Initial Data Management Plan

Deliverable ID	D7.2
Work Package Reference	7
Issue	1.0
Due Date of Deliverable	28/02/2023
Submission Date	11/03/2023
Dissemination Level[1]	PU
Lead Partner	University of Galway
Contributors	-
Grant Agreement No	101071111
Call ID	HORIZON-EIC-2021-PATHFINDERCHALLENGES-01

Co-funded by the European Union

[1] **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services),
RE = Restricted to a group specified by the consortium (including the Commission Services),
CO = Confidential, only for members of the consortium (including the Commission Services)

Prepared by	Reviewed by	Approved by
Muhammad Sohail Riaz	Pau Farràs	Pau Farràs

Issue	Date	Description	Author(s)
1.0	06/03/2023	First issue of the deliverables	Pau Farràs
1.1			
1.2			
1.3			
1.4			
1.5			
1.6			

INDEX

1 DATA SUMMARY

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation Explain the relation to the objectives of the project Specify the types and formats of data generated/collected Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known) Outline the data utility: to whom will it be useful**

The purpose of data collection in the development of an anion exchange membrane electrolyser for saline or seawater splitting for green hydrogen production is to gather information and measurements that can be used to optimize the performance of the electrolysis process. The collected data can help researchers and engineers to identify the best operating conditions, design parameters, and materials for the electrolysis system. Specifically, the collected data can help to:

1. Characterize the properties of the anion exchange membrane and other components of the electrolysis system, such as the electrodes, catalysts, and current collectors.

2. Determine the efficiency and stability of the electrolysis process under different operating conditions, such as current density, temperature, and salt concentration.

3. Identify the factors that affect the performance and durability of the electrolysis system, such as degradation of the membrane and electrode materials over time.

4. Validate the theoretical models and simulations of the electrolysis process by comparing them with experimental data.

5. Guide the design and optimization of the electrolysis system for practical applications, such as green hydrogen production for fuel cells or energy storage.

Based on the objectives outlined in Table 1.1 of the ANEMEL project, data collection will be critical to measuring progress towards achieving each objective.

For SO1, data collection will involve testing the electrocatalysts developed by the project partners for stability, performance, and cost-effectiveness. This will require the collection of data on the performance of the electrocatalysts under various conditions, including in saline electrolytes and at different pH levels. The performance of the non-PGM catalysts will need to be benchmarked against that of PGM catalysts using a standard protocol to verify their efficacy.

For SO2, data collection will involve testing the durability and recyclability of the anion exchange membranes (AEMs) developed by the project partners. This will require the collection of data on the ionic conductivity of the membranes, as well as their resistance to degradation when exposed to oxygenated solutions. The performance of the project's AEMs will need to be compared with commercially available AEMs to verify their breakthrough potential.

For SO3, data collection will involve testing the MEA prototype to ensure it achieves the targeted power output of 2 W/cm2 at mild alkaline operating conditions. This will require the collection of data on the performance of the electrode-catalyst-membrane combination, as well as the development of a testing protocol to validate the results. The cross-laboratory testing platform developed by EPFL will be instrumental in this data collection effort.

For SO4, data collection will involve validating the lab-validated stack operating under low-grade water. This will require the collection of data on the stack's composition, including the number of cells and active area of each cell, as well as the current density achieved. The project partners will need to validate the results in different consortium laboratories to verify their reproducibility.

For SO5, data collection will involve quantifying the sustainability profiles of the baseline and alternative AEM design choices. This will require the collection of data on the materials used to prepare the components, as well as the end-of-life recyclability of the stack. The project partners will need to follow established norms and protocols to ensure their data collection efforts meet international standards.

Overall, data collection will be critical to the success of the ANEMEL project, as it will enable the project partners to measure progress towards achieving their specific objectives and identify areas where improvements are needed.

Based on the given specific objectives, the types and formats of data that could be generated or collected may include:

1. **Electrochemical performance data**: This includes data on the performance of the electrocatalysts and membranes in terms of their ability to conduct and catalyse electrochemical reactions. The data could include current-voltage (IV) curves, electrochemical impedance spectroscopy (EIS) data, cyclic voltammetry (CV) data, and polarization curves.

