



# Data Management Plan

For Project



**Funded by  
the European Union**

**Under  
Horizon Europe Programme  
(HORIZON-TMA-MSCA-PF-EF-2021;  
Grant Agreement ID: 101062601)**



**Universidad de Jaen, Campus Las  
Lagunillas, 23071 Jaen, Spain**



**DOCUMENT DESCRIPTION**

Name	Data Management Plan
Version	1.0
Format	Portable Document Format (.pdf)
Deliverable Type	DMP-Data Management Plan
Deliverable No.	1.1
Work Package No.	WP1
Dissemination level	PU - Public
Due date of deliverable	4/3/2023
Actual submission date	24/2/2023
Author	Dr. Adnan Asad Karim (Marie Skłodowska Curie Postdoctoral Researcher, University of Jaen, Spain, E-mail: aasad@ujaen.es)
Approved By	Prof. Maria Lourdes Martinez Cartas (OliPFUEL Project Co-ordinator & Supervisor, University of Jaen, Spain, E-mail: lcartas@ujaen.es)

**DISCLAIMER**

Project funded by the European Union under Horizon Europe programme (HORIZON-TMA-MSCA-PF-EF-2021; Grant Agreement ID: 101062601). Views and opinions expressed in this document are however those of the authors only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.



## TABLE OF CONTENTS

List of acronyms and abbreviations.....	5
Project information.....	6
Summary of the data management plan.....	7
<b>1. DATA PRODUCTION.....</b>	<b>9</b>
1.1. Goal of data generation.....	9
1.2. Data origin.....	9
1.3. Data types.....	9
1.4. Data size.....	9
1.5. Communities interested in the data .....	9
<b>2. DATA FINDABILITY.....</b>	<b>10</b>
2.1. Data storage.....	10
2.2. Data formats.....	10
2.3. Metadata use to describe data.....	10
2.4. Standardised vocabularies use for data .....	10
2.5 Open free access of data.....	11
2.6. Data harvesting.....	11
2.7. Naming convention of data.....	11
2.8. Clear version numbers for data.....	11
2.9. Persistent identifiers for data.....	11
2.10. Searchable metadata for data.....	12
2.11. Metadata describing the quality of the data.....	12
<b>3. DATA ACCESSIBILITY AND INTEROPERABILITY.....</b>	<b>13</b>
3.2. Open free accessibility of data.....	13
3.3. Data availability.....	13



3.4. Data security.....	13
3.5. Auxiliary data availability that may be of interest to researchers.....	14
<b>4. DATA REUSABILITY.....</b>	<b>15</b>
4.1. Internationally recognised licence for data reuse.....	15
4.2. Data availability time for reuse.....	15
4.3. Documented procedures for quality assurance of the data.....	15
4.4. Data integrity for reuse.....	15
4.5. Support and allocation of resources for long-term data reuse.....	15



## LIST OF ACRONYMS AND ABBREVIATIONS

ANZSRC = Australian and New Zealand Standard Research Classification

CC-BY-4.0 = Creative Commons Attribution 4.0 International

CERN = European Organization for Nuclear Research, Geneva

CORDIS = Community Research and Development Information Service

DOI = Digital object identifier

EOS = CERN's mass storage system for data

ESTAT = European Statistical System

EU = European Union

EuroVoc = Multilingual and multidisciplinary thesaurus covering the activities of the EU

FAIR = Findability, accessibility, interoperability, and reusability

HTTPS = Hypertext Transfer Protocol Secure

ISO 17369 = International Standard for Statistical Data and Metadata Exchange

MSCA-PF-EF = Marie Skłodowska-Curie Actions European Postdoctoral Fellowship

OliPFUEL = Advanced biofuels production from waste olive pomace of olive oil industries

