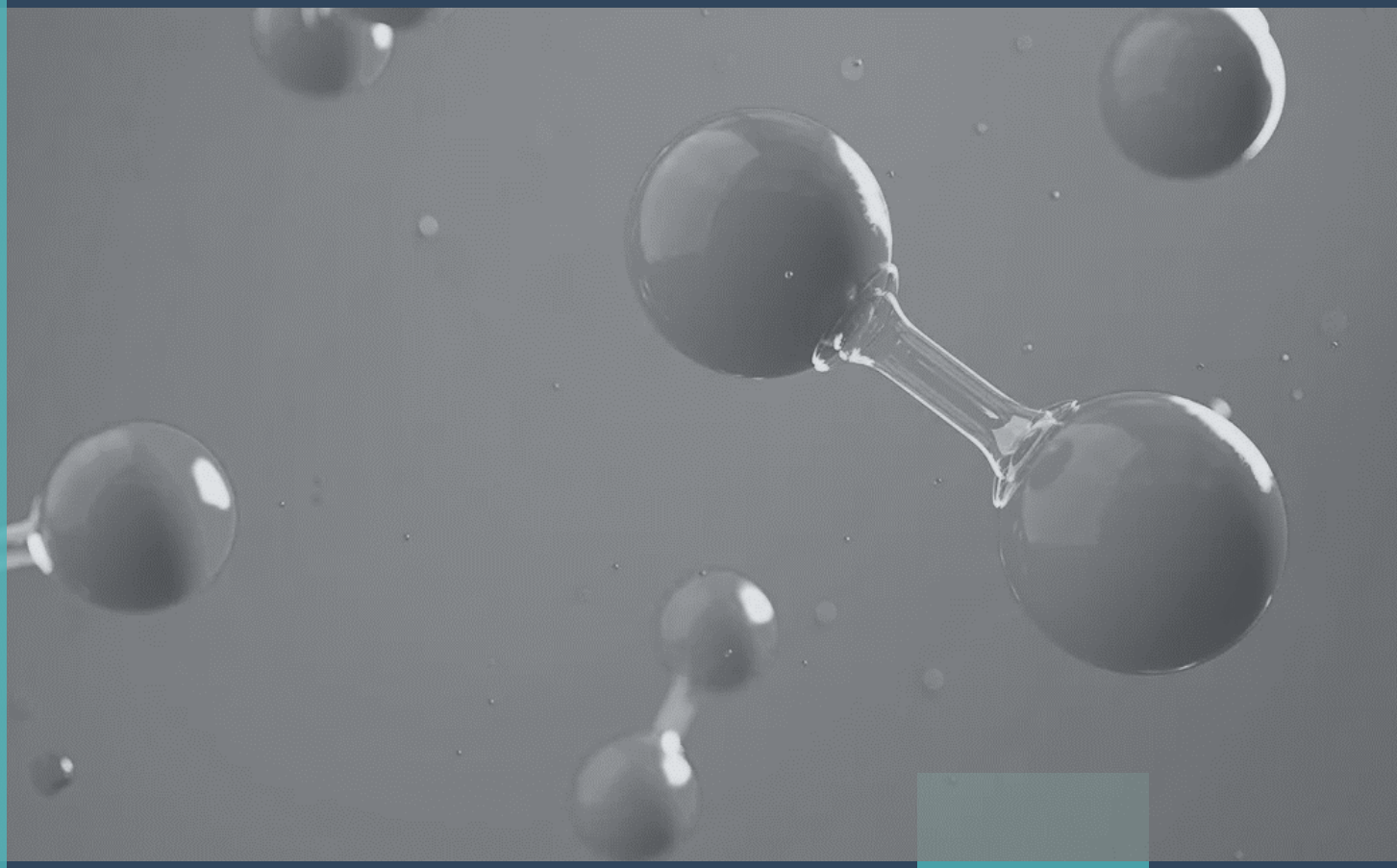
MACGHYVER



2023-02-22

D 1.2

DATA MANAGEMENT PLAN

WP 1 MANAGEMENT AND COORDINATION OF THE PROJECT

This project has received funding from the European Innovation Council (EIC) under grant agreement No 101069981. The EIC receives support from the European Union's Horizon Europe Research and Innovation Program