2. **Material characterization data**: This includes data on the physical and chemical properties of the materials used in the electrocatalysts, membranes, and other components of the electrolysis system. The data could include X-ray diffraction (XRD) patterns, scanning electron microscopy (SEM) images, transmission electron microscopy (TEM) images, and Fourier transform infrared (FTIR) spectra.

3. **Durability and stability data**: This includes data on the durability and stability of the electrocatalysts and membranes under various operating conditions, including in different electrolytes and at different pH values. The data could include long-term stability testing, including measurements of changes in performance over time.

4. **Performance benchmarking data**: This includes data on the performance of the non-PGM catalysts and AEMs compared to commercial benchmarks. The data could include comparisons of IV curves, EIS data, and polarization curves under standardized testing conditions.

5. **Sustainability and life cycle assessment data**: This includes data on the environmental impact of the electrolysis system, including the materials used in the electrocatalysts, membranes, and other components. The data could include life cycle assessments (LCAs), which estimate the

environmental impact of a product over its entire life cycle, and eco-design principles, which aim to minimize the environmental impact of products and processes.

The formats of the data generated could include numerical data in spreadsheets, graphical data in charts and diagrams, image data in various formats, and written data in reports and publications. Here is the possible origin of the data for each objective:

SO1:

- Origin of data for the electrocatalysts could come from the synthesis and characterization of the catalysts by the project partners (NUIG, UNEW and EPFL) using various techniques such as X-ray diffraction (XRD), scanning electron microscopy (SEM), and electrochemical measurements.
- Benchmarking data could come from testing the catalysts against PGM catalysts using a standard protocol.

SO2:

- Origin of data for the anion exchange membranes (AEMs) could come from the synthesis and characterization of the membranes by the project partners (UNEW, TECH, and EPFL) using techniques such as electrospinning, ion exchange capacity measurement, and durability testing.
- Comparison data could come from commercially available AEMs from companies such as Sustanion or Fumatech.

SO3:

- Origin of data for the MEA prototype could come from the characterization of the electrode-catalyst-membrane combination and electrochemical measurements by the project partners (EPFL, NUIG, TUB, and UNEW).
- Testing protocol data could come from the Joint Research Centre (JRC) which could provide support for developing a testing protocol.

SO4:

- Origin of data for the 1 kWe PoC electrolyser could come from the stack components and assembly, and testing of the stack by the project partners (AGFA, IDeN, and EPFL) using techniques such as current-voltage (I-V) curve measurements, electrochemical impedance spectroscopy (EIS), and durability testing.
- Validation data could come from different consortium laboratories (EPFL, TUB, and IDeN).

SO5:

- Origin of data for the eco-design principles could come from the guidelines for the preparation of components and end-of-life recyclability by the project partners (HESSO and LEITAT).

- Sustainability profile data could come from the quantified assessment of sustainability by the project partners (HESSO and LEITAT) using life cycle assessment (LCA) methodology development and application of eco-design principles in electrolysers.

It is difficult to estimate the exact size of the data that will be generated/collected for this project, as it will depend on various factors such as the frequency of data collection, the number of samples, the type of instruments used, and the duration of the project.

However, given the complexity and scope of the project, it is reasonable to expect a large amount of data to be generated/collected, potentially ranging from few to several gigabytes. The size of the data may also increase as the project progresses and more experiments are conducted. It is important for the researchers to plan for appropriate data storage and management strategies to ensure that the data is securely stored and easily accessible for analysis.

The data generated/collected in this project will be useful to various stakeholders involved in the research and development of CRM- free electrolysis technologies for sustainable hydrogen production.

Some potential beneficiaries of the data include:

1. **Researchers and academics**: The data generated/collected in this project can be used by researchers and academics to further advance the knowledge in this field. They can use this data to develop new theories, algorithms, and models related to the electrocatalysts, anion exchange membranes, and stack designs.

2. **Industry partners**: The data can be used by the industry partners to develop and improve the technology for CRM-free hydrogen production. They can use this data to design new products, optimize production processes, and improve the performance of existing technologies.

3. **Government agencies**: The data can be used by government agencies to support policy development related to sustainable energy production. They can use this data to evaluate the performance of different technologies and make informed decisions about which technologies to support and promote.

4. **Environmental organizations**: The data can be used by environmental organizations to evaluate the environmental impact of different hydrogen production technologies. They can use this data to promote the adoption of technologies that are environmentally sustainable and minimize harm to the environment.

