Rare Earth Free e-Drives Featuring Low Cost Manufacturing

ReFreeDrive

Collaborative Project
Grant Agreement Number 770143

Start date of the project: 1st October 2017, Duration: 36 months

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770143

Deliverable no.: D 9.2

Title of the deliverable: Open Data Management Plan

Contractual Date of Delivery:	**31/03/2018**
Actual Date of Delivery:	**04/03/2018**
Lead contractor for this deliverable:	**CIDAUT**
Author(s):	**Blanca Araujo (CIDAUT), Javier Romo (CIDAUT), Alicia Rodríguez (CIDAUT)**
Work package contributing to the deliverable:	**WP9**
Nature:	**PU**
Version:	**04**

REVISION TABLE		
Document version	**Date**	***Modified sections - Details***
V01	23.03.2018	First Full Version
V02	28.03.2018	Changes to new deliverable template
V03	03.03.2018	Reviewed version by Serge Noels (Quality Manager)
V04	04.03.2018	New version with feedback changes

Table of Contents

Abbreviations

API	Application Programming Interface
CERN	European Organization for Nuclear Research (Conseil Européen pour la Recherche Nucléaire)
DoA	Description of the Action
DOI	Digital Object Identifier
FAIR	Findable, Accesible, Interoperable and Reusable
HTTP	Hypertext Transfer Protocol
IPR	Intellectual Property Rights
IT-CDA-DR	IT Department, Collaboration Devices and Applications Group, Digital Repositories Section
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LCA	Life Cycle Analysis
LHC	Large Hadron Collider
NVH	Noise, Vibration, Harshness
OAI-PMH	Open Archive Initiative-Protocol for Metadata Harvesting
ODMP	Open Data Management Plan
OpenAIRE	Open Access Infrastructure for Research in Europe
REST	Representational state transfer
URL	Uniform Resource Locator
WP	Work Package

Executive Summary

This document, D9.2 Open Data Management Plan (ODMP) main objetive is to collect, analyse and share open motor design, characterization and testing data and experience to validate and de-risk future industrial innovations. This objective has been addresed in this document and there have been no deviations in content or time from the deliverable objectives set out in the ReFreeDrive Grant Agreement. The data gathering and management will be a continuous action throughout the duration of the project.

The Consortium strongly believes in the concepts of open science, and in the benefits that the European innovation ecosystem and economy can draw from allowing the reuse of data at a larger scale. Besides, the share and reuse of research data to the electric machine design research community will eliminate barriers and enforece an innovation culture.

The purpose of the ODMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project open research data.

The ODMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse.

Research data linked to exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection, which is a H2020 obligation. The rest of research data will be deposited in an open access repository.

The ODMP is not a fixed document; on the contrary it will evolve during the lifespan of the project. This first version of the ODMP includes an overview of the datasets to be produced by the project, and the specific conditions that are attached to them. The next versions of the ODMP will get into more detail and describe the practical data management procedures implemented by the ReFreeDrive project.

The expected types of research data that will be collected or generated along the project will be discussed following the project work package structure.

1 Introduction

Open access is defined as the practice of providing on-line access to scientific information that is free of charge to the reader and that is reusable. In the context of research and innovation, scientific information can refer to peer-reviewed scientific research articles or research data.

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

Nevertheless, data sharing in the open domain can be restricted as a legitimate reason to protect results that can reasonably be expected to be commercially or industrially exploited. Strategies to limit such restrictions will include anonymising or aggregating data, agreeing on a limited embargo period or publishing selected datasets. It must be duly noted that the automotive industry is highly competitive and ReFreeDrive project aims at providing its industrial partners with added value innovation, which would not be such if made public.

1.1 Purpose

The purpose of the ODMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data.

The ODMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. Figure 1 shows the research data life cycle, taken from [1], which has been used as guideline for this deliverable.

The ODMP is not a fixed document, but will evolve during the lifespan of the project, particularly whenever significant changes arise such as dataset updates or changes in Consortium policies.

