



MIXed plastics biodegradation and UPcycling using microbial communities
MIX-UP

Deliverable D8.3
Final Data Management Plan

Dissemination				
Lead	International partners involved	Level	Type	Delivery Month
RWTH	<input type="checkbox"/> NTU <input type="checkbox"/> BUCT <input type="checkbox"/> IPE <input type="checkbox"/> TU	<input checked="" type="checkbox"/> PU (Public) <input type="checkbox"/> CO (Confidential) <input type="checkbox"/> CI (Classified)	<input type="checkbox"/> R (Document, report) <input type="checkbox"/> DEM (Demonstrator, pilot, prototype) <input type="checkbox"/> DEC (Websites, patent filings, videos etc.) <input checked="" type="checkbox"/> ORDP (Open Research Data Pilot) <input type="checkbox"/> ETHICS <input type="checkbox"/> OTHER	M06

Grant Agreement:	870294
Project acronym:	MIX-UP
Project title:	MIXed plastics biodegradation and UPcycling using microbial communities
Start of the project:	01.01.2020 (48 months)
Work package:	WP8
Deliverable number:	D8.3
Title of deliverable:	Final Data Management Plan
Deliverable due date:	31/12/2021
Date of submission:	31/12/2021
First Author:	Lars M. Blank (RWTH)
Contributors:	Hendrik Ballerstedt (RWTH)
Version:	1.2

DISCLAIMER

This document contains information which is the proprietary to the MIX-UP Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to any third party, in whole or in parts, except with prior written consent of the MIX-UP Coordinator.

The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The document reflects only the authors' view. The Agency is not responsible for any use that may be made of the information it contains. The user uses the information at its sole risk and liability.

VERSION CONTROL

Version	Date	Author (Name, Institution)	Comments
1.0	26/08/21	Lars M. Blank (RWTH)	First draft
1.1	29/12/21	Lars M. Blank (RWTH)	Minor changes 2.2
1.2	31/12/21	Lars M. Blank (RWTH) Hendrik Ballerstedt (RWTH)	Minor changes 1 and 2.3

Table of Contents

1 Introduction 4

2 FAIR data 4

 2.1 Discoverable 4

 2.2 Accessible 5

 2.3 Assessable and intelligible 5

 2.4 Usable beyond the original purpose for which it was collected 6

 2.5 Interoperable to specific quality standards 6

3 Allocation of resources 6

4 Data security 6

5 Ethical aspects 7

1 Introduction

Efficient collaboration relies on efficient and open sharing of data and materials within the consortium, and public sharing of data maximises the impact of the project beyond its four-year runtime. All partners are required to share their data and materials openly. The Consortium Agreement provides a legal framework in this respect. It outlines both the obligations and limitations of sharing data and materials in relation to implementing the project as a whole. In short, the Consortium Agreement determines *if* data or materials should or must be shared, the Data Management Plan (DMP) is a policy on *how* data is to be shared.

A varied spectrum of data results will be achieved from MIX-UP. Main types of data include enzymatic data (e.g., protein structures, kinetic data), physiological data (e.g., growth curves, biomass yields), synthetic biology data (e.g., plasmid maps, genome sequences, metabolic models), and process data (e.g., online fermenter data, production parameters). With their different data structure, these areas of research were identified in the previous EU H2020 project P4SB and were mentioned in the MIX-UP proposal. The handling of each type of data is specified in the accompanying guidelines taken from P4SB if proven helpful and revised if potentially better solutions could be found. The MIX-UP data management plan revolves around publishing FAIR data in public databases and peer-reviewed journals as much as possible. In principle, *all* useful data must be made available in this way, unless valid reasons can be provided against it in accordance with the Consortium Agreement (e.g., IP confidentiality). This ensures that:

- Data is made public through open-access channels.
- Methodologies and standards for data handling are maintained through database standards and peer-review.
- Data is curated and preserved long after the project end.

With the ever-increasing number of scientific publications, the MIX-UP partners show their commitment to the above-outlined logic of data handling in general.

2 FAIR data

2.1 Discoverable

Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable using a standard identification mechanism (e.g. Digital Object Identifier)?

