Deliverable 5.1:

Data Management Plan

Self-Healing Soft Robotics

Call identifier: H2020-FET-2019

Grant Agreement: # 828818

Version Number:	V1.0
Authors:	Joost Brancart, VUB Bram Vanderborght, VUB Input from all partners
Date of preparation:	10/09/2019
Date of last modification:	25/11/2019
Expected delivery date:	30/11/2019
Actual delivery date:	NA
Notes:	NA

TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)

- **Instructions and footnotes in blue must not appear in the text.**
- **For options *[*in square brackets*]*: the option that applies must be chosen.**
- **For fields in [grey in square brackets] (even if they are part of an option as specified in the previous item): enter the appropriate data.**

Introduction

This Horizon 2020 DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. You should develop a single DMP for your project to cover its overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), you should clearly spell this out.

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

FAIR data management

In general terms, your 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.

This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept.

More information about FAIR:

FAIR data principles (FORCE11 discussion forum)

FAIR principles (article in Nature)

Structure of the template

The template is a set of questions that you should answer with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 1 contains a summary of these elements in bullet form.

This template itself may be updated as the policy evolves.

Table of Contents

1. Data Summary

The objective of the Data Management Plan is to describe the systematic approach for the management of the data recorded in all the experimental activities carried out during the project. In particular, this document presents the methodologies and guidelines that will be adopted for managing data collection, storage and dissemination, in view of long-term preservation of the generated data.

The goal of the SHeRo project is to research the fundamentals required for the development of fully autonomous self-healing soft robotic devices, by integrating engineered functional materials, smart sensing, active actuation and control capabilities into soft robots. These soft robotic systems will be able to sense and evaluate loss of performance and repair damage. Three **research objectives** (RO) are defined:

RO.1: Development, characterization and tuning of stimuli-responsive materials with smart, adaptive properties for self-healing robotic and sensory applications. Development and optimization of manufacturing processes for complex geometries and intelligent design.

RO.2: Development of self-healing actuator systems, deformation and damage sensing capabilities and dedicated smart control and response system through artificial intelligence and machine-learning techniques.

RO.3: Development of a mechatronic design, construction of a prototype, control and experimental validation of a fully autonomous self-healing demonstrator with dedicated intelligence, focusing on implementation in robotic applications.

Data generation and collection will happen in the laboratories of the project partners – henceforth referred to as 'beneficiaries' – to complete the tasks and work packages set out in the work plan, as described in the Description of Action (DoA), in view of achieving the defined research objectives. The first work package WP1 focuses on materials innovations into stimuli-responsive materials to satisfy RO.1. WP1 also deals with the processing and manufacturing of the developed materials. These materials and the manufactured parts will then be further employed in the next WP's. WP2 consists of innovations in relation to the three pillars of mechatronics: (i) actuation, (ii) sensing and (iii) control. Developments are foreseen for all three pillars by collaboration of beneficiaries that each excel in at least one of these pillars. The expected developments in WP2 will help achieve RO.2 and support WP3, which deals with the integration of the results of WP1 and WP2 into fully integrated working demonstrators. All results, including all types of data, are owned by the beneficiary that generates them, unless agreed otherwise by the beneficiaries involved in the generation of the data. Joint ownership is possible as detailed in Article 26 of the Grant Agreement. Transfer of ownership of results is possible following the rules set forward in Article 30 the Grant Agreement.

Data will be generated in the R&D labs of the beneficiaries obtained by scientific experimentation. Different types of data can be identified: (i) data generated by standardized and certified measuring devices ('instruments'), (ii) data recorded in custom-made, designated test benches and laboratory setups, (iii) data and observations recorded by the experimentalist. Table 1 gives an overview of the types of experiments that will be conducted during the course of this project, along with a non-exhaustive list of examples of experimental parameters that will be varied and the anticipated data that will be collected. The data will be stored in result files (digital or analogue) with the necessary metadata for reproduction, processing and evaluation. An extensive and heterogeneous amount of data is foreseen to be collected throughout the SHeRo project, which will require consistent and efficient procedures for handling, sharing and disseminating information among the partners and with the scientific community. Most of the data that will be generated during the project will be in a digital format. Non-digital data should ideally be converted to digital formats (picture or text format).