This document is the first version of the ODMP, delivered in Month 6 of the project. It includes an overview of the datasets to be produced by the project. The ODMP will be updated in month 18 if needed and again at the end of the project.

This document has been produced following the EC guidelines for project participating in this pilot and additional consideration described in ANNEX I: KEY PRINCIPLES FOR OPEN ACCESS TO RESEARCH DATA.

Figure 1. Reserach Data Life Cycle

1.2 Research data types

For this first release of ODMP, the data types that will be produced during the project are focused on the Description of the Action (DoA) and on the results obtained in the first months of the project.

According to such consideration, Table 1 reports a list of indicative types of research data that each of the ReFreeDrive work packages will produce. This list may be adapted with the addition or removal of datasets in the next versions of the ODMP to take into consideration the project developments. A detailed description of each dataset is given in the following sections of this document.

Table 1. Work pacakges and expected datasets of the ReFreeDrive project

#	Work Package	Lead Partner	Expected Datasets
1	Project Management	CIDAUT	None
2	Boundary Conditions	PRIVÉ	KPIs, driving cycles and boundary conditions values
3	Induction Machine Design	MDL	Design Simulation results
4	Synchronous Reluctance Machine Design	IFPEN	Design Simulation results
5	e-Drive Design	PRIVÉ	Control Simulation results
6	Prototype Manufacturing	UAQ	Prototype Pictures
7	Powertrain Testing, Vehicle integration and Validation	CIDAUT	Test Results
8	Techno Economic Evaluation and Exploitation	JLR	Environmental assessment (LCA results)

9	Dissemination and Communication		UAQ	None

Specific datasets may be associated to scientific publications (i.e. underlying data), public project reports and other raw data or curated data not directly attributable to a publication. The policy for open access are summarised in the following Figure 2.

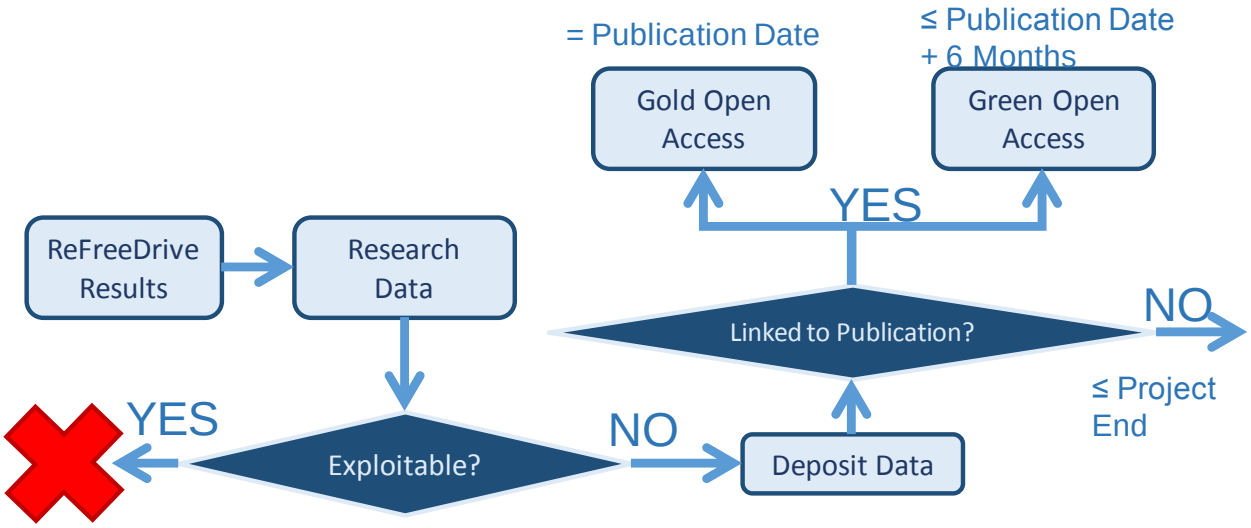

Figure 2. ReFreeDrive timing of different options

Research data linked to exploitable results will not be put into the open domain if they compromise its commercialization prospects or have inadequate protection, which is a H2020 obligation. The rest of research data will be deposited in an open access repository.