OpenAIRE = Open Access Infrastructure for Research in Europe

PostgreSQL = Open-source relational database management system

PU = Public

REA = European Research Executive Agency

ROR = Research Organization Registry

SDMX = Statistical Data and Metadata Exchange Initiative

UIC = Unique identification code

UNESCO = United Nations Educational, Scientific and Cultural Organization

W3C = World Wide Web Consortium



## PROJECT INFORMATION

<b>Project number</b>	101062601
<b>Project name</b>	Advanced Biofuels Production from Waste Olive Pomace of Olive Oil Industries
<b>Project acronym</b>	OliPFUEL
<b>Call</b>	HORIZON-MSCA-2021-PF-01
<b>Topic</b>	HORIZON-MSCA-2021-PF-01-01
<b>Type of action</b>	HORIZON-TMA-MSCA-PF-EF
<b>Service</b>	REA/A/02
<b>Project starting date</b>	5 September 2022
<b>Project duration</b>	24 months
<b>Net EU contribution</b>	€ 165 312,96

## PARTICIPANTS

Number	Role	Short Name	Legal Name	Country	PIC
1	COO	UJA	UNIVERSIDAD DE JAEN	ES	999849423
2.	AP	BIOLIZA	RECURSOS ESTRATEGICOS DE BIOMASA S.L.	ES	905744679



## SUMMARY OF THE DATA MANAGEMENT PLAN

The Data Management Plan (Deliverable 1.1) is a part of the project “Advanced Biofuels Production from waste olive pomace of olive oil industries” that has received funding from the European Union’s (EU) Horizon Europe programme under Marie Skłodowska Curie Postdoctoral fellowship with grant agreement No. 101062601. This document provides a detailed description of the activities and measures to be undertaken for the OliPFUEL project data management on the basis of FAIR principles. A brief summary of the data management plan is tabulated below.

<b>Secure Data Repository</b>	Zenodo, an open-source repository will be used to store project data (expected between 10 and 100 GB size). This repository is developed on the basis of FAIR (findability, accessibility, interoperability, and reusability) principles under the European OpenAIRE program (funded by the European Union to promote open science and open access to research files to everyone). Data will also be stored in the ResearchGate, and Research & Knowledge transfer Portal of the University of Jaen (UJA).
<b>Standardised Data formats</b>	Rich Text Format, Microsoft Word Document, Microsoft Excel Document, Microsoft Powerpoint Document, Portable Document Format, JPEG File Interchange Format, Windows Meta File, MPEG-4 Media File, StatGraphics Data File, Origin Project Format
<b>Standardised Vocabularies</b>	UNESCO Thesaurus, and EuroVoc (Publications Office of the European Union) Thesaurus will be followed.
<b>Naming Convention</b>	The guidelines provided by Standardising Reference Metadata Reporting in SDMX (Statistical Data and Metadata Exchange Initiative) will be followed.
<b>Metadata Schema</b>	The Datacite metadata schema in the Zenodo repository will be used to make the project data files easily findable, accessible, and harvestable. Zenodo will assign a unique persistent digital object identifier (DOI) to each data file to make it easily citable and shareable. The metadata harvesting can be done by using any parameters such as subject discipline, ANZSRC Fields of Research classification code-090499 (Chemical Engineering), researchers' name, ORCID Id, the affiliated institute with a unique identifier (Research Organization Registry-ROR), funding information, project title, data-related description (Methods, Technical Info), etc.



<b>Data Integrity</b>	Data files will contain a unique identification code (UIC), generated by using a checksum calculator, which has been widely used to determine the integrity of shared data files. This code will be given in the metadata of the data files. Any tampering with data files will change the code. This will alert the users that the data file integrity is compromised.
<b>Data Quality Assurance</b>	Data Quality Guidelines, 2021 published by the European Commission will be followed. These guidelines involve the use of aforementioned standardized vocabularies, file formats, metadata, proper documentation, and removing duplicates and repetitions. Also, providing information (open access) on data file versions, sources, methods, procedures, errors, validation, and reproducibility to the users.
<b>Data Reuse</b>	Project data files will be openly available to users in the Zenodo repository prior to research publications under Creative Commons Attribution 4.0 International (CC-BY-4.0). In addition, the auxiliary data (e.g., raw signals from scientific instruments like calibration curves, spectra, and diffraction patterns) will also be provided for easy data reuse by other interested researchers and other users.
<b>Data Management</b>	Research Management Services of the UJA will provide data management support (e.g., storage with backup).
<b>Ethical and Legal Support</b>	The OliPFUEL project doesn't involve any ethical issues. Nevertheless, the UJA has Ethics Commission, which can provide clearance and consultation. The Research Results Transfer Office (OTRI) of the UJA will provide support to protect the intellectual copyright of the project data, and prevent its unethical commercial exploitation.



## **1. DATA PRODUCTION**

### ***1.1. Goal of data generation***

The conventional process for the production of advanced biofuels from lignocellulosic biomass waste (e.g., olive pomace) involves the use of toxic chemicals, expensive enzymes, and *Saccharomyces cerevisiae* yeast, which has certain technological limitations and also poses environmental pollution risks. The major goal of OliPFUEL project is to provide new research data to establish proof of concept for a more environmentally friendly process for the production of advanced biofuels from olive pomace waste. The process consists of thermal and non-conventional yeast fermentation methods. Research activities will generate new primary data related to various experiments, practical observations, and statistical analysis, along with secondary data taken from prior art and industries. The research data will be used for filing the patent, followed by publication, and commercial exploitation.

### ***1.2. Data origin***

In the OliPFUEL project, experimental data will be generated first. On basis of experimental data, observational (research findings) data will be prepared. Derived data will be obtained after analysing the experimental data through different computational and mathematical modelling techniques. For example, the thermal processing of biomass will give experimental data like product yield, observational data like the influence of temperature on product yield, and derived data like a computational prediction of best process parameters to achieve optimum product yield.

### ***1.3. Data types***

- Quantitative data: This mainly includes the quantitative details of the experiments conducted under different process conditions, and characteristics of feedstock, solid and liquid products.
- Qualitative data: This includes the qualitative information acquired from the review of earlier published research papers, and OliPFUEL research findings inferred from the experiments about the influence of process parameters and the identification of optimum process conditions.

### ***1.4 Data size***

The total size of the data produced by the OliPFUEL project could be between 10 and 100GB.

### ***1.5 Communities interested in the data***

- Science communities: Green Chemistry, Materials Science, Environmental Science, Chemical engineering, Environmental Engineering, Biotechnology, Bioengineering, Computing in Natural Sciences and Engineering
- Industries: Olive oil industry, Biofuel industry, Biomass processing industry



## 2. DATA FINDABILITY

### 2.1. Data storage

Zenodo is a widely used open-source repository, developed on basis of FAIR principles under the European OpenAIRE program (funded by the European Union to promote open science and open access to research files to everyone). Its operation and maintenance are done by the CERN (European Organization for Nuclear Research, Geneva).<sup>1</sup> All the relevant research files of the OliPFUEL project such as metadata, experimental design & methodology, data sets, pre-prints, and publications, etc., will be uploaded in the Zenodo. The Zenodo will assign a unique persistent digital object identifier (DOI) to each uploaded research file of OliPFUEL, which will make these files easily citable and shareable.

### 2.2. Data formats

The data formats to be used in the OliPFUEL project are as follows: Rich Text Format (.rtf), Microsoft Word Document (.docx), Microsoft Excel Document (.xlsx), Microsoft Powerpoint Document (.pptx), Portable Document Format (.pdf), JPEG File Interchange Format (.jpg), Windows Metafile (.wmf), Windows Media Video (.wmv), MPEG-4 Media File (.mp4), StatGraphics Data File Format (SF, SF3, or SF6), and Origin Project Format (.opj).

