**PROJECT: SIDERWIN**

**Deliverable:**

**D8.3**

**Title:**

# Data Management Plan Final Version

Date: 25th of April 2023

Version: 2.0

Website: N/A

## Status

| | |
|---|---|
| ☒Final | |
| ☐In Progress. Please explain: | ☐ Iterative Process – This year's results have been 100% achieved.<br>☐ Delay – This year's results were not fully achieved. |

## Tracking Changes

| | | |
|---|---|---|
| Initial Version | March 2020 | |
| Final Version | April 2023 | |

## Level of Dissemination

☒ Confidential
☐ Public

## Author(s)

| | Partner name | Name of the author |
|---|---|---|
| Main Author | **ArcelorMittal** | Valentine WEBER-ZOLLINGER |
| | **ArcelorMittal** | Maxime BALVA |

# Table of Contents

## List of Abbreviations and Acronyms

| Acronym | Full name |
|---|---|
| CA | Consortium Agreement |
| CFD | Computational Fluid Dynamics |
| CSV | Comma-Separated Values |
| D | Deliverable |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| EC | European Commission |
| EP | Exploitation Plan |
| EU | European Union |
| GA | Grant Agreement |
| H2020 | Horizon 2020 |
| IPR | Intellectual Property Rights |
| LCA | Life Cycle Analysis |
| N/A | Non-applicable |
| OA | Open Access |
| ORDP | Open Research Data Pilot |
| T | Task |
| TRL | Technical Readiness Level |
| WP | Work Package |

## Glossary

| | |
|---|---|
| Dataset | Data needed to validate the results presented in scientific publications. |
| Metadata | Metadata is data that describes other data. Meta is a prefix that in most information technology usages means "an underlying definition or description." Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier. |
| Open Access | The practice of providing online access to scientific information that is free of charge to the end-user and reusable. |
| Repository | A digital repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus. |

## Acknowledgement

SIDERWIN project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 768788".

## Disclaimer

"This study reflects only the author's views, and the Commission is not responsible for any use that may be made of the information contained therein"

# 1 Executive summary

SIDERWIN is a H2020 Research and Innovation Action project. Its main objective is to provide a process solution for carbon neutral primary steel production. For this purpose, it relies on the application of electricity.

Today, there is no experience in primary steel production by electrolysis at such high maturity level. The SIDERWIN consortium proposes to deliver, open access and high-quality, data on electrolysis primary steel production to the community of neutral carbon industrial process developers. It is of high importance, in the context of climate change, that research works on neutral carbon processes are not duplicated due to a lack of accessibility of data or because past experiences are not enough discoverable. To these regards, the SIDERWIN consortium is committed to make its results as open as possible and as closed as the necessary exploitation actions require.

The Data Management Plan is a mandatory part of the proposal by joining the Open Research Data Pilot. The SIDERWIN project is characterized by a significant academic activity with production of scientific information and by potential commercial applications in the mid-term that require protection by Intellectual Property Rights. This data management plan presents the obligations of the SIDERWIN project in terms of data management and describes how Open Access and FAIR (Findable, Accessible, Interoperable and Re-usable) principles are practically applied. It also provides the guidelines for the consortium partners to deal with the different types of data generated during the lifespan of the project. These sources of data have been identified as seven datasets that have been described. The publicly open data have been defined and these open data have been uploaded on the chosen repository, the Zenodo platform.

# 2 Technical content

## 2.1 Siderwin project objectives

SIDERWIN project aims at developing an innovative electrochemical process to transform iron oxide into steel metal plates. This process, based on the ULCOWIN technology developed since 2004, produces steel by electrolysis without direct CO2 emissions. In this operation, electrical energy is transformed into chemical energy consisting of separated iron metal from oxygen gas. It is a disruptive innovation that entirely shifts the way steel is presently produced. Electrolysis is the main processing unit of an iron-making route beginning with mineral iron ore and producing liquid steel.

The first objective of SIDERWIN is to develop, build and demonstrate the production of iron metal from its oxide without direct involvement of carbon or fossil fuels and according to the simplest stoichiometry of the reaction of iron oxide decomposition:

$$\tfrac{1}{2}\ Fe_2O_3 \rightarrow Fe + \tfrac{3}{4}\ O_2$$

The second objective of SIDERWIN is to produce iron by electrowinning with a prototype cell equipped with the key components of the final version.

The third objective of SIDERWIN is to interface the electrowinning prototype cell with a communication system to operate it according to electric grid priorities in real time.

The fourth objective of SIDERWIN is to produce iron metal from iron oxide coming from low-grade iron ore incompatible with the conventional process and from residues of non-ferrous metallurgies.

The fifth objective of SIDERWIN is to propose a profitable model that should facilitate the financial support of the next development steps of the SIDERWIN process. Thus, bridging the “valley of death” between TRL 6 and 8 where investment is too high for research programs and too risky for industrial participation.

## 2.2 Siderwin Data Management Plan

### 2.2.1 Introduction

SIDERWIN is a H2020 Research and Innovation Action project. Its main objective is to provide a process solution for carbon neutral primary steel production. For this purpose, it relies on the application of electricity.

