

Data Management Plan in the IOC PAN

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 within proposed research project will include developed synthetic procedures and analytical data of all obtained compounds. None of the synthesized compounds will be considered as manageable data.

All details of the reactions performed during the project will be recorded in electronic lab books. Analytical data will be produced in digital form by specialized software tools used to control scientific equipment. Typical set of analytical data concerning a particular compound will include ¹H NMR, ¹³C NMR, IR, chromatogram, specific rotation and elemental analysis/mass spectroscopy spectra. These data will be collected both on analytical group servers and on the computers of the research team in an assigned folder with the name of the project. Analytical data for each sample will be named according to a specified set of rules –the owner's initials, sample number, research team number.

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 (PDFfiles, 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 hardcopies 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 in a software-specific manner and stored in appropriately labeled folders – with the name of the specific project. Additionally, the data set will be accompanied by information about acquisition and experiment dates and details, sample names, authors, brief descriptions for scanned PDFs or lab books, and others if necessary. The final dataset deposited in the institutional data repository will be accompanied by a README file listing the contents of the other files and outlining the file-naming convention used–the owner's initials, sample number, 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 repetition of experiments. Analytical data will be obtained by independent, qualified team members using precalibrated equipment, validated analytical methods and software delivered by verified

suppliers. Manually entered data will be cross-checked by other team member to avoid mistypes. Consequently, the risk of incorrect data or prejudice during data collection will be significantly reduced.

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 systematically saved (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 be created only when needed, 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 encrypted connection with the server.

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 have access to 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 sent by mail. 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, on the basis of declarations, for the purpose of conducting scientific and research activities, the legitimate interest of the IOC PAS and 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 have personal copyright, in particular the right to authorship of the result, the right to inviolability of the content and form of the result, its use and the right to mark the result with author's name.

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?

In principle, the data generated in the project will not be shared with scientific communities until the scientific publication is accepted or the patent application is approved. The results will be published in high impact journals and analytical data and details regarding reaction conditions will be available free of charge in the form of supplementary information. Additionally, scientific results of the project will be disseminated by Open Access via one of the two possible paths:

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

Open Access to scientific publications will be provided immediately after the date of publication.

Any other data, including raw data, as long as it does not compromise intellectual property interests, will be deposited on an open archive – Digital Repository of Scientific Institutes, and will be available free of charge for all interested parties. Raw data will be stored for at least 3 years after project execution.

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

Any publishable results or materials, as well as all data which underpins patent application or research publications will be long-term stored on the Institute's NAS server. The selection of relevant data will be made by principal investigator. Data in the repository will be stored in accordance with funder and Institute of Organic Chemistry of the Polish Academy of Sciences data policies.

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

All data will be provided in standard formats (doc, pdf, fid, cif files etc.) accessible using standard office software and programs intended for analytical data processing. When possible, data will be exported to formats readable by open-source software, e.g. Spin works, Chem Sketch, Mercury etc. Such software is used in virtually every research group involved in the field of organic synthesis, which is the subject of the submitted research proposal.

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?

It is not envisaged to assign DOI numbers for individual data sets. DOI numbers will be assigned to articles or communications published in scientific journals as part of the project. The procedures and analytical data concerning specific publication will be collected in appropriately named folders, which, after the article is accepted, will be additionally provided with a DOI number.

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 the data produced locally during the implementation of the planned research works lies in principal investigator (PI). Each person responsible for a particular task will be also responsible for proper data management in the scope of their task. The PI will be responsible for the preparation of the final files and their upload into the institutional repository, along with metadata. Data upload into the Digital Repository of Scientific Institutes 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 the back up and final data will be stored.

For collaborative projects, each party will be responsible for the data generated at their institution.

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 PI are already equipped with the required hardware and software necessary for implementation of the data management plan. Funds to cover the final deposit of the dataset in the institutional repository will be provided from indirect cost of the submitted research project.

Compliance with FAIR principles will be supervised by the Director's Attorney for Open Access to Scientific Publications and Research Data.