European Research Council
Executive Agency

Established by the European Commission

European Research Council (ERC)

ERC Data Management Plan

European Research Council
Established by the European Commission

ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

Project Acronym	Project Number
RAMP-UP	101113438

This is version no: 1 of the DMP for the RAMP-UP project. This version is completed on 26.01.2024.

The project website is: https://www.nems.me/microplastics

SUMMARY

In this project, the aim is to use entirely electronic measurements to distinguish microplastic particles from other micro-scale contaminants in water samples. As such three major classes of data are expected to be obtained during the project. These three classes of data will be stored in the repository **Zenodo** which is commonly used in physical sciences. Zenodo platform provides input to *OpenAire* system, meets FAIR standards, and therefore is a *trusted repository*.

The table below summarizes the three classes of project data (and important sub-classes).

Data Class	Sub-Class	Underlying Mechanism	File Extension / Format	Minimal Storage Duration
Class 1	Raw Data	Electronic Sensor	.TXT Text file containing numerical values	5 years
	Derived Data	Electronic Sensor	.MAT Matlab Variable	10 years
Class 2	Raw Data	Raman Spectroscopy	.WIP WITec Software Data File and .TXT Text File	5 years
	Derived Data	Raman Spectroscopy	.MAT Matlab Variable	10 years
Class 3	Raw Data	Optical Microscope	.JPEG Image File	5 years

Data Class 1: Readings from the electronic sensor.

Raw Data Subclass: The main output of the project is the determination of the dielectric permittivity of microparticles by our novel electronic sensors. The output of these electronic sensors is a time-series data. This data output is saved by the electronic instruments as a numerical dataset in .TXT format.

In a typical data run, there are **6 parameters**: microwave sensor amplitude, microwave sensor phase, Coulter sensor amplitude, and 3 time stamps associated with each variable (since time stamps can have slight differences).

In one of the tasks of the project, we are using a multimode measurement approach. In this approach, the number of parameters is increased. Instead of 1, we are using 3 microwave sensor channels: thus instead of 6 parameters, there will be **14 parameters** (6 for microwave sensor readings, 1 for Coulter sensor reading, and 7 for their corresponding time stamps).

For each minute of operation, the system generates approximately 80 MB of data. A single data file typically encompasses 40-60 minutes of operation, therefore takes up approximately 5 GB of storage space. For multimode measurements, this figure increases to 12 GB. These figures are below the 50 GB per dataset for the Zenodo repository. A complete experiment will include 30-50 of such files.

Derived Data Subclass: The raw data described above take a large amount of space (especially since there are many of these). In our technique, we extract information from raw data about microparticles in the MATLAB software. By using a program we wrote, we can obtain the parameters pertaining to each particle. Therefore, this set of parameters serves as a useful summary of the raw waveforms. Since this *derived data set of extracted parameters* is more useful and takes little space (<10 MB per set), we will also provide these files as a derived dataset in conjunction with the project.

Metadata: Each .TXT raw data file also contains **header information** in the first few columns as Metadata. These columns record the date, time of the experiment, specific electronic instruments used, and parameters chosen during data acquisitions. The presence of the header file information will be explained in the Description section of Zenodo forms. More general Metadata includes mapping information between different data classes, so that each microparticle analyzed can be traced across different sensor modalities and classes of data.

Data Class 2: Raman Spectroscopy of Microparticles

In the project, we will be using an optical technique called Raman Spectroscopy for the *cross-validation* of our electronic sensing methodology (described above as Class 1). The data files are obtained from the commercial Raman Spectroscopy instrument in the form of .WIP files. These files can be viewed by WITec Control software suite provided by the Raman Spectroscopy Instrument manufacturer.

Raw Data Subclass: Although WITec software is specialized to a certain commercial company, it is commonly used in the field by other scientists. So, we will deposit the raw data files (with metadata information as described below) as is. We will also convert these to the more accessible .TXT files using the WITec software and deposit these versions as well to enable broader access. Both raw and derived data files take very little space (for each spectrum: <100 kB for TXT data; 1.5 MB for raw data file).

Derived Data Subclass: During our verification studies, we will convert the .TXT files from Raman Spectrometry into Matlab files (.MAT format as the data) to further analyze/plot the data. For completeness, we will also deposit the derived data subclass as well.