Today, there is no experience in primary steel production by electrolysis at high maturity level. The SIDERWIN consortium proposes to deliver open access and high- quality data on electrolysis primary steel production to the community of neutral carbon industrial process developers. It is of high importance, in the context of climate change, that research works on neutral carbon processes are not duplicated due to a lack of accessibility of data or because past experiences are not enough discoverable. To these regards, the SIDERWIN consortium is committed to make its results as open as possible and as closed as the necessary exploitation actions require.

The SIDERWIN project covers a wide range of research and innovation approaches from laboratory electrochemical experiments, CFD simulations, engineering developments, semi-industrial pilot operation, Life Cycle Analysis, simulation of European wide power system and economic modelling. It is a priority of this project that the results of these approaches are disseminated and made accessible to interested stakeholders and communities such as climate policy makers, industrial investment planners, scientific researchers, regulatory communities, potential partners and customers from metal industries.

The SIDERWIN proposal has chosen to participate in the Open Research Data Pilot (ORDP) in Horizon 2020, which aims to improve and maximize access and re- use of research data generated by actions. Thus, contributing to the fast visibility of the SIDERWIN results and their re-use within the research community.

This report is the Data Management Plan as required in ORDP, in its final version. The main purpose of the first version was to identify which data will be generated during the project, among this data which will be open, how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The format of the plan of this document follows Horizon 2020 template described in "Guidelines on Data Management in Horizon 2020" [1].

The other purpose of this document is to provide the main guidelines and elements necessary for all the consortium partners to manage the various data produced during the project execution. In accordance with ORDP, each dataset is described in terms of purpose of data generation, of relation with the objectives of the project, of the types and formats of data generated, of re-use of existing data, of origin of the data, of size of the data and of data utility.

The DMP comes in parallel with the Exploitation Plan D8.4, which have different and to some extend opposite requirements. The consistency between the two is addressed to anticipate possible conflicts between data openness and protection of rights for exploitation. The DMP is also related to the Master dissemination plan D8.2, by contributing to a better impact of the dissemination activities.

This report is the final version of the deliverable D8.3.

### 2.2.2 Basis of Data Management

#### *2.2.2.1 Participation in the Open Research Data Pilot*

As expressed in the GA (Annex 1 – Part B, p32), the SIDERWIN project joins the Open Research Data Pilot (ORDP). As such, all consortium partners embrace the concepts and the principles of open science and acknowledge the benefits of reusing and evaluating already produced data for promoting and supporting research and innovation projects at European level. Some of the data generated during the project activities will be made available in open access for further analysis, so that exploitation is guaranteed. The results should benefit European companies and citizens.

#### *2.2.2.2 Open Access in Horizon H2020*

One of the most important elements of ORDP is the Open Access (OA) policy to scientific information produced by the EU-funded project. The objective is to optimize publicly funded scientific research, both at European level and at the member state level. The practice of OA is defined as providing on-line access to scientific information that is free of charge to the reader and that is reusable without restrictions for researchers, businesses and citizens. In the context of research and innovation, scientific information can refer to peer-reviewed scientific research articles or research data [1].

Regarding the SIDERWIN project, open access of its results is expected to improve the visibility of the project results, firstly in the scientific communities by increasing citations of the research partners and secondly by multiplying opportunities of applications and exploitations. It is of prime importance to demonstrate that metal industry can be conducted as a clean activity contrary to common perception. Furthermore, open access is an opportunity to stimulate knowledge build up on the potentials of electrolytic routes to address carbon neutrality of large industrial sectors.

Research data are essentially measurements from laboratory experiments, from pilot operation, physical simulations, LCA and economical modelling. They are all available in digital form.

Concerning peer-reviewed scientific articles, the project aims at choosing as most appropriate standard towards open access self-archiving, known as "Green Open Access", without exclusion of the "Gold Open Access" route. Green access is granted when the final, peer-reviewed manuscript is deposited by its authors in a repository of their choice. Then open access must be ensured within 6 months. Thus, open access actually follows with some delay due to the so-called “embargo period”.

For the research data needed to validate results presented in scientific publications, the project aims at identifying them as datasets. The preservation for use and reuse of the openly accessible data will be guaranteed by deposition in a repository.

#### *2.2.2.3 FAIR principles*

FAIR relates to a set of principles that have been proposed to enhance the reusability of research data [2]. It is the acronym of Findable, Accessible, Interoperable and Re-use. These principles have been chosen for their concision, their independence regarding the field and

their application to a wide range of scholarly outputs. More explicitly, it proposes to apply the following principles:

1. Making data findable, including provisions for metadata.

   Data are assigned a globally unique and persistent identifier. They are described with rich metadata. They clearly and explicitly include the identifier of the data it describes. They are registered or indexed in a searchable resource.

2. Making data openly accessible.

   Data are retrievable by their identifier using a standardized communications protocol. The protocol is open, free, and universally implementable. The protocol allows for an authentication and authorization procedure, where necessary. Metadata are accessible, even when the data are no longer available.

