

Deliverable 2.1

# Data Management Plan

Deliverable Information			
Title	Data Management Plan		
Deliverable number	2.1		
WP number	2		
Lead beneficiary	University of Galway (short name: GALWAY)		
Contributing partners	All		
Author(s)	Aleksandra Rankov, Milana Todoreskov, Nataša Stojić, Mark Healy, Liam Morrison, Sebastian Primpke, Miloš Rajković		
Type	OTHER		
Dissemination level	PU - Public		
Delivery date	30.06.2023. (M6)		
Version	v1.0	Date	16/06/2023

Project Information	
Title	Twinning Microplastic – Free environment
Acronym	GREENLand
Project No.	101079267
Type of Action	HORIZON-CSA
Start date of the project:	01/01/2023
Duration:	36 months
Website	<a href="https://project-greenland.com/">https://project-greenland.com/</a>

## Contents

1	Introduction .....	4
2	Data Management.....	5
2.1	Research data management	5
2.1.1	Repository	8
2.1.2	Website (GREENLand project)	9
2.2	Project team collaboration data management.....	9
2.3	Internal collaboration at Educons university .....	9
3	FAIR data principles.....	9
3.1	Making data findable, including provisions for metadata .....	9
3.2	Making data accessible.....	10
3.3	Making data interoperable .....	11
3.4	Increase data re-use .....	12
4	Other research outputs .....	12
5	Allocation of resources.....	12
6	Data security .....	13
7	Ethics .....	14
8	Other issues.....	14
9	Conclusions .....	14
10	List of Abbreviations .....	14

# 1 Introduction

Data management plan (DMP) outlines the approaches in which data (collected, generated, processed, and re-used) will be managed during the GREENLand project lifetime and upon its completion. It is a public live document which will be updated and provide all relevant details about specific characteristics of the GREENLand project in terms of data management.

The document structure is aligned with the template recommended for Horizon Europe beneficiaries and represents a mandatory element of all projects involving data.

Data should comply with the FAIR data principles which aim to make data more valuable by making it easily findable, accessible, interoperable, and reusable.

The DMP indicates what data sets will be collected and created as part of the project and how these data will be stored and made easily findable and accessible. It describes the measures that ensure the data collected and generated during the project are interoperable i.e. available for exchange in standardised format and understood by different systems or applications. Approaches to increasing the data reusability through good documentation and preservation are also described, so data can be easily understood and utilised in different contexts and by different users/apps. The measures undertaken to secure and protect sensitive data are also included.

The data acquired and generated by this project will be stored in secure storage resources in a Cloud-Based System set up by the Project Coordinator and will be accessible anywhere from any device to anyone having assigned access rights.

The data management plan within the GREENLand project is created in line with its objectives that are summarised here:

- O1: Develop science, innovation, and technology (SIT) roadmaps of microplastics in soil, fresh water and microorganisms
- O2: Significantly improve education and administrative workflow by developed e-Tools
- O3: Improve capability to compete for EU funding
- O4: Build capacity through the new generation of highly skilled staff
- O5: Significantly improve dissemination, exploitation and collaboration

The first objective (O1) is related to the research activities and manipulation of the research data and research results.

The second objective (O2) sets the activities that will build and increase capacities for teaching and knowledge transfer between teachers and students, and at the same time improve the operational processes and procedures at university (Educons university) through deployment of e-Tools.

The third (O3) and fourth (O4) objective should raise the research capacities and capabilities for EU funds through knowledge exchange, networking and staff training.

The last objective (O5) should lead towards better, more impactful communication and dissemination of research results.

To support realisation of all these objectives, the data management plan should consider three types of data: (1) research data and research results, (2) project internal (operating) data, and (3) Educons University data.



## 2 Data Management

### 2.1 Research data management

Here is provided important information on the type of data that will be used within the GREENLand project in the process of creating the research outputs. Each project partner (AWI<sup>1</sup>, EDU<sup>2</sup>, GALWAY<sup>3</sup>) specifies the type of collected/generated/used data, their purpose, data formats, re-use of data and their purpose if applicable, the origin of data, expected size and data utility outside the project. Collected information on data will be used to further specify requirements for FAIR data.

