DATA MANAGEMENT PLAN

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

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

- data produced by the project personnel, e.g., laboratory notes and data concerned the protocol stored as (xlsx files);

- data produced during laboratory experiments – NMR, IR, UV spectra, X-ray diffraction data;

- data calculated during computational experiments.

The types of data will vary depending on the experiment or computational analysis performed but mostly it will consist either numerical or graphic files. The data of the performed measurement will be recorded by the qualified personnel according to appropriate standards and scientific procedures. If it will be necessary for receive reliable data, the measurements will be repeated several times.

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

Types of received data:

- physical samples of chemical compounds

- data of computational experiments (recorded in xlsx and jpg format)

- data of synthetic procedures of samples (recorded in xlsx format);

- laboratory experiments data dedicated to synthetic procedures – e.g., NMR, IR, UV spectroscopic analysis and X-ray diffraction analysis of samples (recorded in xlsx, csv, fid, cif formats);

- laboratory experiments data dedicated to biological assays (recorded in xlsx format);

- laboratory notes stored in the laboratory notebooks (hardcopy).

The predicted total volume of achieved and analyzed data will be in the range 10-100 GB.

# 2. Documentation and data quality

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

The data will be produced on the way of measurements and experimental procedures performed recorded by the qualified personnel according to appropriate scientific standards. The naming of the files and catalogs will be standardized. Selected data will be facilitated by open research data repository system MOST Wiedzy Open Research Data Catalog (common name Bridge of Data) with general metadata standards. Metadata description will be stored in xlsx or txt format. Authors will be identified and authorized by ORCID number.

## What data quality control measures will be used?

The data will be achieved by reliable standard experimental methods with established protocols. The data will be recorded by the qualified personnel according to appropriate standards and scientific procedures. The purity of the reagents and products will be controlled using spectroscopic methods, mainly using NMR spectroscopy. The obtained spectra will be compared with the standard spectra. Spectrometers will be calibrated at least once a week using reference samples provided by the manufacturer. Reactive substrates for syntheses will be stored at low temperatures (depending on the stability of the compound from -80°C to +4°C) in an atmosphere of inert gas (Ar).

# 3. Storage and backup during the research process

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

The data obtained during realization of the project will not exceed the storage of standard hard drives available, therefore, the most of the data will be stored on desktop computers of the research team members and shared via the secured network connection. The additional copy of all data will be stored on offline external drive/s by the project principal investigator (PI). The data received by any of team member will be shared with other investigators. The backup of all data will be done regularly during the whole project duration. Data will be backed up once every three months.

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

The data will be available and accessible only for t research team members from Gdansk University of Technology (GUT). The desktop computers of the research team members will be protected with a strong passwords, which will be changed once a month. Only the main investigators have access to all data and know the passwords to the data disks. Each data disc is protected with a different password. Technical staff and students or other collaborators participating only in the implementation of a part of the project do not have access to data discs. Selected data is provided to them directly by the main investigators.

# 4. Legal requirements, codes of conduct

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

Nie dotyczy,

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

The ownership and management of any intellectual property relating to the project remain in the rights of Gdansk University of Technology accordingly to the Polish law and institutional regulations (Resolution of the Senate of the Gdańsk University of Technology No. 117/2021/XXV of 19 May 2021 https://link.pg.edu.pl/GdańskTech\_intprop).

The detailed conditions for the industrial application of obtained results of development works, or the possible sale of rights to use obtained results and subject of intellectual property will be regulated in a separate agreement, taking into consideration the intellectual property rights and the rules of public aid.

The data and results published in open-access will have one of the Creative Commons licenses (CC BY or CC0).

# 5. Data sharing and long-term preservation

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

The part of the data will be published by the institutional Bridge of Data repository. The part of the data will be published in the scientific journals which may also require raw data publication (e.g., tiff, csv, fid, cif files). In case that for the obtained data the protection of the intellectual property will be considered, their publishing and dissemination will occur after the preparation of patent applications and effective achievement the protection of intellectual property.

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

Bridge of Data will be also the main service for data preservation purposes. The repository is CoreTrustSeal certified, which means that it has established good preservation and dissemination practices. Data deposited there will be automatically categorized for long term storage, without expiration date Selected data will be chosen on the basis of its scientific quality and good exemplarity. Processed data in open formats (csv/txt/tiff) will be accessible for the general public. The published data will include data, which are necessary to prepare scientific articles or patent applications. The data provided in the repository will fulfill FAIR requirements and will be categorized and labeled according to the standard file formats.

What is more, all data not selected for sharing will be stored by PI for at least 10 years after the project is finished and access to them will be possible on direct request with the PI consent.

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

Depending on the dataset the software used to its processing will be either closed license software or open-source. The data selected to share via repository will be deposited in open formats (e.g., txt, csv). The data in raw formats will be provided on direct requests. Data constituting the basis for publication will be collected in electronic supplementary information available free of charge on the journal's website.

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

The datasets provided in the repository will have the DOI assigned.

# 6. Data management responsibilities and resources

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

The project PI will be responsible for DMP , data storage and dissemination and procedures assessment and overall data quality.

## 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)?

Two percent of the direct costs of the projects will be allocated to open access publication costs and the purchase the necessary backup tools.