3. Making data interoperable.

   Data use a formal, accessible, shared and broadly applicable language for knowledge representation. They use vocabularies that follow FAIR principles. They include qualified references to other data.

4. Increase data re-use.

   Metadata are richly described with a plurality of accurate and relevant attributes. They are released with a clear and accessible data usage license. They are associated with detailed provenance. They meet domain-relevant community standards.

Application of the FAIR principles contributes to the sound data management of H2020 projects [3].

#### *2.2.2.4 Grant and Consortium Agreements for provisions regarding Data Management*

The Grant Agreement (GA) was signed on 24/08/2017. It addresses data related issues in the Articles 28 for exploitation and Articles 29 for dissemination [4]. An amendment of this agreement has been signed on 04/03/2021.

Excerpts of Articles 28:

Each beneficiary must — up to four years after— take measures aiming to ensure 'exploitation' of its results (either directly or indirectly, in particular through transfer or licensing;) by:

(a) using them in further research activities (outside the action);
(b) developing, creating or marketing a product or process; or marketing a product or process;

(c) creating and providing a service, or
(d) using them in standardization activities.

Unless it goes against their legitimate interests, each beneficiary must — as soon as possible 'disseminate' its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium).

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of — unless agreed otherwise — at least 45 days, together with sufficient information on the results it will disseminate.

Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests.

Excerpts of Articles 29:

Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.

In particular, it must:

(a) as soon as possible and at the latest on publication, deposit a machine- readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;
(a) moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.
(b) ensure open access to the deposited publication — via the repository — at the latest:
    a. on publication, if an electronic version is available for free via the publisher, or within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.
(c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.

The bibliographic metadata must be in a standard format and must include all of the following:

- the terms "European Union (EU)" and "Horizon 2020";
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable,
- a persistent identifier.

The Consortium Agreement (CA) has been signed and came into force on 22/11/2018 [5]. It forms the legal basis in dealing with IPR issues and covers clear rules for dissemination or exploitation of project data. It explains pre-existing background of each partner. It is to

be referred to for dissemination by Article 8 and for Access Rights for exploitation to Article 9.

#### *2.2.2.5 Objectives of the Data Management Plan*

The Data Management Plan is a mandatory part of the proposal joining the ORDP. It describes how, in the SIDERWIN project, OA and FAIR are practically applied. It covers the overall approach of the project on data management. It is presented to facilitate partners in addressing issues related with data.

It reflects provisions regarding Data Management expressed in the GA and CA. It is conducted in parallel with the Exploitation Plan of the project, to insure their consistencies.

It describes the research data that are generated by the WP in terms of origin, standards, preservation and sharing.

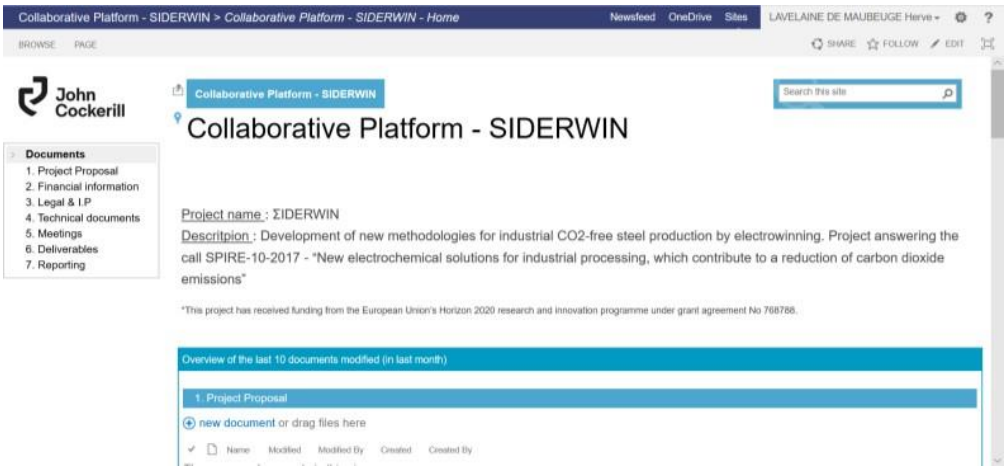
### 2.2.3 Practice of Data Management

#### *2.2.3.1 Project collaborative Platform and public webpage*

The project's management and dissemination activities contribute to data management with a Collaborative Platform and a public webpage.

##### 2.2.3.1.1 Collaborative Platform

A Collaborative Platform has been created to facilitate sharing of information between partners, it corresponds to the Deliverable D1.3 [6]. It is maintained by one of the partners, John Cockerill, and updated by the coordinator.



#### 2.2.3.1.2 Webpage

The public webpage presents the objectives and the main characteristics of the project. It gives updated information with a twitter account https://twitter.com/siderwin_spire . It gives access to project documents such as project NewsLetter, conference presentations, scientific publications, and public deliverables. It has been created and is maintained by one of the partners Tecnalia. It corresponds to the deliverable D8.2 [7].


