



Next REnewable multi-GENeration technology enabled by TWO-phase fluids machines

## D8.6 Data Management Plan

Dissemination level:	PU
WP/Task/Deliverable	8/8.5/8.6
Contractual delivery date:	28/02/2021
Actual date of delivery	01/03/2021
Version:	2
Document status	Final version
Total number of pages:	35
Keywords:	#data #metadata #management #findable #accessible #interoperable #re-usable #open #repository #security



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 851541

Project Data			
<b>Project Acronym</b>	REGEN-BY-2		
<b>Project Title</b>	Next RENEwable multi-GENeration technology enabled by TWO-phase fluids machines		
<b>Grant Agreement n.</b>	851541		
<b>Call identifier</b>	H2020-LC-SC3-2019-RES-TwoStages		
<b>Topic identifier</b>	LC-SC3-RES-1-2019-2020 - Developing the next generation of renewable energy technologies		
<b>Funding Scheme</b>	RIA - Research and Innovation action		
<b>Project duration</b>	1 September 2020- 31 august 2024		
<b>Coordinator</b>	UNIVERSITA DI PISA		
<b>Website</b>	<a href="https://www.regen-by-2.eu/">https://www.regen-by-2.eu/</a>		
Deliverable Document Sheet			
<b>Deliverable No.</b>	8.6		
<b>Deliverable title</b>	Data Management Plan		
<b>Description</b>	Plan for the curation and management of data		
<b>WP No.</b>	8		
<b>Related task</b>	8.6		
<b>Lead Beneficiary</b>	R2M		
<b>Author(s)</b>	Laura Pérez (R2M), Tatiana Loureiro (R2M)		
<b>Contributor(s)</b>	Raymond Sterling (R2M)		
<b>Type</b>	ORDP: Open Research Data Pilot		
<b>Dissemination L.</b>	PU		
<b>Language</b>	English		
<b>Due Date</b>	28/02/2021	<b>Submission Date</b>	01/03/2021

Version	Date	Author	Action
0	29/01/2021	Laura Pérez (R2M)	First draft
1	22/02/2021	Tatiana Loureiro (R2M), Raymond Sterling (R2M), Laura Pérez (R2M)	Reviewed draft (internal)
2	26/02/2021	Laura Pérez (R2M), Tatiana Loureiro (R2M)	Final draft, incorporating feedback from partners

## **DISCLAIMER**

The opinion stated in this deliverable reflects the opinion of the authors and not the opinion of the European Commission.

All intellectual property rights are owned by REGEN-BY-2 consortium members and are protected by the applicable laws. Reproduction is not authorised without prior written agreement.

The commercial use of any information contained in this document may require a license from the owner of that information.

## **ACKNOWLEDGEMENT**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **Nº 851541**.

# Table of contents

<b>List of Acronyms</b>	<b>6</b>
<b>Executive Summary</b>	<b>7</b>
<b>1 Introduction</b>	<b>8</b>
<b>2 Methodology</b>	<b>9</b>
2.1 Data Availability and Open Access	9
2.1.1 Classification of available data	10
2.1.2 Open Access to publications	11
2.1.3 Open data and FAIR principles	12
2.1.4 Copyright licenses	13
2.2 Data storage and sharing	13
2.3 DMP template	13
<b>3 REGEN-BY-2 Data Management Plan</b>	<b>15</b>
3.1 Data Summary	15
3.2 FAIR data	17
3.2.1 Making data findable, including provisions for metadata	17
3.2.2 Making data openly accessible	18
3.2.3 Making data interoperable	18
3.2.4 Increase data re-use (through clarifying licences)	18
3.3 Allocation of resources	19
3.4 Data security	19
3.5 Ethical aspects	20
<b>4 Conclusions</b>	<b>21</b>
<b>Annex 1: Detailed datasets</b>	<b>22</b>

## List of Figures

Figure 1: Open access to scientific publications and research data in the wider context of dissemination and exploitation	10
---	----

## List of Tables

Table 1: Classification of available data	10
Table 2: FAIR principles	12
Table 3: DMP template	14
Table 4: Data Summary	16

## List of Acronyms

Acronym	Meaning
EC	European Commission
EU	European Union
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Re-usable
GA	Grant Agreement
H2020	Horizon 2020
IPR	Intellectual Property Rights
M	Month
OA	Open Access
ORD	Open Research Data
RP	Reporting Period
WP	Work Package

## Executive Summary

This document contains the Data Management Plan (DMP) for the Horizon 2020 REGEN-BY-2 project (under GA No. 851541). This DMP outlines how data collected, processed and/or generated during the REGEN-BY-2 project life cycle will be organised, stored and shared. As part of making research data findable, accessible, interoperable and re-usable (FAIR data), the DMP provides information on 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 and how data will be curated and preserved after the end of the project. Therefore, this plan will ensure that REGEN-BY-2 data and results are managed suitably, in line with all the requirements set by the European Commission in the “Guidelines on FAIR Data Management in Horizon 2020”. This DMP, scheduled for month 6 of the project, is the first of four submissions (in M6, M18, M30 and M48) and establishes the baseline, with 20 datasets identified.

# 1 Introduction

This document portrays the Data Management Plan (DMP) for the Horizon 2020 project “REGEN-BY-2: Next REnewable multi-GENeration technology enabled by TWO-phase fluids machines”. The main aim of this document is to define a strategy for the management of all the data collected, processed and generated during the REGEN-BY-2 project, in compliance with the EC guidelines<sup>1,2</sup>, including:

- the handling of research data during and after the end of the project;
- the identification of the data to be collected processed and/or generated, including research data, environmental and economic evaluations;
- the methodology and standards that will be applied;
- whether the data will be shared and made open access, and how;
- how the data will be curated and preserved;
- the management of databases to collect data from the project; and
- the management of internal requirements of the partners.

REGEN-BY-2 will follow a proactive strategy for internal data and knowledge management. The DMP is deliberately scheduled for M6; however, it is intended to be a living document, evolving throughout the project duration. Thus, updates to the DMP are planned for M18, M30 and M48, which will reflect any important changes to the project that may occur due to inclusion of new datasets, changes in consortium policies or external factors.

As established in Article 29 of the GA, REGEN-BY-2 project falls under the requirements of Open Research Data Pilot (ORD pilot)<sup>1,3</sup> and will therefore seek to openly disseminate its research results, except in those cases where it goes against the legitimate interests of the project partners (i.e. exploitation and/or protection of results), or there will be a high administrative burden for the disclosure of a dataset or limited worth to other users. The aim of participating in the ORD pilot is to “improve and maximise access to and re-use of research data generated by the project”<sup>1</sup>, taking into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data preservation questions.

The data resulting from the research and implementation of the project is what will validate the impact of REGEN-BY-2 and the conclusions drawn in the scientific publications arising. Many benefits may come with the sharing of knowledge and data from publicly funded research to the whole research community, the commercial sector and the civil society, and thus it is intended that, where possible, these data will be made available through open access repositories. 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.

---

<sup>1</sup> EC - DG RTD. Guidelines on FAIR Data Management in Horizon 2020, version 3.0, 26 July 2016, in:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)  
accessed on 11/01/2021

<sup>2</sup> EC - DG RTD. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, version 3.2, 21 March 2017, in:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)  
accessed on 11/01/2021

<sup>3</sup> OpenAIRE. What is the EC Open Research Data Pilot? <https://www.openaire.eu/what-is-the-open-research-data-pilot>  
accessed on 12/01/2021



## 2 Methodology

According to the “Guidelines on FAIR Data Management in Horizon 2020”<sup>1</sup>, the EC is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot) with the aim of maximising the access to data generated by H2020 projects, and encouraging its re-use. Prior to 2017, the scope of this pilot was limited to selected areas; however, it was reviewed and extended to cover all the thematic areas of H2020 Framework Programme. Open access to research data thereby becomes applicable by default in REGEN-BY-2, although the EC also recognises that there are good reasons to keep some or even all research data generated in a project closed, thus providing opt-out possibilities at any stage. The ORD pilot follows the principle “*as open as possible, as closed as necessary*” and focuses on encouraging sound data management as an essential part of research best practices<sup>4</sup>.