When the research data is linked to a scientific publication, the provisions outlined in the Grant and Consortium agreements will be followed. Research data needed to validate the results presented in the publication should be deposited at the same time for "Gold" Open Access[1] or before the end of the embargo period for "Green" Open Access[2]. Underlying research data will consist of selected parts of the general datasets generated, and for which the decision of making that part public has been made.

Other datasets will be related to any public report or be useful for the research community. They will be selected parts of the general datasets generated or full datasets and be published as soon as possible.

[1] **"Gold" Open Access:** Authors make a one-off payment to the publisher so that the scientific publication is immediately published in open access mode.

[2] **"Green" Open Access:** Due to the contractual conditions of the publisher, the scientific publication can undergo an embargo period up to six months since publication date before the author can deposit the published article or the final peer-reviewed manuscript in open access mode.

1.3 Responsibilities

Each ReFreeDrive partner has to respect the policies set out in this ODMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

The Project Coordinator has a particular responsibility to ensure that data shared through the ReFreeDrive website are easily available, but also that backups are performed and that proprietary data are secured. CIDAUT, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Validation and registration of datasets and metadata will be done by CIDAUT in close collaboration with the Work Package Leader generating the respective datasets. Metadata constitutes an underlying definition or description of the datasets, and facilitate finding and working with particular instances of data.

Backing up data for sharing through open access repositories will be done by CIDAUT.

Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publically available data. WP1 will provide naming and version conventions.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result, as outlined in the Grant and Consortium Agreements of this project.

2 Data Sharing

Relevant datasets will be stored in ZENODO [2] , which is the open access repository of the Open Access Infrastructure for Research in Europe, OpenAIRE [3]. ZENODO builds and operates a simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities. ZENODO enables researchers, scientists, EU projects and institutions to: easily share the long tail of small research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science. display their research results and get credited by making the research results citable and integrate them into existing reporting lines to funding agencies like the European Commission. easily access and reuse shared research results.

Data access policy will be unrestricted since no confidentiality or Intellectual Property Rights (IPR) issues are expected regarding the environmental monitoring datasets. All collected datasets will be disseminated without an embargo period unless linked to a green open access publication. Data objects will be deposited in ZENODO under:

- Open access to data files and metadata and data files provided over standard protocols such as HTTP and Open Archive Initiative-Protocol for Metadata Harvesting (OAI-PMH).
- Use and reuse of data permitted.

- Privacy of its users protected.

2.1 Findable, Accessible, Interoperable, Reusable (FAIR) Principles

FAIR Principles definition as referenced from FAIR principles description [4].

2.1.1 To be Findable:

- **F1**: (meta)data are assigned a globally unique and persistent identifier
  - A Digital Object Identified (DOI) is issued to every published record on Zenodo.
- **F2**: data are described with rich metadata (defined by Reusable R1 principle below)
  - Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichements. The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions.
- **F3**: metadata clearly and explicitly include the identifier of the data it describes
  - The DOI is a top-level and a mandatory field in the metadata of each record.
- **F4**: (meta)data are registered or indexed in a searchable resource
  - Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing.
  - Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

2.1.2 To be Accessible:

- **A1**: (meta)data are retrievable by their identifier using a standardized communications protocol
  - Metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name.
  - Metadata is also retrievable through the public Representational state transfer (REST) Application Programming Interface (API) API.
- **A1.1**: the protocol is open, free, and universally implementable
  - See point A1. OAI-PMH and REST are open, free and univesal protocols for information retrieval on the web.
- **A1.2**: the protocol allows for an authentication and authorization procedure, where necessary
  - Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.
- **A2**: metadata are accessible, even when the data are no longer available
  - Data and metadta will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
  - Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

2.1.3 To be Interoperable:

- **I1**: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
    - Zenodo uses JavaScript Object Notation (JSON) Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML. The Dublin Core Schema is a small set of vocabulary terms that can be used to describe digital resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks. The full set of Dublin Core metadata terms can be found on the Dublin Core Metadata Initiative website. MARCXML is an XML schema based on the common MARC21 standards. MARCXML was developed by the Library of Congress and adopted by it and others as a means of facilitating the sharing of, and networked access to, bibliographic information. Being easy to parse by various systems allows it to be used as an aggregation format, as it is in software packages.
- **I2**: (meta)data use vocabularies that follow FAIR principles
    - For certain terms Zenodo refers to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE).
- **I3**: (meta)data include qualified references to other (meta)data
    - Each referrenced external piece of metadata is qualified by a resolvable URL.

2.1.4 To be Reusable:

- **R1**: (meta)data are richly described with a plurality of accurate and relevant attributes
    - Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.
- **R1.1**: (meta)data are released with a clear and accessible data usage license
    - License is one of the mandatory terms in Zenodo's metadata, and is referring to a Open Definition license.
    - Data downloaded by the users is subject to the license specified in the metadata by the uploader.
- **R1.2**: (meta)data are associated with detailed provenance
    - All data and metadata uploaded is tracable to a registered Zenodo user.
    - Metadata can optionally describe the original authors of the published work.
- **R1.3**: (meta)data meet domain-relevant community standards
    - Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

2.2 Archiving and Preservation

Zenodo is hosted by CERN which has existed since 1954 and currently has an experimental programme defined for the next 20+ years. CERN is a memory institution for High Energy Physics and renowned for its pioneering work in Open Access. Organisationally Zenodo is embedded in the

IT Department, Collaboration Devices and Applications Group, Digital Repositories Section (IT-CDA-DR).

Zenodo is offered by CERN as part of its mission to make available the results of its work (CERN Convention, Article II, §1 [5]).

Data files and metadata are backed up nightly and replicated into multiple copies in the online system.

2.2.1 Data storage

All files uploaded to Zenodo are stored in CERN's EOS service[3] in an 18 petabytes disk cluster. Each file copy has two replicas located on different disk servers.

For each file Zenodo stores two independent MD5[4] checksums. One checksum is stored by Invenio[5][6], and used to detect changes to files made from outside of Invenio. The other checksum is stored by EOS, and used for automatic detection and recovery of file corruption on disks.

Zenodo may, depending on access patterns in the future, move the archival and/or the online copy to The CERN Advanced STORage manager (CASTOR) [7] in order to minimize long-term storage costs.

EOS is the primary low latency storage infrastructure for physics data from the Large Hadron Collider[6] (LHC) [8] and CERN currently operates multiple instances totalling 150+ petabytes of data with expected growth rates of 30-50 petabytes per year. CERN's CASTOR system currently manages 100+ petabytes of LHC data which are regularly checked for data corruption.

Invenio provides an object store like file management layer on top of EOS which is in charge of e.g. version changes to files.

3 Datasets Description

The Table 2 refers to each of the datasets that will be produced during the project, their description and importance to the project.

[3] EOS is a disk-based, low-latency storage service. Having a highly-scalable hierarchical namespace, and with data access possible by the XROOT protocol, it was initially used for physics data storage. Today, EOS provides storage for both physics and user use cases. Instances of EOS include EOSUSER, EOSPUBLIC, EOSATLAS, EOSCMS.

[4] The MD5 algorithm is a widely used hash function producing a 128-bit hash value. Although MD5 was initially designed to be used as a cryptographic hash function, it has been found to suffer from extensive vulnerabilities. It can still be used as a checksum to verify data integrity, but only against unintentional corruption.

[5] Invenio is an open source software library management package that provides the tools for management of digital assets in an institutional repository. The software is typically used for open access repositories for scholarly and/or published digital content and as a digital library.
Invenio is developed by the CERN Document Server Software Consortium and is freely available for download.

[6] The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. It first started up on 10 September 2008, and remains the latest addition to CERN's accelerator complex.

