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 will be collected and generated from two main sources. Firstly, by using the benchmark simulation models (BSM) from the IWA scientific and technical report. The BSM are standard models in wastewater treatment research and provide reliable and widely accepted simulation data. These models will be used as a starting point for analysis and as a reference for validation of results. Secondly, a new database of cases will be created using models developed in MATLAB Simulink. This data will be stored in .mat and .xlsx formats for easy processing and analysis in MATLAB and other tools. All .mat and .xlsx files will be properly labelled and described for easy identification and future use.

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

Types and formats of data generated or collected during the project:  
Text (.docx, .doc, .txt, .pdf, .pptx). Papers, tutorials, reports and presentations less than 10 Gb;  
Images (.fig, .png, .jpg, .eps .svg .pdf) no more than 10 Gb;  
Software files (MATLAB (.m, .slx, .mat .mlx), Python (.py), Jupyter Notebook (.ipynb)). up to 10 Gb;   
Numerical files - generated database (.json, .csv, .xls, .xlsx, .mat), possibly over 200 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 folders will be organised primarily by research task and secondarily by date. Selected data will be managed using the MOST Wiedzy Open Research Data Catalogue (commonly known as the Bridge of Data), a repository provided by Gdańsk Tech. The data will be described with attributes conforming to common metadata standards, and the metadata descriptions will be stored in JSON-LD format. In addition, a GitLab server hosted by Gdańsk Tech will be used to facilitate version control, collaboration and secure data storage. README file will also be created to accompany the research data. The author will be identified and authorized by ORCID number.

## What data quality control measures will be used?

Simulations will be carried out on proven models that have been successfully described in scientific papers (available at open access). Input data and initial conditions will be derived from open-source models prepared by the IWA task group on benchmarking of control strategies for WWTP. The new data will be collected and archived according to FAIR standards. Each dataset from the simulations will be validated by comparison with the BSM data. The validation will include error analysis and deviations from reference results, allowing us to assess the accuracy of models. The data available in an open repository will have DOI assigned and will be positioned to ensure 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 the research will be stored in three places: on the Gdańsk Tech GitLab server, on the Gdańsk Tech Cloud Drive, where the backup is automatically created and on the disk of the local computer.

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

The data generated and pre-processed, as well as collected within the project will be stored in a secure environment, e.g., lockable computer systems with passwords, firewall system in place, virus protection, and controlling access to digital files with password protection. Access to the institutional server is only for users with individual accounts. Every operation performed by the user will be monitored to identify who made the changes. The hard drive and laboratory notes will be kept in a locked cabinet supervised by project PI. Sensitive data will not be processed in the project.

# 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?

In the project no personal, or other requiring protection, data will be gathered or processed.

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

According to the 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 ownership and management of any intellectual property related to the project remain rights of the Gdańsk University of Technology. Selected data and results will be published upon the open-access model under CC BY or CC0 license with the consent of all their authors to allow other researchers to interpret and reuse them. Metadata descriptions for these datasets will be always 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?

Part of the data obtained in the project will be published in the open research data repository after publication of articles derived from these data, Bridge of Data. Some of the data will be published in scientific journals, which may also require the publication of the raw data. The selected author's code will be published on the GitLab platform to ensure it is accessible and compliant with the principles of open access to science. Only verified data will be published, once the verification process has been completed. The project will utilise data from benchmark models created by the IWA group or other publicly available papers or reports. Data derived from benchmarks models or available in open access repositories will not be duplicated.

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

The selection of data to be stored will be correlated with the main results obtained from the particular research tasks. The main repositories will include GitLab. and Bridge of Data, which is CoreTrustSeal certified (as the only one in Poland). This certificate confirms the repository's trustworthiness and sustainability. The data provided in the repository will fulfil FAIR requirements and will be categorized and labelled according to the standard file formats. The project does not assume the presence of data requiring to be destroyed for contracts, legal or regulatory reasons.

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

All facilitated data will be saved in .txt, .csv, .pdf, .png, .py formats, thus publicly available freeware programs are sufficient to use them. The software used to the data processing will include: Microsoft Office (Word, Excel, Power Point), MATLAB, MATLAB Simulink, Visual Studio Code, Python. The data stored in open repositories will be in open formats (e.g., .txt, .xlsx, .docx). This will enable simple and clear access to the necessary data.

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

In the repository where data will be stored, the deposited datasets will be given unique DOI identifiers. In addition, GitLab provides a unique identification number for each project called Project ID.

# 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 Principal Investigator will be responsible for data management plan, as well as data storage and dissemination. The Principal Investigator will be responsible for the 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)?

Any data or metadata processing or labelling required for it to comply with FAIR rules will be supervised by the PI. To fulfil obligations, PI will need to purchase an external drive for backup and a label printing machine and special boxes for storing samples.