Project	Microfluidic Wastewater Treatment and Creation of Green Hydrogen Via Electrochemical Reactions (MACGHYVER)
Grant number	101069981
Author(s)/Organisation(s)	Wei Zhao and Abhilash Venkateshaiah (EDEN)
Work Package	WP 1 Management and coordination of the project
Delivery Date (DoA)	23.02.2023
Actual Delivery Date	28.02.2023
Abstract:	This deliverable contains the first iteration of the data management plan developed for the MACGHYVER project. It provides an overview of the main elements of the data management strategy and policy used for all the data sets generated during the project. This deliverable describes how MACGHYVER data is gathered, generated, processed, and shared. Furthermore, it highlights the established methods and procedures for data storage and dissemination.

Document Revision History			
Date	**Version**	**Author/Contributor/ Reviewer**	**Summary of main changes**
27.01.2023	0.1	Abhilash Venkateshaiah (EDEN)	Draft version 1
03.02.2023	0.2	Wei Zhao (EDEN)	Data Summary, Draft Version 2
06.02.2023	0.3	Abhilash Venkateshaiah (EDEN)	General review and editing
07.02.2023	0.4	Wei Zhao (EDEN)	Quality control
20.02.2023	0.5	All Partners	Quality control
22.02.2023	vF	Abhilash Venkateshaiah & Wei Zhao (EDEN)	Final review and editing

Dissemination Level		
PU	Public	☑
CO	Confidential, only for members of the consortium (including the EC)	□
PP	Restricted to other programme participants (including the EC Services)	□
RE	Restricted to a group specified by the consortium (including the EC Services)	□

MACGHYVER Consortium			
Participant Number	**Participant organisation name**	**Short name**	**Country**
1	Eden Tech	EDEN	France
2	Universidad De Castilla - La Mancha	UCLM	Spain
3	Technische Universiteit Delft	DELFT	Netherlands
4	Gottfried Wilhelm Leibniz Universitaet Hannover	LUH	Germany
5	Technische Universitat Clausthal	TUC	Germany
6	Politechnika Poznanska	POZ	Poland

Contents

1. Introduction 4
2. Data Summary 4
   2.1. MACGHYVER types of data 5
      2.1.1. Type 1 6
      2.1.2. Type 2 7
      2.1.3. Type 3 9
      2.1.4. Type 4 10
      2.1.5. Type 5 12
      2.1.6. Type 6 13
      2.1.7. Type 7 14
   2.2. Public Deliverables 16
   2.3. Data Management Strategy 16
      2.3.1. Making Data Findable, Including Provisions for Metadata (F) 16
      2.3.2. Making Data Openly Accessible (A) 17
      2.3.3. Making data interoperable (I) 17
      2.3.4. Ensuring data reuse (R) 18
      2.3.5. Data quality and review 18
3. Allocation of Resources 18
4. Data Security 18
5. Ethical aspects 19

1. Introduction

The present deliverable D1.2 describes the first iteration of the data management plan (DMP) for the MACGHYVER project, funded by the European Innovation Council (EIC) under grant agreement number 101069981. The work described here was carried out in the framework of WP1 Management and coordination of the project. The document briefs on the data management strategy that the partners will implement in relation to the datasets produced as part of the project. The DMP, in particular, describes the primary data that will be produced inside MACGHYVER, outlines how research data will be handled throughout the project, and specifies how and which portions of the data sets will be made publicly available.

This document is designed for the consortium's internal use and aims to offer recommendations to project partners on data management. The DMP enables the creation of common practices and understanding concerning the processing of data within the MACGHYVER project. In order to make research data more easily Findable, Accessible, Interoperable, and Reusable (FAIR) the DMP seeks to define the management plan for data collected throughout the project.

The DMP is a dynamic document which will be updated as the project progresses and upon significant changes.

