serves as a data steward responsible for personal (sensitive) data management and law enforcement. A dedicated person from the Institute's library staff is responsible for managing the online manuscript repository. The Principal Investigator will act as a data steward and will be responsible for managing raw data storage and flow, granting access to the data, and storing scientific data long-term for the reproduction of published data. The principal investigator may designate another person to act as a data steward in the project. The Institute's Data Officer will support the Principal Investigator in solving any data-related issues.

Środki (np. finansowe i czasowe) przeznaczone do zarządzania danymi i zapewnienia możliwości odnalezienia, dostępu, interoperacyjności i ponownego wykorzystania danych
What resources will be dedicated to data management and ensuring that data will be FAIR?

Online manuscript repository and data repository management is the duty of the Institute's library staff. Long-term support and management of computers and hard drives will be provided by the Institute's IT department.

No additional costs are needed for data management.

OR

Additional resources up to PLN/EUR will be required for cloud data storage exceeding the limit provided by the Institute during the project and after project completion for external hard drives for long-term preservation.