### *2.2.3.2 Relations with the Exploitation Plan*

The SIDERWIN project is expected to deliver a technological development with a strong innovation content. It addresses a medium-TRL technology with the aim to provide market-oriented solutions in the mid-term. Particularly, the electrowinning cell design combines multiple original aspects that may contribute to an industrial application.

The project involves academic partners and institutions, research organizations, as well as commercial companies. It is also part to the project to accelerate the exploitation of these innovations, as described in the tasks of WP7. The research data related to the identified exploitable results will not be put into the open domain. Conforming to the H2020 obligations, protection by IPR will be taken firstly to secure commercialization prospects and secondly to facilitate open communication on the underlying validating results of these innovations. An embargo period separates protection by IPR, typically during patent filing, from publication of scientific results.

For all the data produced, the consortium has examined possible conflicts with the EP.

#### 2.2.3.3 Publicly disseminated deliverables

Among the results that are already declared as Open Access are the deliverables in the GA which will be disseminated at public level, cf. Table 1.

**Table 1: Deliverables disseminated at public level declared in the GA**

| Deliverable | Description | Lead Beneficiary | WP |
|---|---|---|---|
| D5.7 | Public Final test report | ArcelorMittal | WP5 |
| D7.4 | Environmental LCA assessment final report, v1 | Quantis | WP7 |
| D7.5 | Environmental LCA assessment final report, v2 | Quantis | WP7 |

#### 2.2.3.4 Choice of a data repository

The project needs a repository to preserve the scientific data generated to produce the scientific publications. The repository chosen is the Zenodo platform.

The repository has been chosen according to the following criteria:

1. Gives visibility to the results.
2. Gives referencing to the submitted datasets.
3. Helps to facilitate citation of the results.
4. Allows linking of the datasets to other websites.
5. Standard of the repository compatible with the project datasets.
6. Produces users behaviour statistics to improve presentation of data.
7. Policy of long-term preservation and curation of the data.
8. Possibility to apply an embargo period to confidential data.
9. Provides sustainable data formats and metadata standards.
10. Capacity to export data from repository to any point.
11. Indexed by search engines.
12. Home of communities close to the research conducted during the project such as process engineering.
13. Reliable backup practices.
14. Curated by a trusted institution.

The metadata scheme applied to the datasets of the project combine repository practice and the obligations described in Article 29.2 of the GA. These last obligations make mandatory to include the following information acknowledging of EU funding: European Union; H2020; SPIRE; SIDERWIN; GA768788.

### 2.2.3.5 Data generation in the SIDERWIN project

The SIDERWIN project combines different approaches to upscale a process from laboratory to commercializing application. The first approach is the laboratory study of the process and dedicated experimental set ups. These produce data as measurements time series, pictures from observations, chemical analysis. The second approach is the engineering sizing and drawing of a feasible solution. The third approach is a pilot operation which involves 150 probes. They will generate results as time series measurement of voltage, current, temperature, etc. along with pictures and analysis. The fourth approach is numerical simulation of physical phenomena that will generate the most significant amount of data due to the multiplicity of phenomena and the size of the pilot to be simulated. The fifth is the modelling of the future Europe continent wide power system. The sixth is the environmental assessment of the overall processing route. The seventh is the economical modelling of a plant at the horizon of 2050. The partners and WP related to these data generations are presented on Table 2.

**Table 2: Data generation by Partners and WP**

| | Lead Partner | Description of DATA generation | Related WP |
|---|---|---|---|
| 1 | AMMR | Measurements from pilot operation, pictures, analysis results | WP5 |
| 2 | JohnCockerill | Technical drawings of pilot and cell | WP4 |
| 3 | EDF | Modelling results of power infrastructure | WP7 |
| 4 | CFD-Numerics | Simulation results from CFD | WP3 |
| 5 | Quantis | Results of life cycle analysis modelling | WP7 |
| 6 | Tecnalia | Data analysis from pilot operation | WP5 |
| 7 | UAVR | Measurements and analysis from dedicated laboratory setups | WP3 &6 |
| 8 | Mythilineos | Measurements, chemical analysis | WP6 |
| 9 | NTUA | Measurements and analysis from dedicated laboratory setups | WP6 |
| 10 | Recoy | Modelling results | WP7 |
| 11 | Dynergie | N/A | |

### 2.2.3.6 Data management process in the SIDERWIN project

The different steps of the data management are presented in the flow diagram below, cf.

Figure 1. Raw data have been produced by the experimental studies conducted in laboratory or on pilot facilities and by the simulation tools. This data has been processed and analyzed into communicable forms such as reports, publications or presentations. The open access policy of the project has been applied to the scientific data from academic partners and to the committed public deliverables. Exploitation sensitive data have been treated by the commercializing partners to file patents or licenses, according to the rules explained in the CA.

The research results of the SIDERWIN project have been examined in order to maximize openness, dissemination and exploitation at the same time. Openness is expected by the publication of open access scientific article in Green OA. Care will be taken to incorporate the possible supplementary materials that could contribute to better sharing of data. Exploitation is an obligation for each beneficiary under Article 27 of the Grant Agreement. It is expected from the commercializing partners. It assumes an intermediate step of patent or license filing. During this time an embargo will be applied on the related results. Care will be taken during the reviewing meetings to limit this time of embargo by starting as early as possible the discussions on patenting. The publicly sharable results will be made open access by preservation in a repository to be usable beyond their original purpose. The reusability of these data will contribute to the continuity of development of new industrial processes.