2. Data Summary

The MACGHYVER DMP aims to establish a plan for managing critical data created and gathered during the project, as well as to maximize access to and re-use of scientific data. The DMP will be reviewed and modified on a regular basis because it is designed to be a "living" document that describes how the MACGHYVER research data will be handled both during and after the project.

Every EU-funded project is required to work on information dissemination with the goal of reaching a broader audience and researchers. MACGHYVER aims at establishing best practices with regard to Open Science by following the recommendations of the "Open Research Data Pilot" (ORDP, for Horizon 2020) and "Open Research Data and Data Management Plans" (for ERC grants). The data generated within the project will be assessed and classified as open, embargoed, or restricted by the partners who generated the content or the data set. All the generated data, irrespective of its classification, will be stored in each

of the partners' databases and in the highly secure Dropbox cloud service folder, which has been created as the internal database of the partners. Additionally, the data sets classified as open or embargoed (after the embargo period is over) will be publicly shared through the project website and in a trusted repository provided by DELFT (https://repository.tudelft.nl/). The repository provides open access to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC-BY). The exception will be made for data that would be against the beneficiary's legitimate interests, including regarding commercial exploitation; be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations.

MACGHYVER will ensure that open access will be provided to peer-reviewed scientific publications relating to the project-generated results. Furthermore, by the time of publication, MACGHYVER will ensure that a machine-readable electronic copy of the published version, or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications for immediate open access, under the most recent available version of the Creative Commons Attribution International Public License (CC-BY) or an equivalent license. We will also ensure that information is given via the repository about any research output, or any other tools and instruments needed to replicate/validate the conclusions of the scientific publication. The metadata of the deposited publication will be under a Creative Common Public Domain Dedication (CC 0) or equivalent.

2.1. MACGHYVER types of data

During the course of its R&D activities, MacGyver will generate a vast amount of data. This section will provide an overview of the many types of datasets that MACGHYVER consortium partners will develop, as well as the structure of their content. This list may be modified as the project develops in subsequent iterations of the DMP, based on project developments. The data will be stored in the following formats: documents, images, videos, numerical codes and other forms of data.

The information will be separated into the following categories for convenience:

TABLE 1 SUMMARY OF DATA TYPES AND THEIR RELATED PARTNERS AND WPS

Data type	Main partners	WPs
Type 1	POZ, TUC	WP2, WP6

Type 2	EDEN, TUC, DELFT	WP2, WP3, WP6
Type 3	LUH, EDEN	WP2, WP4, WP6
Type 4	POZ, EDEN	WP2, WP3, WP6
Type 5	LUH, DELFT, EDEN	WP3, WP4, WP5, WP6
Type 6	UCLM, EDEN	WP5
Type 7	EDEN	WP1, WP7, WP8

2.1.1. Type 1

Datasets concerning an exhaustive list of non-Critical Raw Materials (Non-CRMs) for the electrodes, can include:

- Raw data of potential electrode candidates for the Microfluidic Electrolyzer (ME).
- Raw data of potential electrode candidates for the Electrochemical Hydrogen Compressor (EHC).
- Raw data of potential electrode candidates for the catalytical hydrogen peroxide production
- Raw data and metadata generated and collected during the material characterization.

TABLE 2 DATASETS TYPE 1 CONCERNING AN EXHAUSTIVE LIST OF NON-CRITICAL RAW MATERIALS (NON-CRMS) FOR THE ELECTRODES

Type 1	
Purpose and relation to the objectives of the projects	The datasets include information for the priority list of non-CRM electrodes benchmarked by the market overview. The datasets also contain the requirement list followed by the electrochemical characterizations that will deliver a list of the best-performing electrodes in alkaline potassium hydroxide solutions, electrodes for the EHC, and catalytical electrodes for hydrogen peroxide production. Depending on other parameters in the requirement list, especially the amount of critical raw materials, a prioritized list of different electrode materials will evolve.
Data Types	Documents, images
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml.) Images: JPEG (.jpeg, .jpg, .jp2) if originally created in this format, RAW image format (.raw), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .svg, .psd, .eps, .webp)
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project