REGEN-BY-2, as a grant beneficiary, is thus required to develop (and keep up-to-date) a DMP, which will specify whether the data will be made open access; deposit its data in a research data repository (e.g. Zenodo<sup>5</sup> online OA repository); take measures to enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data; and provide information about the tools needed to validate the research results. For this purpose, all the REGEN-BY-2 project’s partners have been requested to fill a survey including the questions stated in the template provided by the EC guidelines<sup>1</sup> (summarized in Section 2.3). The partner’s answers to the survey have served for the development of this first version of the DMP. As the execution of the project progresses and new information and data are generated, the DMP will be updated in the context of the periodic reviews envisaged within the GA (M18, M30, M48).

### 2.1 Data Availability and Open Access

Open access (OA)<sup>6</sup> is defined as the practice of providing on-line, free-of-charge access to ‘scientific information’, which in the context of research and innovation can refer to peer-reviewed scientific research articles (published in academic journals), or research data (data underlying publications, curated data and/or raw data).

As outlined in the “Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020”<sup>2</sup>, broader access to scientific publications and research data serves a number of purposes:

- i) Improve the quality of research by building on a stronger body of existing work.
- ii) Increase efficiency of research by encouraging collaboration and reducing duplication of effort.
- iii) Speed up innovation (faster progress to market) by reducing barriers to information flow.
- iv) Enhance the transparency of scientific progress, easing the involvement of citizens and society.
- v) Allow other researchers, industry or citizens to access and re-use information that has been paid for with public money and hence should not have to be paid for again each time it is accessed.

---

<sup>4</sup> EC. H2020 Online Manual - Data Management.

[https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm) accessed on 11/01/2021

<sup>5</sup> Zenodo - open access repository. <https://zenodo.org/> accessed on 14/01/2021

<sup>6</sup> EC. H2020 Online Manual - Open Access.

[https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm) accessed on 11/01/2021

The EU wants to improve access to scientific information and boost the benefits of public investment in research funded under Horizon 2020. This means making publicly-funded scientific information available online, at no extra cost, to European researchers, innovative industries and the public, while ensuring that it is preserved in the long term<sup>6</sup>.

These OA principles are translated into specific requirements in the model GA for H2020 projects. As outlined in Figure 1, one of the first decisions to be made in regard to research dissemination is whether to publish the research findings of the project or to protect some aspects for commercial exploitation. This process creates a series of interdependencies between REGEN-BY-2’s technical work program (WP2-WP6), the strategic development roadmap and the communication, dissemination and exploitation work packages (WP7-WP8), and the project management procedures (WP1).

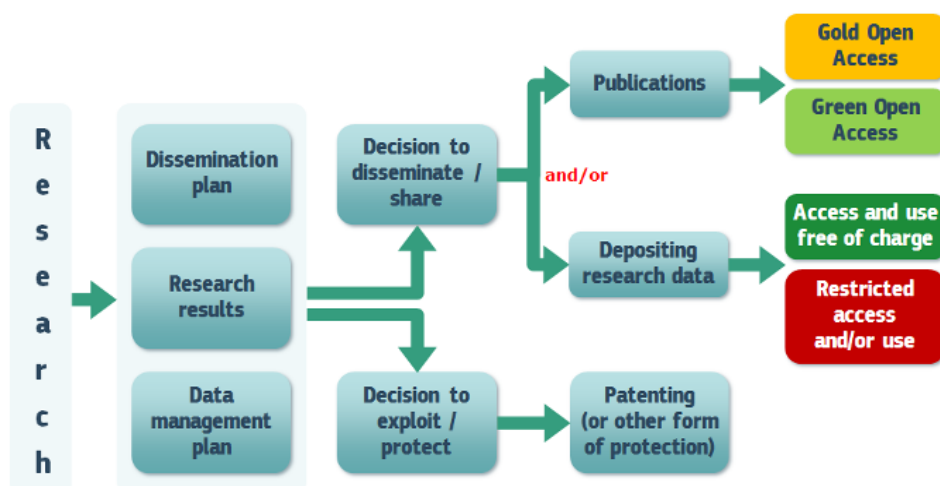


Figure 1: Open access to scientific publications and research data<sup>6</sup>

For each research result of the project there should be an evaluation whether to disseminate/share or to exploit/protect it. As specified in the GA, “the beneficiaries do not have to ensure open access to specific parts of their research data, if the achievement of the action’s main objective would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access”. The Communication and Dissemination strategy for REGEN-BY-2 (D8.2) will also be defined within WP8, in particular T8.1 and T8.2, whilst the exploitation and identification of suitable protection mechanisms will be identified as part of T8.3 and T8.5 (D8.7 IPR & Exploitation Plan).

### 2.1.1 Classification of available data

All available data, datasets and project information are therefore categorised at this stage in one of three ways (Table 1):

Data availability	
<b>Open</b>	Data that is shared for re-use or that underpins a scientific publication.
<b>Consortium</b>	Confidential data that is accessible to all partners, but retained within the consortium and subject to the project Non-Disclosure Agreement (NDA).
<b>Private</b>	Data that is maintained by an individual partner for their own purposes.

Table 1: Classification of available data

The REGEN-BY-2 project will seek to openly disseminate its research findings, except in cases where there are defined exploitable outcomes, privacy concerns or there will be a high administrative burden for a dataset to separate open/consortium/private sections leading to limited worth to other users. The two main aspects of this dissemination approach are OA to scientific publications and OA to research data, which are presented in the following sections of this document.

## 2.1.2 Open Access to publications

Open access publishing is defined as the free availability of peer-reviewed scientific publications for any user. There are two different paths to follow this approach, the “green” and “gold” OA routes:

- **Green OA** (self-archiving) is the release/deposition of a final peer-reviewed manuscript or article in an online repository before, alongside or after its publication. Repository software usually allows authors to delay access to the article during an 'embargo period', but OA must be ensured within a maximum of 6 months.
- **Gold OA** (open access publishing) is the immediate publication of an article in open access mode through a publisher/journal website. OA must be granted at the latest on the date of publication, and a machine-readable electronic copy must also be deposited in a repository. In the gold route, the payment of publication costs is shifted away from subscribing readers. This model is often based on one-off payments by authors, but these costs (Article Processing Charges, APCs) are eligible for reimbursement during the duration of the project as part of the overall project budget.

Nevertheless, the EC has recently launched the Open Research Europe<sup>7</sup> publishing platform, where H2020-funded researchers can submit scientific papers for peer reviewed publication, fully compliant with all the GA obligations in terms of embargo periods, repositories and provision of OA in a single process, and free of charge. This alternative will be considered by the REGEN-BY-2 project's partners.

The requirement to make OA the scientific publications derived from the project do not imply an obligation to publish results. The decision to publish is entirely up to the H2020 grant beneficiaries. Thus, OA becomes an issue only if publication is chosen as a means of dissemination. Moreover, the decision on whether to publish through OA comes after the more general decision on whether to publish directly or to first seek protection, OA hence does not affect the decision to exploit research results commercially.

The REGEN-BY-2 consortium will encourage OA publishing of scientific publications. Considering the technological focus and the academic composition of REGEN-BY-2, the consortium partners expect to deliver up to 5 publications (articles) in high-impact journals. Additionally, the project may provide other types of scientific publications (not peer-reviewed), such as monographs, books, conference proceedings and grey literature (informally published written material not controlled by scientific publishers, e.g. reports, press releases). In order to comply with the OA requirements, REGEN-BY-2's dissemination leader (R2M) will establish and maintain a Zenodo<sup>5</sup> online repository for self-archiving of publications and research results, ensuring that they can be found, read, downloaded and printed.

The EC guidelines<sup>2</sup> indicate that OA should also be provided, through the repository and in a standard format, to the bibliographic metadata<sup>8</sup> that identify the deposited publication. The purpose is to make it easier to find publications and ensure that EU funding is acknowledged.

<sup>7</sup> EC. Open Research Europe. <https://open-research-europe.ec.europa.eu/> accessed on 19/02/2021

<sup>8</sup> Metadata is "data that provides information about other data". In other words, it is "data about data". Metadata is used to summarize basic information about data which can make tracking and working with specific instances of data easier.

Article 29 of the GA also mentions that H2020 grant beneficiaries (ORD pilot<sup>3</sup> participants) must aim to deposit, at the same time as the publication, the research data needed to validate the results presented in the deposited scientific publications ('underlying data'), other curated and/or raw data not directly attributable to a publication, including the associated metadata. This requirement is not related to the openness of the data but to data management.

### 2.1.3 Open data and FAIR principles

Open access to research data refers to the right to access and reuse digital research data under the terms and conditions set out in the GA; research data consisting of information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation.

