

Project acronym: **PASSENGER**

Project title: “**Pilot Action for Securing a Sustainable European Next Generation of Efficient RE-free magnets**”

Grant agreement number: **101003914**

Start date of project: **1 May 2021**, Duration: 48 months



Deliverable No. 7.6

Final Data Management Plan

Due date of deliverable	31/10/2021
Submission date	26/11/2021
File Name	D7.6 PASSENGER_name of deliverable
Organisation Responsible of Deliverable	IMDEA Nanociencia
Author name(s)	Mark Davies, Alberto Bollero, Ignacio Torres
Revision number	03
Status	Draft/Final ¹
Dissemination Level	PU ²

¹ Document will be a draft until it is approved by the coordinator

² 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)



PARTNERS

Fundación IMDEA Nanociencia
Metalpine GmbH
Fundación ICAMCYL
Ingeniería Magnética Aplicada SL
MBN Nanomaterialia SPA
MATRES Material Research & Design
Kolektor Group d.o.o.
Centro Ricerche FIAT SCPA
EIT Raw Materials GmbH
Technical University Darmstadt
European Science Foundation
Industrie Ilpea SPA
OSLV Italia SRL
Asociación Española de Normalización
Balrog Plastics GmbH
MNLT Innovations PC
Tizona Motors SL
Jožef Stefan Institute
Smart Waste Engineering SRL
WILO SE
Less Common Metals Ltd



REVISION HISTORY

Version	Date	Modified by	Comments
Draft1.0	22/07/2021	Ignacio Torres & Mark Davies	
Version1	29/11/2021	Ignacio Torres	Sent for review to all partners

ACKNOWLEDGEMENT AND DISCLAIMER



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101003914.

The contents of this publication are the sole responsibility of its author and do not necessarily reflect the opinion of the European Union.

LIST OF ACRONYMS

Abbreviation	Meaning
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Re-usable
OA	Open Access
RE	Rare Earths
PM	Permanent Magnets
DOI	Digital Object Identifier
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
ISBN	International Standard Book Number
TRL	Technology Readiness Levels
LCA	Life Cycle Analysis
LCC	Low-Cost Carrier



SUMMARY

The purpose of the DMP is to provide an overview of the main elements of the data management policy that are used by the Consortium with regard to the project research data. The DMP is not a fixed document but has been evolving during the lifespan of the project.

The DMP covers the complete research data life cycle of the PASSENGER project. It describes the types of research data that were generated during the project, the strategies on research data preservation and the provision on access rights. The research data should be “FAIR”, that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation solution.

The repository ZENODO has been chosen by default to store, classify and provide Open Access (OA) to the data objects originated within the PASSENGER project frame, but this choice is not exclusive and the Partners can choose their own Open Access repositories to store their data. We summarise all the repositories the Partners have chosen to make openly accessible their results in Section 2.2.



TABLE OF CONTENTS

SUMMARY.....	4
TABLE OF CONTENTS.....	5
1. DATA SUMMARY.....	6
Purpose of data collection	6
Relation to the objectives of the project	6
Types and formats of data.....	6
Re-using of data.....	6
Origin of data.....	6
Expected size.....	7
Data utility	7
2. FAIR DATA.....	7
2.1. MAKING DATA FINDABLE.....	8
Metadata provision	8
Clear versioning	8
Standard identification mechanism.....	9
Naming conventions.....	9
Keywords	9
2.2. MAKING DATA OPENLY ACCESSIBLE	9
Which data will be made openly available	9
How the data will be made available	10
Methods or software tools needed to access the data	10
Where the data documentation and code are deposited.....	10
Restrictions	11
2.3. MAKING DATA INTEROPERABLE	12
2.4. DATA RE-USE.....	12
Licence to the data.....	12
When the data will be made available for re-use	13
Third parties and re-usability	13
Data quality assurance processes.....	13
Length of time for which the data will remain re-useable	13
3. ALLOCATION OF RESOURCES	13
Estimated costs for making the data FAIR.....	13
Responsibilities for data management.....	14
Value of long term preservation.....	14
4. DATA SECURITY	14
5. REFERENCES.....	14



1. DATA SUMMARY

Purpose of data collection

The purpose for data collection is to capture qualitative evidence for their analysis to lead to the formulation of answers to the questions that concern the PASSENGER project. Accurate data collection will give researchers the ability to answer relevant research questions and facilitate the repetition and validation of the study to other researchers.