Data production methods	The dataset will be generated collectively by partner laboratories through experimental trials, measurements, and bibliographical and marketing studies. The dataset will also include summaries of project meeting minutes between partners, and relevant publications in scientific journals.
Expected size of the data	500 GB – 1 TB
Data utility	The collected data will be used for further electrolyzer development in WP3, for EHC development which will result in an appropriate size and number of membrane electrode assemblies (MEA) in WP4, and for the micropollutant removal process from wastewater in WP2.
Potential for reuse	In addition to the project itself, the dataset will be useful for other projects inside the EIC Green Hydrogen portfolio or other research groups working on the related subjects
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme. University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.1.2. Type 2

Datasets concerning the Design of the ME, can include:

- Computer-Aided Design (CAD) files of the ME device's design together with its parameters and descriptions.
- Numerical and custom code implemented to perform the simulation.
- Numerical and custom code implemented to elaborate and post-treat acquired data.
- Raw data and metadata generated and collected during the ME device's design and its main characteristics.
- Raw data collected during the implemented experiment.

TABLE 3 DATASETS TYPE 2 CONCERNING THE DESIGN OF THE ME

Type 2	
Purpose and relation to the objectives of the projects	The datasets contain technical information and detailed characteristics of the core design of the ME as well as its operational testing results, which will include: • The interconnection of thousands of microchannels on a single stackable CD, where the electrolysis occurs. • The detailed smart architecture in each microchannel for the sake of a membraneless electrolyzer without the risk of gas crossover • The optimization of the network on a single stackable CD to ensure maximal coverage of the CD surface with

	microchannels while minimizing pressure and ohmic losses. • The design and optimization of electrodes network by adapting to the microfluidic network, while minimizing the total electrical resistance. • The production methods and parameters, the chosen materials for both microfluidic and electrodes' network, which will be compatible for mass production and operation conditions. • The experimental testing metadata for the operational electrolyzer and its characteristics (gas production, energy use, electrolyte flow rate and usage). These datasets will be crucial and the guidance for the final optimal prototype for the use of the generation of green hydrogen.
Data Types	Documents, Images, Videos, Data, and Numerical codes
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml) Images: JPEG (.jpeg, .jpg, .jp2) if original created in this format, RAW image format (.raw), Photoshop files (.psd), CAD formats (.dwg, .stp, .igs, .dxf, .svg), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .psd, .eps, .webp), Videos: MPEG-4 (.mp4), OGG video (.ogv, .ogg), motion JPEG 2000 (.mj2), AVCHD video (.avchd). Data: text format tables that are readable by common analysis software, or encrypted for specific data treatment software (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .txt, .rtf) Numerical codes: Matlab/Octave (.m, .mat), COMSOL Multiphysics (.mph), programming languages such as Python, Fortran, C/C++, etc.
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project
Data production methods	The dataset will be generated collectively by partner laboratories through experimental trials, measurements, and bibliographical and marketing studies. The dataset will also include summaries of project meeting minutes between partners, and relevant publications in scientific journals.
Expected size of the data	500 GB – 1 TB
Data utility	The collected datasets will be used for the prototype realization and the generation of several intellectual patents (IP) and high impact factor scientific articles. It will also be

	used for the life cycle sustainability (LCA, LCC) of the overall production, according to WP6.
Potential for reuse	As no existing approaches to current proposed solution, in addition to the project itself, the datasets will be useful for other projects inside the EIC Green Hydrogen portfolio or other research groups working on the related subjects
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme. University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.1.3. Type 3

Datasets concerning the Design of EHC, can include:

- Computer-Aided Design (CAD) files of the EHC device's setup together with its parameters and description.
- Numerical and custom code implemented to perform the simulation.
- Numerical and custom code implemented to elaborate and post-treat acquired data.
- Raw data and metadata generated and collected during the EHC device's setup and its main characteristics.
- Raw data collected during the implemented experiment.