“Open data” is normally defined as openly accessible research data that the user can access, mine, exploit, reproduce and disseminate free of charge. As stated in the EC guidelines<sup>1,4</sup>, all the data that REGEN-BY-2 will make publicly available (open) will comply with the FAIR principles<sup>9,10</sup> (Table 2) in order to ensure sound data management, as an essential part of research best practice.

<b>F</b>	<b>Findable</b>	Discoverable with metadata, identifiable and locatable by means of a standard identification mechanism.
<b>A</b>	<b>Accessible</b>	Always available and obtainable; even if the data is restricted, the metadata is open.
<b>I</b>	<b>Interoperable</b>	Both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organisations or countries.
<b>R</b>	<b>Re-usable</b>	Sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources.

Table 2: FAIR principles<sup>9,10</sup>

FAIR is a set of principles, not a standard. Rather, the EC supports FAIR data as a framework to follow when designing a DMP. It should be noted that participating in the ORD pilot<sup>3</sup> and following the FAIR principles does not mean opening up all the project’s data. Data can be FAIR but not open (e.g. data could meet the FAIR principles, but be private or only shared under certain restrictions), whilst open data may not be FAIR (e.g. publicly available data may lack sufficient documentation to meet the FAIR principles, such as licensing for clear reuse)<sup>11</sup>.

Datasets for dissemination of the REGEN-BY-2 project will be open access by default, at the very least to validate scientific publications. However, not all the project work packages will produce datasets that are intended for public dissemination; much of the data created and stored during the project is to be kept confidential.

<sup>9</sup> Force 11. The FAIR Data Principles, in: <https://www.force11.org/group/fairgroup/fairprinciples> accessed on 08/01/2021

<sup>10</sup> Wilkinson, M. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. <https://doi.org/10.1038/sdata.2016.18> accessed on 08/01/2021

<sup>11</sup> OpenAIRE. How to make your data FAIR. <https://www.openaire.eu/how-to-make-your-data-fair> accessed on 12/01/2021

## 2.1.4 Copyright licenses

When material is widely shared, copyright licences protect the authors of the work and grant specific rights to publishers and others to use this work. The EC encourages authors to retain their copyright and grant adequate licences to publishers. Creative Commons (CC)<sup>12</sup> provides legal tools and licensing solutions to enable OA in these circumstances, with CC-BY and CC0 enabling re-use by third parties.

When a project's research findings (i.e facts, data) are published in a journal or other scientific outlet, there should be consideration of the copyright agreement with the publishers, which may involve an embargo period. Submission to the chosen repository requires the authors to agree to a non-exclusive distribution licence, and a Creative Commons licence may be added at this stage. This type of licence is a good legal tool for providing open access in its broadest sense.

## 2.2 Data storage and sharing

Data will be stored and shared based on the different categorisation defined in Section 2.2.1 (private, consortium, open):

- **Private data** will be stored locally by the individual owners on their own servers. This type of data storage and sharing process is not in the scope of REGEN-BY-2 DMP; however, each partner will ensure the implementation and compliance with their internal data protection processes/policies and national/European regulations.
- **Consortium (confidential) data** will be stored and shared through the project's internal archive, which has been set up using the space provided by the project coordinator in Google Drive<sup>13</sup>, as stated in D1.4 "Project Management Plan". This online repository is used for storing and sharing all project related documents/information via the Internet, allowing all the project partners to download, archive and exchange project-related data during the whole project execution.
- **Open data** will be made available through Zenodo<sup>5</sup> open data online repository, to be established and maintained by REGEN-BY-2's partner R2M. This storage will ensure that the publicly available (open) scientific publications and research results from the REGEN-BY-2 project can be searched, read online, distributed, downloaded and printed.

## 2.3 DMP template

The following template (Table 3) provides a summary of the DMP issues to be addressed, as outlined in the EC guidelines<sup>1</sup>, which will be applied to REGEN-BY-2's datasets as they are developed across the project's lifetime.

---

<sup>12</sup> Creative Commons. About The Licenses. <https://creativecommons.org/licenses/> accessed on 13/01/2021

<sup>13</sup> Google Drive - file storage and synchronization service. <https://www.google.es/drive/>

DMP component	Issues to be addressed
<p><b>1. Data summary</b></p>	<ul style="list-style-type: none"> <li>• State the purpose of the data collection/generation</li> <li>• Explain the relation to the objectives of the project</li> <li>• Specify the types and formats of data generated/collected</li> <li>• Specify if existing data is being re-used (if any)</li> <li>• Specify the origin of the data</li> <li>• State the expected size of the data (if known)</li> <li>• Outline the data utility: to whom will it be useful</li> </ul>
<p><b>2. FAIR Data</b>  <b>2.1. Making data findable, including provisions for metadata</b></p>	<ul style="list-style-type: none"> <li>• Outline the discoverability of data (metadata provision)</li> <li>• Outline the identifiability of data and refer to standard identification mechanisms. Do you make use of persistent and unique identifiers such as Digital Object Identifiers (DOI)<sup>14</sup>?</li> <li>• Outline naming conventions used</li> <li>• Outline the approach towards search keyword</li> <li>• Outline the approach for clear versioning</li> <li>• Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how</li> </ul>
<p><b>2.2 Making data openly accessible</b></p>	<ul style="list-style-type: none"> <li>• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so</li> <li>• Specify how the data will be made available</li> <li>• Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?</li> <li>• Specify where the data and associated metadata, documentation and code are deposited</li> <li>• Specify how access will be provided in case there are any restrictions</li> </ul>
<p><b>2.3. Making data interoperable</b></p>	<ul style="list-style-type: none"> <li>• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.</li> <li>• Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?</li> </ul>
<p><b>2.4. Increase data re-use (through clarifying licences)</b></p>	<ul style="list-style-type: none"> <li>• Specify how the data will be licenced to permit the widest reuse possible</li> <li>• Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed</li> <li>• Specify whether the data produced and/or used in the project is</li> </ul>

<sup>14</sup> A DOI is a persistent identifier for a document that can be handled by a resolution service to direct communications to the correct server. Developed by the International DOI Foundation ([www.doi.org](http://www.doi.org)).

	<p>usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why</p> <ul style="list-style-type: none"> <li>• Describe data quality assurance processes</li> <li>• Specify the length of time for which the data will remain re-usable</li> </ul>
<b>3. Allocation of resources</b>	<ul style="list-style-type: none"> <li>• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs</li> <li>• Clearly identify responsibilities for data management in your project</li> <li>• Describe costs and potential value of long term preservation</li> </ul>
<b>4. Data security</b>	<ul style="list-style-type: none"> <li>• Address data recovery as well as secure storage and transfer of sensitive data</li> </ul>
<b>5. Ethical aspects</b>	<ul style="list-style-type: none"> <li>• To be covered in the context of the ethics review, ethics section of Description of the Action and ethics deliverables. Include references and related technical aspects if not covered by the former</li> </ul>
<b>6. Other</b>	<ul style="list-style-type: none"> <li>• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)</li> </ul>

Table 3: DMP template

### 3 REGEN-BY-2 Data Management Plan

This section presents the baseline for REGEN-BY-2’s DMP, outlining how the data collected, processed and/or generated during the project life cycle will be organised, stored and shared following the FAIR principles. The aim of this plan is to ensure that REGEN-BY-2’s information, data and results will be managed suitably. To this end, all the project partners have filled a survey in answer to the questions stated in the template provided by the EC in the “Guidelines on FAIR Data Management in Horizon 2020”<sup>1</sup> (summarized in Table 3). The answers to the survey have been collected and analysed by R2M and are compiled and described in the following sections of this document.

In this first version of the DMP (M6), not all the template’s questions have been answered in detail. The DMP is intended to be a living document, in which information and additional data will be made available on a finer level of granularity through updates as the implementation of the REGEN-BY-2 project progresses, and when significant changes occur. This plan will be updated in the context of the periodic reviews envisaged within the GA (M18, M30, M48).

#### 3.1 Data Summary

REGEN-BY-2’s project information (“I”) and datasets (“D”) are reported in a summarized form below (in Table 4). This initial list of datasets has been identified based on the data collected during the first six months of the project and the datasets identified as part of the proposal and project planning.