The validating data mention in the flow chart represents the results needed to validate the information presented in their final form in public reports, scientific publications, conferences presentations and PhD defenses. Their volumes are generally considered excessive to be incorporated into a communication article. Digital storage in dedicated servers helps to share these data, improve scientific practice and increase credibility of scientific publications.

The list of scientific publications is available in the updated versions of the Deliverable D8.2 Master dissemination and communication [8].

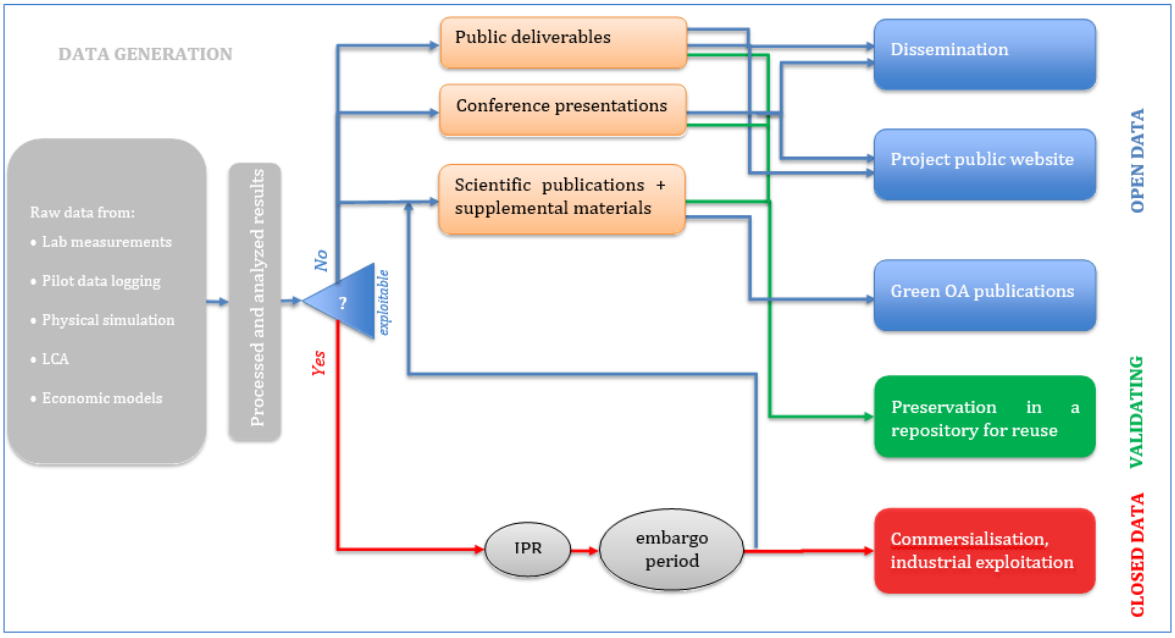



**Figure 1: Flow diagram of the data management**

## 2.3 Identified Datasets

According to the current progress and the foreseen developments of the project, seven datasets have been identified.

Five of the seven datasets identified are the pilot operation at normal conditions and under variable load, the CFD simulations, the Life Cycle Assessment results, and the economic modelling. Due to the confidentiality of the raw data produced, only some treated data will be made accessible, via public reports, publications, or presentations.

The two last datasets have been made publicly available on the Zenodo platform, with their corresponding publications, and correspond to different laboratory studies performed by UAVR in the framework of the WP3 and WP6, to evaluate the impact of other elements on the iron electrolysis.

The validation and registration of these datasets and their accompanying metadata are the responsibility of the partner that generates them.

### 2.3.1 Dataset #1 summary

The Dataset#1 refers to the data generated by monitoring pilot operation conducted in WP5. The treated results will be presented in the final public report (D5.7), however the corresponding raw data will not be openly accessible as they will be linked to protected IP concerning the pilot configuration.

| | DATASET#1 DESCRIPTION |
|---|---|
| *Purpose of the data collection / generation* | The purpose of the data collection is to develop a tool in order to:<br>i. analyzes the effect of a parameter in the final performance of the pilot<br>ii. define the optimal process parameters (current density, flow rate, concentration of iron oxide particles…) in order to maximize the performance of the system.<br>A data gathering campaign of the ΣIDERWIN cell operation will be carried out so:<br>Correlation between the process parameters and the performance of the system will be analyzed, making use of statistical analysis techniques.<br>A data-driven predictive model will be developed based on machine learning techniques, to predict the performance depending on the process parameters. This model would also allow analyzing the effect of changing the value of a process parameter in the performance of the system. |
| *Relation to the objectives of the project* | This purpose is aligned with the second objective of SIDERWIN project, that is, to produce iron by electrowinning with a prototype cell equipped with the key components of the final version. |
| *Types and formats of data generated /c ollected* | Most of the data collected here will be numeric readable text format, such as CSV time series, coming from sensors and PLC. |
| *If existing data is being re-used (if any)* | It is not planned to reuse existing data because it is important to consider the interrelation between process parameters and not all the sensors are currently available. |
| *Origin of the data* | The data will be acquired through the 150 probes and sensors monitoring the cell operation, essentially voltage, current, temperature, flow rate etc.... As the system is expected to be composed of a wide variety of sensors from different brands, data acquisition will be performed using a universal protocol, such as OPC DA (Data Automation) or OPC UA (Unified Architecture). |
| *Expected size of the data (if known)* | Considering that it will be necessary to record about 100 simple precision floating point variables (4 bytes) with an acquisition interval of 10 seconds, the daily information will occupy (24 * 60 * 60) / 10 * 100 * 4 = 3456,000 bytes,<br>i.e. about 3,3 Mb. For analysis purposes it will be performed |