Table 1 Non-exhaustive list of experiment types, related experimental parameters, data results and associated metadata.

Type of experiment	Parameters	Data and metadata
Material synthesis	Synthesis protocol Reactivity and reaction mechanism Reaction parameters - Temperature - Reaction time - Solvent, catalyst … - Mixing ... Extraction and purification	Purity and quantities of substances Reaction kinetics Reaction conditions & protocol In-situ analysis during reaction Extraction and purification protocol Product yield
Material characterization	Spectroscopy - Measuring mode - Spectral range - Number of scans - Resolution Thermal analysis	Instrument calibration and certification Sample preparation and concentration Material spectra - Structure confirmation - Extent of functionalization Thermal transitions

	- Heating/cooling rate - Temperature range Mechanical testing - Temperature range - Stress, strain, strain rate Microscopy - Magnification - Resolution Functional properties - Electrical conductivity - Magnetic properties	Viscoelastic properties Mechanical behaviour Microscopy images - Morphology and topology - Structural composition Data post-processing - Background/baseline correction - Data analysis
Material processing and manufacturing	Processing parameters - Temperature profile - Residence time - Heating/cooling rate Manufacturing parameters - (Extrusion) rate - Laser power and speed Post-processing - Post-curing, annealing - Surface treatment	Processing conditions - Homogeneity and uniformity - Effect on properties Product properties
Actuator design and testing	Actuator design - Actuator type - Mode of actuation - Design parameters Actuator performance testing - Actuator control - Dedicated test benches	Actuator concept ideas and designs Actuator simulations Actuator test setup and conditions Real-time performance data - Actuator input/output data Video or image footage
Sensor design and testing	Sensor design - Sensor type - Sensor geometry - Sensor morphology Sensor performance testing - Electromechanical testing	Sensor concept ideas and designs Sensor Sensor test setup and test conditions Real-time performance data Video or image footage
Actuator control	Control algorithms - Type of control (force, position …) - Actuator input control	Controller performance: Controller bandwidth, accuracy, stability and robustness Source Code Video or image footage
Machine learning and data processing	Learned Models Learning algorithms Training data	Raw data Pre-trained models Performance data Learning architecture and training algorithm source code

The data will be collected in various **types of data formats**, depending on the experimental method and the used laboratory equipment and instrumentation. The **size of the data** result files may vary from a few kb to hundreds of Mb, depending on the amount of data points and the size of information per data point and the data format. By the end of the project about 10 Gb of data is expected per researcher working on the project. These data will be stored in data storage platforms of the beneficiaries during the course of the project and will later archived in repositories, as described in section 2.

Re-use of data, both raw and processed, is expected when new insights and knowledge require further or more advanced processing of existing data. Existing published and unpublished (background) data, as well as foreground data, from beneficiaries will be used by other beneficiaries due to the high level of complementarity of the project partners and the expected high level of collaborative research. The Consortium Agreement contains the background descriptions of the partners and the rules for the other partners regarding the implementation and exploitation of these types of data. All aforementioned data (Table 1) are anticipated to be necessary at some point to ensure smooth progress of the project and for dissemination purposes.

The **data utility** is expected to be very wide within the fields of research associated to this project. The data recorded in WP1 consists of synthesis, characterization and processing of self-healing materials and the improvement thereof in view of robotic ad sensor applications. The data recorded in WP2 consists of manufacturing of robotic parts and sensors and the testing thereof. The raw data from WP1 and WP2 are useful to scientists and researchers in the same or similar fields of study, provided they have the software necessary to read out the data

and the knowledge to interpret the data. In general, these data need processing to be correctly interpreted. In WP3 control algorithms are developed and data are collected for the demonstrators, where sensors are incorporated in controlled actuator systems. **Treated data** and data reports will be more widely useful to a broader audience of scientists and researchers and even non-experts (see sections 2.2 and 2.3 for the accessibility and interoperability of the data. Publications, along with their supporting data, will provide the most complete story and have the widest data utility.