At the time of this deliverable submission, **15 technical datasets and 5 project information datasets** are foreseen. Although the REGEN-BY-2 consortium is committed to OA and public dissemination, not all identified datasets have been classified as OA. There are 12 datasets classified as Consortium (CO), to be kept confidential among the project partners and the EC, although part of these datasets may be made OA at some point (e.g. for validation of the results presented in the public deliverables).

No.	Dataset description	WP/Task	Class	Data manager
I1	Project information not for dissemination purposes: Periodic reports, confidential deliverables, meeting minutes, meeting presentations and like items.	WP1	CO	UNIPI
I2	Dissemination materials: public deliverables and scientific publications.	WP1, WP8	OA	UNIPI, R2M
I3	Communication materials: newsletters, brochures, posters, videos, and like items	WP8	OA	R2M
I4	Personal Data <sup>15</sup> : of any single person or group of people, either within (i.e. project participants) or external to the Consortium (i.e. stakeholders database), who have in some way actively provided REGEN-BY-2 with useful information that could potentially be of some value.	WP1, WP8	CO	UNIPI, R2M
I5	Personal Data <sup>14</sup> : of persons and institutions interested in passively receiving information from or about REGEN-BY-2, such as the periodic newsletter or specific queries received through the website contact form.	WP8	CO	R2M
D1	Datasets, data files, models, algorithms and source codes for working fluid selection/screening and thermodynamic cycle modelling	WP2/ T2.1-2.2	OA	TIFEO, CNRS
D2	Models for CFD/FEM (whose results should be shared in anonymized/specific unit of measurement way)	WP2/ T2.2-2.4	OA	RINA-C
D3	Sensors data (Inputs), Actuators setpoints (Outputs), Control Algorithms, Electrical Diagrams (schemas)	WP2/T2.4	OA	CARTIF
D4	P&ID, isometric drawings, plants, 3D views, exploded-view drawings, item specification tables, , component lists (pdf), text documents, images, python files, mex files, in addition to eventual codes, functions, algorithms, etc. in order to lead the project to a fully functioning lab-scale prototype.	WP2/ T2.1-2.5	CO	TIFEO
D5	State-of-the-art, modelling and experimentation data from measurements, source codes (simulation model of expander developed in Python)	WP3/ T3.1-T3.6	OA	NTUA
D6	Ansys Fluent® CFD files from the numerical validation (simulation model) of the prototype expander and compressor	WP3/T3.4 WP4/T4.4	CO	LSTME, NTUA, ULIEGE
D7	Prototype CAD (STEP) files	T3.3 & T3.5 T4.3 & T4.5	CO	EXOES

<sup>15</sup> The treatment of personal data in REGEN-BY-2 will be in compliance with the EU General Data Protection Regulation (GDPR 2016/679 <https://eur-lex.europa.eu/eli/reg/2016/679>).



<b>D8</b>	State-of-the-art, modelling and experimentation data from measurements, source codes (simulation model of compressor developed in Python/Matlab)	WP4/ T4.1-4.6	OA	ULIEGE
<b>D9</b>	Data gathered for the engineering and development of the prototype	WP5/T5.2	CO	HYS
<b>D10</b>	Dataset related to risk assessment and Health and Safety aspects (ATEX compliance).	WP5/T5.2	CO	RINA-C
<b>D11</b>	Data from the experimental campaign (csv files); safety procedures of each partner involved in experimental activities; photos and videos of the test rig; text files for deliverables.	WP6/ T6.1-6.4	CO	UNIPI
<b>D12</b>	Standards/regulation/market assessment and analysis data	WP7/ T7.1-7.5	CO	RINA-C
<b>D13</b>	Results of the Cost Benefit Analysis of REGEN-BY-2 technology	WP7/T7.2	OA	NTUA, JER
<b>D14</b>	Datasets related to standardization landscape and applicable standards	WP7/T7.5	CO	UNE
<b>D15</b>	Exploitable results description; value proposition map; SWOT analysis; IPR and exploitation plan; business model canvas	WP8/ T8.2-8.4	CO	R2M

Table 4: Data Summary (more details in the Annex)

### 3.2 FAIR data

This section describes how REGEN-BY-2 research outputs will be organised so they can be more easily accessed, understood, exchanged and reused. The EC promotes FAIR data to maximise the integrity and impact of their research investment. However, not all the project’s data can be open, as there are protection, confidentiality and security obligations, reflected in the GA, which apply. For instance, the REGEN-BY-2 project will design, construct and experiment a first-of-its-kind lab-scale prototype of a recent near-worldwide patented thermodynamic cycle and related plant. Most of the data handled during the project is protected by this patent, and must therefore be kept confidential. Nevertheless, the REGEN-BY-2 consortium will follow the ORD pilot’s approach described as "*as open as possible, as closed as necessary*", having in mind that providing other researchers with access to your research data facilitates knowledge discovery and improves research transparency<sup>10</sup>.

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

REGEN-BY-2 will aim to make its research data findable by ensuring it:

- has a **persistent identifier** (PIDs), which allow to unambiguously identify data and facilitate data citation. An example of a PID is a Digital Object Identifier (DOI)<sup>13</sup>. Zenodo<sup>5</sup> repository has been selected for depositing REGEN-BY-2 data because it assigns PIDs.
- has **rich metadata**, which supports findability, citation and reuse, also providing important context for the interpretation of the data and making automated analysis easier. REGEN-BY-2 will aim to follow standard metadata schemes, when possible.
- is **searchable and discoverable online**.

More details about the discoverability measures expected for each of the datasets are presented in the Annex. At this stage of the project, several partners could not specify which data and metadata standards will be used during their research activities, which have not yet started. Throughout the project execution, R2M will provide guidance to the rest of the partners on how to make their data findable with the help of data and metadata standards.

### 3.2.2 Making data openly accessible

REGEN-BY-2 partners intend to make their data accessible by ensuring it is **retrievable online using standardised protocols**; and has **restrictions in place** if necessary.

As previously mentioned, not all data will be made openly accessible in REGEN-BY-2, but it will be as FAIR as possible, even when restricted. However, if open access is allowed, REGEN-BY-2 research data will be retrievable without the need for specialised protocols. Table 4 indicates whether each dataset will be shared and made open access. The Annex shows more detailed information about how the project partners will make their data accessible. R2M will ensure that the project's open data can be accessed online through Zenodo<sup>5</sup> repository.

### 3.2.3 Making data interoperable

In REGEN-BY-2, data will be made interoperable by using **common formats and standards, controlled vocabularies**, community-agreed schemas, keywords or ontologies, where possible. This will allow the data generated within REGEN-BY-2 to be integrated with other data, applications and workflows. When possible, data will not be created with proprietary software, making it available in OA formats.

The answers provided by the partners in regard to the interoperability of their data are gathered in the Annex. The survey specifically requested information about any specific tools (including software) related to the data, and whether these respected data accessibility and interoperability. R2M will evaluate the formats, standards and tools that will be used by the partners during the execution of their activities in order to provide tailored guidance to the partners and determine if any actions need to be taken (i.e. change of format) to improve the interoperability of the REGEN-BY-2 project's data. These actions, if required, will be specified in the upcoming versions of the DMP.

### 3.2.4 Increase data re-use (through clarifying licences)

REGEN-BY-2 will aim to make its data reusable by ensuring it:

- is **well-documented**, by creating documentation (e.g. a README file) to help ensure that the data can be correctly interpreted and analyzed by others. For example, such documentation should contain a short description of what data it includes (for each filename); definition of column and row labels, data codes, and measurement units (for tabular data); any data processing steps that may affect interpretation of results; if applicable, a description of what associated datasets are stored elsewhere; or whom to contact with questions.
- has **clear licence and provenance information**. It is important to have clarifying licenses<sup>16</sup> to govern the terms of the data reuse. The OA guidelines under H2020<sup>2,10</sup> recommend CC-0 or CC-BY as an effective way to make it possible for others to mine, exploit and reproduce the data.

---

<sup>16</sup> Ball, A. (2014). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <https://www.dcc.ac.uk/guidance/how-guides> accessed on 27/01/2021

REGEN-BY-2 partners were not able to provide very detailed information on data licensing at this early stage. Nevertheless, it should be noted that most of the data generated within REGEN-BY-2 will be consortium-internal data (proprietary license); and that the project is driven by a patented technology exclusively licensed to the partner TIFEO. Any licenses to be granted to the project's openly accessible data will have to be agreed within the consortium, and will be specified in upcoming submissions of the DMP. The expected re-usability of some of the identified datasets is described in the Annex.