Table 2. Datasets generated by the ReFreeDrive Project

Who (WPs generating the dataset)	What (Dataset description)	Why (Importance of this dataset)	How (use of this dataset in the project)
WP2	KPIs: Project targets at vehicle levels	These figures set the design space for the project electric motors	These values will drive the different designs
WP2	Full Subsystems Technical specifications	Assigned boundary conditions at the subsystem level	These values will drive the different designs and lead the in vehicle integration activities
WP3	Simulation Results: Electromagnetic, mechanical, thermal, Noise Vibration, Harshness (NVH)	Induction Machine design expected result will help comparisons with other technologies	This dataset will be the basis for at least one scientific publication.
WP4	Simulation Results: Electromagnetic, mechanical, thermal, NVH	Synchronous Reluctance design expected results will help comparisons with other technologies	This dataset will be the basis for at least one scientific publication.
WP3 & WP4	Material characterization values	Grain oriented and non grain oriented materials magnetic and mechanical performance will help other designers reuse this knowledge	This information will be used for design purposes. This dataset will be the basis for at least one scientific publication.
WP5	Control algorithm design results	Implementation of control strategies or innovation in control strategies	This dataset will be the basis for at least one scientific publication. It will drive the power electronic configuration.
WP6	Prototype Manufacuring Pictures	Comparison with other technologies, technology demonstration feasibility	The project will use these pictures for communication purposes
WP7	Motor Test Results: test results for the integrated e-Drive	These data will enable a comparison with other technologies and help designers set ambitious targets in future designs	This dataset will be the basis for at least one scientific publication. The project will use these data for the techno economic evaluation and exploitation strategies
WP7	Powertrain test results: results in the powertrain test	These data will enable a comparison with other technologies and help	This dataset will be the basis for at least one scientific

	bench	designers set ambitious targets in future designs	publication. The project will use these data for the techno economic evaluation and exploitation strategies
WP7	Vehicle driving test results	These data will enable a comparison with other technologies at the vehicle level	This dataset will be the basis for at least one scientific publication. The project will use these data for the techno economic evaluation and exploitation strategies
WP7	Vehicle integration pictures	Technology demonstration feasibility	The project will use these pictures for communication purposes
WP8	Life Cycle Analysis (LCA) results: environmental assessment and comparatives of the studied technologies	LCA data are used throughout the electric vehicle market for marketing, communication, and new designs comparative evaluations.	LCA will be key to demonstrate the rare earth free environmental advantages poised by the ReFreeDrive technologies

4 ANNEX I: KEY PRINCIPLES FOR OPEN ACCESS TO RESEARCH DATA

These principles can be applied to any project that produces, collects or processes research data. As indicated in Guidelines on Data Management in H2020 [9], scientific research data should be easily:

1. Discoverable

The data and associated software produced and/or used in the project should be discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)

2. Accessible

Information about the modalities, scope and licenses (e.g. licensing framework for research and education, embargo periods, commercial exploitation, etc.) in which the data and associated software produced and/or used in the project is accessible should be provided.

3. Assessable and intelligible

The data and associated software produced and/or used in the project should be assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. the minimal datasets are handled together with scientific papers for the purpose of peer review, data are provided in a way that judgments can be made about their reliability and the competence of those who created them).

4. Useable beyond the original purpose for which it was collected

The data and associated software produced and/or used in the project should be useable by third parties even long time after the collection of the data (e.g. data are safely stored in certified repositories for long term preservation and curation; they are stored together with the minimum software, metadata and documentation to make it useful; the data are useful for the wider public needs and usable for the likely purposes of non-specialists).

5. Interoperable to specific quality standards

The data and associated software produced and/or used in the project should be interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins).

5 Bibliography

[1] OPERA ”Open Sea Operating Experience to Reduce Wave Energy Costs” Deliverable D8.5 Data Management Plan (first version), July 2016

[2] https://zenodo.org

[3] https://www.openaire.eu/

[4] FAIR principles description: Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

[5] https://council.web.cern.ch/en/content/convention-establishment-european-organization-nuclear-research#2

[6] http://invenio-software.org/

[7] http://castor.web.cern.ch/

[8] https://home.cern/topics/large-hadron-collider

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