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?

The data produced in the project may be categorized as:

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 of 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 is necessary to receive reliable data, the measurements will be repeated several times.

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

The predicted total volume of achieved and analyzed data will be in the range of 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 and recorded by qualified personnel according to appropriate scientific standards. The naming of the catalogs, files, and samples catalogs will be standardized across all indexes and lab notebookes to allow quick information searching and identification. Selected data will be facilitated by the MOST Wiedzy Open Research Data Catalog managed by Gdańsk

Unversity of Technology. Every dataset will have, stored in JSON format,.metadata compatible with generally used standards and schemas, such as DataCite. Repository metadata descriptions will contain also persistent identifiers (PIDs) where possible, i.e. ORCID for authors..

## 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 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 the realization of the project will not exceed the storage of standard hard drives available, therefore, most of the data will be stored on the desktop computers of the research team members and shared via the secured network connection. The additional copy of all data will be stored on an offline external drive/s by the project principal investigator (PI). The data received by any of the team members 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.

OSF, SONATA-20 Strona 34 ID: 632900, 2024-12-03 16:06:

The data will be available and accessible only to Gdańsk Tech research team members. The desktop computers of the team members will be protected with 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. The One Drive cloud storage service from Microsoft (Office 365 A1 for faculty license) will be additional backup.

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

# 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?

## 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 the Gdansk

University of Technology and the research team members according to the Polish law and institutional regulations .

Due to project requirements and with the Gdańsk Tech authority's consent, the data and results will be published in the open-access model under one of the Creative Commons licenses (CC BY or CC0 where possible). Metadata will always be available without any restrictions (CC0). No embargo or any other restrictions are necessary.

# 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 Gdańsk Tech research data repository – MOST Wiedzy Open Research

Data Catalog. Part of the data will be published in scientific journals which may also require raw data publication (e.g., tiff, csv, fid, cif files). Data will be shared in accordance with the publication dates of the articles that are derived from this data. 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 of 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)?

repozytorium lub archiwum danych)

The main repository of data will be the MOST Wiedzy repository, which is CoreTrustSeal certified. This certificate confirms the repository's trustworthiness and sustainability. The data to be deposited in the repository will be chosen on the basis of its scientific quality and exemplarity. Moreover, all data not selected for sharing and preservation in the

MOST Wiedzy repository will also be stored for at least 10 years after the project is finished and access to them will be possible only 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 for 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. 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)?

OSF, SONATA-20 Strona 35 ID: 632900, 2024-12-03 16:06:

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)?

OSF, SONATA-20 Strona 36 ID: 632900, 2024-12-03 16:06: