- IDEALFUEL -

Lignin as a feedstock for renewable marine fuels

GRANT AGREEMENT No. 883753

HORIZON 2020 PROGRAMME - TOPIC LC-SC3-RES-23-2019
"Development of next generation biofuel and alternative renewable fuel technologies for aviation and shipping"

Deliverable Report

D7.1– Data Management Plan

Deliverable No.	IDEALFUEL D7.1	
Related WP	WP7	
Deliverable Title	Data Management Plan	
Deliverable Date	30-10-2020	
Deliverable Type	ORDP: Open Research Data Pilot	
Dissemination level	Public (PU)	
Written By	Roy Hermanns (TUE)	22-10-2020
Checked by	Eva Bogelund (UNR)	28-10-2020
Reviewed by	Eva Bogelund (UNR)	28-10-2020
Approved by	Project Coordinator	28-10-2020
Status	Version 1.0	28-10-2020

Disclaimer/ Acknowledgment

Publishable summary

The EU H2020 project IDEALFUEL aims to develop an efficient and low-cost chemical pathway to convert lignocellulosic biomass into a Biogenic Heavy Fuel Oil (Bio-HFO) with ultra-low sulphur levels that can be used as drop-in fuel in the existing maritime fleet. The document presents the first version of the project "Data Management Plan", describing the implementation measures envisioned to efficiently manage the research data gathered in the project. This Data Management Plan is part of Work Package 7 – "Communication, Dissemination and Exploitation of Results". It will be reviewed and updated regularly. This deliverable outlines how the collected or generated research data will be handled during and after the IDEALFUEL project is finished. It describes which standards and methodology for data and generation of research data will be followed, and whether and how data can be shared. There are no deviations from the description of this deliverable as given in Annex 1 of the Grant Agreement.

Contents

Tables

Abbreviations

Symbol / short name	
Bio-HFO	Biogenic Heavy Fuel Oil
DMP	Data Management Plan
FAIR	Findable, accessible, interoperable and reusable

1   Introduction

1.1 Scope

The IDEALFUEL Data Management Plan (DMP) constitutes one of the outputs of the work package dissemination, communication and exploitation, dedicated to raising awareness and promoting the project and its related results, achievements. The present deliverable is prepared at an early project stage (Month 6), in order to commence on a strategy on data management from the project onset. It is also envisaged that the Data Management Plan will be implemented during the entire project lifetime and updated on a regular basis.

The main focus of the IDEALFUEL data management framework is to ensure that the project's generated and gathered data can be preserved, exploited and shared for verification or reuse in a consistent manner. The main purpose of the Data Management Plan (DMP) is to describe *Research Data* with the metadata attached to make them *discoverable*, *accessible*, *assessable*, *usable beyond the original purpose* and *exchangeable* between researchers. The definition of Research data is defined in the "Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020" (2015) as:

> "*Research data* refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form."

According to the EC provided documentation[1] for data management in H2020 aspects like research data access, sharing and security should also be addressed in the DMP. This document has been produced following these guidelines and aims to provide a policy for the project partners to follow.

1.2 Objectives

The generated and gathered research data need to be preserved in-line with the EC requirements. They play a crucial role in exploitation, verification of the research results and should be effectively managed. This Data Management plan (DMP) aims at providing a timely insight into facilities and expertise necessary for data management both during and after the project is finished, to be used by all IDEALFUEL.

The most important reasons for setting up this DMP are:

- Embedding the IDEALFUEL project in the EU policy on data management. The rationale is that the Horizon 2020 grant consists of public money and therefore the data should be accessible to other researchers;.
- Enabling verification of the research results of the IDEALFUEL project;
- Stimulating the reuse of IDEALFUEL data by other researchers;
- Enabling the sustainable and secure storage of IDEALFUEL data in repositories;

This first version of the Data Management plan is submitted to the EU in October 2020. It is important to note however that the document will evolve and further develop during the project's life cycle. It can be identified by a version number and a date. Updated versions will be uploaded by the coordinator TUE, which is the primary responsible for data management.

---

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2 Findable, Accessible, Interoperable and Reusable (FAIR) Data

This document takes into account the latest "Guidelines on FAIR Data Management in Horizon 2020". The IDEALFUEL project partners should make their research data **findable, accessible, interoperable and reusable** (**FAIR**) and ensure that is soundly managed. Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse[1].

2.1 Data Management Plan

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on[2]:

- the handling of research data during and after the end of the project;
- what data will be collected, processed and/or generated;
- which methodology and standards will be applied;
- whether data will be shared/made open access;
- how data will be curated and preserved.

3 IDEALFUEL Implementation of FAIR Data

3.1 Data Summary

It is a well-known phenomenon that the amount of data is increasing while the use and re-use of data to derive new scientific findings is more or less stable. This does not imply, that the data currently unused are useless - they can be of great value in the future. The prerequisite for meaningful use, re-use or recombination of data is that they are well documented according to accepted and trusted standards. Those standards form a key pillar of science because they enable the recognition of suitable data. To ensure this, agreements on standards, quality level and sharing practices have to be defined. Strategies have to be fixed to preserve and store the data over a defined period of time in order to ensure their availability and re-usability after the end of the IDEALFUEL project. Data considered for open access would include items as fuel properties, energy flows and balances, modelling calculations etc. For example, the consortium expects that the following data will be obtained and made available:

- Physico-chemical characterization of the Bio-HFO (WP2/3);
- Data underpinning the efficiency of the conversion steps (WP2/3);
- Emission data and actual measurements obtained during combustion of Bio-HFO (WP5);
- Data on storage and material interaction of the Bio-HFO (WP4).

The data will be documented in 4 types of datasets:
1. **Core datasets** – datasets related to the main project activities.
2. **Produced datasets** – datasets resulting from IDEALFUEL applications, e.g. sensor data.
3. **Project related datasets** – datasets resulting from the documentation of the progress of the IDEALFUEL project. They are a collection of deliverables, dissemination material, training material and scientific publications.
4. **Software related datasets** – datasets resulting from the development of the combustion reaction mechanisms. These can be used for various purposes in the combustion area including research tasks or the development of new appliances.

Generally, the datasets will be stored in file formats which have a high chance of remaining usable in the far future (see Annex 1). Especially the datasets which will be available for open access will be stored in these selected file formats. In principle the 4TU.ResearchData[3] platform is selected to insure open access of the datasets, persistent

---

[2] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[3] https://data.4tu.nl/info/en/

identifiers, data discovery and preservation of data for a long term. The open access data is useful for different stakeholder groups from the scientific community, industry as well as socio-economic actors. For example:

- **Industry and potential end-users of the Bio-HFO.** To implement Bio-HFO in the marine sector, the potential end-users need to be aware of their options. The end users will have certain demands, such as cost and comfort levels, which the industry needs to accommodate. This will be addressed by the datasets generated in WP6 and WP7.
- **Social and Environmental impacts of the IDEALFUEL value chain to the Regulatory Framework.** To allow commercial use of Bio-HFO in the marine sector, both the fuel as well as the heating systems need to comply with numerous regulations. In WP6 the regulatory framework on different levels will be documented.

3.2 FAIR Data

3.2.1 Making data findable, including provisions for metadata

In order to support the discoverability of data the 4TU.ResearchData platform has been selected. This platform support multiple using unique identifiers (doi, arxiv, isbn, issn, etc) which are persistent for a long time. The coordinator TUE uses this platform already for quite some time in all of its European Projects. In order to use this platform in an optimal manner the following best practices need to be taken into account:

- the discoverability of data (metadata provision);
- the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- the naming conventions used;
- the approach towards search keyword;
- the approach for clear versioning;
- specification of standards for metadata creation (if any). If there are no relevant standards available a documentation of what type of metadata will be created and how it is being created will be performed.

3.2.2 Making data openly accessible

The consortium will store the research data in a format which is suited for long-time preservation and accessibility. To prevent file format obsolescence, some precautions have been taken. One such measure is to select file formats which have a high chance of remaining usable in the far future (see Appendix A).

Furthermore, in a future update of this deliverable the following issues will be addressed:

- Specification of the data which will be made openly available. If some data is kept closed, a rationale for doing so will be given;
- Specification where the data and associated metadata, documentation and code are deposited;
- Specification how access will be provided in case there are any restrictions.

3.2.3 Making data interoperable

In order to support the interoperability of the IDEALFUEL project data a list of standard and metadata vocabularies need to be defined. Additionally it will be checked whether the data types present in our data set allow inter-disciplinary interoperability. If necessary mapping to more commonly ontologies will become available. The present version of the Data Management plan does not include the actual metadata about the data being produced in the IDEALFUEL project. Access to this project's related metadata will be provided in an updated version of the DMP.

- when data will become available for re-use. If applicable, it is mentioned whether a data embargo is necessary;
- the data produced and/or used in the project which is useable by third parties, in particular after the end of the project is listed. If the re-use of some data is restricted, it is explained why this is necessary;
- data quality assurance processes;
- the length of time for which the data will remain re-usable.

3.3 Allocation of resources

Lead for this data management task will be with TUE, though all partners are involved in the compliance of the DMP. The partners deliver datasets and metadata produced or collected in IDEALFUEL according to the rules described in the Annex 1. The project coordinator will lead the implementation of the DMP and track the compliance of the rules as documented in this DMP. The IDEALFUEL project partners have covered the costs for data FAIR in their budget estimations.

3.4 Data security

In this project various types of experimental and numerical data will be generated. The raw data will be stored by each partner according to their own standard procedures minimum for ten years after ending of the project. For example, results of smaller simulation runs and experiments will be stored on network drives and local drives, that are backed-up on external hard drives. Furthermore, smaller data sets and postprocessed data will be stored on cloud services, e.g. SURFdrive that is convenient for sharing data in a secure manner. Processed data will be made available to the community in three ways: a) available in the form of project reports and open access publications b) as supplementary material to publications, and c) through the managed repository environment 4TU.ResearchData. The processed data will become available in the form of project reports and open access publications. This data will be further exploited in webinars, articles in professional journals, and by conference presentations. The 4TU.ResearchData platform[4] has been selected for secure long-term storage and access of these datasets. The research data used for communication, dissemination and exploitation will be stored also on internal communication platform METT (http://www.mett.nl). This internal platform is only accessible for the project partners. Access to research data which is not marked as confidential will be granted via a repository.

3.4.1 Rights to access and re-use of research data

Open access to research data refers to the right to access and re-use digital research data under the terms and conditions set out in the Grant Agreement. Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated free of charge for the user.

Building on the proposed Consortium Agreement of the IDEALFUEL partnership the present data management plan is setup. The Consortium Agreement described general rules how data will be shared and/or made open, and how it will be curated, preserved and the proper licenses to publish, e.g. Creative Commons license. In an updated version of this DMP the right to access and re-use of research data will be documented in detail.

3.5 Ethical aspects and Legal Compliance

The legal compliance related to copyright, intellectual property rights and exploitation has been agreed on in the Consortium Agreement, which is also applicable to access to research data. It is unlikely that the IDEALFUEL project will produce research which is sensitive to personal and ethical concerns.

4 Discussion and Conclusions

This Data Management Plan (DMP) is focussed on the support of use and re-use of research data to validate or derive new scientific findings. The prerequisite for meaningful use, re-use or recombination of data is that they are well documented according to accepted and trusted standards. Those standards form a key pillar of science because they enable the recognition of suitable data. To ensure this, agreements on standards, accessibility and sharing practices have been defined. Strategies have to be fixed to preserve and store the data over a defined period of time in order to ensure their availability and re-usability after the end of IDEALFUEL. Especially, the metadata vocabularies and licences to permit the widest reuse possible need to be addressed more in detail in a future update of this deliverable.

---

[4] https://data.4tu.nl/info/en/

Risk Register

No risks foreseen related to this deliverable.

Acknowledgement

The author(s) would like to thank the partners in the project for their valuable comments on previous drafts and for performing the review.

Project partners:

#	Partner short name	Partner Full Name
1	TUE	Technische Universiteit Eindhoven
2	VERT	Vertoro BV
3	T4F	Tec4Fuels
4	BLOOM	Bloom Biorenewables Ltd
5	UNR	Uniresearch B.V.
6	WinGD	Winterthur Gas & Diesel AG
7		(Formerly SeaNRG, is now GOODFUELS #12)
8	TKMS	Thyssenkrupp Marine Systems GMBH
9	OWI	OWI – Science for Fuels gGmbH
10	CSIC	Agencia Estatal Consejo Superior De Investigaciones Cientificas
11	VARO	Varo Energy Netherlands BV
12	GOOD	GoodFuels B.V.

Appendix A - Preferred formats

All formats of digital files stand the risk of becoming obsolete in the future. As a general guideline, IDEALFUEL considers that the file formats best suited for long-time preservation and accessibility:

1. are commonly used;
2. have open specifications;
3. are independent of specific software, developers or suppliers.

Preferred and Acceptable formats

It is not always possible to select formats that meet with all of these ideal attributes. A number of file formats has been listed and considered as preferred format and acceptable format in Table 1. This list will change over time as new formats will be developed and others will fall into disuse.

- Preferred formats are the file formats which can be trusted to offer the best long-term guarantees for usability, accessibility and robustness.

- Acceptable formats are file formats which are commonly used besides the preferred formats; have average to reasonable scores regarding their usability, accessibility and robustness in the long term. A strong preference exists for the use of preferred formats but the use of acceptable formats will be allowed in to the archive as well.

Table 1: Preferred and acceptable formats for data storage

	Preferred format(s)	Acceptable format(s)
Text documents	• PDF/A (.pdf) • OpenDocument Text (.odt)	• MS Word (.doc, .docx) • Rich Text File (.rtf) • PDF (.pdf)
Text file	• Unicode TXT (.txt, …)	• Non-Unicode TXT (.txt, …)
Spreadsheets	• OpenDocument Spreadsheet (.ods) • Comma Separated Values (.csv)	• MS Excel (.xls, .xlsx) • PDF/A (.pdf) • OOXML (.docx, .docm)
Presentations	• PDF/A (.pdf) • OpenDocument presentation (.odp)	• OpenDocument presentation (.odp) • MS Powerpoint (.ppt, pptx) • PDF (.pdf)
Images (raster)	• JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png)	• JPEG 2000 (.jp2)
Images (vector)	• Scalable Vector Graphics (.svg)	• Adobe Illustrator (.ai) • PostScript (.eps)