Overall, the data generated/collected in this project has the potential to advance the development of sustainable hydrogen production technologies, which can have a positive impact on the environment and support the transition to a low-carbon economy.

2 FAIR DATA

2.1 Making data findable, including provisions for metadata: Outline the discoverability of data (metadata provision)

- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

Discoverability of data (metadata provision): The data will be made discoverable through the provision of metadata. Metadata will include information such as the title, author, date created, description, and keywords.

Identifiability of data and standard identification mechanism: Data will be identified using standard identification mechanisms, such as Digital Object Identifiers (DOIs), to ensure persistent and unique identification.

Naming conventions used: Data files will be named according to a standard naming convention that includes information such as the project name, date created, and type of data.

Approach towards search keywords: Keywords will be chosen based on the content of the data and will be designed to make the data easily searchable.

Approach for clear versioning: A clear versioning system will be used, with each version of the data file labelled with a unique version number and date.

2.2 Making data openly accessible

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **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)? Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

All data that has been collected, processed, and analysed during the course of the research will be made openly available for public access, except for any data that contains sensitive or personal information, which will be kept confidential.

The data will be made available through an online repository such as Zenodo, Dryad, or Figshare. The repository will be selected based on the specific needs of the research project.

The data will be made available in a machine-readable format such as CSV, JSON, or XML. Users will need software tools to access and analyse the data, such as statistical software packages like R or Python. The necessary software and documentation will be provided alongside the data in the repository.

The data, associated metadata, documentation, and code will be deposited in an online repository. The specific repository will be selected based on the needs of the research project.

Access to the openly available data will be provided without any restrictions. For data that is kept closed due to sensitive or personal information, access will be granted only to authorized personnel who have obtained the necessary permissions and approvals. Access to the confidential data will be provided through a secure and controlled system that ensures confidentiality and data protection.

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **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?**

To ensure interoperability of the data generated by the ANEMEL project, the consortium will assess the compatibility of their data with established vocabularies, standards, and methodologies. The Data Management Plan (DMP) will specify the data and metadata vocabularies, standards, or methodologies to be followed to facilitate interoperability.

To enable inter-disciplinary interoperability, the consortium will use standard vocabularies for all data types present in their data set wherever possible. In cases where standard vocabularies are not used, the consortium will provide mappings to more commonly used ontologies. This will help to ensure that the data generated by the ANEMEL project can be easily accessed, shared, and reused by other researchers, and will promote collaboration and knowledge exchange within the renewable energy sector.

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**

- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

To increase data re-use, ANEMEL will ensure that the data generated by the project will be licensed to permit the widest possible reuse. The licenses will be chosen carefully to ensure that the data can be reused by third parties. The data will be made available for re-use as soon as possible following EU rules. If a data embargo is necessary for specific reasons, it will be specified in the Data Management Plan (DMP) and justified in detail.

The data produced and/or used in the project will be useable by third parties, in particular after the end of the project, unless there are legal, ethical, or confidentiality issues that prevent its release. In such cases, the reasons will be clearly stated in the DMP. To ensure data quality, ANEMEL will establish rigorous quality assurance processes, including data validation, verification, and cross-checking. The quality assurance processes will be documented and included in the DMP.

The length of time for which the data will remain re-usable will be specified in the DMP. The data will be kept accessible for at least 10 years following the project's completion unless there are legal or ethical constraints that prevent its release. If data needs to be preserved beyond this period, it will be transferred to an appropriate data repository or archive, where it will be available for re-use according to the repository or archive's policies and standards.

3 ALLOCATION OF RESOURCES

- **Explain the allocation of resources, addressing the following issues:**
- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

The online repository chosen here, Zenodo, is a free-to-use platform. Therefore, there is no charge expected for the deposition or storage of the data. For the scientific publications, a certain amount of fees are payable to the publishers for open access, which will give free, public access for the publications.

Funds have been allocated for open access fees in the budget of the project in order to ensure data openness and dissemination. In addition, it is possible under certain circumstances that the host institutions/organizations within the consortium could offer funding to cover the fees for open access publications or expenses of patenting.