TABLE 4 DATASETS TYPE 3 CONCERNING THE DESIGN OF EHC

Type 3	
Purpose and relation to the objectives of the projects	The datasets contain an electrochemical model for fluid and energy description as well as the optimization part for the improvement of the hardware design as sealings and flow field design. It will also create a database of current commercialized solutions and their performance will be compiled. These datasets will help to build a calculator that permits comparison between the current solution and the MACGHYVER solution to instantly provide energy saving with the MACGHYVER solution.
Data Types	Documents, Images, Videos, Data, and Numerical codes
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml) Images: JPEG (.jpeg, .jpg, .jp2) if original created in this format, RAW image format (.raw), Photoshop files (.psd), CAD formats (.dwg, .stp, .igs, .dxf, .svg), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .psd, .eps, .webp), Videos: MPEG-4 (.mp4), OGG video (.ogv, .ogg), motion JPEG 2000 (.mj2), AVCHD video (.avchd).

	Data: text format tables that are readable by common analysis software, or encrypted for specific data treatment software (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .txt, .rtf) Numerical codes: Matlab/Octave (.m, .mat), COMSOL Multiphysics (.mph), programming languages such as Python, Fortran, C/C++, etc.
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project
Data production methods	The dataset will be generated collectively by partner laboratories through experimental trials, measurements, and bibliographical and marketing studies. The dataset will also include summaries of project meeting minutes between partners, and relevant publications in scientific journals.
Expected size of the data	500 GB – 1 TB
Data utility	The collected data will be used for the hydrogen compression system to facilitate the separation and collection of hydrogen and oxygen gases. These datasets will also be used to complete the calculation of the life cycle sustainability (LCA, LCC) of the overall production, in order to evaluate the efficiency of the system solution, according to WP6.
Potential for reuse	In addition to the project itself, the dataset will be useful for other projects inside the EIC Green Hydrogen portfolio or other research groups working on the related subjects
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme. University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.1.4. Type 4

Datasets concerning the setup for hydrogen peroxide production and micropollutant treatment, can include:

- Computer-Aided Design (CAD) files of the hydrogen peroxide production setup's design together with its parameters and descriptions.
- Numerical and custom code implemented to elaborate and post-treat acquired data.
- Raw data collected during the implemented experiment.

TABLE 5 DATASETS TYPE 4 CONCERNING THE SETUP FOR HYDROGEN PEROXIDE PRODUCTION AND MICROPOLLUTANT TREATMENT

Type 4

Purpose and relation to the objectives of the projects	The datasets contain the experimental setup's characteristic parameters and hydrogen peroxide production analysis, according to the chosen electrode for anode in WP2.
Data Types	Documents, Images, Videos, Data, and Numerical codes
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml) Images: JPEG (.jpeg, .jpg, .jp2) if original created in this format, RAW image format (.raw), Photoshop files (.psd), CAD formats (.dwg, .stp, .igs, .dxf, .svg), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .psd, .eps, .webp), Videos: MPEG-4 (.mp4), OGG video (.ogv, .ogg), motion JPEG 2000 (.mj2), AVCHD video (.avchd). Data: text format tables that are readable by common analysis software, or encrypted for specific data treatment software (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .txt, .rtf) Numerical codes: Matlab/Octave (.m, .mat), COMSOL Multiphysics (.mph), programming languages such as Python, Fortran, C/C++, etc.
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project
Data production methods	The dataset will be generated collectively by partner laboratories through experimental trials, measurements, and bibliographical and marketing studies. The dataset will also include summaries of project meeting minutes between partners, and relevant publications in scientific journals.
Expected size of the data	500 GB – 1 TB
Data utility	The outcome of these datasets will be a concept that defines the type of wastewater feeds that are applicable for water electrolysis and which electrodes perform best under these circumstances in terms of efficiency and pollutant removal, which are related with the environmental impact of the system and hence with the WP6.
Potential for reuse	In addition to the project itself, the dataset will be useful for other research groups working on the related subjects, such as micropollutant removal process in wastewater treatment applications.
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme.