| | |
|---|---|
| | a 3-4 month gathering campaign (90-120 days). Therefore, for a campaign of 120 days, the necessary storage will be 120*3,3 Mb = 396 Mb |
| *Data utility: to whom will it be useful* | The data can be used for next scale up development toward TRL 7, since the data of the pilot cell could be compared with cells installed in other facilities and thus, to generalize the case of use. It is of paramount importance to validate CFD and experimental simulations conducted during the first phase of the project. The treated results will be presented in a final public report, however the corresponding data will not be openly accessible as they will be linked to protected IP concerning the pilot configuration. |

### 2.3.2 Dataset #2 summary

The Dataset#2 refers to the data generated with CFD simulations conducted in WP3 and WP5. The technical information contained in this Dataset is commercially sensitive and

requires protection to guarantee industrial exploitation. The expected protection will be by filing patent for IPR. These data will then not be openly available.

| | DATASET#2 DESCRIPTION |
|---|---|
| *Purpose of the data collection / generation* | CFD simulations have been used to design the cell. Each simulation is saved and recorded in different files on our servers. Those files contain all the results that were used to take design decisions. |
| *Relation to the objectives of the project* | The CFD results were mainly used to help to design the cell and also to understand flow physics (bubble and electrolyte). |
| *Types and formats of data generated / collected* | Most of the results are stored in ANSYS-Fluent format files: case file containing the geometric description of the model and data files containing the results. Both are needed to reread the results. After post-processing all the results have been presented in powerpoint and word files. |
| *If existing data is being re-used (if any)* | All the data are saved and backed-up. They might be re-used in the future to understand some technical choices |
| *Origin of the data* | Data have been generated with the CFD software ANSYS-FLUENT |
| *Expected size of the data (if known)* | For the WP3, the size is expected to be > 2 To (TeraOctet) |
| *Data utility: to whom will it be useful* | Those data could be used by any user who wants to understand how the technical choice have been made.<br>These data will be closed data. |

### 2.3.3 Dataset #3 summary

The Dataset#3 refers to the data generated during operation of the pilot when studying the ability of the electrowinning cell to operate “on demand”. This study is conducted in

WP5 by T5.4. Some of the treated data will be available in the Public Final Test report (D5.7), but the raw data will not be openly accessible as they refer to some parts of the design of the pilot which are protected by IP.

| | DATASET#3 DESCRIPTION |
|---|---|
| *Purpose of the data collection / generation* | The data will be generated to validate the ability of the pilot to operate flexibly according to imposed load variations. This corresponds to variations currently demanded by the Transmission System Operators to balance the electrical grid. |
| *Relation to the objectives of the project* | The data collected during these conditions will indicate to what extend the ΣIDERWIN process can contribute to manage the grid with ample and steep variations of load as expected by the integration of more intermittent renewable energy sources. |
| *Types and formats of data generated / collected* | The data are generated by the operation of the pilot. They are logged by interfaced instruments and recorded as CSV .txt format files. These raw data are analyzed and treated into communicable form such as reports and presentations. |
| *If existing data is being re-used (if any)* | This data set corresponds to conditions that have never been investigated until now. |
| *Origin of the data* | The pilot is operated with a multiplicity of probes to control production and to follow its response in terms of energy consumption. |
| *Expected size of the data (if known)* | Today, the final size of this datasets is in the order of 100GB |
| *Data utility: to whom will it be useful* | They will provide the first insight on flexible steel production. They will be used inside the project to estimate the potential of the technology to participate in Demand Side Response and to evaluate the corresponding economic benefice. These data will not be publicly shared. |

### 2.3.4 Dataset #4 summary

The Dataset#4 refers to the data generated with the LCA activity conducted in WP7. The treated data will be publicly available in the Environmental LCA assessment final reports (D7.3 and D7.4), but the raw data will not be openly accessible as they refer to confidential information.

