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 through experimental research: microscopic imaging, cytometric analysis, western blotting, chromatography (LC-MS/MS, HPLC-DAD, GC-MS), mass spectrometry, Light Cycler (PCR), dynamic light scattering (DLS), scanning and transmission electron microscopes (SEM and TEM). The collected data will include mainly: microscopic images, cytometric, chromatograms, mass spectra, PCR, size distribution and zeta potential (DLS), SEM and TEM images files. All data will be collected from at least 3 independent experiments, then it will be processed according to standard procedures and analyzed. Data will be stored electronically in formats available in MS Office and Windows Photos. Data will be compiled into formats that are publicly available.

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

The stored data will mainly include files in .xlsx, .pdf, .jpg, .doc format, in particular: text files (.doc, .txt, .pdf), microscopic images (.jpg, .tif, .czi) ~10 000 files, western blot (.tif) ~ 200 files, spreadsheets (.exe, .xlsx), cytometric files (.c6 – a file type generated from BD AccuriTM C6 Software Version 1.0.264.21), ~3000 files, chromatograms (xlsx, csv, arw, .Icd files, .qgd formats) 1000 files, mass spectra (xlsx, csv, fid, Mnova) ~200 files, PCR (xlsx) ~500 files, SEM and TEM images (.jpg, .tif,) ~100 files, size distribution and zeta potential (.dts and .del, Malvern Zetasize) ~ 1000 files. The data will be stored in appropriate files and folders. The size of above-mentioned dataset should not exceed 2000 gigabytes (2000GB). The size is high according to the size of images taken by fluorescence and confocal microscopes (migration assays, spheroids growth changes and drug delivery). The calculated theoretical descriptor, based on the structures, will also be transferred to common datasheets. To prepare the data in a computer-readable format, CAS and SMILES notations are used.

# 2. Documentation and data quality

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

Collected data will be organized in a software-specific manner and stored in appropriately marked folders with the name of the specific project. Moreover, information about the date of the experiment, its author, method used and other significant details (such as incubation times, doses etc.) will be attached to the dataset. If necessary, the files will be accompanied by short notes in text documents. Selected data will be facilitated by MOST Wiedzy Open Research Data Catalog (Bridge of Datawith metadata standard DataCite. Metadata descriptions will be stored in JSON-LD format and will include, whenever possible, Persistent Identifiers (PID), such as ORCID or GRID.

## What data quality control measures will be used?

All laboratory protocols will be based on methods presented in the scientific literature, manuals provided by the equipment and reagents producers, and good research practice. The data will be collected from a device which would be calibrated according to the manufacturer’s instructions and it will be checked by us for errors in accordance with manufacturer’s recommendations. The correctness and repeatability of all tests will remain verified by repeating the experiments at least three times. Manually entered data will be validated by another member of the team to avoid accidental mistakes. The data will be cataloged in a standardized manner that meets the requirements of the FAIR standards. 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 the tests and the processed data will be stored on the drives of at least two local password-protected computers, to which only members of the research team will have access and in the Gdańsk Tech Repository or on Gdańsk Tech Drive. Backups will be saved systematically (at least once a month). The additional copy of all data will be stored on offline external drive/s by the project principal investigator (PI) – Ewa Augustin. 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 Gdańsk Tech and partner University of Warsaw (UW) and Medical University of Gdańsk (MUG).

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

Sensitive data will not be collected or processed in our project.

All computers used to collect, store, or analyze data have appropriate security software and anti-virus protection. Computers used and backup servers are all password protected and will be accessible only by the research team. Moreover, all USB keys used for data storage and transfer will also be protected with a  password. Data will be transferred from portable devices as soon as possible. The backup of all the data will be done during the whole project duration. The data will be available and accessible only for trained personnel – members of research team from Gdańsk University of Technology (Gdańsk Tech) and University of Warsaw (UW) and Medical University of Gdańsk (MUG). Paper documents with research data will be kept securely in drawers locked with a key, to which only the research team will have access. Paper data will not contain the names of research team.

# 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?

No personal data will be collected in our project.

## 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 developed in collaboration relating to the project remain in the equal rights of the Gdańsk Tech and partner of consortium (UW and MUG) which will be proportional to their contribution. The percentage of share of each institution will be defined in a separate agreement. The data and results published in open-access will have CCBY or CC0 licensce.  Metadata in the repository will always be available without any restrictions (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?

Project data will not be shared with the research community until the time of acceptance of the scientific publication. The results will be published in high impact journals and its content will be available immediately after the publication date in accordance with the NCN’s Open Science policy. Alongside the article, data underlying this article will be published as supplementary materials or as the dataset in the Bridge of Data repository or another repository proposed by consortium partners that maintains the same service quality standards. 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)?

Raw experimental data will be stored at least 10 years after project completion. Any publishable results and materials, as well as all data underlying scientific publications will be stored for a long time in the Bridge of Data  from Gdańsk Tech. The repository is CoreTrustSeal certified, which means that it has established good preservation and dissemination practices. All data will be prepared according to the FAIR principle and will be categorized and labeled according to the standard file formats.

## 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 standard formats (e.g. .doc, .pdf, .jpg, .exe). 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?

DOI numbers will be given to article(s) published in scientific journals as part of the project. If the dataset is shared in the Bridge of Data repository it will also be assigned a DOI number.

# 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 (pg.edu.pl/openscience) established by the Gdańsk University of Technology will be responsible DMP and the quality of metadata descriptions in the Bridge of Data repository. Ewa Augustin, the project PI will also be responsible for the proper management and storage of data on individual experiments as well as preparation of final files and sending them to the Gdańsk Tech repository.

## 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)?

A significant overhead associated with research data is the time and additional work that the PI will have to undertake to ensure data integrity and consistency alongside all research teams, as well as proper preparation of data and their descriptions before depositing them in the repository.