	University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.1.5. Type 5

Datasets concerning the integration of the entire hydrogen generation-separation-storage system, can include:

- Computer-Aided Design (CAD) files of the setup structure together with its description.
- Numerical and custom code implemented to perform the simulation.
- Numerical and custom code implemented to elaborate and post-treat acquired data.
- Raw data and metadata generated and collected for the main characteristics of the setup.
- Raw data collected during the implemented experiment.

TABLE 6 DATASETS TYPE 5 CONCERNING THE INTEGRATION OF THE ENTIRE HYDROGEN GENERATION-SEPARATION-STORAGE SYSTEM

Type 5	
Purpose and relation to the objectives of the projects	The datasets contain the model describing the total MACGHYVER system including EHC, ME device, and peripherals. The datasets also conclude the optimization approaches and operational management regarding the reliability, efficiency, and cost reduction.
Data Types	Documents, Images, Videos, Data, and Numerical codes
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml) Images: JPEG (.jpeg, .jpg, .jp2) if original created in this format, RAW image format (.raw), Photoshop files (.psd), CAD formats (.dwg, .stp, .igs, .dxf, .svg), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .psd, .eps, .webp), Videos: MPEG-4 (.mp4), OGG video (.ogv, .ogg), motion JPEG 2000 (.mj2), AVCHD video (.avchd). Data: text format tables that are readable by common analysis software, or encrypted for specific data treatment software (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .txt, .rtf) Numerical codes: Matlab/Octave (.m, .mat), COMSOL Multiphysics (.mph), programming languages such as Python, Fortran, C/C++, etc.
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project

Data production methods	The dataset will be generated collectively by partner laboratories through experimental trials, measurements, and bibliographical and marketing studies. The dataset will also include summaries of project meeting minutes between partners, and relevant publications in scientific journals.
Expected size of the data	500 GB – 1 TB
Data utility	The outcome datasets will be applied for establishing the overall energy cost, the amount of gas produced, electrolyte consumption and overall size. It will help to monitor the device performance characteristics and energy managements strategies in WP6. The ultimate data utility will be to validate the POC.
Potential for reuse	In addition to the project itself, the dataset will be useful for other projects inside the EIC Green Hydrogen portfolio or other research groups working on the related subjects
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme. University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.1.6. Type 6

Datasets concerning the Life Cycle Analysis (LCA) and Life Cycle Cost (LCC) calculations, can include:

- Raw data collected during the design describing all the applied raw materials.
- Numerical and custom code implemented to perform the calculation and analysis.
- Files containing the metadata of the parameters of the designed system.

TABLE 7 DATASETS TYPE 6 CONCERNING THE LIFE CYCLE ANALYSIS (LCA) AND LIFE CYCLE COST (LCC) CALCULATIONS

Type 6	
Purpose and relation to the objectives of the projects	The datasets contain information about mass and energy requirements, type of materials, and operation conditions collected from the different devices (electrolyzer, separator, compressor) that are part of the system for hydrogen production. LCA and LCC models will be applied to generate the sustainability characteristics of the POC system.
Data Types	Documents, Images, Data, and Numerical codes
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml)

	Images: JPEG (.jpeg, .jpg, .jp2) if original created in this format, RAW image format (.raw), Photoshop files (.psd), CAD formats (.dwg, .stp, .igs, .dxf, .svg), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .psd, .eps, .webp), Videos: MPEG-4 (.mp4), OGG video (.ogv, .ogg), motion JPEG 2000 (.mj2), AVCHD video (.avchd). Data: text format tables that are readable by common analysis software, or encrypted for specific data treatment software (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .txt, .rtf) Numerical codes: Matlab/Octave (.m, .mat), COMSOL Multiphysics (.mph), programming languages such as Python, Fortran, C/C++, etc.
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project
Data production methods	The dataset will be generated collectively by partner laboratories through experimental trials, measurements, and bibliographical and marketing studies. The dataset will also include summaries of project meeting minutes between partners, and relevant publications in scientific journals.
Expected size of the data	Less than 500 GB
Data utility	The dataset will be used for assessments of the sustainability of the project and to perform the rest of deliverables of the WP6, will also be the inventory for the sustainability and economic assessment of the global system. Social and sociological aspects (Health, Safety, socio-economic impact…) of the system for hydrogen production will be assessed by these datasets, along with the main environmental impacts such as carbon and water footprint, fresh water ecotoxicity, human toxicity, etc.
Potential for reuse	In addition to the project itself, the dataset will be useful for other projects inside the EIC Green Hydrogen portfolio or other research groups working on the related subjects
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme. University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.1.7. Type 7