### 3.3 Allocation of resources

This section of the DMP addresses the issues related to the costs of making the project's data FAIR and its long-term preservation, and the identification of responsibilities for data management. The partners responsible for each dataset are identified in Table 4. Although R2M will provide guidance on how to follow the FAIR principles (through the DMP), it will be the responsibility of each partner to take measures in order to ensure a sound management of their data. Even though most partners are planning to use their own IT server and network resources for data preservation, all the project data will be also stored and shared (within the consortium) through the project's internal archive (Google Drive), which has been set up by the UNIPI. No additional costs will be incurred for this purpose.

Besides, REGEN-BY-2's publications and open data will be made accessible through Zenodo<sup>5</sup> online OA repository, being the costs related to open access eligible as part of the H2020 grant. The potential value and the duration of REGEN-BY-2's data preservation in the long term is an issue still to be discussed among the partners and to be specified in future updates of the project's DMP. Regarding the costs of making the REGEN-BY-2 project's data FAIR, the upcoming updated versions of the DMP will provide more detailed information on whether the project's partners needed to allocate additional resources to comply with the FAIR principles.

### 3.4 Data security

This section addresses data recovery as well as secure storage and transfer of sensitive data. In order to keep data safe for the long term, REGEN-BY-2 has selected a trustworthy repository (Zenodo<sup>5</sup>) that stores the data safely, ensures the data is findable, describes the data appropriately (metadata) and adds license information<sup>10</sup>. Moreover, the partners will store their data in the project's internal archive (Google Drive) as well as in their own servers and data storage systems. The providers of cloud storage and cloud computation resources and tools employed by the partners (e.g. Google, Microsoft, etc.) guarantee data security, keeping it protected against unauthorized access. The hardware (i.e. laptops) where data is stored are also password protected. Overall, REGEN-BY-2 partners will rely on the data security guaranteed by their service providers and their internal security procedures for data storage.

Special attention will be paid to the management of sensitive data throughout the research process. Sensitive data must be protected against unwanted disclosure and its accessibility should hence be safeguarded. The protection of sensitive data<sup>17</sup> may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations. In particular, REGEN-BY-2 will comply with the strong EU regarding personal (GDPR<sup>15</sup>) and sensitive non-personal data (e.g. REGEN-BY-2's confidential data). Sensitive data can still meet the FAIR principles and be processed in a way that the needed protection is guaranteed also in the future.

---

<sup>17</sup> OpenAIRE. How to deal with sensitive data. <https://www.openaire.eu/sensitive-data-guide> accessed on 27/01/2021

Sensitive data should be archived under a closed license in a trustworthy repository. However, the metadata could be made public, without making the data itself openly accessible, so it can be findable for other researchers while placing conditions around access to the data. Nevertheless, it should be noted that it is not always necessary to keep all the data, as it will depend on criteria such as its uniqueness, long-term value, reuse potential and the necessity to validate results in publications<sup>16</sup>.

### 3.5 Ethical aspects

This section of the DMP addressed any potential ethical issues that may impact data sharing. The ethical aspects of the REGEN-BY-2 project and its data have been covered in the context of the ethics review and the deliverables D9.1 and D9.2 on WP9 “Ethics requirements”. The research done by the REGEN-BY-2 project’s partner LSTME in the Republic of South Korea (non-EU country) does not raise any ethical issues (D9.1); while health and safety procedures and safety risk management plans have been developed, considering that there will be experiments performed in REGEN-BY-2 (D9.2).

If any ethical issues concerning the management of the project’s data are identified during the project execution, they will be included in the following scheduled submissions of the DMP.

## 4 Conclusions

This document is the first version of REGEN-BY-2's DMP, to be updated in M18, M30 and M48 of the project; and it establishes the baseline for data management, with 20 datasets identified. This DMP and its subsequent updates fulfil the function of ensuring a sound management of the REGEN-BY-2 project's data during the project's lifetime and the compliance with the FAIR principles, in accordance with the EC guidelines, the GA and best practices. For this purpose, the project partners were asked to fill a survey in order to identify the data that they will be handling during their research activities, what standards and softwares they will use, whether and how it will be shared and stored, among other questions related to data management and security. Besides, Zenodo has been selected as the OA repository for the project's publications and data. As the project advances, additional details on the already identified datasets, as well as potential new datasets, will be included in the following updates of the REGEN-BY-2 project's DMP.

## Annex 1: Detailed datasets

The following tables show a compilation of the answers provided by the REGEN-BY-2 project partners to the DMP survey.

UNIFI - Datasets: I1, I2, I4, D11	
<b>Data description</b>	<p><b>Types and formats:</b> List of participants (including email address); safety procedures of each partner involved in experimental activities (pdf and word files); csv files from experimental campaign; photos and videos of the test rig; ansys CFD files about expanders and compressors; word and pdf files for deliverables.</p> <p><b>Purpose:</b> Execution of the tasks.</p> <p><b>Re-use:</b> no for the time being.</p> <p><b>Expected size:</b> 10 GB</p> <p><b>Utility:</b> experimental data will be used by all the partners; CFD files will be used by LSTME, ULIEGE, NTUA, EXOES, TIFEO, RINA-C</p>
<b>WPs &amp; Tasks</b>	CFD (Task 3.4, 4.4); experimental (WP6); project management (WP1, WP9)
<b>Responsible partner(s)</b>	Experiments and project managements (UNIFI)CFD (LSTME),
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	csv, pptx, docx, xlsx, pdf, Ansys CFD file format, jpg, mp4
<b>Standards for any related metadata</b>	Spreadsheet files
<b>Specific tools/software (accessibility and interoperability)</b>	Microsoft Excel, Microsoft Word, Microsoft PowerPoint, Ansys CFD
<b>Accessibility</b>	In the Project sharepoint
<b>Deposition of data and associated metadata</b>	In the Project sharepoint, in the EU portal (only deliverables)
<b>With whom is the data shared during the project?</b>	Project partners
<b>Data licensing</b>	-
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. 5 years. UNIFI
<b>Ethical and/or legal requirements</b>	-
<b>Provisions for data security</b>	Internal repository

Table 5: Datasets from UNIFI

RINA-C - Datasets: D2, D10, D12	
<b>Data description</b>	WP2: models for CFD/FEM whose results should be shared in an anonymized/specific unit of measurement. WP5: Health and safety aspects (confidential, or not fully open) WP7: standards/ regulation/market assessment and analysis data that in some cases could be confidential (not fully open)
<b>WPs &amp; Tasks</b>	WP2 - WP5 - WP7
<b>Responsible partner(s)</b>	RINA-C
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	It will depend on the analysis to be performed. So far it looks quite hard to define. Nevertheless, we're used to using "standard" extensions.
<b>Standards for any related metadata</b>	No
<b>Specific tools/software (accessibility and interoperability)</b>	We usually use/share data in Excel/word
<b>Accessibility</b>	Project Repository
<b>Deposition of data and associated metadata</b>	Data not to be disclosed
<b>With whom is the data shared during the project?</b>	Consortium only
<b>Data licensing</b>	Proprietary
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. In RINA-C servers as confidential.
<b>Ethical and/or legal requirements</b>	-
<b>Provisions for data security</b>	Yes - in RINA-C servers

Table 6: Datasets from RINA-C

CARTIF - Datasets: D3	
<b>Data description</b>	Sensors data (Inputs), Actuators setpoints (Outputs), Control Algorithms, Electrical Diagrams (schemas)
<b>WPs &amp; Tasks</b>	T2.4, T3.6, T4.6, T5.2, T5.3, T5.4, T6.2, T.6.4
<b>Responsible partner(s)</b>	CARTIF, TIFEO, UNIPI, ULIEGE, NTUA, EXOES, HYSYTECH
<b>Dissemination level</b>	OPEN - Data that is shared for re-use or that underpins a scientific publication
<b>Data standards</b>	.docx, .xlsx, .pptx, .pdf, .csv, SQL data, ladder and SCL programming language
<b>Standards for any related metadata</b>	PLC programming according to IEC 61131-3 standard, version numbers for control algorithms programs
<b>Specific tools/software (access and interoperab.)</b>	Microsoft Office, Step7 and WinCC (TIA Portal), See Electrical Expert
<b>Accessibility</b>	Project data repository
<b>Deposition of data and associated metadata</b>	Datasets will be deposited in the corresponding project repository
<b>With whom is the data shared during the project?</b>	Data will be shared with the rest of the partners and scientific papers will be open access.
<b>Data licensing</b>	Scientific papers will be open access and the rest of data no licence required (to be reviewed through project lifecycle)
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. At least 5 years. CARTIF through its quality management platform. Where: A possibility could be the AI4EU platform. Costs: this would go within CARTIF indirect costs (if any)
<b>Ethical and/or legal requirements</b>	Defined datasets doesn't have any ethical and/or legal requirements so far (to be reviewed through project lifecycle)
<b>Provisions for data security</b>	Datasets will be stored in the corresponding project repository and in the CARTIF quality management platform repository as backup. If AI4EU platform or any other certified platform is going to be used as a project repository, this will be another dataset backup.