<b>Data Summary: Beneficiary 2 – EDU</b>		
<b>What is the purpose of the data collection/generation and its relation to the objectives of the project?</b>	<p>Analysis of soil samples (sampled from the selected pilot areas) on organic and inorganic pollutants like PCBs, PDDEs, PAHs, Pesticides and heavy metals, their physical and chemical characteristics, as well as the characterization of the pilot area.</p> <p>The presence of microplastics in the selected samples will also be analysed.</p> <p>Obtained results will be used for the development of a database, as well as for the spatial display of the Resilience Index at different locations.</p> <p>This will enable us to get a visual innovative roadmap that will help us in accepting and implementing ideas and conclusions in practice.</p>	
<b>What types and formats of data will the project generate/collect?</b>	<p><b>Types of data</b></p> <ul style="list-style-type: none"> <li>Binary/Numeric data</li> <li>Tabular data</li> <li>Textual data</li> <li>Image data</li> <li>Documentation</li> <li>Papers and Articles</li> <li>Geographic data (ARCGIS) - raster and vector</li> </ul>	<p><b>Formats of data</b></p> <ul style="list-style-type: none"> <li>.0, .1, .2, .3</li> <li>CSV, XML</li> <li>TXT</li> <li>PNG, JPG, JPEG, BMP</li> <li>ODT, PDF, HTM, XLSX</li> <li>TXT/DOC/DOCX/PDF</li> <li>ARCGIS</li> </ul>

<sup>1</sup> Alfred Wegener institute (AWI)

<sup>2</sup> Educons University (EDU)

<sup>3</sup> University of Galway (NUIG)



<b>Will you re-use any existing data and how? State the reasons if you are not planning to re-use it.</b>	We will select the pilot area based on our previous research. We have about 5,000 results on the concentration of phthalates in soil from the territory of Vojvodina. Since phthalates are mainly of plastic origin, in locations where the concentration of total phthalates is higher than legally prescribed, we will sample the soil and perform analyses.
<b>What is the origin of the data?</b>	New experiments and existing results from analyses conducted at EDU
<b>What is the expected size of the data?</b>	max 4 GB
<b>To whom might your data be useful ('data utility') outside the project?</b>	Researchers, Educators, Governmental agencies, Policymakers, General Public

#### Data Summary: Beneficiary 1- AWI

<b>What is the purpose of the data collection/generation and its relation to the objectives of the project?</b>	Analysis of soil samples sampled in the selected pilot area on organic and inorganic pollutants like PCBs, PDDEs, PAHs, Pesticides, and heavy metals, physico-chemical characteristics, as well as characterization of the pilot area. The presence of microplastics in the selected samples will also be analysed. Obtained results will be used for development of the database and the spatial display of the Resilience Index in different locations. This will enable us to get a visual innovative roadmap that will help us in accepting and implementing ideas and conclusions in practice.	
<b>What types and formats of data will the project generate/collect?</b>	<b>Types of data</b> Binary/Numeric data Tabular data Textual data Audio data Image data Documentation Papers and Articles Special	<b>Formats of data</b> .0, .1, .2, .3 CSV TXT SPE JPG, TIFF, BMP ODT, PDF, XLSX TXT/DOCX/PDF PRP

<b>Will you re-use any existing data and how? State the reasons if you are not planning to re-use it.</b>	We will use existing reference libraries for microplastics analysis. AWI will analyse previously collected samples generated by Educons.
<b>What is the origin of the data?</b>	New experiments and existing results from analyses conducted at EDU
<b>What is the expected size of the data?</b>	max. 30 TB
<b>To whom might your data be useful ('data utility') outside the project?</b>	Researchers, Educators, Governmental agencies, Policymakers, General Public

