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?

- software for high-performance liquid chromatograph coupled to a mass spectrometer (Shimadzu, LCMS8060);

- LabSolutions CS Liquid Chromatography (LC) Data Integrity software (Shimadzu, LabSolution 5.99 SP2 software),

- Minitab (Minitab, LLC, USA)

- Orange Data Mining (Bioinformatics Laboratory, Faculty of Computer and Information Science, University of Ljubljana, Slovenia)

- Microsoft Excel 2016 (Microsoft Corporation, Redmond, WA, USA);

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

- MS and MS/MS spectra of compounds detected in the analyzed samples; .pdf - below 100 MB,

- chromatograms obtained during the analyses with the use of high-performance liquid chromatography technique; .lcd - approx. 2000 GB

- numerical data in tables; .cef and .xls - below 300 MB,

# 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 obtained in the project will be catalogued in separate folders and subfolders.

The name of each folder will uniquely identify its contents.

A separate folder will be created containing the data obtained during the method development (catalogue 1 containing: subfolders relating to the selection of chromatographic separation conditions and mass spectrometer operating parameters, and subfolders with MS and MS/MS spectra), validation (catalogue 2 containing information on the determination of validation parameters ), sample analysis of the contaminants (BPs, PFAS) (catalogue 3).

Relevant data will be deposited in the Open Catalog of Research Data MOST Wiedzy - a repository provided by the Gdańsk University of Technology, in accordance with the FAIR rules. The metadata description will be stored in JSON-LD format.

The author will be identified and authorized by the ORCID number.

## What data quality control measures will be used?

The data will be catalogued in a way that meets the requirements of the FAIR rules.

Data available in an open repository will be assigned with a unique identifier (DOI number), which will ensure their visibility and general

availability. All analyzes using MS/MS techniques will be performed by experienced scientists using specialized software.

All analyzes will be performed in accordance with good laboratory practice. The analytical procedures used will be validated and will be described in detail.

The equipment used will be regularly calibrated and used in accordance with the instructions provided by the manufacturers. The measurement parameters will be saved in the memory of the apparatus used.

# 3. Storage and backup during the research process

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

Laboratory data will be saved on computers connected to the laboratory equipment used and then transferred to a PC via an external drive dedicated solely to this purpose. There will be a dedicated partition on the PC for storing project-related data, which will be periodically copied to a removable drive or to a PG cloud dedicated solely to storing project data.

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

The data will be stored on university employees' computers and secured with a strong password changed regularly and known only to those involved in the project implementation. All data saved on computers connected to the laboratory equipment used will be deleted immediately after saving their copy on a computer with a separate partition for design data.

# 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?

The population study is not planned within this project, therefore no personal data will be processed or stored.

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

The ownership and management of all intellectual property related to this research project remain the rights of the Gdańsk University of Technology and of the persons involved in the implementation of the project in accordance with Polish law and institutional regulations.

# 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 data will be deposited in the Open Catalog of Research Data MOST Wiedzy - a repository provided by the Gdańsk University of Technology, at the latest when articles based on this data are published. In addition, some of the obtained data may also be published in scientific journals that require the provision of input data.

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

The main data repository will be MOST Wiedzy - a repository provided by the Gdańsk University of Technology. Due to the small amount of space necessary to archive the generated data, there is no need to select the data intended for storage. In addition, some of the research results will also be available in the open-access system, which does not have a release end date.

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

Data for potential users will be deposited in generally available file formats (.xls, .txt), which will ensure access using standard software tools.

## 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 made available in the MOST Wiedzy repository will have DOI assigned to them.

# 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 Open Science Competence Center established by PG (pg.edu.pl/openscience) will be responsible for DMP and the quality of data deposited in the MOST Wiedzy repository. The person carrying out the research activity inż. Dominika Osiecka will be responsible for evaluating the procedures and the overall quality of the 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)?

No additional measures are needed to ensure data quality and compliance with FAIR principles. The costs of purchasing removable drives necessary for data transfer and backup are planned to be financed from the indirect costs of the project.