MY Data Management PLAN (DMP)

DMP TITLE

DMP: RecyClear - Data- & AI-driven sorting for efficient rigid plastics recycling

ADMIN DETAILS

Project Name: RecyClear - Data- & AI-driven sorting for efficient rigid plastics recycling
Grant Title: C3/23/019
Principal Investigator / Researcher: Christian Steuwe
Project Data Contact: Christian.steuwe@kuleuven.be
Description: In Europe, 29.1 million tons (Mton) of plastic waste is collected every year. Of that collected fraction, only 4 Mton (13.7% of the total waste volume) is converted into new products by mechanical recycling. Recycling rates have been increasing by only a few percentage points per year over the last 2 decades. The reason for this slow increase is that recycling comes at a cost which makes the competition with virgin plastic prices difficult. Unfortunately, this pace of progress is far too slow to achieve the goals set in pledges such as the one by the Circular Plastics Alliance (i.e., to reach 10 Mton of recycled plastics in 2025; meaning that in just 3 years we would have to increase the recycled fraction by 6Mton!) and less voluntary commitments such as the European Recycling targets for packaging plastics of 55% by 2030. It is clear that radical innovations are needed. Current sorting installations only operate cost-efficiently at (very) high volumes. Therefore the goal of this C3 project is to improve the situation by enabling distributed sorting closer to the source, empowering new, more efficient value chains for mechanical recycling. Our proposal is to have consumers sort their waste using a self-scan system (similar to self-checkout in supermarkets or baggage scanners) when they visit the recycling center. Instead of throwing a plastic item in a container, a visitor would drop it on a transport belt or a similar installation. The recycling center then receives pre-sorted fractions of plastics which could be sold for winst rather than disposing the plastics at a cost.

1. GENERAL INFORMATION

Name applicant
Christian Steuwe

Project Number & Title
C3/23/019 - RecyClear - Data- & AI-driven sorting for efficient rigid plastics recycling
Affiliation
- KU Leuven

2. DATA DESCRIPTION

Will you generate/collect new data and/or make use of existing data?
  ● Generate new data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume, ideally per objective or WP of the project. You might consider using the table in the guidance.

The expected data volume generated within the C3 project is 10-20 Gb per month.

Objective 1: Development and validation of NIR, MIR, and VIS optical sensing modules for sorting.
Data related to this objective are mostly acquired for the characterization of plastics. The details of raw and exported data formats are listed below.

equipment	raw format	exported format
Developed NIR spectrometer	CSV	CSV
Varian MIR spectrometer	CSV	CSV
Bruker NIR spectrometer	Bruker Opus	CSV
Camera	Uncompressed 8-12 bit	Jpeg compressed

Objective 2: Demonstrator of an artificial intelligence architecture & prototype.
To achieve this objective, we will mainly use the data from objective 1. We will set up a database to automatically acquire data from spectral sensors and cameras. The database architecture has still to be defined. Storage will be arranged in a cloud-based solution.

Objective 3: Demonstrator of a mechanical sorting architecture & prototype.
To achieve this objective, we will use the database infrastructure from objective 2. During tests, we will further feed the database from objective 2 with data.
.

3. LEGAL & ETHICAL ISSUES

Will you use personal data? If so, shortly describe the kind of personal data you will use (add the reference to your file in your host institution's privacy register - not relevant yet)
- No

Privacy Registry Reference: /
Short description of the kind of personal data that will be used: /

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)
- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?
- Yes

If the proposed concept works, a meeting will be planned with the KU Leuven tech transfer office to investigate how to protect the resulting database and to investigate patentability. In case a patent application will be filed directly resulting from this project, the related data will not be freely shared at least until the patent filing.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?
- No

4. DOCUMENTATION & METADATA

What documentation will be provided to enable reuse of the data collected/generated in this project?