2. FAIR data

In the following sections, distinctions will be made between the storage and curation of data during the project (work in progress) and the long-term preservation of data (archiving).

2.1. Making data findable, including provisions for metadata

The following provisions are applicable to digital data, unless mentioned otherwise. A distinction is made between data storage during the project and long-term storage after the project completion.

The data obtained from instrumental experimentation have digital metadata attached to the generated data files. This **metadata** includes time stamps, location of generation, identification of the experimentalist, information about the experimental procedure and test conditions. These metadata allow to identify the data and the experimental conditions for the conception of the data. The metadata should preferably also contain information about the standardization or calibration of the instrument or testing equipment or this information should be otherwise easily accessible. Data that are generated otherwise and that do not possess such digital metadata should have a digital accompanying document, such as a text file (.txt) or .csv file or equivalent, that includes such information about the performed tests. Where possible, laboratory experiments with robotic actuators and sensors should be supported by image and video files, to support digital data results. A guiding document (.txt or .doc) will be associated to each data folder to detail the contents and hierarchical structure of that data folder. For **long term storage** (archiving) of data sets, a unique object identifier, information about the author and keywords will be associated to the data set.

The data should be findable and easily identifiable. The **name of the data set** should include at least a string to identify the studied material or object, a reference to the type of test that was carried out and the main test conditions or parameters varied in the test. For digital data from standardized instrumentation, all this information is also present in the metadata, however consistent name-giving is essential. The naming convention serves to facilitate the identification, handling and processing of the data. Data acquired through duplication of experiments performed under the same conditions for reproducibility or statistical analysis will be numbered at the end of the name of the data. An example could be material1_test2_condition3_004.

Documents containing newly generated data by **processing of raw data** (e.g. Excel documents, MATLAB files) should be named using the same naming convention as described above for easy identification. These documents should be numbered with **version numbers** when major changes are made to the way the raw data or previously processed data are further processed with information explaining the most important changes between the different versions and identifying the person who made the changes and specifying the reason for the changes made. The information provided with the document should allow other users to understand the way the raw data were treated and allow replication of the data treatment.

Software files or other document types used to process or generate data (e.g. Excel documents, MATLAB files, Lab view projects) should be named in a descriptive manner for easy identification and should be provided with the data if necessary to access, treat or interpret the data.

2.2. Making data openly accessible

As stated in Article 29 of the Grant Agreement, the beneficiaries must deposit the digital research data, including associated metadata, needed to validate the results presented in scientific publications, as soon as possible in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate free of charge for any user. Open-access publications and supporting, openly accessible data will be deposited in the universities' repositories and made openly accessible by default (see Table 2).

Table 2 List of data repositories for data storage and management during the project (closed), open-access publication of data and for archiving after completion of the project.

Data storage (closed)	Host	Comment
MO Sharepoint	Microsoft Corporation (host) Vrije Universiteit Brussel (client)	Internal data management for all beneficiaries (closed)
MO OneDrive	Microsoft Corporation (host) Vrije Universiteit Brussel (client) University of Cambridge (client)	Personal, internal data storage for researchers (closed)
Cambridge Research Data Management website	University of Cambridge (host)	Personal, internal data storage for researchers (closed)

