Functionalized Soft robotic gripper for delicate produce harvesting powered by imitation learning-based control

D1.3. SOFTGRIP DATA MANAGEMENT PLAN

Deliverable Number	3
Work package Number and Title	WP1 Technical and Functional Requirements Analysis
Lead Beneficiary	THL
Version - Status	1.2
Due Date	30 June 2021
Deliverable Type[1]	ORDP
Dissemination Level[2]	PU
Internal Reviewer	TEAGASC
Filename	D1.3_SoftGrip_Data_Management_Plan_v1.2

[1] **Please indicate the type of the deliverable using one of the following codes**:
R = Document, report
DEM = Demonstrator, pilot, prototype
DEC = Websites, patent fillings, videos, etc.
ETHICS = Ethics requirement
ORDP = Open Research Data Pilot
DATA = data sets, microdata, etc.

[2] **Please indicate the dissemination level using one of the following codes**:
PU = Public
CO = Confidential, only for members of the consortium (including the Commission Services)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101017054

DOCUMENT INFO

Authors

Surname	First name	Organization	E-mail
Editor			
Koutsoumpa	Christina	THL	christina.koutsoumpa@twi.gr
Authors			
Maselli	Martina	SSSA	martina.maselli@santannapisa.it
Retsinas	George	ICCS	georgeretsi@gmail.com
Marougkas	Isidoros	ICCS	ismarougkas@gmail.com

Internal Reviewers

Surname	First name	Organization	E-mail
Grogan	Helen	TEAGASC	helen.grogan@teagasc.ie

Document History

Date	Version	Editor	Change	Status
15/06/2021	1.0	THL	§Initial version	Sent to internal reviewer
23/06/2021	1.0	TEAGASC	Feedback received: version approved	Under review
25/06/2021	1.1	THL	MITSUI table added	Sent to consortium for review
30/06/2021	1.2	THL	Minor edits	Submitted

TABLE OF CONTENTS

INTRODUCTION

SoftGrip is a Horizon 2020 project participating in the Open Research Data Pilot. This pilot is part of the Open Access to Scientific Publication and Research Data programme in H2020. The goal of the program is to foster access to data generated in H2020 projects.

The current document provides detailed information about the datasets that are planned to be captured by the partners of the SoftGrip project. The foreseen datasets are those agreed by the partners as of month 6 of the project. A more complete list of datasets and the data management aspects will be included in the future, as the project progresses. The DMP will be updated continuously and reviewed internally every semester.

SUMMARY OF FORESEEN SOFTGRIP DATASETS

The following table provides a list of the expected datasets, whereas the detailed description of each dataset, in accordance with the H2020 DMP template is provided in the following section. At this stage (M6), there are 8 foreseen datasets for the project, covering a series of research dimensions on the skills the SoftGrip system should develop. In the course of the project more datasets will be added in the Data Management Plan.

TABLE 1. SUMMARY OF FORESEEN SOFTGRIP DATASETS (AS OF MONTH 6)

NO	DATASET NAME
1	DS1_SSSA_Material_characterization
2	DS2_SSSA_Actuator_design_and_testing
3	DS3_SSSA_Soft_gripper_design_and_testing
4	DS4_ICCS_Mushroom_cultivation_visual
5	DS5_ICCS_SSSA_Soft_gripper_visual
6	DS6_UESS_Expert_demonstration_visual
7	DS7_UESS_Expert_demonstration_sensing
8	DS8_UESS_Simulated_mushroom_picking
9	DS9_MITSUI_Material_formulation

1. DATASETS DESCRIPTION

This section provides detailed information about the datasets that are planned to be captured by the partners of the SoftGrip project. These are the foreseen datasets as of month 6 of the project, which are expected to be updated in the course of the project. The DMP will be updated every semester.

In order to meet the requirements of the DMP according to the Pilot Open Access of the Horizon 2020, each partner provided the description of their datasets in a template table formed according to the EC guidelines of the dataset aspects that should be reported in DMPs of the H2020 projects. Based on this information, the SoftGrip partners compiled the following tables.