An overview file that contains references to the raw data files will be kept. Regular reports based on the data will be generated using Microsoft Word and on GIT. PowerPoint files will be used for presentation at regular internal meetings between the WP leaders and researchers involved in the project. In both the Word reports and Powerpoint presentations, the file names of the raw data files will be included.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

General experiments:

The details of each experiment will be kept in an electronic lab notebook. In this notebook, also the names of the raw and processed datafiles will be mentioned. Files will be named according to a pre-agreed convention. This working method obviates the need for a separate INFO.txt file in each directory yet ensures that the data can be understood by other team members and can be reused in the future.

For published papers, the subset of the raw and processed data discussed in that manuscript will be copied and organized according the paper structure. Likely, this is the data subset that will be most frequently revisited and shared afterwards.

Database:

The centralized database (Objective 2) will consist of data that is registered using a pre-discussed protocol in an automated manner. The conventions are to be stored on GIT.

5. DATA STORAGE & BACK UP DURING THE PROJECT

Where will the data be stored?
The data will be stored via a cloud storage solution that allows sharing with the WP leaders and researchers involved in the project.
The database will be stored on a cloud platform.

How is backup of the data provided?
The data on the cloud storage server are automatically backed up. Unlimited versioning is included in the selected plan so that accidental erasing or modifying does not pose a risk.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.
   ● Yes
The total amount of data generated during the project should not exceed several TB and is therefore compatible with the selected cloud storage solution.

What are the expected costs for data storage and backup during the project? How will these costs be covered?
The costs for saving the data to the cloud storage server (including regular backup) should not exceed a few hundred euros. These costs will be covered by the project.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?
The data generated during the project will be systematically transferred to the cloud storage server. Only the WP leaders and researchers involved in this project will have access to the shared folders where the data, reports and presentations will be stored.

6. DATA PRESERVATION AFTER THE PROJECT

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All the generated data will be stored on the cloud storage server for a period of 5 years after the end of the project.

Where will the data be archived (= stored for the longer term)?

All the generated data will be stored on the cloud storage server for a period of 5 years after the end of the project.

Beyond 5 years after the end of the project, one of the following options will be picked (1) continuation of storing the data on the cloud storage server or (2) transferring the data to the KU Leuven central servers for archiving.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

The annual cost for long-term storage of the data, either through a cloud storage service or the university's central servers, is estimated at a few hundred euro. Since the budget of the current project will no longer be available, creative solutions will have to be found.

7. DATA SHARING AND REUSE

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?
   ● Yes. Specify:
The data will remain accessible among the WP leaders and the researchers involved in the project. Access to the data can be granted to other persons, upon request and agreement among the leaders. Imminent commercial interests or a patent application can be a limiting factor in granting access to the data.

Which data will be made available after the end of the project?
In case no restrictions apply, the written reports and Powerpoint presentations summarizing the results obtained can be made available. The (raw) data used in publications can be made available on a repository, if e.g. requested by the Editors or Publisher of a scientific journal.

Where/how will the data be made available for reuse?
   ● Upon request by mail
The data that will not have been uploaded in a resository such as 'Scientific data' (https://www.nature.com/sdata/) to accompany a publication can be requested via email.

When will the data be made available?
   ● Upon publication of the research results
Description of the full scientific method and results will be made available with journal publications.

Who will be able to access the data and under what conditions?
Upon request, access to the samples and data can be granted, upon agreement of the project leaders. Commercial reuse will not be allowed.

What are the expected costs for data sharing? How will the costs be covered?
Because of the choice for a cloud storage solution for the data, no additional costs will be booked for data sharing.

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?
The project leaders and the researchers involved in this project will be jointly responsible.

Who will be responsible for data storage & back up during the project?
The project leaders and the researchers involved in this project will be jointly responsible. Because of the choice for a cloud storage solution, no additional action is needed for data backup.

Who will be responsible for ensuring data preservation and reuse ?
The project leaders and the researchers involved in this project will be jointly responsible.

Who bears the end responsibility for updating & implementing this DMP?
The project leaders bear the end responsibility of updating & implementing this DMP.