Relation to the objectives of the project

Data collection is essential for the achievement of the objectives of the project, which in first place are: maintaining the integrity of research, making informed decisions and ensuring quality assurance. In a second stage, the objectives of disseminating and communicating the outcomes will be benefited by the correct and systematic acquisition of data.

Types and formats of data

The types and formats of data acquired within the PASSENGER project frame are the following:

- Laboratory data: datasets (*.txt), images (*.jpg)...etc.
- Research data: statistics (*.xlsx), graphs (*.ogg), bibliography (*.enl)...etc.
- Scientific texts: manuscripts (*.docx), presentations (*.pptx), books (*.pdf)...etc.
- Dissemination material: fact-sheets, images, videos...etc.
- Management documents: deliverables, patents.

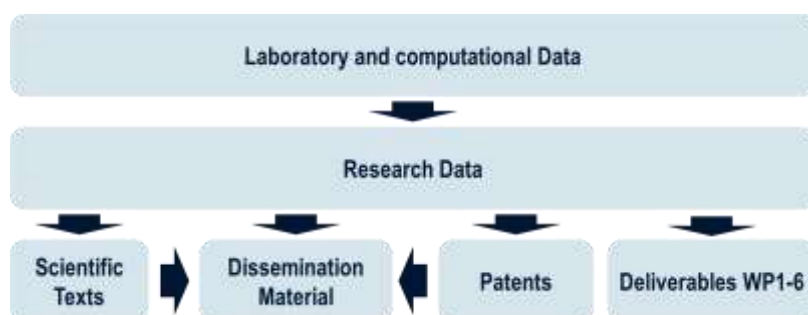
A complete list of all data types and formats is listed in Table 1 of the Annex.

Re-using of data

The PASSENGER Consortium has, to date, no plan to re-use any previous data other than the expertise and knowledge defined from the scientific background of the partners. Whenever any set of data is being re-used, it will be appropriately notified.

Origin of data

There is a hierarchical origin of the data, schematized in the flowchart below.



- **Laboratory and computational data** are the keysets of data obtained directly from the research setups (digital or analogic sources). The sources will be identified in the deposited datasets.
- The **research data** is processed data from the laboratory and computational data and is composed of charts, diagrams, statistics...etc.
- The **scientific texts** and **dissemination material** is prepared from the research data. Author-accepted manuscripts, media (images, videos, gifs), leaflets etc. are included in this category.
- **Patents** will be prospectively a product of the PASSENGER research data.
- **Deliverables** of the work packages (WP) 1 to 4 are reports for the European Commission and are based on the research data.

Expected size

The size of the data will be attempted to be as low as possible with the purpose of facilitating its storage and exchange. In any case, a single file should not exceed the upload limit of each repository (50 GB for Zenodo). The expected size of the data is listed in Table 2 of the Annex.

Data utility

The data collection within the frame of the PASSENGER project will be useful during the lifespan of the project, as well as in the posterior period. The data will be valuable for different types of audiences, depending on the kind of data:

- Laboratory and computational data is the core of the investigation and the source of posterior research data such as statistics and graph plots. Researchers of the PASSENGER project will make use of this type of data.
- The research data is processed data and is readily available for the scientific community. The researchers of the PASSENGER project will use the research data to elaborate scientific texts, disseminating material, patents (if any) and deliverables. Also, the research data will be useful for validating the results and re-use in future research projects, beyond the PASSENGER project.
- The dissemination material is intended to reach all kind of audiences (including general public, scientific community and stakeholders), with the purpose of enhancing and spreading information, and establishing communication. In a reciprocal exchange, the general public and the scientific community will benefit from this material, as well as the PASSENGER Consortium.
- Deliverables are reports aimed to the European Commission to track the progress of the project, as well as to the general public in the case of public deliverables.
- The benefactors of a prospective patent product of the PASSENGER project research are the prospective stakeholders and, as a last resort, the society as a whole, and more specifically industrial sectors that use permanent magnets and their users, and, as a last resort, the environment due to reduced consumption of rare earths.

2. FAIR DATA

In compliance with the European Commission guidelines, the data generated by the PASSENGER project must be FAIR, that is findable, accessible, interoperable and re-



usable. The decision to be taken by the project on how to publish its documents and data sets will come after the more general decision on whether to go for an academic publication directly or to seek first protection by registering the developed Intellectual Property.