#### Data Summary: Beneficiary 3 - GALWAY

<b>What is the purpose of the data collection/generation and its relation to the objectives of the project?</b>	Analysis of soil samples sampled in the selected pilot area on organic and inorganic pollutants like PCBs, PDDEs, PAHs, Pesticides and heavy metals, physico-chemical characteristics, as well as characterization of the pilot area. The presence of microplastics in the selected samples will also be analysed. Obtained results will be used for development of the database and the spatial display of the Resilience Index in different locations. This will enable us to get a visual innovative roadmap that will help us in accepting and implementing ideas and conclusions in practice.																
<b>What types and formats of data will the project generate/collect?</b>	<table border="0"> <thead> <tr> <th><b>Types of data</b></th> <th><b>Formats of data</b></th> </tr> </thead> <tbody> <tr> <td>Binary/Numeric data</td> <td>.0, .1, .2, .3</td> </tr> <tr> <td>Tabular data</td> <td>CSV, XML</td> </tr> <tr> <td>Textual data</td> <td>TXT</td> </tr> <tr> <td>Image data</td> <td>PNG, JPG, JPEG, TIFF, BMP</td> </tr> <tr> <td>Documentation</td> <td>PDF, HTM, XLSX</td> </tr> <tr> <td>Papers and Articles</td> <td>TXT/DOC/DOCX/PDF</td> </tr> <tr> <td>Geographic data (ARCGIS) - raster and vector</td> <td>ARCGIS</td> </tr> </tbody> </table>	<b>Types of data</b>	<b>Formats of data</b>	Binary/Numeric data	.0, .1, .2, .3	Tabular data	CSV, XML	Textual data	TXT	Image data	PNG, JPG, JPEG, TIFF, BMP	Documentation	PDF, HTM, XLSX	Papers and Articles	TXT/DOC/DOCX/PDF	Geographic data (ARCGIS) - raster and vector	ARCGIS
<b>Types of data</b>	<b>Formats of data</b>																
Binary/Numeric data	.0, .1, .2, .3																
Tabular data	CSV, XML																
Textual data	TXT																
Image data	PNG, JPG, JPEG, TIFF, BMP																
Documentation	PDF, HTM, XLSX																
Papers and Articles	TXT/DOC/DOCX/PDF																
Geographic data (ARCGIS) - raster and vector	ARCGIS																

<b>Will you re-use any existing data and how? State the reasons if you are not planning to re-use it.</b>	n/a. The University of Galway will analyse previously collected samples generated by Educons.
<b>What is the origin of the data?</b>	New experiments and existing results from analyses conducted at EDU
<b>What is the expected size of the data?</b>	max 4 GB
<b>To whom might your data be useful ('data utility') outside the project?</b>	Researchers, Educators, Governmental agencies, Policymakers, General Public

### 2.1.1 Repository

Research teams from three partner institutions (AWI, EDU, GALWAY) focus on research and development of science, innovation, and technology (SIT) roadmaps of microplastics in soil, fresh water and microorganisms.

Educons University provides the data (historical and new ones) from pilot areas that is further analysed and processed together with project partners GALWAY and AWI, the experts in microplastic detection methodologies, in their designated laboratories. In addition to the knowledge exchange and creation of the training and education material, the research work will be published in high-impact journals and presented at scientific conferences, workshops, and symposia. Academic networking structures will be leveraged as a "multiplier" to transfer knowledge and information to the primary industry sectors, stakeholders, and decision-makers for their effective use and benefit. It is expected that the project encourages additional applied R&D, innovation and exploitation activities within the science community. Each partner will promote and deliver presentations at targeted conferences, as well as publications in peer reviewed journals as specified in the project proposal.

All peer-reviewed scientific publications during the 3-year duration of the project will be published via Gold Open Access. With such a publishing model, articles will be freely available online to anyone accessing the internet immediately upon publication without any fees. Academic/research partners have included costs in their partner budgets, so that fee can be paid to the publisher to cover all the costs. By making the research available to a wider scientific community it will facilitate collaborations, knowledge exchange and contribute to the public good.

