

Data Management Plan in the IOC PAS

1. Data description and collection or re-use of existing data

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

Data collected in the proposed research project will include synthetic procedures and analytical data for all obtained compounds. These will be recorded in electronic lab books, with analytical data generated digitally using specialized software. Typical data sets include ¹H NMR, ¹³C NMR, IR, chromatograms, specific rotation, and mass spectra. Data will be stored on analytical group servers and team computers, organized in project-specific folders. Each sample's data will follow a naming convention: owner's initials, sample number, and team number. Existing data, including published spectroscopic datasets, reaction protocols, and previously collected internal data from related studies, will be reviewed and integrated. Public databases such as Reaxys and SciFinder will be used to compare results and refine analytical methods.

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

The data obtained during the project implementation will fall into two categories:

1. Reaction parameters required for execution of the chemical transformation (PDF files, 100 MB).
2. The spectroscopic and general characterization data of all compounds produced during the project (raw FIDs etc., PDF files, 5-10 GB).

All generated data will be stored in electronic format with additional hard copies of lab notes or spectra collections if necessary.

2. Documentation and data quality

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

Raw instrumental data will be organized according to the specific software used and stored in clearly labeled folders, named after the project. The dataset will include metadata such as acquisition dates, experiment details, sample names, authors, and brief descriptions of scanned PDFs or lab books. Additional information may be included if necessary to clarify the dataset.

When the final dataset is uploaded to the institutional data repository, it will be accompanied by a README file. This file will provide an overview of the dataset contents and describe the file-naming convention used, which includes the owner's initials, sample number, and research team number.

2.2. What data quality control measures will be used?

The correctness and reproducibility of all developed procedures and reaction conditions will be verified by repeating experiments. Analytical data will be obtained by independent, qualified team members using pre-calibrated equipment, validated analytical methods and software delivered by verified suppliers. Manually entered data will be cross-checked by another team member to avoid mistypes. These measures will significantly reduce the risk of incorrect data or bias during data collection.

3. Storage and backup during the research process

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

All data will be stored in electronic format on hard drives belonging to the researchers involved in the project and analytical group servers (applies to analytical data). Backups of these storage devices will be performed systematically (at least once a month) to the Institute's NAS server. The research team will have access to up to 1 TB of data storage on the Institute's NAS server, which can be expanded if necessary. The IT team will be responsible for data backup and recovery in case of any damage. Paper copies of lab notes or analytical data will only be created when necessary, and each document will be scanned to ensure proper storage. In the case of remote work, all generated data will be sent to the Institute's NAS server via an encrypted connection.

3.2. How will data security and protection of sensitive data be taken care of during the project?

All computers used to record, store or analyze data have proper security software and antivirus protection. Password-protected and encrypted computers and backup systems will be used. Each member of the project team will access the data using a password. To ensure data protection, all files will be shared via private groups in the cloud storage and sharing system or on external hard drives, rather than via email. For collaborative projects, all relevant data will be uploaded to the cloud storage and sharing system and access via password protected accounts will be available to the involved parties.

4. Legal requirements, codes of conduct

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

The administrator of the personal data of the research group members is the IOC PAS. The Institute will process their data, with their consent, based on declarations, for the purpose of conducting scientific and research activities, in accordance with the legitimate interests of the IOC PAS and relevant legal regulations, depending on the circumstances. More information is available on the IOC PAS website (www.icho.edu.pl/rodo).

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

The applicable regulations are provided in "Regulations for the protection and use of intellectual property at the IOC PAS". The creator of the results holds personal copyright, in particular the right to authorship of the result, the right to the inviolability of the content and form of the result, the right to use the result, and the right to attribute the result to the author.

5. Data sharing and long-term preservation

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

Data generated in the project will not be shared with the scientific community until a publication is accepted or a patent application is approved. Results will be published in high-impact journals, with analytical data and reaction conditions provided free of charge as supplementary information. Scientific results will be disseminated through Open Access via one of two routes:

- The Gold Path – publishing directly in an Open Access journal;
- The Green Path – depositing a copy of the published work in an open archive or repository.

The Institute uses OPEN, the Repository of Open Scientific Publications, at the University of Warsaw. Open Access to publications will be available immediately after publication. Other data, including raw

data that doesn't compromise intellectual property, will be deposited in RepOD, an open repository for research data at the same institution. Raw data will be stored for at least 3 years post-project.

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

Publishable results and materials, as well as all data underpinning patent applications or research publications, will be preserved long-term on the Institute's NAS server. The selection of relevant data for preservation will be made by the Principal Investigator. Data in the repository will be stored in compliance with the data policies of the funder and the Institute of Organic Chemistry of the Polish Academy of Sciences.

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

All data will be provided in standard formats (e.g., .doc, .pdf, .fid, .cif) that are accessible using standard office software and programs designed for processing analytical data. When possible, data will be exported into formats that are compatible with open-source software, such as SpinWorks, ChemSketch, and Mercury. These software tools are widely used by research groups involved in organic synthesis, which is the focus of the proposed research project.

5.4. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

DOI numbers will not be assigned to individual datasets. Instead, DOI numbers will be assigned to articles or communications published in scientific journals as part of the project. The procedures and analytical data related to each publication will be organized in appropriately named folders, and once the article is accepted, these folders will be assigned a DOI number.

Publications deposited in the **OPEN** repository will be identified by the same DOI as the corresponding publication. Datasets deposited in the **RepOD** repository will have a unique DOI assigned.

6. Data management responsibilities and resources

6.1. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The responsibility for managing data produced locally during the implementation of the planned research work lies with the Principal Investigator (PI). Each individual responsible for a specific task will also be responsible for managing the data within the scope of their task. The PI will oversee the preparation of final files and their upload to the institutional repository, along with the necessary metadata. Data uploads to open archives and repositories will be supervised by the Director's Attorney for Open Access to Scientific Publications and Research Data. The Institute's IT staff will manage the NAS server, where backups and final data will be stored.

For collaborative projects, each partner institution will be responsible for managing the data generated within their organization.

6.2. What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The IOC PAS and the Principal Investigator (PI) already possess the necessary hardware and software resources to implement the data management plan effectively. Financial resources for the final deposit of the dataset in the institutional repository will be allocated from the indirect costs of the submitted research project.

In terms of time, the Principal Investigator and relevant project staff will dedicate time to ensure data is managed according to the FAIR principles. Compliance with these principles will be overseen by the Director's Attorney for Open Access to Scientific Publications and Research Data, who will ensure that the data is properly stored, accessible, and adheres to standards for findability, interoperability, and reusability.