Metadata: At the dataset level, each .WIP file contains metadata information describing instrument settings for individual measurements. Thus, sharing .WIP files will enhance interoperability. At a more general level, for each dataset, the individual spectrum for each microparticle will possess a systematic filename enumerating the microparticles, such as
Particle001.txt
Particle002.txt
The relevant metadata of these particles will be posted as a .TXT file so that matching between this data class (Raman Spectra) and the other data classes (electronic sensor measurement of the same particle as Class 1, and the microscope image of the particle as Class 3) can be established.

Data Class 3: Optical Microscopy Images of Microparticles

During the project, microscopic images of microparticles will be obtained during two separate occasions:
1) in Raman Spectroscopy, in addition to the spectrum, the image of a particle can, and will, be saved.
2) in our electronic measurements, conducted in a microfluidic channel and under an optical microscope, the images of the electronically measured particles will be obtained.

The optical microscopy images complement the other types of data classes. For instance, the small size or dark color of a microparticle can obfuscate the Raman Spectroscopy measurements (Data Class 2), so linking the spectrum with the image of a particle could play an important role for such particles. Similarly, the shape of a particle can induce extra uncertainty in the electronic measurements (Data Class 1). However, the knowledge of the particle shape as obtained from optical microscopy can enable us to remove such uncertainties. For these reasons, optical microscopy images will be stored and linked to the other data classes of the same microparticle.

The images from our microscope are typically saved in .CZI format. This is a rather specialized format, but we also can save these images in more general formats such as in .JPG, .JPEG, .PNG etc. The typical image would be around 35 MB (.CZI format), 5 MB (.JPEG format), 10 MB (.PNG

format) in size. Due to the size of the file and general accessibility, we will deposit JPEG format images for each microparticle.

Metadata: As metadata, a .TXT file will be deposited which will link the image filenames with the enumeration of particles used in other data classes. This way, all three types of data for the same particle can be traced.

Organization of Data Sets
For datasets used in publications, to achieve coherence and transparency, a separate and complete dataset will be deposited (which may include data files from earlier entries). These datasets will contain all applicable data classes. In addition to data curation and metadata information, a systematic scheme will be used for the filenames, so that different measurements of the same particle can be linked together, and a machine can access and process the files smoothly, if desired.

Any dataset not used in publications will be considered for uploading to the repository whenever there is a scientifically valid reason (such as verification of a different study, calibration of standard samples etc.).

1. MAKING DATA FINDABLE

As a general measure, we will also publish the data management plan on the website of the Action:

https://www.nems.me/microplastics

We will also provide a link to the data repository from this page, when datasets are submitted, for instance following a publication. For each publication, the data repository will be cited within the paper to provide access, typically found in the Data Availability Statements.

The Zenodo repository provides text entry areas for the dataset description. As discussed above, for each dataset class, the text entry areas will be used for providing the important metadata information and identification of specialized metadata files. The most critical metadata is the mapping between different data classes of the same particle. This will be described in detail and endorsed by the systematic choice of filenames.

The Zenodo platform also provides a DOI number which can be used as a persistent and unique identifier of the dataset. We have already created a username with this platform and reviewed the way it works. As systematic datasets for publications are completed, these will be uploaded there together with the corresponding metadata.

We will further mark the datasets for inclusion in the **Microplastics in the Environment** community of Zenodo to make it further findable within the context of microplastics problem.

2. MAKING DATA OPENLY ACCESSIBLE

Since the data generated in this project have no dimensions involving confidential or personal information, all data will be made openly available. There could be two exceptions. First, if a dataset can be used as a trade secret (a potential outcome of a PoC project), then the decision to release the data will be assessed within the research team and in consultation with the funding agency. The second exception is to protect the IP rights. In this case, we will postpone uploading the data until **3 months after** the corresponding patent application is filed. The duration of 3 months is chosen, because any communications with the patent office (e.g. about the form and completeness of a patent application) is received by the inventors typically within 3 months.

The repository we will use **both for metadata and the data** will be Zenodo which is a free repository managed by CERN. Zenodo is a well-established and trusted platform. In the table below, we list how the data can be accessed depending on the type of the data.