Research data will be stored at NUIG file servers, subject to the institute’s data security measures and with a current policy of weekly backups, copies are stored on duplicate on local and cloud drives. Items deposited in Zenodo, including all the scientific publications, will be archived, and retained for

the lifetime of the repository, which is currently the lifetime of the host laboratory CERN (at least for the next 20 years).

Finally, the ANEMEL project considers the costs and potential value of long-term preservation of research data. Long-term preservation ensures that research data remains accessible and usable beyond the project lifetime, allowing for further analysis, reuse, and validation. The costs of long-term preservation may include storage, backup, migration, and format conversion. However, the potential value of preserved data includes increased citation rates, improved reproducibility and transparency, and enhanced scientific collaborations. Therefore, the ANEMEL project considers the potential benefits of long-term preservation when developing its data management plan and budget.

4 DATA SECURITY

Address data recovery as well as secure storage and transfer of sensitive data

ANEMEL aims to generate and validate new data and knowledge in AEMWE components for renewable energy in Europe. To ensure good research practice and planning, a Data Management Plan (DMP) will be implemented, specifying how data will be handled during and after the project. The DMP will address the secure storage and transfer of sensitive data, as well as data recovery in case of loss or damage. The consortium will follow the FAIR rules and Open Science practices to ensure the management and protection of knowledge and intellectual property.

Within the consortium, datasets will be shared via private data repositories, which run at the coordinator facilities (e.g. NUIG which has automated backup and mirroring, with support from Microsoft). The coordinator of the project, Dr Pau Farràs, will inform the consortium on how to safely upload datasets to such a shared data repository to ensure that all of the collected datasets are correctly stored and backed up. All data and backups are stored in EU servers.

The data generated by the project will be of various types, including electronic copies of synthetic methodologies, characterisation files, electrocatalysis responses, and established protocols and internal reports. The data will be stored securely in each partner's facilities with a backup copy in ANEMEL's internal servers.

Following EU rules, ANEMEL is committed to making as much data available in an open-access repository with minimum delay while maintaining data security and intellectual property rights. The precise details of what data will be made available and when will be determined during the project. ANEMEL will take every care to ensure that the data is complete and correct, subject to data protection legislation.

Overall, ANEMEL will prioritize the secure storage and transfer of sensitive data, data recovery in case of loss or damage, and adherence to data protection legislation and intellectual property rights.

5 ETHICAL ASPECTS

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

The ANEMEL project involves the preparation of catalysts and membranes, including the use of chemical reagents and nanoparticles, and testing of catalytic reactions. This activity may pose risks to the safety of research staff and the environment. Therefore, the project will comply with the ethical principles and relevant legislations, such as Directive 91/322/EEC on chemicals agents at work and Regulation EC No 1907/2006 Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), and the "European Commission Recommendation of 07/02/2008 on a Code of Conduct for Responsible Nanosciences and Nanotechnologies Research."

All partners will ensure a safe work environment by following the European and national Health & Safety rules adopted in each chemical laboratory, and researchers will prepare a project risk assessment detailing the hazards for health and the environment, and appropriate ways to dispose of wastes. The project will comply with the Second Regulatory Review on Nanomaterials COM/2012/0572, and all workers in the organization will be fully trained and prepared to work with chemicals and nanomaterials.

Furthermore, there will be transfer of experimental materials from partners in the EU to third countries, and from these third countries to the EU, which will comply with national and EU legislation. The materials which will be imported to/exported from the EU will be submitted as a deliverable, along with copies of import/export authorizations.

The project will appoint an external advisor with expertise on the use of chemicals to develop safety protocols for researchers dealing with dangerous chemicals. All relevant authorizations and approvals will be obtained before the beginning of the relevant research activities and kept on file. The ethics standards and guidelines of Horizon Europe will be rigorously applied, regardless of the country in which the research is carried out.

The above information should be covered in the context of the ethics review, ethics section of the DoA, and ethics deliverables. Technical aspects related to the ethical dimension of the objectives, methodology, and likely impact of the project should also be addressed in these documents.

References:

- Directive 91/322/EEC on chemicals agents at work
- Regulation EC No 1907/2006 Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) "European Commission Recommendation of 07/02/2008 on a Code of Conduct for Responsible Nanosciences and Nanotechnologies Research"
- Second Regulatory Review on Nanomaterials COM/2012/0572