Datasets concerning the hydrogen marketing study, can include:

- The acquired raw metadata (Documents, Images, Videos)
- Text files describing the marketing study.
- Analyzed and treated data for the main characteristics.

TABLE 8 DATASETS TYPE 7 CONCERNING THE HYDROGEN MARKETING STUDY

Type 7	
Purpose and relation to the objectives of the projects	The datasets contain the market study and relevant materials for hydrogen energy and strategies for the project transition to market
Data Types	Documents and Images
File formats	Documents: All common electronic document formats (.docx, .csv, .xls/xlsx, .tab, .por, .sav, .dta, .mdb/.accd, .dbf, .ods, .odt, .pdf, .txt, .rtf, .hml, .html, .xhtml.) Images: JPEG (.jpeg, .jpg, .jp2) if originally created in this format, RAW image format (.raw), other acceptable formats (.tif, .tiff, .png, .gif, .bmp, .svg, .psd, .eps, .webp)
Reuse of existing data	Processed and aggregated data will be shared by partners not collecting data for the advancement of the project
Data production methods	The dataset will be generated by the coordinator mainly and with the collective help from partners through public/private communications and marketing studies. The dataset will also include summaries of Portfolio meeting minutes between projects, partners, and relevant publications in scientific journals.
Expected size of the data	Less than 500 GB
Data utility	The dataset will be used for determining the key factors affecting the penetration of the proposed technologies in each market segment and the relative competitiveness of each solution in different end-user applications.
Potential for reuse	In addition to the project itself, the dataset will be useful for other projects inside the EIC Green Hydrogen portfolio or other research groups working on the related subjects
Diffusion principles	The dataset generated will be shared inside the consortium through the private, highly secured cloud service platform Dropbox. The consortium will determine which data shall be made publicly available to the Open Access Decision scheme. University public data repositories (DELFT) will be used along with open-access publications in scholarly journals

2.2. Public Deliverables

The MACGHYVER project has classified some of its outputs among the project's datasets and deliverables as "PU" and these will be made available to the public as "Open Access" without any charge. These outputs include:

- Open access articles published in scientific journals;
- Conference and workshop abstracts or articles;
- Project deliverables D2.1, D2.2, D2.3, D3.2, D4.1, D4.2, D4.4, D5.1, D5.2, D6.4.

TABLE 9 SUMMARY OF PUBLIC DELIVERABLES

Deliverable	WP	Lead Partner	Type
D2.1 Priority list of non-CRM electrodes	2	TUC	Data
D2.2 Membrane electrode assembly for electrochemical hydrogen compressor (EHC) in WP4	2	TUC	Data
D2.3 Concept: wastewater in electrolysis	2	TUC	Report
D4.1 Model for design of EHC ready	4	LUH	Other
D4.2 Final design and dimensioning of EHC	4	LUH	Demonstrator
D4.4 Separation solution established	4	DELFT	Demonstrator
D5.1 System model is ready to use, energy management strategies fixed	5	LUH	Demonstrator
D5.2 Optimal system design and operation strategy is fixed	5	LUH	Other
D6.4 Social Life Cycle Assessment	6	UCLM	Report

2.3. Data Management Strategy

MacGyver's approach to data management is centered on following FAIR data management guidelines.

2.3.1. Making Data Findable, Including Provisions for Metadata (F)

A naming convention is established for the files and folders within data repositories, which consists of project name, document type, document name, version, author, and institution. For Example: *MACGHYVER_DX.x_Document1_Author's initials (institution)*.