Table 7: Datasets from CARTIF

CNRS & UL - Datasets: D1	
<b>Data description</b>	<p><b>Types and formats:</b> datasets, data files, models, algorithms, source codes, dissemination materials (presentations, articles), methodologies.</p> <p><b>Purpose:</b> the analysis of existing datasets and data files is required to screen the working fluid and to define the thermodynamic model of all considered working fluids; user-defined datasets are created starting from the available and analysed datasets; models, algorithms and source codes are produced to treat data contained in datasets, to enable the thermodynamic description of fluid properties and to model the thermodynamic cycle of REGEN-BY-2; to understand what is the impact of the presence of the oil on the thermodynamic properties of the working fluid.</p>



	<p><b>Re-use:</b> yes, in future fluid screenings, for applications other than REGEN by 2</p> <p><b>Origin:</b> Existing databases acquired in the context of the project (from NIST, DDB, DIPPR) or the results of our work.</p> <p><b>Expected size:</b> Depending on the data, from very low size to the order of terabytes</p> <p><b>Utility:</b> to any person who wish to select a fluid for a thermodynamic cycle, according to specific criteria; to researchers who need to determine the operating conditions of REGEN-BY-2; the developed thermodynamic model will be useful to anyone who wish to model the properties of a pure fluid in an accurate way or of binary mixtures of working fluid + an oil.</p>
<b>WPs &amp; Tasks</b>	Task 2.1 and 2.2
<b>Responsible partner(s)</b>	CNRS, TIFEO
<b>Dissemination level</b>	OPEN - Data that is shared for re-use or that underpins a scientific publication.
<b>Data standards</b>	<p>Data produced in this project by CNRS are interoperable and compliant with open software applications.</p> <p>Data will be produced and shared in Excel format, Fortran format, .txt format, PowerPoint and Word format, images in .jpeg, .png, .PDF.</p> <p>We will use at most standardized vocabularies.</p> <p>We will provide mappings to more commonly used ontologies.</p>
<b>Standards for any related metadata</b>	No
<b>Specific tools/software (accessibility and interoperability)</b>	<p>Excel, Word, Fortran, Matlab, PowerPoint.</p> <p>No documentation about the software is needed to access the data.</p>
<b>Accessibility</b>	Currently, data is shared within the project by email. A repository will be created if demanded by Tifeo.
<b>Deposition of data and associated metadata</b>	Project repository
<b>With whom is the data shared during the project?</b>	With partners that participate to tasks in which CNRS is involved
<b>Data licensing</b>	We do not plan to bound the data to any licence.
<b>Preservation after the end of the project? How long? Who is responsible?</b>	<p>Yes. No time limits.CNRS. On the Intranet of CNRS and on the project repository.</p> <p>No costs are expected for data preservation</p>
<b>Ethical and/or legal requirements</b>	<p>Data used by CNRS have been either bought (databases by NIST, DDB, DIPPR) or produced by CNRS. Data has been bought in the first semester of the project.</p> <p>Acquired data are stored on the PC of the researchers that work on the project.</p> <p>The researchers that work on the project will be able to access the data. It is not personal data.</p>
<b>Provisions for data security</b>	Yes, our sensitive data are stored in a repository internal to our CNRS laboratory and, thus, protected by its proper security system.

Table 8: Datasets from CNRS & UL

R2M - Datasets: I2, I3, I4, I5, D15	
<b>Data description</b>	Exploitable Results description; Value proposition map; SWOT analysis; IPR and exploitation plan; Business canvas; online communication (website and social media); communication materials; dissemination activities; OA publications; stakeholders database.
<b>WPs &amp; Tasks</b>	WP8 - 8.3; 8.3; 8.4; 8.5; 8.4; 8.1; 8.1; 8.2; 8.2; 8.2.
<b>Responsible partner(s)</b>	R2M
<b>Dissemination level</b>	R2M will handle both Open and Confidential data: OPEN - Data that is shared for re-use or that underpins a scientific publication. CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	source code; images (.jpg, .png, .ai); presentations (.ppt, .pdf); audio-visual (.mp3, .mp4); documents (.doc, .csv).
<b>Standards for any related metadata</b>	HTML; scripts; spreadsheets. DOI for OA publications.
<b>Specific tools/software (accessibility and interoperability)</b>	Google Drive; Adobe suite; Wordpress; Twitter; LinkedIn.
<b>Accessibility</b>	Consortium datasets: project's repository. Public datasets: OA repository; project's website. All data will also be stored in R2M's Google Drive.
<b>Deposition of data and associated metadata</b>	Public datasets: Zenodo or other OA repository.
<b>With whom is the data shared during the project?</b>	Public datasets; Consortium datasets: all people working in the project.
<b>Data licensing</b>	OA publications: CC 4.0
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. At least 3 years. R2M. R2M's internal repository; website and OA repository. Costs have already been covered; the rest of the costs will be assumed by R2M.
<b>Ethical and/or legal requirements</b>	Stakeholder database: personal data, it must be gathered with informed consent complying with GDPR
<b>Provisions for data security</b>	-

Table 9: Datasets from R2M

TIFEO - Datasets: D1, D4	
<b>Data description</b>	<p><b>Types and formats:</b> office documents, pdf (P&amp;ID, isometric drawings, plants, 3D views, items specification tables, exploded views, components list), text documents, images, python files, mex files. In addition, eventual codes, functions, algorithms.</p> <p><b>Purpose:</b> the purpose of the produced data is to lead the project to a fully functioning lab-scale prototype;</p> <p><b>Re-use:</b> Possibly yes, we have been referencing to existing literature sources in R&amp;D activities;</p> <p><b>Origin:</b> Any reliable source</p> <p><b>Expected size:</b> From few KB to some GB</p> <p><b>Utility:</b> Surely to other consortium partners (especially those involved in technical development). In future, also to potential plant users.</p>
<b>WPs &amp; Tasks</b>	All tasks, in particular WP2, WP3 and WP4
<b>Responsible partner(s)</b>	TIFEO
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	.pdf, .jpeg, .png, .odp, .ppt, .ods, .xlsx, .txt, .odt, .pdf, .mp3, .mp4, .mexw64, .py
<b>Standards for any related metadata</b>	Scripts, spreadsheets
<b>Specific tools/software (accessibility and interoperability)</b>	Open Office, Phyton, Pdf readers
<b>Accessibility</b>	All data related to the consortium deliverables will be available. The data needed for the success of the project will be shared with other partners. On the other hand, data covered by patent will be kept private
<b>Deposition of data and associated metadata</b>	Internal consortium platform
<b>With whom is the data shared during the project?</b>	Some data will be accessible to the entire consortium, other data will be private for Tifeo itself.
<b>Data licensing</b>	-
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. Not defined yet.
<b>Ethical and/or legal requirements</b>	-
<b>Provisions for data security</b>	We are using common online folders and personal physical storage