1.1. MATERIAL CHARACTERIZATION

Dataset name: Material Characterization	
Data description: mechanical testing (stress, strain, strain rate), functional properties (electrical conductivity, magnetic properties), data post-processing (background/baseline correction, data analysis)	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	SSSA
Partner in charge of the data collection	SSSA
Partner in charge of the data analysis	SSSA
Related WP(s) and task(s)	WP2 (T2.2 and T2.3)
Standards	
Info about metadata (production and storage dates, places) and documentation.	Data will be generated in the R&D labs of SSSA, obtained by scientific experimentation. Different types of data can be identified: data generated by standardized and certified measuring devices ('instruments'), and data recorded in custom-made, designated test benches and laboratory setups.
Standards, format	The data will be stored in result files (digital or analogue). Non-digital data should ideally be converted to digital formats where possible (picture or text format). The data will be collected in various types of data formats, depending on the experimental method and the used laboratory equipment and instrumentation. Example of data formats for this dataset are: .xls, .csv, .txt, .mat
Estimated data size	1 Gb

Data exploitation and sharing	
Purpose use of the data analysis	Data collected within this dataset consists of characterization of materials to be used to improve the design of soft gripper. This data should be used also to validate the results presented in scientific publications.
Data access policy/dissemination level	1st level: Personal, internal data storage for SoftGrip researchers (closed) 2nd level: Public availability also through dissemination activities
Embargo periods.	12 months
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	These data will be temporarily stored in data storage platforms of SSSA during the project. As soon as the common portal for the SoftGrip project is decided, data will be uploaded, stored, and shared on it during and after the project. Data storage platforms of SSSA are: MO OneDrive - Microsoft Corporation (host), Scuola Superiore Sant'Anna (client). SSSA Repository (archiving): CINECA IRIS Institutional Research Information System. Data will be stored in the long term after completion of the project (closed).

1.2. ACTUATOR DESIGN AND TESTING

Dataset name: Actuator design and testing	
Data description: actuator type, mode of actuation, design parameters, actuator simulations, actuator test setup and conditions, actuator performance testing (actuator control, dedicated test benches)	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	SSSA
Partner in charge of the data collection	SSSA
Partner in charge of the data analysis	SSSA
Related WP(s) and task(s)	WP3 (T3.1 and T3.2)
Standards	

Info about metadata (production and storage dates, places) and documentation.	Data will be generated in the R&D labs of SSSA, obtained by scientific experimentation. Different types of data can be identified: data generated by simulations, data obtained by standardized and certified measuring devices ('instruments'), and data recorded in custom-made, designated test benches and laboratory setups.
Standards, format	The data will be stored in result files (digital or analogue). Non-digital data should ideally be converted to digital formats where possible (picture or text format). The data will be collected in various types of data formats, depending on the experimental method and the used laboratory equipment and instrumentation. Example of data formats for this dataset are: excel documents, MATLAB files, CAD files.
Estimated data size	1 Gb
Data exploitation and sharing	
Purpose use of the data analysis	Data collected within this dataset consists of design, conceptualization and manufacturing of actuation parts of the soft gripper and the testing thereof. This data should be used also to validate the results presented in scientific publications.
Data access policy/dissemination level	1st level: Personal, internal data storage for SoftGrip researchers (closed) 2nd level: Public availability also through dissemination activities
Embargo periods.	12 months
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	These data will be temporarily stored in data storage platforms of SSSA during the project. As soon as the common portal for the SoftGrip project is decided, data will be uploaded, stored, and shared on it during and after the project. Data storage platforms of SSSA are: MO OneDrive - Microsoft Corporation (host), Scuola Superiore Sant'Anna (client). SSSA Repository (archiving): CINECA IRIS Institutional Research Information System. Data will be stored in the long term after completion of the project (closed).

1.3. SOFT GRIPPER DESIGN AND TESTING

Dataset name: Soft Gripper design and testing	
Data description: gripper design, gripper simulations, gripper test setup and test conditions, real-time performance data	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	SSSA
Partner in charge of the data collection	SSSA
Partner in charge of the data analysis	SSSA
Related WP(s) and task(s)	WP3 (T3.3)
Standards	
Info about metadata (production and storage dates, places) and documentation.	Data will be generated in the R&D labs of SSSA, obtained by scientific experimentation. Different types of data can be identified: data generated by simulations; data recorded in custom-made, designated test benches and laboratory setups; and data and observations recorded by the experimentalist.
Standards, format	The data will be stored in result files (digital or analogue). Non-digital data should ideally be converted to digital formats where possible (picture or text format). The data will be collected in various types of data formats, depending on the experimental method and the used laboratory equipment and instrumentation. Example of data formats for this dataset are: excel documents, MATLAB files, CAD files.
Estimated data size	1 Gb
Data exploitation and sharing	
Purpose use of the data analysis	Data collected within this dataset consists of design, conceptualization and manufacturing of the soft gripper and the testing thereof. This data should be used also to validate the results presented in scientific publications.

Data access policy/dissemination level	1st level: Personal, internal data storage for SoftGrip researchers (closed) 2nd level: Public availability also through dissemination activities
Embargo periods.	12 months
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	These data will be temporarily stored in data storage platforms of SSSA during the project. As soon as the common portal for the SoftGrip project is decided, data will be uploaded, stored, and shared on it during and after the project. Data storage platforms of SSSA are: MO OneDrive - Microsoft Corporation (host), Scuola Superiore Sant'Anna (client). SSSA Repository (archiving): CINECA IRIS Institutional Research Information System. Data will be stored in the long term after completion of the project (closed).

1.4. MUSHROOM CULTIVATION VISUAL DATA

Dataset name: Mushroom Cultivation Visual Data	
Data description: Multiple Views of RGB-D images picturing real mushroom cultivations. Diversity is required: Capture multiple cultivations from multiple views under different conditions (e.g. illumination). A complementary feed of synthetic mushroom data would be generated.	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	TEAGASC - realistic dataset ICCS, THL - synthetic dataset
Partner in charge of the data collection	TEAGASC - realistic dataset ICCS, THL - synthetic dataset
Partner in charge of the data analysis	ICCS, THL
Related WP(s) and task(s)	WP4
Standards	

Info about metadata (production and storage dates, places) and documentation.	*Real data:* Data will be captured in a suitable mushroom farming facility, supervised by Teagasc. Data will comprise photos/videos recorded by RGB-D camera sensors, located at a designed topology, according to certain coverage criteria. Also, data will include image-based annotations of mushrooms' 2D locations (in terms of Bounding Boxes) and classes that will be provided with the use of State-of-the-Art software tools. *Synthetic data:* Rendered 3D scenes of mushroom cultivations, with the use of Blender graphics engine, in RGB-D image pairs. Accompanied by 6D Rigid Poses, Mushroom 3D models, Semantic/Instance Segmentation masks, 2D Bounding Boxes.
Standards, format	*Real data:* Proposed Data Collection Protocol: video footage per cultivation (3 phases) x N cultivations (assuming different mushroom configuration and, if needed, lighting randomization) x M camera views x L Mushroom species (white/brown Agaricus mushrooms) x RGB-D pairs HDF5 File Format Raw: .mp4 videos of cultivations / final: .jpg/.png images with metadata stored in .json format. *Synthetic data:* Collection of generated .jpg images or sequence of .jpg/.png images (using Blender) with .json/.jpg metadata files.
Estimated data size	100-200GB (~10000 RGB-D image pairs from different camera views)
Data exploitation and sharing	

Purpose use of the data analysis	- Detection of Mushrooms' locations - Mushroom Quality Classification - Mushroom Instance Segmentation - Mushroom's Cap size estimation - Mushroom 6D Rigid Pose Estimation - Mushroom 3D Shape Reconstruction - Reduce Sim-to-Real gap (facilitating the algorithms' transition from simulation to realistic conditions)
Data access policy/dissemination level	1st level: All Softgrip partners 2nd level: Public availability also through dissemination activities e.g. Organize mushroom segmentation competition if possible - depends on the data/annotation quality
Embargo periods.	1.5 year provisionary
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	Personal, internal data storage for researchers (closed) Data are temporarily stored in separate repositories of ICCS/THL preference i.e. ICCS server, Google Drive, SoftGrip website. As soon as the common portal for the SoftGrip project is decided, data will be uploaded stored and shared on it during and after the project.

1.5. SOFT GRIPPER VISUAL DATA

Dataset name: Soft Gripper Visual Data	
Data description: RGB-D videos of soft gripper motion (multiple views). If possible, gripper motion should be captured along mushroom cultivation, in order to gather real data from gripper-mushroom interaction.	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	SSSA and ICCS
Partner in charge of the data collection	SSSA and ICCS

Partner in charge of the data analysis	ICCS, THL
Related WP(s) and task(s)	WP4
Standards	
Info about metadata (production and storage dates, places) and documentation.	Real data would be acquired using a finalized (or close to final) camera setup, gripper design and materialization. Location: to be announced Metadata Documentation: - marker annotation (if they exist)
Standards, format	HDF5 File Format Stored as (.mp4) videos of the different phases of the gripper motion
Estimated data size	10-30 GB (~50 RGB-D videos of multiple camera views)
Data exploitation and sharing	
Purpose use of the data analysis	- Finger detection of the soft gripper - Gripper 3D Shape Reconstruction - Reduce Sim-to-Real gap from Gripper simulations (e.g. tune hyperparameters) - Relative Gripper-Mushroom pose (*angle of attack*), if mushroom cultivation is included in the setting.
Data access policy/dissemination level	All Softgrip partners (private). Can be made public after the completion of the SoftGrip project, if needed.
Embargo periods.	1.5 year provisionary
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	Data are temporarily stored in separate repositories of ICCS/THL preference i.e. ICCS server, Google Drive, SoftGrip website. As soon

	as the common portal for the SoftGrip project is decided, data will be uploaded stored and shared on it during and after the project.