Outside the duration of the project, i.e., after three years, papers will still be published via Open Access but the author(s) will choose to use either the Gold or Green Open Access route. Green open access refers to self-archiving where authors deposit their published or even unpublished work/articles in a digital repository at no fee after obtaining permission from the publisher (after the embargo period). Green Open Access allows researchers to make their work more widely available without violating copyright agreements. Research results can be made available to everyone including researchers, students, educators, governmental



agencies, policymakers, and the general public. Compliance with open access mandates is important for researchers as it promotes transparency, accountability, and knowledge-sharing.

Educons University established a new repository called REDUN that will be used for depositing all the published work resulting from the project. In addition to publications, project partners will decide what supplementary materials can be added as supporting material to published work. This can include raw datasets, additional results and data representations that can help the research community to re-use the data and reproduce the published results.

### 2.1.2 Website (GREENLand project)

All scientific publications, public deliverables and reports will be available through the project's website, i.e. project information portal The Microplastics-free Hotspot. The portal will be developed by EDU and offer a detailed source of information in the field of Microplastics for the academic community and stakeholders.

## 2.2 Project team collaboration data management

Collaboration between project partners

MS Teams and SharePoint applications are used for collaboration between project partners. They include presentations/photos/visual/template and other documents.

## 2.3 Internal collaboration at Educons university

There is a newly deployed cloud-based system, Google Workspace, at Educons University which aims to:

- a. facilitate and improve collaboration and communication between the employees
- b. provide support to eLearning and exchange of learning materials
- c. enable a document management system (at least for the backup purposes)

## 3 FAIR data principles

The project data will comply with the FAIR data principles – making the data Findable, Accessible, Interoperable and Reusable.

This is in alignment with the commitment to early and open sharing of research results and outputs involving all relevant knowledge actors - Researchers, Educators, Governmental agencies, Policymakers, General Public that take part in co-creation of SIT roadmaps with the new approaches for the purpose of microplastic-free environment.

### 3.1 Making data findable, including provisions for metadata

Data to be shared with the research community and/or specific stakeholders for which the data will be useful should be easy to find (by both human and machine) and therefore will have to meet the following criteria:

- Each data type will use an internationally accepted file format for its effective sharing, reuse, and preservation. This will improve interoperability of data between different

systems. For example, csv and xml are standard formats for tabular data that can easily be used by different software.

- Data will be accompanied with relevant documentation that will describe context, content and structure of data. That will include all the important information on data origin, collection methods and any manipulation or processing applied to data.

With REDUN, the Educons University repository, each dataset will have a Digital Object Identifier (DOI) and Persistent URLs so that it can be easily and unambiguously found and permanently cited. DOI will provide a permanent link to a digital object, such as a dataset or a publication, even if its format or location changes over time.

Dublin Core Metadata schema will be used.

For better traceability of each dataset and its subsequent use, search keywords will be provided once uploaded to the REDUN repository.

Specific types of metadata that will be created for digital objects depend on the needs and priorities of the domain and community in which they are being used. Rich and standardised metadata will be created as it is essential for enabling discovery, understanding, and reuse of digital objects. These include descriptive (title, author, date subject), structural (file format, size), administration (access rights, preservation strategy, history of objects), technical (hardware), preservation (migration, backup, recovery) metadata.

Each deposited item (document) will be individually tagged with an appropriate license.

The metadata for individual records, as well as for record collections, are harvestable using the OAI-PMH protocol by the record identifier and the collection name. Metadata are made available to harvesters and aggregators under the Creative Commons [CC0 Public Domain Dedication](#) licence. Data can be harvested by robots and aggregators for full-text indexing or analysis in line with the assigned licence. Metadata are freely accessible to all, and freely reusable by all, under the terms of the CC0 licence.

The naming convention should follow the following format: [GREENLand\_Date\_WP#T#]

## 3.2 Making data accessible

To make data openly accessible, GREENLand will use the following open repositories: REDUN and the project website.