2.1. MAKING DATA FINDABLE

All the records deposited in the Zenodo repository are indexed immediately in OpenAIRE, which is the Open Access Infrastructure for Research in Europe. OpenAIRE does this by aggregating European funded research output from nearly 1000 repositories from all over the world and makes them available via the OpenAIRE portal. Records indexed in OpenAIRE will be immediately available in the European Commission Participant Portal.

Metadata provision

The data generated under the PASSENGER frame will be discoverable, identifiable and locatable by means of suitable metadata. Descriptive metadata refers to the information about the objects content. Zenodo and other OA repositories offer the possibility to assign several tags (metadata) to all uploads, in order to make the content findable, in compliance with the OAI-PMH protocol. The data of PASSENGER will be recorded using the following tags:

Publication type (journal article, presentation, book, thesis...etc.).

- Title, authors, affiliation.
- Description of the content. Link to the publication or other associated research datasets.
- Communities that the data belongs to.
- Grants which have funded the research.
- Identifiers (DOI, ISSN, PubMed ID, URLs...etc.).
- Contributors.
- References.
- Journal name, volume, issue and pages; in the case of a manuscript.
- Conference title, place, session...etc.; in the case of a conference proceeding.
- Publisher, place, ISBN, pages; in the case of parts of books and reports.
- Awarding university and supervisors; in the case of theses.

All metadata is stored in Zenodo repository internally in JSON-format according to a defined JSON schema. Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the OpenAIRE Guidelines). Access to metadata and data files is provided over standard protocols such as HTTP and OAI-PMH.

Clear versioning

The version of the data which is aimed to be deposited will be the final version. The digital object identifiers (DOIs) are automatically generated upon deposition on the repository. If necessary, posterior versions will be deposited; these posterior versions will be identified by their own DOI and identifiable by the date of deposition and file name.

Including semantic information such as the version number in a DOI will not be encouraged, because this information may change over time, while DOIs must remain persistent. Most



importantly, version suffixes are not machine readable. Zenodo DOI versioning is linear, which means that the Zenodo version number may in fact not be the real version number of the resource.

The approach of PASSENGER to the versioning is the one Zenodo provides: two versions (two DOIs) are semantically linked in the metadata of a DOI. This ensures that discovery systems have a machine readable way to discover that two DOIs are versions of the same resource.

Standard identification mechanism

All deposited data will be uniquely identifiable through the standard identifier DOI. Additionally, other identifiers, such as Handle, ARK, PURL, ISSN, ISBN, PubMed ID, ORCID, PubMed Central ID, ADS Bibliographic Code, arXiv, Life Science Identifiers (LSID), EAN-13, ISTC, URNs and URLs may be used.

Data generated under the PASSENGER frame will acknowledge the grant in the following way:

“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101003914”

so can automatically be indexed in the OpenAIRE portal.

Naming conventions

Zenodo DOI versioning is linear. Currently DOIs registered by Zenodo follow the pattern “10.5281/zenodo.”, where 10.5281 is the Zenodo DOI prefix and is a sequentially assigned integer. The word “zenodo” is semantic information, and, as mentioned in the previous section, to include semantic information in DOIs it is not supported by PASSENGER as it may change over time. The current practice was introduced when Zenodo was launched, and while it is not ideal Zenodo did not want to change the existing practice.

Keywords

All deposited data will have an associated group of keywords to facilitate identification.

2.2. MAKING DATA OPENLY ACCESSIBLE

Open Access must be granted to all scientific publications resulting from Horizon 2020 actions. This will be done in accordance with the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (21 March 2017).

Which data will be made openly available

Any dissemination data linked to exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection.

Categories of outputs that PASSENGER will give Open Access (free of charge) include:

- Scientific publications (author accepted manuscripts, supplementary files, and conference proceedings).



- Research Data (key datasets accompanying publications that are needed to validate the results).
- Deliverables (public).
- Dissemination material.

We will provide restricted access to the members of the Consortium only for templates (e.g. Power Point templates) and documents concerning internal meetings (e.g. kick-off).

Dissemination and outreach material will be openly available via the PASSENGER website and the network social sites.