Table 10: Datasets from TIFEO

NTUA - Datasets: D5, D6, D13	
<b>Data description</b>	<p><b>Purpose:</b> Data that will be generated/processed and collected within our Tasks involves the development of a design script for the two-phase expanders and the respective data results of the analysis (Task 3.1). In Task 3.2, additional experimental data will be collected and processed to calibrate the model from Task 3.1. Task 3.3 will process the experimental data and generate additional design data to provide and share with the involved partners. Finally, additional data will be extracted from the CFD (Task 3.4) and the testing of the expander prototypes (Task 3.6). Finally, data will be collected by manufacturing companies with respect to the CBA and processed to generate the LCA/LCC/s-LCA in Task 7.2.</p> <p><b>Re-use:</b> Data will be re-used in terms of development of project-related publications. Data generated by the results of our analysis will be re-used in the future for expansion of work -in particular for the LCA/LCC inventories- , without sharing/publishing any of the collected within the project data.</p> <p><b>Expected size:</b> Expected size of data cannot be estimated at this point.</p> <p><b>Origin:</b> Within WP3, data will be generated by the experiments on the test rig and the executions of the developed code. With respect to WP7, data will be collected by partners involved in manufacturing and assumptions from literature. LCA/LCC/s-LCA results will be generated via use of dedicated software.</p> <p><b>Utility:</b> Data from the experiments will be useful to the expander related parties and companies. LCA/LCC/s-LCA results are useful to anyone related with environmental studies</p>
<b>WPs &amp; Tasks</b>	<p>Task 3.1 data generated with respect to the design model. Task 3.2 data collected from the experiments of the two-phase commercial expander and processed to update the design model and finally generate updated data for the design of the expander prototype along with standard design procedures in Task 3.3. Task 3.4 will generate CFD results on the designed prototype and provide potential guidelines for improvement. Finally, Task 3.6 will generate experimental data on the two-phase expander prototype. With respect to Task 7.2, data will be collected by, involved in the REGEN-By-2 prototype, partners and the LCA/LCC/s-LCA results will be generated.</p>
<b>Responsible partner(s)</b>	<p>With respect to data of Tasks 3.1, 3.2 and 3.3 NTUA, ULIEGE, EXOES and TIFEO. Task 3.4 involves LSTME, UNIPI, RINA and NTUA. Task 3.6 NTUA, CARTIF, EXOES and TIFEO, while Task 7.2 involves RINA and NTUA and JER.</p>
<b>Dissemination level</b>	<p>OPEN - Data that is shared for re-use or that underpins a scientific publication.</p>
<b>Data standards</b>	<p>With respect to Task 3.1 the script is developed in source code data standard for Python. The results (Tasks 3.1, 3.3 and 3.4) will be available and analyzed in spreadsheets (.xlsx). Currently is not definitive which type of data standard will be used for the experiments, most probable is the use of .txt files that will be processed (Task 3.2, Task 3.6). With respect to Task 7.2 , results will be generated in .xlsx files as is the only output from the respective software. In this perspective, all data is in commonly used data standards and therefore no additional action is needed to make the data interoperable.</p>
<b>Standards for any related metadata</b>	<p>At this point, no planning for development of metadata has been conducted by NTUA within the works of the project.</p>

<p><b>Specific tools/software (access. and interop.)</b></p>	<p>The design scripts are developed in Python software and therefore are attributed as FAIR. Experimental results will most probably be collected, as already mentioned above, in .txt format hence there is no need for dedicated software. However, the LCA analysis (Task 7.2) will be conducted in Simapro software, which is a licensed software and therefore not accessible by everyone. Both the input data as well as the output will be in .xlsx files. These files can be also saved in ods format to ensure interoperability. However, as this data is confidential and only the results will be communicated by means of a publication, it is considered that interoperability is ensured.</p>
<p><b>Accessibility</b></p>	<p>Data related to WP3 will be shared among partners, but apart from the already defined in the CA publications is not expected to be made publicly available, unless otherwise decided by the consortium. Data from Task 7.2 will be used to generate a publication as stated in the CA, however, without any communication of the inventory. Conducted publications can be stored in the involved NTUA laboratory's website.</p>
<p><b>Deposition of data and associated metadata</b></p>	<p>Conducted publications can be stored in the involved NTUA laboratory's website. Data can also be stored at NTUA's official repository.</p>
<p><b>With whom is the data shared during the project?</b></p>	<p>Data, apart from the publications, will not be shared publicly within the contents of NTUA's work. With respect to Task 7.2, data will be shared by manufacturing partners to NTUA for the development of the inventories. For Task 3.1 and 3.2, data will be shared with TIFEO, EXOES and ULIEGE. Task 3.3 data will be shared with TIFEO and EXOES, while in Task 3.5 interaction will be between NTUA and EXOES. Finally, in Task 3.6 data will be shared with EXOES, TIFEO and CARTIF.</p>
<p><b>Data licensing</b></p>	<p>At this point cannot be answered.</p>
<p><b>Preservation after the end of the project? How long? Who is responsible?</b></p>	<p>Yes. Data will only be preserved in terms of repositories and publications. The uploading in repositories of the university is free of charge and therefore the duration can be agreed. Use of the repository from UNIFI and therefore NTUA is not responsible.</p>
<p><b>Ethical and/or legal requirements</b></p>	<p>No such issues apply in this case.</p>
<p><b>Provisions for data security</b></p>	<p>Data developed so far is stored in certified by the NTUA repository, that is available under credentials login only relevant to the project members of the involved NTUA laboratory. All the MoMs and the reports are stored in the UNIFI repository.</p>

Table 11: Datasets from NTUA

ULIEGE - Datasets: D6, D8	
<b>Data description</b>	<p><b>Types and formats:</b> Source codes (simulation model of compressor). Reporting state of the art and modelling, data from measurements.</p> <p><b>Purpose:</b> Simulation model used to design two-phase scroll compressors.</p> <p><b>Re-use:</b> Simulation code based on PDSim model (published in open-source), but significantly adapted/improved/completed.</p> <p><b>Origin:</b> code developed by researchers in ULiege/NTUA.</p> <p><b>Expected size:</b> Size negligible.</p> <p><b>Utility:</b> Useful for ULiege, NTUA, Exoes. Useful to other research groups if published in open-source during or after the project. To the next WP to build the final machine.</p>
<b>WPs &amp; Tasks</b>	WP4
<b>Responsible partner(s)</b>	ULIEGE
<b>Dissemination level</b>	OPEN - Data that is shared for re-use or that underpins a scientific publication.
<b>Data standards</b>	<p>Simulation code of compressor developed in Python/Matlab. The use of Python and Matlab allows for easy sharing of the code with other partners or other research groups. Python is open and can be freely used.</p> <p>The modifications of PDSim simulation tool will respect the conventions regarding the way to write thermodynamic/mechanical variables. That should also ease the interoperability with other developers.</p>
<b>Standards for any related metadata</b>	No standard used for any related metadata.
<b>Specific tools/software (access. and interoperab.)</b>	Python, Matlab, CoolProp, GitHub
<b>Accessibility</b>	The source code will be made accessible. However, some values of parameters in the code could be not shared in order to protect the IP of Exoes. The modeling of some specific technical features of the compressor could also be not shared in order to protect the IP of partners (if patents are submitted).
<b>Deposition of data and associated metadata</b>	Probably GitHub
<b>With whom is the data shared during the project?</b>	All partners from WP4. Scientific collaborators of ULiege. Visiting scholars in ULiege working on the project?
<b>Data licensing</b>	Open Source code, Public domain (scientific papers, PhD manuscripts), except developed technical IP that could be patented (to be decided with other partners)
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. 5 years. ULIEGE. Preservation in ULiege cloud. Costs covered by our institution.
<b>Ethical and/or legal requirements</b>	No specific ethical aspects.
<b>Provisions for data security</b>	Use of ULiege cloud ( <a href="https://dox.uliege.be">https://dox.uliege.be</a> )

Table 12: Datasets from ULIEGE

UNE - Datasets: D14	
<b>Data description</b>	Datasets related to standardization landscape and applicable standards
<b>WPs &amp; Tasks</b>	WP7 - T7.5
<b>Responsible partner(s)</b>	UNE
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	UNE will use excel sheets, as this is the common format to deliver the results in the databases to be used. Those files could be potentially used by other researchers, though there is a little interest foreseen on the topic and some of the files will be generated from public databases.
<b>Standards for any related metadata</b>	<p>Usually metadata for standards corresponds to "fields" in a normal database. Examples of metadata are:</p> <pre> &lt;Noun&gt;UNE 112011:2011&lt;/Noun&gt; &lt;Code&gt;N0047865&lt;/Code&gt; &lt;Committee Code="CTN 112" Noun="CORROSION AND PROTECTION AGAINST CORROSION OF METALLIC MATERIALS" /&gt; &lt;Status&gt;Published&lt;/Status&gt; &lt;Publication_Date&gt;2011-09-07&lt;/Publication_Date&gt; &lt;Withdrawal_Date /&gt; &lt;Title_EN&gt;Corrosion of concrete reinforcement steel. Determination of the carbonatation dept for in-service concrete.&lt;/Title_EN&gt; - &lt;ICSS&gt; &lt;ICS code="91.080.40" noun="Concrete structures" /&gt; &lt;ICS code="77.060" noun="Corrosion of metals" /&gt; &lt;ICS code="77.140.15" noun="Steels for reinforcement of concrete" /&gt; &lt;/ICSS&gt; - &lt;Cancellations&gt; &lt;Cancellation code="N0008029" noun="UNE 112011:1994" action="Cancel" /&gt; &lt;/Cancellations&gt; &lt;InternationalRelations /&gt; - &lt;Reaffirmations&gt; </pre>
<b>Specific tools/software (access. and interoperab.)</b>	Basically Excel sheets.
<b>Accessibility</b>	<p>Most data will be generated by filtering public databases, e.g. standards under development in a certain European Technical Committee, or standards published on safety requirements for boilers of a certain power, standards cited in pressure equipment Directive...</p> <p>Based on those data and combining the results from different searches and research, a deliverable will be developed and distributed amongst the participants. Besides that, there is no intention to make those data accessible.</p>
<b>Deposition of data and associated metadata</b>	No deposit foreseen