### ***REDUN repository***

The EDUCONS University repository, REDUN (<https://redun.educons.edu.rs>), has been established and it will support realization of the Data Management Plan.

REDUN is the institutional digital repository of EDUCONS University. It provides open access to publications and other research outputs resulting from projects implemented by this institution.

The software platform of the repository is adapted to the modern standards applied in the dissemination of scientific publications, and is compatible with international infrastructure in this field. It is provided and maintained by the Computing Centre of the University in Belgrade (RCUB).

The external application Authors, Projects, Publications (APP) can be used to browse and search authors and funding information. APP also enables metadata export and displays Altmetric scores and Dimensions, Scopus and Web of Science citation counts.

Bibliographic metadata are publicly available under the CC0 1.0 Universal license (CC0 1.0), meaning transfer to the public domain to all users, without prior registration, both through the user interface and through the OAI-PMH protocol.

Certain documents can only be accessed by registered users, after logging into the system.

**The project website** - <https://project-greenland.com/>

The project website provides important information about the project and it will be also used to provide access to public deliverables and dissemination materials.

When sharing results, we will follow the principle "as open as possible, as closed as necessary", and data that can be shared will be shared under Creative Commons licenses. Data will be shared upon publishing the results based on these data.

Depositing publications into the repository does not require additional documentation. If depositing other types of data, users can provide additional, supplementary documentation that is important for understanding, interpreting, and reusing the data. This will include a context of data collection, methods for data collection, file structure, data organisation and reliability, any changes to raw data, data processing methods, data confidentiality, access rights, software required to open files if needed. All this will be provided in README file.

Deposited documents will be made available to the public except in cases when embargoed for a limited period of time which can only apply to the period upon the project completion if no Gold open access applies.

Metadata will be guaranteed to remain available even after data is no longer available.

### 3.3 Making data interoperable

To achieve data interoperability, data will be collected and shared using a standardised approach, using a common format for the specific data type referenced above.

Information about the software needed to run the data will also be provided, where applicable.

Obstacles to data access due to interoperability problems are not likely.

The format of metadata will follow the convention of the REDUN repository.

For (meta) data REDUN uses a formal, accessible, shared, and broadly applicable language for knowledge representation. The repository uses JSON Schema as an internal metadata representation and offers export to other popular formats, such as Dublin Core or MARCXML. The data use vocabularies that follow FAIR principles. For specific terms, REDUN refers to open, external vocabularies, e.g., license (Open Definition), funders (FundRef), and grants (OpenAIRE). All (meta) data include qualified references to other (meta)data. Each referenced external piece of metadata is qualified by a resolvable URL.

Metadata is mapped in accordance with the Dublin Core, which is an official, accessible and widely accepted standard that ensures machine readability.

Different types of attachments/documents (research data, surge versions) can be linked to the base record through the dc.relation + qualifier field.

### 3.4 Increase data re-use

Public data will be available for reuse. To maximise data reuse and facilitate data retrieval by a broader user base, it is favourable to adopt a protocol that is free and open-source.

The consortium will attach specific licences to the stored data that define all the conditions under which the work is made available under open or restricted access.

REDUN automatically offers five different licensing options among **Creative Commons (CC) licences**, all of which require attribution in order for authors to be properly credited for their original work (attribution, licence reference, and modification information).

In case the partner would like to further limit access to the uploaded data, alternative licences will also be selected through the Creative Commons licence chooser among the REDUN offered options.

All **Personal Identifiable Information** will be restricted to internal usage and will not be shared with third parties.

As mentioned above, when sharing results, the consortium will generally follow the principle, "as open as possible, as closed as necessary"

Documentation (e.g. readme files) needed to validate data analysis and facilitate data re-use will be provided with all necessary information on data, data cleaning, analyses, variable definitions, units of measurement, etc.