Empa file server	Swiss Federal Research Institutes for Material Science and Technology (Empa)	Personal, internal data storage for researchers (closed)
ODS MyCore	Centre National de la recherche scientifique (CNRS, host) L'École supérieure de physique et de chimie industrielles de la ville de Paris (ESPCI, client)	Personal, internal data storage for researchers (closed)
Transferts de Fichiers	L'École supérieure de physique et de chimie industrielles de la ville de Paris (ESPCI, host)	Secured data management for identified recipients (closed)
Local file server (two local drives)	SupraPolix	Personal, internal data storage for researchers (closed)
Open access publishing	**Host**	**Comment**
Open access journals	Publisher of open access journal	Open access publishing
Pure Database	Elsevier (host) Vrije Universiteit Brussel (client)	Open access publishing
Digital Object Repository Empa	Swiss Federal Research Institutes for Material Science and Technology (Empa, host)	Open access publishing
HAL	Centre National de la recherche scientifique (CNRS, host) L'École supérieure de physique et de chimie industrielles de la ville de Paris (client)	Open access publishing
Repository (archiving)	**Host**	**Comment**
Zenodo	CERN Data Centre & Invenio	Open access publication of data sets after completion of the project Creative Commons licenses
Universiteitsarchief	Vrije Universiteit Brussel (host)	Long term storage of data after completion of the project (closed)
Github	University of Cambridge (host)	Open access long term storage of data and source code
ILL data portal	Institut Laüe Langevin (host) L'École supérieure de physique et de chimie industrielles de la ville de Paris (client)	Private access during the project Open access publication of data sets after completion of the project
Soleil data Retrieval	Synchrotron Soleil (host) L'École supérieure de physique et de chimie industrielles de la ville de Paris (client)	Long term storage of data after completion of the project (closed)
Github	University of Cambridge (host)	Open access long term storage of data and source code

Other data, including associated metadata, will be protected (Article 27 and 37 of the Grant Agreement) and handled confidentially (Article 36 of the Grant Agreement) and will not be openly accessible as long as deemed necessary to support publication or until the deadline set forward by the beneficiaries has been reached. In the former case the data will be made openly available following the procedure for supporting data and information for open access publication (see above paragraph). The deadline for open-access publication of all the project data is 1 year after the completion of the project. This time window should give ample time to finish peer reviewed publications that are still pending.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the project's main objective (as described in Annex 1) would be jeopardised by making those specific parts of the research data openly accessible. In addition, the beneficiaries can **opt out from open access** publication of specific parts of their research data in view of potential valorisation of the intellectual property. Three reasons for opting out are identified:

(i) data supporting peer reviewed publications. The data supporting the publication shall remain closed and protected until the date of publication of the peer reviewed journal article. Upon publication, the supporting data will be made openly accessible, as previously described.

(ii) data detrimental for the application of a patent. The data sets supporting the patent application shall remain closed and protected until the date of publication of the patent.

(iii) data comprising a substantial amount of protected background information, such that the data cannot reasonably be disclosed without disclosing the incorporated background information. In this case, the beneficiary can choose to publish part of the data set that does not jeopardise the protection of the incorporated background information. If the extent of the data set that cannot be disclosed is too large – decided in good conscience by the involved beneficiary – the data set is exempt from open access publication.

Note that in the case that multiple beneficiaries are involved with the generation of the data, and a such share the ownership of the intellectual property, a consensus needs to be reached between the beneficiaries. In case no consensus is reached, the abovementioned data will not be made openly available if one of the involved beneficiaries opts not to disclose the data.

The raw data obtained from materials characterization (e.g. spectroscopy and thermal analysis) will be very heterogeneous and dedicated software is required to open and treat the data. The data can be exported in commonly available formats (.txt, .csv, .xls) or treated by Matlab programs. For open access publication of data and for depositing data in repositories for long term preservation (archiving), the data would be made available in open source formats. In the exceptional case where it is not possible to make the data available in open source formats, the beneficiaries will provide information about the tools and instruments at the disposal of the beneficiaries and necessary for validating the results and provide information about the tools and software necessary to access the data (e.g. open source data processing software). This information will be deposited in the form of instruction notes (.txt or .doc). The instruction notes will include the basic instructions on how to use the software to correctly interpret and validate the data. Where necessary, the software itself will be made available.

The closed data storage platforms for data storage during the project are only accessible using login identification upon granted access rights. The repository for open access publishing of the data after completion of the project will require registration and identification to gain access to the deposited data. See section 2.4 for more information on access rights and licenses allowing data re-use.