| | DATASET#4 DESCRIPTION |
|---|---|
| *Purpose of the data collection / generation* | The LCA model and subsequent results are relying on an inventory of inputs (e.g. material and energy flows) as well as outputs (e.g., wastes, emissions) which serve as a reference to calculate the environmental impact of the ΣIDERWIN technology vs. the reference Blast Furnace and Basic Oxygen Furnace technology. |
| *Relation to the objectives of the project* | The LCA model provides the quantitative basis to calculate the carbon footprint and other environmental indicators as well as to demonstrate the environmental benefits and disadvantages of the ΣIDERWIN technology vs. the reference technology. |
| *Types and formats of data generated / collected* | Model as .csv file which can be imported in the software Simapro. |
| *If existing data is being re-used (if any)* | No |
| *Origin of the data* | Compilation of different literature sources that will be listed in the deliverable D7.4 ad D7.5. |
| *Expected size of the data (if known)* | Less than 10 MG |
| *Data utility: to whom will it be useful* | LCA practitioners.<br><br>The treated data will be publicly available in the Environmental LCA assessment final reports (D7.4 and D7.5), but the raw data will not be openly accessible as they refer to confidential information. |

### 2.3.5 Dataset #5 summary

The Dataset#5 refers to the data generated with the economic modelling of the ΣIDERWIN process conducted in WP7. Some of the treated data will be available in the Public Final Test report (D5.7), but the raw data will not be openly accessible as they refer to confidential information.

| | DATASET#5 DESCRIPTION |
|---|---|
| *Purpose of the data collection / generation* | In order for the techno-economic model to perform its intended purpose it requires input-data related to the market such as energy-prices, materials-prices and data related to the plant-technology such as capex, opex, materials required etc. It will generate as outputs various KPIs such as Plant NPV, levelized manufacturing costs and a comparison to the reference technology of Blast Furnace and Basic Oxygen Furnace technology |
| *Relation to the objectives of the project* | The techno-economic model has as its objective to assess the economic viability of the ΣIDERWIN technology. It does so by running various scenarios of input parameters to assess the robustness against different plant set-ups and economic circumstances |
| *Types and formats of data generated / collected* | Input data and output data is numeric |
| *If existing data is being re-used (if any)* | Some data is existent before the project started (e.g. EU Reference data), the rest is particular to the project and not re-used. |
| *Origin of the data* | The origin of the data is both public (such as EU reference scenarios) as well as proprietary to the project such as plant / materials data |
| *Expected size of the data (if known)* | Based on current insights this will amount to around 10 MB |
| *Data utility: to whom will it be useful* | It could be useful to professionals in the Steel industry as well as Policy makers.<br>Some of the treated data will be available in the Public Final Test report (D5.7), but the raw data will not be openly accessible as they refer to confidential information. |

### 2.3.6 Dataset #6 summary

The Dataset#6 refers to the data generated during laboratory experiment performed in the study of the impact of the possible presence of aluminium in the composition of iron oxide on the electrochemical deposition of Fe from powder suspensions under alkaline conditions. This work was realized by the University of Aveiro.

| | DATASET#6 DESCRIPTION (DOI: 10.1149/1945-7111/AB9A2B) |
|---|---|
| *Purpose of the data collection / generation* | Study of the impact of the possible presence of aluminum in the composition of iron oxide on the electrochemical deposition of Fe from powder suspensions under alkaline conditions. |
| *Relation to the objectives of the project* | The presented data allows one to understand if alternative iron oxide feedstocks rich in -Al content can be used as a raw material for Fe electrodeposition. In this case, the major components of red mud residue (Fe2O3 and Al2O3) were tested for the Fe electrowinning. |
| *Types and formats of data generated / collected* | The data was obtained by electrochemical tests performed by a potentiostat and recorded as .xlsx files. This data was recalculated, taking into account to the working electrode area. XRD data (.xlsx) and SEM (.jpg) images were added according to each type of analysis. |
| *If existing data is being re-used (if any)* | The corresponding experimental conditions have never been tried so far. |
| *Origin of the data* | The data was obtained by experimental tests of Fe electrodeposition from $Fe_2O_3$, $Fe_{1.8}Al_{0.2}O_3$ and $Fe_{1.4}Al_{0.6}O_3$ ceramic suspensions. |
| *Expected size of the data (if known)* | Dataset #6 has a total size of 4297 KB. |
| *Data utility: to whom will it be useful* | The presented data is useful for screening the prospects for electrodeposition from various potential iron oxide feedstocks. |

### 2.3.7 Dataset #7 summary

The Dataset#7 refers to the data generated during laboratory experiment performed in the Studies of potential impacts provided by the presence of magnesium in iron oxide on iron electrodeposition from ceramic powder suspensions under alkaline conditions. This work was realized by the University of Aveiro.