### 2.3. Metadata use to describe data

The DataCite Metadata Schema 4.4 is widely used by researchers. It contains all standard metadata properties essential for consistent identification and accurate data retrieval by other researchers. Therefore, for OliPFUEL the Datacite metadata schema will be used, which will contain all relevant information i.e., subject discipline, ANZSRC Fields of Research classification code-090499 (Chemical Engineering), researchers' full name with ORCID ID, the affiliated institute with a unique identifier (Research Organization Registry-ROR), funding information, project work title, data-related description (Abstract, Methods, Technical Info), etc.

Common sample nomenclature/coding will be also used for the metadata, data, and publications. Furthermore, the registered identifier and depository link for data will be cited in the research publications.

### 2.4. Standardised vocabularies use for data

Vocabularies define concepts and relationships that describe an area of concern (e.g. a subject or problem). In the semantic web, they enable the integration of data allowing inferences to be made. In

---

<sup>1</sup> <https://zenodo.org/>



the OliPFUEL project, UNESCO thesaurus,<sup>2</sup> and EuroVoc thesaurus<sup>3</sup> will be followed for preparing data. The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences, social and human sciences, communication, and information. EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It is managed by the Publications Office of the European Union, which moved forward to ontology-based thesaurus management and semantic web technologies conformant to W3C recommendations as well as the latest trends in thesaurus standards.

### ***2.5. Open free access of data***

The metadata of OliPFUEL project files will be published in Zenodo, an open-access repository, and thus the research files will be freely available to other researchers. Furthermore, the persistent DOI (Digital Object Identifier) link for metadata will also be provided in the research publications of the OliPFUEL project, which will maximize its visibility and traceability to other researchers.

### ***2.6. Data harvesting***

The data of OliPFUEL project will be stored in the Zenodo repository and will be harvestable with OAI-PMH (DataCite v3) protocol.<sup>4</sup> The metadata harvesting can be done by using any parameters such as community Identifier name, abstract, authors, DOI, issue date, title, etc.

### ***2.7. Naming convention of data***

The guidelines provided by Standardising Reference Metadata Reporting in SDMX (Statistical Data and Metadata Exchange Initiative,<sup>5</sup> will be followed for metadata naming conventions of OliPFUEL documents. The SDMX has also been published as an International Standard (ISO 17369) and used by the European Commission projects for European Statistical System (ESTAT).

### ***2.8. Clear version numbers for data***

The OliPFUEL data files will have clear version numbers, which inform users of older data from any changes, thus making data more reliable.

### ***2.9. Persistent identifiers for data***

Persistent identifiers are independent of the physical location of the data. Using persistent identifiers means having the freedom to move the data to new storage without impacting your users. All OliPFUEL data files deposited to the Zenodo repository will have DOIs (via high-quality DataCite Metadata

---

<sup>2</sup> <http://vocabularies.unesco.org/thesaurus>

<sup>3</sup> <http://publications.europa.eu/resource/dataset/eurovoc>

<sup>4</sup> <http://zenodio.lsst.io/en/latest/harvest.html>

<sup>5</sup> [https://sdmx.org/?page\\_id=4345](https://sdmx.org/?page_id=4345)



Schema). The data will be issued with DOI identifiers once the data has reached an approved level of maturity for consumption by interested parties.

### ***2.10. Searchable metadata for data***

The Zenodo repository to be used for OliPFUEL data is recommended by FAIRsharing.org. In Zenodo, by using the advanced search options users can search the metadata and get the refined search results in different ways such as simple search (one or multiple terms), phrase search, field search, combined simple, phrase or field search, time range search, Sorting based on publication date, data resources type, identifiers schemes, data access rights, and contributors' types.

### ***2.11. Metadata describing the quality of the data***

The provenance information (including location, creation and changes history, etc.) about the data will be included in the metadata for the OliPFUEL project. This will help users to use data correctly.



### **3. DATA ACCESSIBILITY AND INTEROPERABILITY**

#### ***3.1. Ethical or legal restrictions on sharing data***

In the OliPFUEL project, some data deposition and sharing will be withheld till the patent filing and obtaining clearance from the Intellectual Property Rights Office, University of Jaen (UJA), Spain.