2.3. Making data interoperable

Most raw data will be collected in standard **file formats** associated to the instrument the data is generated on and can be read out by anyone who is in possession of the dedicated software. To facilitate exchange and re-use of data within the consortium, data sets will be exported and stored in formats that are more generally accessible (e.g. .txt, .csv or .xls), including the required metadata and additional comments to aid the interpretation of the data. Similarly, raw data collected using non-standardized tests and experimental observations will be stored in commonly accessible data formats.

To ensure the interoperability of the data and metadata, the 'best practices in the field' will be applied, *i.e.* standard **vocabularies** will be used when applicable (e.g. IUPAC Goldbook for chemistry and physics) and vocabularies from commonly accepted literature within the field. Where necessary a clarifying vocabulary will be associated to the data set for interdisciplinary interoperability. This is especially true in case new vocabularies or generated or uncommon vocabularies are used. Abbreviations will always be written in full and clarified at least the first time they are used.

2.4. Increase data re-use (through clarifying licences)

Materials, technologies and information that are under protection because they were conceived outside of the frame of the project may be made available during the project under certain restrictions to aid the progress of the project. Data generated using such materials, technologies or information may be subject to non-disclosure agreements or transfer agreements, limiting the information and data dissemination. Access to these data may be provided by way of **data use agreements** between the involved parties.

Data sets that are deposited into repositories for long-term open-access storage will be associated with **Creative Commons licenses** to protect the authorship of the data. The Attribution ShareAlike (CC-BY-SA) license will be the preferred type for most open-access data as it allows third parties to reuse and build further upon the deposited data sets, even for commercial purposes, as long as the authors of the data sets get credit and license their new creations under the identical terms. In particular cases restrictions of sharing or commercial restrictions may apply to certain data sets.

As defined in section 2.2 the data will be deposited in open-access repositories after an **embargo period** of 1 year, with the exception of data sets that support unfinished peer review publications, patent applications or information that cannot legally be made openly accessible. After this embargo period the data will remain **openly accessible for at least 10 years** in the open-access repository and preserved in the institution's data storage platforms for at least 10 years for re-use.

3. Allocation of resources

The deposition of the data and making the data FAIR are the responsibility of the researchers who generate the data. Data storage facilities hosted by the beneficiaries' institutions are used free of additional charge. The Principle Investigators (PI) of the beneficiaries are responsible to ensure long-term preservation of the data.

Green open access publication and bronze (delayed) open access publication come are typically free of charge, while gold open access publication comes at a high cost per article. Costs for gold open access publishing are covered as budgeted in the Description of Action and as agreed upon in the signed Grant Agreement.

Long term storage of data in the open access repository Zenodo is free of charge. The PI of the project is in charge of the long-term preservation of the data in the project community in Zenodo. The PI of each beneficiary may decide

to extend the preservation period of their data for a longer period than set forward by the consortium and warranted by the Pi of the project.

4. Data security

Raw data will be stored on the permanent memory of the computer where the data was generated. All data will be stored in at least 2 separate physical locations (e.g. personal work computer of researcher and external back-up memory) to prevent loss of data. Back-ups of data should be made at least every week to the institute's data storage facility. Similarly, on a weekly basis the data storage and management facility (MO Sharepoint) of the consortium should be updated. This platform serves for data transfer within the consortium and as an additional backup for sensitive the data.

Processed and sensitive data should only be stored on **password protected** computers and be removed from the transfer medium upon completion of the data transfer. All data storage platforms used during and after completion of the project are **certified** and **system encrypted**.

5. Ethical aspects

No ethical or legal issues are anticipated that would impact data sharing.

6. Other issues

No other issues need to be addressed.

7. Further support in developing your DMP

The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools.

The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

Registry of Research Data Repositories

Some repositories like Zenodo, an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them.

Other useful tools include DMP online and platforms for making individual scientific observations available such as ScienceMatters.

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

HISTORY OF CHANGES		
Version	**Publication date**	**Change**
v0	n.a.	First draft sent to project partners for input
v1	25.11.2019	Initial version, including input from all beneficiaries