| | DATASET#7 DESCRIPTION (DOI: 10.1149/1945-7111/AC1490) |
|---|---|
| *Purpose of the data collection / generation* | Studies of potential impacts provided by the presence of magnesium in iron oxide on iron electrodeposition from ceramic powder suspensions under alkaline conditions. |
| *Relation to the objectives of the project* | The presented data allows one to understand if alternative iron oxide feedstocks rich in -Mg content can be used as a raw material for Fe electrodeposition. In this case, the iron oxide-based magnesium spinel allows one to study typical iron-rich waste found in the metallurgical industries. |
| *Types and formats of data generated / collected* | The data was obtained by electrochemical tests performed by a potentiostat and recorded as .xlsx files. This data was recalculated, taking into account the working electrode area. XRD data (.xlsx) and SEM (.jpg) images were collected according to each type of analysis. |
| *If existing data is being re-used (if any)* | The corresponding experimental conditions have never been tried so far. |
| *Origin of the data* | The data was obtained by experimental tests of Fe electrodeposition using $Fe_{2.3}M_{0.7}O_4$ bulk cathodes and suspensions. |
| *Expected size of the data (if known)* | Dataset #2 has a total size of 5112 KB. |
| *Data utility: to whom will it be useful* | The presented data is useful for screening the prospects for electrodeposition from various potential iron oxide feedstocks. |

## 2.4 FAIR principles in the SIDERWIN project

The SIDERWIN consortium is committed to make its data reusable in the conditions presented in this report. From the FAIR principles, guidelines can be derived to be applicable to all datasets.

### 2.4.1 Findable

The first condition of reusability of the project datasets is the capacity to identify and discover them whatever the motivation or the point of entry. Then, the first step is to assign a unique DOI to the public part of the datasets when they will be stored in a repository.

A naming convention and standard identification mechanism of the datasets for persistent identifiers will be decided among the project consortium partners in a later stage of the project. Similarly, the keywords to define the metadata accompanying the datasets will be chosen to be meaningful and assist in their discovery. They will be selected firstly to describe, contextualize the content of the dataset but also to cover the wide range of fields and communities that may be interested in the project results.

To increase their findability, the public part of the datasets will be associated with links to their related scientific articles, public deliverables, or conference presentations.

### 2.4.2 Accessible

When identified and located, the public part of the datasets should be accessible to any external users. The visibility of the datasets is firstly addressed in the dissemination actions of the project. For example, by the multiplicity of links to websites such as commercial publishers, patent offices, social media, partners webpages, academic social networks, or researchers' personal libraries such as ResearchGate (http://www.researchgate.net) and Mendeley (https://www.mendeley.com).

The accessibility of the public part of the datasets is maximized by open access practices endorsed by the project consortium and described in this report. Free-access to peer-reviewed and Open data are the most significant contributions to accessibility.

### 2.4.3 Interoperable

When accessible, the datasets should be readable, and the data be processed with ordinary computer software. In the SIDERWIN project, concerning data generation from experimental activities, the data format is Comma-Separated Values (CSV) time series written as .txt open files. Provided properly descriptive column headers and physical dimensions indications, these data are easily treated by common spreadsheet software. The interoperability of the data from modelling activities will rely on the associated documentation so that the data are understood by non-experts.

### 2.4.4 Reusable

The datasets will be reusable if there are fundable, accessible, and interoperable in the long term. The question of long-term storage will be addressed during the choice of the repository. It is current practice to guarantee a 5-year access after the end of the project. At this stage of the project, embargo period and other restriction for re use is not foreseen.

In case of third-party interest in the data, the licensing policy described in the GA and CA will be applied.

### 2.4.5 Allocation of resources

The costs related to the data generation and processing are covered by the project budget.

### 2.4.6 Data security

The selection of the repository has been based on its high data security standards.

### 2.4.7 Ethical aspects

There is no privacy sensitive information produced or collected during the project execution.

### 2.4.8 Other

Regarding confidentiality, all partners keep any data, documents or other material that is identified as confidential during the implementation for the project and for four years after the period, cf. Article 36 of GA.

# 3 Conclusions

This DMP has presented the obligations of the SIDERWIN project in terms of data management. It has provided the guidelines for the consortium partners to deal with the different types of data which have been generated during the lifespan of the project.

The SIDERWIN project is characterized by a significant academic activity which has produced scientific articles in Open Access and by potential commercial applications in the mid-term that has been protected by IPR.

The main sources of data generation and the corresponding datasets have been described in the first version of the DMP and actualized in this final version. The final version of the DMP has presented the chosen Data repository, Zenodo, used to preserve the publicly available data generated during the project execution, to assign Digital Object Identifier to existing project open access documents such as conference presentations, and to update datasets and to undertake protection by IPR to limit the time of embargo on the results generated by the project.

# References


[1] European Commission Directorate-General for Research & Innovation. 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

[2] The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data | 3:160018 | DOI: 10.1038/sdata.2016.18

3 European Commission Directorate-General for Research & Innovation. Guidelines on FAIR Data Management in Horizon 2020 Version 3.0 26 July 2016

[4] SIDERWIN (2017). Grant Agreement No. 768788. Project Title: Development of new methodologies for industrial CO2-free steel production by electrowinning

[5] SIDERWIN Consortium Agreement (2017). Research and Innovation action No. 768788. Project Title: Development of new methodologies for industrial CO2-free steel production by electrowinning

[6] Cédric Raoul, SIDERWIN Deliverable: D1.3 Title: Collaborative Portal

[7] https://www.siderwin-spire.eu/

[8] Monica Serna Ruiz, SIDERWIN Deliverable: D8.2 v1.0 Title: Master Dissemination and Communication Plan and Updates