It is the intention to publish all relevant data and its interpretation in peer-reviewed publications. Manuscripts and data will be published in journals that make them identifiable with a DOI and are indexed in generally accepted scientific search engines such as PubMed and web of science. Large datasets such as genomes and transcriptomes will be uploaded to well-curated and publically accessible databases such as NCBI Genbank and the Gene Expression Omnibus. They will be identifiable through standardised systems governed by these databases (e.g., NCBI accession.version numbers). The data presentation in peer-reviewed journals is at the moment changing. With the reproduction crisis in everyone's mind, scrutiny during peer review hopefully increases, while new data presentation formats are evaluated. From our proposal, data will generally be made accessible and a publication, which will clarify the data and its significance, making it intelligible. This also ensures that the data will undergo

thorough, external assessment through the peer-review system. We encourage MIX-UP partners to use data journals that support FAIR data management.

2.2 Accessible

Are the data and associated software produced and/or used in the project accessible, and in what modalities, scope, licenses?

The general aim of publication is golden open access, ensuring accessibility in both the long and short term. Whenever possible, raw data associated with publications must be made accessible through supplemental files within the same journal. In case this is not possible, for instance, when a journal doesn't allow supplemental data, the raw data should be 1) added into FAIRDOM before publication to include the generated DOI into the paper, or 2) submitted to a data-dedicated journal such as Data in Brief, or 3) the data should be posted on another online repository such as Researchgate or the company/university website. Data-dedicated journals like Data in Brief (Elsevier) or Scientific Data (Nature), although purely presenting data, without interpretation, are peer-reviewed, open access, and have therefore a very high findability. We will use these possibilities for data of general interest, which goes beyond data presented in a scientific article, for example, data usually partially hidden in the appendix (e.g., transcriptomic data (already published in a database), with the contextual data of, e.g. the fermentation). A further alternative are journals that support directly the presentation of data used in graphs and figures (e.g., at Elsevier, some journals offer "Interactive Plots", connecting graphs and figures with the data from which these presentations were generated). The same goes for large omics datasets, which are uploaded to well established and publically accessible databases that have proven their longevity.

sss

2.3 Assessable and intelligible

Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in scientific scrutiny and peer review contexts?

Data will generally be made accessible along with a publication, which will clarify the data and its significance, making it intelligible. This system also ensures that the data undergo thorough, external assessment through the peer-review system.

While we proposed to sort all data by defined and standardised metadata, we now organise the data by tasks. This allows flexible data exchange. The task will be mentioned in the file title to confirm correct sorting. Metadata will be included at the beginning of each document to make the data easily accessible. We identified early four distinct groups of data within MIX-UP that each have their own terminology, which will be mentioned at the beginning of the file:

1. Plastic hydrolysis/enzymology
2. Microbial physiology
3. Synthetic biology
4. PHA production process

2.4 Usable beyond the original purpose for which it was collected

Are the data and associated software produced and/or used in the project useable by third parties even long after collecting the data?

The use of well-established journals and public databases ensures the long-term accessibility of the data. Proper annotation by metadata and associated publications (as explained above) will enable third parties to use the data well after project end.

2.5 Interoperable to specific quality standards

Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc.?

Yes, the data management plan ensures that the exchange is possible. The common language is English. Widely accepted data standards are defined for the four individual research fields. Examples are genomic data storage in Genbank according to NCBI rules and transcriptome data storage in GEO according to NCBI rules.

3 Allocation of resources

All EU partners have been allocated resources to cover open-access fees for publications. In general, all partners are responsible for data management, and all partners have been allocated resources to enable proper data management.

4 Data security

Pre-publication data will be stored in-house by the party that generated the data. Each partner is required to maintain proper in-house data handling and storage. This includes indexing and file naming to enable efficient recovery of data and data storage and archiving, keeping at least one backup in a secure location. FAIRDOM is the platform of choice for sharing data with the consortium as a whole. Files that are shared on FAIRDOM should be named according to the following convention: MIX-UP_[task number]_[partner short name]_[YY-MM-DD]_[short description].[extension]. Example: “*MIX-UP_D8.3_RWTH_31-12-21_final data management plan.pdf*”.

The establishment of a large systematic internal data repository was discussed within the consortium and rejected. The heterogeneity of data generated in the project and the generally good communication between partners ask for an as-needed exchange of data through more traditional means (i.e. e-mail, Gigamove – a file sharing service from German universities), as long as the privacy policy of the transferring medium does not pose a data security threat.

Post-publication, the data is stored in a publically accessible publication and/or database. In addition, partners are required to archive a secure in-house copy as well, to be made available upon request from both internal and external parties.

5 Ethical aspects

We foresee no ethical aspects related to the data generated within MIX-UP. Workplace including laboratory security measures for the co-workers generating the data is the obligation of every MIX-UP partner. Where applicable, gene technology law has also been followed.