Data will be made freely available in the public domain to enable the widest reuse possible.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

## 4 Other research outputs

Beneficiaries will also consider **management of other research outputs** that will be generated or re-used throughout their projects. Such outputs will be either **digital** (e.g. **workflows, protocols, models**, etc.) or **physical** (e.g. **new materials, samples**, videos, presentations etc.).

The same principles will be applied to the digital research outputs as to the published research results and will be available for re-use.

## 5 Allocation of resources

### ***Estimation of Costs***

The costs associated with open access to research data are eligible as part of the Horizon Europe grant (as long as the costs are in line with the Grant Agreement and the rules of the Horizon Europe programme).

The sources for long-term data preservation, the associated costs, as well as the means on how the data will be preserved after the end of the project and for how long, will be decided by the Project Advisory Board during the project.

Estimated person-months: 2.

Long-term preservation of data: free of charge.

### ***Responsibilities for Data Management***

Responsibility for data management and all issues related to open access or data security belongs to the Project Coordinator (EDU), which has allocated a budget and person months for this activity under WP1.

**Costs related to research data/output management are eligible** as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

In addition, the Google Workspace that is used for data storage and archiving is free for educational institutions.

The responsible data manager for the GREENLand project is: Aleksandra Rankov

## **6 Data security**

The Data Protection Officer (DPO) of the project will be Nikola Plavšić from EDU, and data protection will comply with the General Data Protection Regulation (GDPR).

The initial plan is to store videos on a YouTube channel of the project.

Other data will be stored on a Cloud Based System (CBS). Different CBS will be used for different purpose (internal organisational data/published data/ project team collaboration data).

Two backups will be taken every 1st day of the month on external disks.

The folder named **GREENLand** project will include subfolders with the names of work packages that will be administered by corresponding WP Leaders. All members of the team will have access to view files on a cloud-based system, but only leaders of WPs will administrate the folder of their WP, and the Project Coordinator (PC) will administer all data in all folders.

Furthermore, all members of the team will have access to the utilised data on cloud-based folders after the implementation of the project, but the raw data that is not used will remain the property of the team member who discovered them and also to the PC.

Computers will have antivirus protection and firewall software. All protocols and measurement data will be stored on a Cloud Based System in Excel files. Excel files will have name in the following format: "GREEN" as abbreviation for **GREENLand** project; year in which the activity was conducted; "WP" as abbreviation for Work Package; number of WP to which the document belongs; "T" as abbreviation for Task; number of task; e.g. "GREEN2023WP2T2" – 2nd task in work package 2 in 2023.

List of all project activities will be placed in the main folder as well as the instructions for storing data in folders. All materials will be also stored in subfolders of WP to whom they belong.

All deliverable documents (scientific publications, pictures, strategies and guides) will be stored in the EDU repository called REDUN. Files and metadata deposited onto repository will stay there permanently and may be removed only in exceptional cases. In such cases withdrawn items will only be removed from public view.

## 7 Ethics

Research activities in microplastics do not include any personal data collection.

In addition, the REDUN repository that will be used for depositing research data and results guarantees privacy protection so any information gathered upon visiting the repository does not identify the user personally.

## 8 Other issues

Data management process will not make use of other national/funder/sectorial/departmental procedures.

## 9 Conclusions

Data Management Plan is supposed to be a “living document” which will be updated during the project implementation when significant changes occur.

The document (1) introduced the data management plan that will guide the project, (2) identified the data sets that will be collected or created, and (3) described how they will be stored and shared.

It also specified which data would be made freely available and which would be kept confidential within the consortium, if possible at this stage.

Finally, the repositories and resources for sharing and storing data are identified.

## 10 List of Abbreviations

APP	external application Authors, Projects, Publications
CC licences	Creative Commons licences
DMP	Data management plan
DOI	Digital Object Identifier
EC	European Commission
EDU	Educons
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
O1 to O5	Objectives
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
PC	Project Coordinator
SIT	Science, innovation, and technology
WP	Work Package
XML	Extensible Markup Language