The project data is made searchable using metadata, including tags, keywords, unique identifiers, and digital object identifiers (DOI), and can be easily located using standard identification methods. The metadata, including information about data type and location, will be recorded in a standardized form, and kept separate from the raw data.

2.3.2. Making Data Openly Accessible (A)

To enhance the reach and impact of the MACGHYVER research data, the generated data will be disseminated within the consortium as well as beyond. The selected data and outcomes will be communicated with the scientific community and relevant stakeholders through journal publications, conference presentations and open access data repositories. As a result, the project's scope is maximized in terms of dissemination and commercial exploitation via multiple channels. All data will be made publicly available and easily retrievable for verification and reuse, with the exception of any data safeguarded by partners due to the beneficiary's legitimate interests, including commercial exploitation. The public deliverables for MACGHYVER will be made accessible via the project's official website, open access journals, and after the end of the project via the DELFT repository (https://repository.tudelft.nl/). The DELFT repository is certified by CoreTrustSeal and assigns DOIs to the datasets. All data uploaded on the abovementioned platforms is freely accessible to the public without any restriction; however, in order to acquire access to additional data, potential users must contact the data owner or the IPR team. If necessary, appropriate IPR procedures (e.g. NDAs) will be followed.

2.3.3. Making data interoperable (I)

MACGHYVER aims to ensure effective interoperability by making its data accessible in the most popular and common standard formats, facilitating seamless data exchange, and allowing for efficient re-use among various academic and organizational entities. Future iterations of the DMP will provide more information on data interoperability as the project progresses and data is collected. This will include specifications on standards, techniques, and data and metadata vocabularies to aid interoperability, as well as whether a common vocabulary is employed for all data types to support cross-disciplinary collaboration.

2.3.4. Ensuring data reuse (R)

To facilitate broad data reuse by the global scientific community and society as a whole, open access data will be made available to users under standard creative licenses, in accordance with the Consortium Agreement and Grant Agreement terms and conditions. To ensure ease of reuse an internal quality review process will be implemented throughout the project's duration to evaluate both the project data/products and its processes. However, restrictions will be applied for the use of protected data.

2.3.5. Data quality and review

To ensure that the primary project deliverables are created with a high degree of quality, an internal peer review is conducted. The data produced by the MACGHYVER project will remain accessible for use by third parties, even after the project has ended. The DELFT data repository is intended to remain operational even after the end of the project. Protected data is subject to a restricted reuse policy, limiting its use to only within the project's partner organizations.

3. Allocation of Resources

Below are some of the expenses associated with rendering data FAIR:

- Cost of the secure Dropbox data repository service
- Costs incurred in publishing scientific articles with project research data in "Gold" open-access journals.
- Costs associated with the operation of the project website: Coordinator (EDEN)
- Data archiving at the DELFT repository
- Copyright licensing with Creative Commons: free of charge

Each partner bears responsibility for the data they produce, and the cost of publishing generated data in Open Access through scientific journals will be borne by the data owner or author. When there are multiple authors/partners, the allocation of costs will be established on a case-by-case basis by the authors.

4. Data Security

For the data created throughout the project, MACGHYVER will employ strategies emphasizing both security and quick access. The data generated will be backed up frequently

to avoid any data loss. All data and the project deliverables will be stored in Dropbox with exclusive access only to the project partners. The Dropbox service provides full control/monitoring and high security on all operations performed by any user and provides regulated access according to user rights granted and overseen by the only administrator (project coordinator, Eden Tech). Additionally, the following guidelines will be followed in order to ensure the security of the data:

- To prevent data loss, the generated data will be saved in at least two different locations.
- Data will be encrypted if deemed necessary by the partners.
- The usage of USB flash drives will be limited or if possible compleetely avoided.
- To ensure the coherence of the final data collection, the files and folders will be named according to a systematic naming scheme.

5. **Ethical aspects**

No Ethical issues have been identified.