1.6. EXPERT DEMONSTRATION VISUAL DATA

Dataset name: Expert Demonstration Data – RGB-D streams	
Data description: RGB-D image sets of experts carrying out mushroom picking	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	UESS
Partner in charge of the data collection	UESS
Partner in charge of the data analysis	UESS
Related WP(s) and task(s)	WP5
Standards	
Info about metadata (production and storage dates, places) and documentation.	Real data would be acquired using the RGB-D camera setup. Metadata Documentation: Basic annotations (binary contours and/or masks) in part of the dataset
Standards, format	RGB images: 16bit jpg or png Depth images: 16bit jpg or png Videos: .mp4

Estimated data size	10-30 GB (2-5 hours of RGB-D videos)
Data exploitation and sharing	
Purpose use of the data analysis	· Extraction of Movement Primitives · Learning embeddings for applying Imitation and Reinforcement Learning · Identifying optimal strategies to adapt simulated scenarios
Data access policy/dissemination level	During the project: All Softgrip partners (private) After the project: Publicly available for research purposes only
Embargo periods.	6-12 months
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	During project: OneDrive and/or Google drive After the project: Public repositories (OpenAIRE, EU Open Data Portal)

1.7. EXPERT DEMONSTRATION SENSING DATA

Dataset name: Expert Demonstration Data – Force profile streams	
Data description: RGB-D image sets of experts carrying out mushroom picking	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	UESS

Partner in charge of the data collection	UESS
Partner in charge of the data analysis	UESS
Related WP(s) and task(s)	WP5
Standards	
Info about metadata (production and storage dates, places) and documentation.	Real data would be acquired using a glove-like device featuring tactile sensing Metadata Documentation: Basic annotations (outrooting phases) in part of the dataset
Standards, format	CSV
Estimated data size	10-100MB (2-5 hours of force profile recording)
Data exploitation and sharing	
Purpose use of the data analysis	· Extraction of Movement Primitives · Learning embeddings for applying Imitation and Reinforcement Learning · Identifying optimal strategies to adapt simulated scenarios
Data access policy/dissemination level	During the project: All Softgrip partners (private) After the project: UESS
Embargo periods.	6-12 months
Archiving and preservation (including storage and backup)	

Data storage. Where? For how long?	During project: UESS-operated OneDrive and/or Google drive After the project: UESS-operated OneDrive and/or Google drive for up to two years

1.8. SIMULATED MUSHROOM PICKING DATA

Dataset name: Simulated Mushroom Picking Data	
Data description: Synthetic RGB-D image sets of mushroom picking including force profiles	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	UESS
Partner in charge of the data collection	UESS
Partner in charge of the data analysis	UESS
Related WP(s) and task(s)	WP5
Standards	
Info about metadata (production and storage dates, places) and documentation.	Real data would be acquired using the RGB-D camera setup. Metadata Documentation: Basic annotations (binary contours and/or masks) in part of the dataset
Standards, format	RGB images: 16bit jpg or png Depth images: 16bit jpg or png Videos: .mp4 Force profiles: CSV

Estimated data size	1-5 GB (1 hour of synthetic RGB-D videos and force)
Data exploitation and sharing	
Purpose use of the data analysis	· Development of Sim2Real methodologies · Refinement of learned policies · Experimentation with different Reinforcement Learning algorithms
Data access policy/dissemination level	During the project: All Softgrip partners (private) After the project: Publicly available for research purposes only
Embargo periods.	6-12 months
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	During project: OneDrive and/or Google drive After the project: Public repositories (OpenAIRE, EU Open Data Portal)

