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 – datasets recorded during measurements;

- 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 (e.g., in case of biological experiments).

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

Types of received data:

- physical samples;

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

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

- laboratory experiments data dedicated to synthetic procedures – e.g., spectral analysis of samples (recored in xlsx, csv, fid formats);

- laboratory experiments data dedicated to biological assays (recored 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 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 Gdańsk Tech repository - the MOST WIEDZY Open Research Data Catolog. Metadata descriptions will be compatible with commonly used metadata standards and schemas, such as DataCite and stored in JSON-LD format. Persistent identifiers, such as ORCID for authors or ROR for institutions will be applied where available.

## What data quality control measures will be used?

The data will be achieved by reliable standard experimental methods with established protocols, therefore, the method chosen will not alter the achieved datasets. The data 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 (e.g., in case of biological experiments). The data available in an open repository will have DOI assigned and the data will be positioned in the way to ensure its accessibility.

# 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 project PI will be responsible for coordination of actions dedicated to analysis and management of data obtained as a result of cooperation between team members.

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

The backup of all data will be done regularly during the whole project duration. The data will be available and accessible only for trained personnel of research team from Gdansk University of Technology (Gdańsk Tech) and University of Gdańsk. The desktop computers of the research team members will be protected with a strong passwords, which will be changed once a month. No sensitive data will be collected during the experiments.

# 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 developed in collaboration relating to the project remain in the rights of Gdańsk Tech and University of Gdańsk, which will be proportional to their contribution. The percentage share of each consortium member in the law will be defined in a separate agreement. 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 of the consortium members and their institutional regulations. In the case of Gdańsk Tech it is Regulations for the Management and Commercialization of Intellectual Property at the Gdańsk University of Technology (Resolution of the Gdańsk Tech No. 117/2021/XXV, .) and for the University of Gdańsk …..

# 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 institution (Gdańsk Tech) open research data repository – the MOST WIEDZY. Data will be shared no later than the release of the articles based on these data. Due to project requirements and with the Gdańsk Tech authorities consent the data and results will be published in the open-access model under CC0 or CCBY license. Metadata will always be available without any restrictions (CC0). 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)?

The MOST WIEDZY will be the main data repository. The repository is the CoreTrustSeal certified service, which means that it has established good preservation and dissemination practices. Selected data will be chosen on the basis of its scientific quality and good exemplarity. Processed data in open formats (csv/txt/jpg) will be accessible for the general public. Data deposited in the MOST WIEDZY repository will be automatically categorized for long term storage, without expiration date. Moreover, 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 stored in open-repositories will be in open formats (e.g., txt, csv). The data in raw formats will be provided on direct requests.

## 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)?

Open Science Competence Center (https://pg.edu.pl/openscience) – established by Gdańsk Tech – will be responsible for DMP as well as data storage and dissemination through the MOST Wiedzy repository. The project will be responsible for the procedures assessment and overall data quality. Furthermore, the project PI will be responsible for coordination of actions dedicated to analysis and management of data.

## 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 main cost associated with this project will be the time investment required from the PI to ensure proper data consistency and integrity. This will involve verification processes, regular quality checks, and coordination of the data transfer between research team members.