#### ***3.2. Open accessibility of data***

Data will be openly accessible with no restrictions (for example, allowing anonymous download and with a permissive access license), which will greatly enhance data accessibility.

#### ***3.3. Data availability***

The data will be available with DOI identifiers, which will be accessed by different users through different channels such as Zenodo, ResearchGate, Research article Publishers, the EU CORDIS web platform, and the Research Portal of the UJA.

#### ***3.4. Data security***

Zenodo open access repository (to be used for OliPFUEL data files) is embedded in the IT Department, Collaboration Devices and Applications Group, Digital Repositories Section of CERN. It is powered by CERN Data Centre and the Invenio digital library framework and is fully run on open-source products. Zenodo servers are managed via OpenStack and Puppet configuration management system which ensures that servers always have the latest security patches applied. Servers are monitored via CERN's monitoring infrastructure based on Flume, Elasticsearch, Kibana, and Hadoop. Application errors are logged and aggregated in a local Sentry instance.

All files uploaded to Zenodo are stored in CERN's EOS service in an 18 petabytes disk cluster. Each file copy has two replicas located on different disk servers. Metadata and persistent identifiers in Zenodo are stored in a PostgreSQL instance operated on CERN's Database on Demand infrastructure with a 12-hourly backup cycle with one backup sent to tape storage once a week. Metadata is in addition indexed in an Elasticsearch cluster for fast and powerful searching. All changes to metadata records on Zenodo are versioned, and happening inside database transactions. The CERN Security Team runs both host and network-based intrusion detection systems and monitors the traffic flow, pattern, and contents into and out of CERN networks in order to detect attacks. Access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages. Zenodo stores user passwords with strong cryptographic password hashing algorithms.<sup>6</sup>

---

<sup>6</sup> <https://zenodo.org/>



### ***3.5. Auxiliary data availability that may be of interest to researchers***

Auxiliary data (e.g., settings of analytical instruments) will provide information on the quality of the data. After the publication of the project's research findings, the project will make available auxiliary data that will be useful to researchers looking to use the data.



## **4. DATA REUSABILITY**

### ***4.1 Internationally recognised licence for data use***

The OliPFUEL data will be openly available to users for reuse under Creative Commons Attribution 4.0 International (CC-BY-4.0), as recommended by the European Commission. This license will make it easier for other researchers worldwide to use the data without any restrictions.

### ***4.2 Data availability time for reuse***

The data will be made available in the Zenodo repository prior to submission of research article for publication in the open access journals. The DOI link of the data will be provided in the publications. The data will be made available and supported for 10 years or more.

### ***4.3 Documented procedures for quality assurance of the data***

Data Quality Guidelines, 2021 published by the European Commission will be followed for the OliPFUEL project. These guidelines involve the use of standardized vocabularies, file formats, metadata, proper documentation, and removing duplicates and repetitions. Also, providing information (open access) on data file versions, sources, methods, procedures, errors, validation, and reproducibility to the users.

### ***4.4 Data integrity for reuse***

In the OliPFUEL project, metadata of research data files will also contain a unique identification code (UIC). This code for each file will be generated using a checksum calculator, which has been widely used to determine the integrity of shared data files.<sup>7</sup> Any changes/tampering with the original data file will change the checksum-generated UIC and users will know that the data file is compromised and not reliable.

### ***4.5. Support and allocation of resources for long-term data security and reuse***

The OliPFUEL project data will be stored in the Zenodo repository (OpenAIRE), hosted by CERN large-scale digital repositories. Documentation and the contact details of the "Project Investigator " will be provided in the metadata to the users to increase its usefulness. The project data will also be stored in an institutional archive of the UJA.

---

<sup>7</sup> <https://apps.microsoft.com/store/detail/simple-checksumcalculator/9NBLGGH5L6T9?hl=en-us&gl=us>