How the data will be made available

The data will be available via Zenodo repository and other repositories from the Consortium Members' choice. The data which is owned by the Consortium will be deposited as soon as possible, in a repository with open access rights. In the case the data needs to be protected (for example, in the case of a future publication or patent), the data will be deposited with embargoed access. The embargo will be lifted after the data has been disclosed.

The author accepted manuscripts will be deposited as soon as possible, and at the latest after 6 months after publication. The preferred route to deposit the manuscripts will be the 'Green' model: the corresponding author or representative deposits the published article or final peer-reviewed manuscript in an online repository. Whenever the 'Green' route cannot be assured, the Consortium will provide open access following the 'Gold' route: an article is immediately released in Open Access mode by the scientific publisher upon publication; the payment of publication costs is shifted away from subscribing readers. Details on the publication routes followed in the PASSENGER project will be described in the Dissemination and Exploitation Plan (Deliverable 6.3).

All data from the PASSENGER project is available through the OpenAIRE portal:

https://explore.openaire.eu/search/project?projectId=corda_h2020::0e376418f930226f3ae57ba79d2f4389

Methods or software tools needed to access the data

The software tools necessary to access the data will be standard software tools and free of charge, such as Open Office or Adobe Acrobat Reader. Other standard programs could be also used, such as the Microsoft Office package. This well-used standard software is likely to be readable in the future.

In the case of pre-processed documents (LaTeX), the software necessary to run the programs is open and can be freely downloaded. Raw confocal microscopy image files (.LIF) could be visualized by using ImageJ and Fiji programs, free-of-charge software readily available. Processed microscopy images (.JPG, .TIFF) could be opened by using standard image viewer software.

In the case of research data files that cannot be opened with standard programs, the software necessary to access the data will be indicated in the section "Additional notes", upon deposition in the repository.

Where the data documentation and code are deposited



The Consortium has chosen Zenodo as the default central scientific publication and data repository for the project outcomes. This online repository has been created through the European Commission's OpenAIREplus project and is hosted at CERN.

The Zenodo community PASSENGER has been specifically created to gather all data under the frame of the PASSENGER project:

<https://zenodo.org/communities/>

Additionally, the public data will be included in the European Commission Funded Research (OpenAIRE) Zenodo community (<https://zenodo.org/communities/ecfunded/>), which is curated by Zenodo.

Also, the Consortium members have deposited their results in other Open Access repositories, such as:

- IMDEA Nanociencia institutional repository. The institutional Repository of IMDEA Nanociencia is the digital archive that collects the scientific production of the institute, with the aim of preserving and disseminating research results in open access. <https://repositorio.imdeananociencia.org/>

The manuscripts can also be collected from external services such as:

- Microsoft Academic Graph. It is a heterogeneous graph containing scientific publication records, citation relationships between those publications, as well as authors, institutions, journals, conferences, and fields of study.
- Datacite.
- Crossref.
- Pubmed Central.
- DOAJ Articles. The Directory of Open Access Journals.

Restrictions

The deposited data could be accessible under four types of **access rights**:

- Open Access.
- Embargoed Access.
- Restricted Access.
- Closed Access.

The data which is owned by the Consortium will be deposited as soon as possible, in the repository with **open access** rights. In case there is any **embargo** period on the deposited data, the access to the data will be granted after the embargo period, and at the latest after 6 months after publication. In the case of author-accepted manuscripts with an embargo period longer than 6 months, the Gold Open Access route will be followed.

Restricted and closed accesses are not considered under the PASSENGER project frame. We will provide restricted access to the members of the Consortium only for templates (e.g. Power Point templates) and documents concerning internal meetings (e.g. kick-off), which are not public.



2.3. MAKING DATA INTEROPERABLE

Data from PASSENGER project will be provided, whenever is possible, in open software formats. To facilitate the interoperability, the data will be aimed to be provided in standard formats such as .pdf, .doc, .xls, .jpg, .txt or any other machine-readable format.

Data sets provided by the partners will contain standard vocabulary to facilitate interdisciplinary interoperability. If abbreviations are used, they will be properly defined in the data files to allow complete understanding of the results. Abbreviations conventionally used in this type of data set will include measurement units (e.g., weight, volume, length), PE (permanent magnets), RE (rare earths), technological abbreviations (e.g., TLR, LCA, LCC).

2.4. DATA RE-USE

Data re-use is subject to the **license** under which the data objects are deposited.