Data Class	**Sub-Class**	**File Extension / Format**	**Data Access Method and Software**
Class 1 Electronic Sensor	Raw Data	.TXT Text File with numerical values	Data can be downloaded to a local computer and opened by a myriad of programs such as Notepad, Python, Matlab, Microsoft Excel, and Apple Numbers.
	Derived Data	.MAT Matlab Variable	Data can be downloaded. While Matlab program is required to access MAT files, there is a MATLAB Online service which provides **basic access to the program for free.** Thus, any interested user can convert these MAT files to any other format as desired for free and without having to install any software.
Class 2 Raman Spectroscopy	Raw Data	.WIP WITec Software Data File and .TXT Text File	Data can be downloaded. While the WIP files require the specialized WITec software, the same data in TXT format will also be provided, which can be opened by any text editor software such as the free programs of Notepad (Microsoft) or Notes (Apple).
	Derived Data	.MAT Matlab Variable	Data can be downloaded. While Matlab program is required to access MAT files, there is a MATLAB Online service which provides **basic access to the program for free.** Thus, any interested user can convert these MAT files to any other format as desired for free and without having to install any software.
Class 3 Optical Microscopy Images	Raw Data	.JPEG Image File	Data can be downloaded. Any image processing software can open JPEG files including the free software Image-J.

3. MAKING DATA INTEROPERABLE

For the field of research undertaken in this project, there are not yet well-established standards for data interoperability in general.

For data class 1 (electronic sensor data), there are no standards or field-specific conventions. As such, other than providing the complete data and metadata, no further actions are necessarily foreseen.

For data class 2 (Raman Spectroscopy), the data will be provided both in TXT format (to be accessible by anyone), as well as the WIP format which enables other practitioners of the field and users of Raman Spectroscopy Instruments to import and visualize the same data on their computers seamlessly. Since WIP file contains metadata information describing instrument settings, all the relevant parameters while taking the data will be accessible to the other scientists, and this information ensures the interoperability of the Raman Spectroscopy data.

Finally, for data class 3, in addition to providing the image of microparticles, the metadata field will include **the field of view information (including scale bars)** so that accurate size information from the images can be obtained.

4. INCREASE DATA RE-USE

The raw and derived data files will be stored in repositories indefinitely. Per the policy of Zenodo, in case of a closure, measures will be taken (by Zenodo and us) to migrate the data to other suitable repositories.

We believe that storing data for 5-10 years will be sufficient since as the field matures the conclusions and findings published in curated form in scientific journals will become more relevant than the raw data, given that the experiments outlined will become more reproducible and accessible by the community. That is why we believe that the minimal storage durations of 5 years for **raw data** and 10 years for the more useful **derived data** are suitable values. Indeed, Zenodo platform does not currently put any limit on the duration of the data set, and with CERN being its supporting institution, a storage duration longer than 10 year is reasonably expected.

Embargo period: For any publication, as long as the paper is published online, the data will be uploaded without any embargo period. For any patent applications (which will be processed before any publications), we foresee an embargo period of 3 months after filing the data, since any initial communication from a patent office (e.g. the control of the form and figures of the application) typically takes as long as 3 months.

Licensing: For this project, we will select **CC BY-NC** as the licensing. This licensing type **prevents the licensed data from being used for commercial purpose.** Indeed, since the rationale behind

the Proof-of-Concept project is the valorization of research results, the data produced can contain commercially valuable information and processing-techniques (such as the matching between the electronic sensor information and the Raman spectra). To prevent a third-party from using this information for commercial purposes and protect the rights of the IP holders for future commercialization, we will select CC BY-NC licensing.

5. ALLOCATION OF RESOURCES and DATA SECURITY

Zenodo repository is free to use, so we do not expect any costs for storing and sharing data. The Zenodo repository does not cite any disruption to its services for the foreseeable future. The website for the action is supported by the internal resources of the host institute (Bilkent University) and we do not foresee any disruption for the foreseeable future.

Data backup and recovery: any data generated in the laboratory is copied to the internal FTP server of Bilkent University at the end of each day: this way, different users in the research team can access and analyze the data. This mechanism automatically provides a means for data backup. To prepare an extra level of protection, **each data set uploaded to the repository will also be saved in the Dropbox account of the Principal Investigator.** This way, the data will be migrated to a Cloud service, and can be recovered even if the repository or the local machine fails.

As mentioned earlier, the project does not contain any dimensions for personal information or security. On the other hand, it may contain data with commercialization value. In addition to protecting IPR by patenting and using embargo periods, the release of any datasets which could be used as a trade-secret will be performed in consultation with ERC, first.