1.9. MATERIAL FORMULATION AND CHARACTERIZATION

Dataset name: Material Formulation and Characterization	
Data description: Material formulation - Substances, composition, process condition Mechanical properties - Tensile strength, modulus, elongation, Visco-elasticity, hardness Sensing properties - Force, displacement, time, voltage	
Partner responsibility	**Partner name**
Partner owner of data; copyright holder (if applicable)	Mitsui
Partner in charge of the data collection	Mitsui
Partner in charge of the data analysis	Mitsui
Related WP(s) and task(s)	WP3 (T3.1, T3.2, T3.3, T3.4, T3.5)
Standards	
Info about metadata (production and storage dates, places) and documentation.	Production site: Mitsui R&D lab in Japan. Testing site: Mitsui R&D lab in Japan. Equipment: Instron, Rheometer (DMA), Hardness meter. Data storage: Mitsui R&D lab in Japan
Standards, format	Test standard: ISO, ASTM, JIS Data format: doc, xls, csv, txt, mat, jpg, stp, step, stl Data folder: Under Mitsui R&D database
Estimated data size	1 Gb
Data exploitation and sharing	
Purpose of the data usage	For development and improvement of soft gripper and sensor materials. For internal and external reports.
Data access policy/dissemination level	1st level: High confidential data and know-how in Mitsui which becomes invention through SoftGrip project. / Limited in Mitsui. 2nd level: Personal, internal data through SoftGrip project. / Limited in SoftGrip partners. 3rd level: Public and non confidential data through SoftGrip project / Open.
Disclosable timing	1st level: Do not disclose until patent submission.

	2nd level: After consensus with partners. 3rd level: After Mitsui internal check.
Archiving and preservation (including storage and backup)	
Data storage. Where? For how long?	Data storage place: Mitsui R&D laboratory database. Storage period: More than 10 years.

2.FAIR DATA

In general terms, the SoftGrip research data should be 'FAIR' that is Findable, Accessible, Interoperable, and Re-usable.

2.1. DATA PROVISION AND PORTAL

As a temporary solution, preliminary datasets are currently stored in *CINECA IRIS, Bitbucket, Google Drive* managed by each partner internally. Current partner-specific data management policies will be fused to a common SoftGrip policy for data storage, sharing, publishing and archiving. To facilitate sharing and public accessibility, the SoftGrip available datasets will be stored in a common portal such as the European open science cloud *OpenAire* or the official portal for European data *data.europa.eu*. The portal will be decided within the course of the project. The data will be stored in folders organised and named according to the work packages. Initially all and only SoftGrip partners will have access to the data folders. The data that will become publicly available will be stored in separate folders and will first be approved by all partners. An embargo period specific to each dataset is possible to be applied prior to data publishing.

2.2. NAMING CONVENTIONS

The name for each dataset, template, and files will consist of the following parts:
(a) a prefix indicating the project name, along with its unique identification number, e.g. 'SG1'
(b) the name(s) of the partner(s) responsible to provide it, e.g. THL
(c) a short title of the dataset summarizing its content and purpose
(d) a suffix indicating the date of the last upload into the portal in YYYYMMDD format, for all documents destined to be shared externally.

Each of these elements is separated by an underscore: "_", to ensure interoperability with different Operating Systems. For programming source code no strict naming conventions are required. The repository naming should indicate clearly the purpose and function for the code.

2.3. CLEAR VERSIONING

All programming source codes will be automatically version controlled using a common Git type repository i.e. as Github and Bitbucket. Bitbucket is currently used in SoftGrip.

3. ALLOCATION OF RESOURCES

TWI-Hellas (THL) will be responsible for the SoftGrip data management plan. The cost for making data FAIR as well as the resources for the long preservation of the data (costs and potential value, what data will be kept and for how long) will be identified in the course of the project.

4. DATA SECURITY

With respect to Privacy and Data Protection, the EU-legislation - the General Data Protection Regulation - imposes several new obligations upon the consortium partners being data processors. Moreover, several new rights are granted to data subjects, and significant fines are introduced in case of a data breach. Apart from this legislation, the consortium partners regard privacy and data protection as a fundamental principle and hence apply a strict policy on this matter.

4.1. DATA CONFIDENTIALITY AND INTEGRITY

Data confidentiality and integrity are implemented at various levels:

- Data will be stored at the decided portal and kept protected against unauthorised access by means of standard Login. Appropriate access levels will be granted by the creation of groups according to the defined work packages of the SoftGrip project.
- Data in transit is secured by means of secure data transfer mechanisms
- Consortium partners will impose a strict policy on all employees, co-workers, subcontractors having access to the data.

4.2. DATA AVAILABILITY

Data availability is guaranteed by TWI, for which the specifics will be announced in the next version of the deliverable.

5. ETHICAL ASPECTS

At present, there are no ethical or legal issues that can have an impact on data sharing and no personal data will be stored.

6. OTHER ISSUES

At present, there are no additional issues, such as other national/funder/sectorial/departmental procedures, which have to be considered.