<b>With whom is the data shared during the project?</b>	The reports derived from data are shared to all the partners. Data themselves are not shared.
<b>Data licensing</b>	NA
<b>Preservation after the end of the project? How long? Who is responsible?</b>	No. NA.
<b>Ethical and/or legal requirements</b>	No ethical or legal issues
<b>Provisions for data security</b>	Excel sheets generated whilst preparing the report will be stored in a PC and in a server in the cloud.

*Table 13: Datasets from UNE*

EXOES - Datasets: D7	
<b>Data description</b>	Prototype CAD files
<b>WPs &amp; Tasks</b>	WP3: T3.3, T3.5. WP4: T4.3, T4.5.
<b>Responsible partner(s)</b>	EXOES, NTUA and ULIEGE
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	The data produced is prototype CAD files
<b>Standards for any related metadata</b>	No metadata is used
<b>Specific tools/software (access. and interoperab.)</b>	CAD software Solidworks, it respects interoperability between CAD softwares with STEP files
<b>Accessibility</b>	Not accessible
<b>Deposition of data and associated metadata</b>	Nowhere, the CAD files stay at Exoes.
<b>With whom is the data shared during the project?</b>	Part of the data will be share with NTUA, ULIEGE, LSTME and TIFEO
<b>Data licensing</b>	The data can not be re-used because STEP files can not be modified.
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. 20 years. EXOES. Preserved In Exoes internal servers. No costs expected.
<b>Ethical and/or legal requirements</b>	No
<b>Provisions for data security</b>	Our internal server is protected and secured.

*Table 14: Datasets from EXOES*



LSTME - Datasets: D6	
<b>Data description</b>	<p><b>Types and formats:</b> datasets (Ascii file), data contents(Plots, and videos).</p> <p><b>Purpose:</b> Numerical Validation for the WorkPackage(WP) 3 and 4.</p> <p><b>Re-use:</b> Not sure, we will need to use the experimental results from WP 5 and 6 for comparison purpose.</p> <p><b>Origin:</b> The data from the postprocessing of numerical simulations, ascii file.</p> <p><b>Expected size:</b> From the previous experience, for one case is around 1 TB, it depends on the cases we simulate.</p> <p><b>Utility:</b> for the validation of the experimental results in the REGEN-BY-2 project (WP5 and 6) and also for the publications for the project.</p>
<b>WPs &amp; Tasks</b>	We will perform the numerical simulation in WP3 and 4, the corresponding data will be collected and processed in these two packages.
<b>Responsible partner(s)</b>	When we get the results, we need to provide all our results to the Lead of WP 3 and 4, i.e., NTUA and ULIEGE.
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	The data will be exchanged and re-use inside of the project. Standards: the ascii file, images, and videos. LSTME will use standard vocabularies to allow inter-disciplinary interoperability. LSTME can provide mappings to more commonly used ontologies.
<b>Standards for any related metadata</b>	The data in the project will be post-processed into plots and videos for validation, if our internal partners required for the metadata, we also would like to provide. But our results are mainly used for validation of the experimental data.
<b>Specific tools/software (access. and interoperab.)</b>	We will mainly use python to postprocess the data. Python is open source code, so the corresponding code is possible to be provided.
<b>Accessibility</b>	For this we will follow the instruction of the leaders of WP 3 and 4, the data will be open for sure between the researchers in WP3 and 4. And for the validation of model tests, we will follow the instruction of the WP3 and 4 to provide the results for WP 5 and 6 (model tests).
<b>Deposition of data and associated metadata</b>	The platform will be the same as other research groups in the WP 3 and 4.
<b>With whom is the data shared during the project?</b>	Internal partners with WP 3, 4, 5, and 6.
<b>Data licensing</b>	Open source. When: we need to discuss with our internal partners. The data from the numerical results is mainly used for the validation, therefore it would be internal use.
<b>Preservation after the end of the project? How long? Who is responsible?</b>	It depends on the decision of WP 3 and 4. Responsibility of the leader of WP3 and 4 and LSTME BB. Initially, the data will be sent to our partners in REGEN-BY-2 and also in LSTME BUSAN BRANCH. Our financial supporters will cover these costs.
<b>Ethical and/or legal requ.</b>	No
<b>Provisions for security</b>	We will discuss with our internal partners of REGEN-BY-2.

Table 15: Datasets from LSTME

JER - Datasets: D13	
<b>Data description</b>	Data of cost benefit analysis of REGEN-BY-2 technology.
<b>WPs &amp; Tasks</b>	WP7 - T7.2
<b>Responsible partner(s)</b>	JER, NTUA
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	Spreadsheets
<b>Standards for any related metadata</b>	Spreadsheets
<b>Specific tools/software (access. and interoperab.)</b>	Microsoft Excel
<b>Accessibility</b>	Project file sharing platform
<b>Deposition of data and associated metadata</b>	Project file sharing platform
<b>With whom is the data shared during the project?</b>	NTUA, RINA-C, R2M, CARTIF, UNE, LSTME, TIFEO
<b>Data licensing</b>	n/a
<b>Preservation after the end of the project? How long? Who is responsible?</b>	Yes. 5 years. Responsibility of JER, NTUA, RINA-C. On the server. No costs.
<b>Ethical and/or legal requirements</b>	No
<b>Provisions for data security</b>	GDPR

*Table 16: Datasets from JER*

HYSYTECH - Datasets: D9	
<b>Data description</b>	The data formats produced are: .xlsx, .docx, .pdf. The data is gathered for the engineering and development of the prototype. The data will be developed during the project by Hysytech in collaboration with the relevant partners. The final size of the dataset will be in the neighbourhood of hundreds of MB. This data will be useful for the consortium.
<b>WPs &amp; Tasks</b>	WP5 - Task 5.2
<b>Responsible partner(s)</b>	HYSYTECH
<b>Dissemination level</b>	CONSORTIUM - Confidential data that is accessible to all partners, but retained within the consortium and subject to the project NDA.
<b>Data standards</b>	Hysytech will collect and document the data using easily usable standards (i.e. .pdf, .xlsx, .docx) for the datasets to ensure that those can be correctly understood, interpreted and reused.
<b>Standards for any related metadata</b>	An example for the naming of the dataset could be: aaaa.mm.dd_Regen-by-2_Dataset short name_versionX
<b>Specific tools/software (access. and interoperab.)</b>	Microsoft office suite. Eventual engineering documentation will be shared as .pdf files to make it accessible.
<b>Accessibility</b>	The dataset created by Hysytech will be reported in deliverable D5.1 and will be available to the consortium. Those datasets will regard block flow diagram, process flow diagram, P&ID, mass balance and plant layout.
<b>Deposition of data and associated metadata</b>	The datasets are firstly stored in Hysytech's servers and then will be made available in the sharepoint of the consortium.
<b>With whom is the data shared during the project?</b>	Only with the consortium.
<b>Data licensing</b>	The data will be confidential within the consortium. The re-use after the project must be decided with the involved parties during the project.
<b>Preservation after the end of the project? How long? Who is responsible?</b>	No
<b>Ethical and/or legal requirements</b>	No
<b>Provisions for data security</b>	Hysytech servers are daily backuped in order to preserve the data if anything goes wrong.

Table 17: Datasets from HYSYTECH