Licence to the data

The Zenodo repository offers five types of licence types for the files under Open Access right:

Licence type - Creative Commons Attribution	The licensor allows to copy, distribute and communicate publicly the work, as well as create and disseminate works derived from the former	The licensor allows the reproduction, dissemination and public communication of the work only for commercial purposes
4.0	✓	✓
Share-Alike 4.0	✓, only under the same or similar licence.	✓
NoDerivatives	✓, but it is not allowed to disseminate works derived from the original work.	✓
NonCommercial 4.0	✓	✗
NonCommercial-NoDerivatives	✓, but it is not allowed to disseminate works derived from the original work.	✗

The license chosen by PASSENGER will be a Creative Commons **Attribution-NonCommercial 4.0**. The characteristics of this licence are:

- The licensor allows to copy, distribute and communicate publicly the work, as well as create and disseminate works derived from the former.
- The licensor allows the reproduction, dissemination and public communication of the work only for non-commercial purposes.

By uploading content, no change of ownership is implied and no property rights are transferred to the repository owners.

Metadata will be licensed under CC0 (Creative Commons), except for e-mail addresses.



When the data will be made available for re-use

The data will be immediately available for re-use upon deposition. As we have stated in Section 2.2., the data deposited will be conferred Open Access rights on deposition and at the latest after 6 months.

Third parties and re-usability

The data produced may be used by third parties, since will be openly available during the lifetime of the repositories.

Data quality assurance processes

The cornerstone of digital preservation is data integrity: data is complete and unaltered as it was when it was originally recorded. All data files in Zenodo will be stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.

Length of time for which the data will remain re-useable

The PASSENGER project aims to make available the research outputs as long as possible after the end of the project (30.04.2025) and for, at least, the lifetime of the repositories in which the data has been deposited.

All data files deposited in Zenodo are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis. Items 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. In case of closure of Zenodo repository, CERN will make best efforts to integrate all content into suitable alternative institutional and/or subject based repositories.

Records can be retracted from public view; however, the data files and records are preserved and can be no longer changed or retrieved from Zenodo repository. The record's metadata can be always modified.

3. ALLOCATION OF RESOURCES

Estimated costs for making the data FAIR

The estimated cost of the article processing charges is in average 3 500 € per publication [3]. Considering an estimate of 6 publications per year, the total cost of making the data openly accessible for the PASSENGER project is about 84 k€.

The associated costs are covered by the author and/or co-authors of the publication as agreed in the PASSENGER Grant Agreement (eligible costs in Horizon 2020 projects).

PASSENGER is a Horizon 2020 project, eligible to publish in Open Research Europe, an open research publishing option that allows researchers to publish quickly, at no cost, and



in compliance with the H2020 Open Access mandate. Whenever is appropriate, authors will publish their manuscripts using this option.

Responsibilities for data management

Any member of the Consortium can upload content in any repository. The content in Zenodo and IMDEA Nanociencia Institutional Repository will be curated and approved by the coordinator of PASSENGER, IMDEA Nanociencia.

Uploaded items cannot be deleted in Zenodo. New versions of the content can be uploaded together with previous versions; all versions are simultaneously available.

Value of long term preservation

The value of long-term preservation is on ensuring and facilitating the accessibility and usability of the preserved data. It involves planning, resource allocation and application of the preservation methods that have been described in Section 2. The goal is the accurate reordering of authenticated content over time, so it remains usable as technological advances render original software obsolete.

4. DATA SECURITY

In this section, data recovery, secure storage and transfer of sensitive data is addressed.

All data files are stored in CERN data centres. CERN has considerable knowledge and experience in building and operating large scale digital repositories and a commitment to maintain this data centre to collect and store 100s of PBs of LHC data as it grows over the next 20 years. In the highly unlikely event that Zenodo will have to close operations, CERN guarantees the migration of all content to other suitable repositories, and since all uploads have DOIs, all citations and links to Zenodo resources will not be affected.

5. REFERENCES

- [1] Guidelines to the rules of Open Access to scientific publications and Open Access to research data in Horizon 2020. Version 3.2. (21 March 2017).
- [2] <http://help.zenodo.org/> (Consulted on August 1st, 2021).
- [3] R. van Noorden, Open access: The true cost of science publishing. *Nature* **495**, 426 (2013).
- [4] www.en.wikipedia.org/wiki/Digital_preservation (June 2017).

