Flexible Assembly Manufacturing with Human-Robot Collaboration and Digital Twin Models
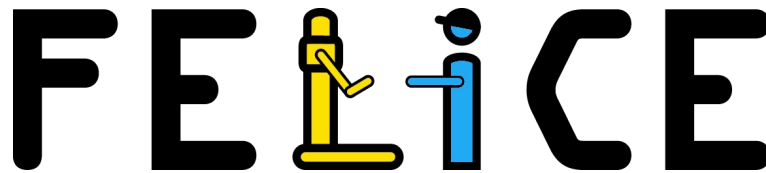
FELiCE

D1.2: Data management plan[1]

Abstract: This deliverable provides the initial version of the FELICE Data Management Plan (DMP). Following the FAIR Data Management guidelines in Horizon 2020[2], this document describes how research data will be handled during and after the end of the project, what data will be collected, processed and/or generated, which methodology and standards will be applied, whether data will be shared/made open and how data will be curated and preserved. Since the DMP will evolve over the course of the project, it will be updated to reflect significant changes that may arise in the data sources exploited in the project and on policies and methodologies to be used as the project progresses.

Contractual Date of Delivery:	30/06/2021
Actual Date of Delivery:	30/06/2021
Security Class:	Public (PU)
Editor:	*Maria Pateraki (ICCS)* *Alexandra Papadaki (ICCS)*
Contributors:	ICCS, FORTH, FHOOE, TUD, UNISA
Quality Assurance:	*Manolis Lourakis (FORTH)* *Maria Crociani (EUNL)*
Deliverable Status:	Final

The *FELICE* Consortium

Participant	Participant Organisation Name	Part. Short	Country
1	Institute of Communication & Computer	ICCS	Greece
2	Centro Ricerche FIAT S.C.p.a.	*CRF*	Italy
3	FH OÖ Forschungs & Entwicklungs GmbH	*FHOOE*	Austria
4	AEGIS IT RESEARCH GmbH	*AEGIS*	Germany
5	Leibnitz Research Centre for Working Environ-ment and Human Factors	*IFADO*	Germany
6	Foundation for Research and Technology	*FORTH*	Greece
7	CAL-TEK S.r.l.	*CALTEK*	Italy
8	Technical University Darmstadt	*TUD*	Germany
9	University of Salerno	*UNISA*	Italy
10	Fraunhofer-Gesellschaft zur Foerderung der Angewandten Forschung e.v.	*Fraunhofer*	Germany
11	ACCREA Engineering	*ACC*	Poland
12	PROFACTOR GmbH	*PRO*	Austria
13	Eunomia Ltd	*EUNL*	Ireland

Document Revisions

Version	Date	Editor	Overview
1.0	30/06/2021	Maria Pateraki (ICCS) Alexandra Papadaki (ICCS)	Final version
0.7	29/06/2021	Maria Pateraki (ICCS) Alexandra Papadaki (ICCS)	Third draft – reviewer comments addressed
0.6	29/06/2021	Maria Crociani (EUNL)	Second review
0.5	29/06/2021	Maria Pateraki (ICCS) Alexandra Papadaki (ICCS)	Second draft – reviewer comments addressed
0.4	29/06/2021	Manolis Lourakis (FORTH)	First review
0.3	28/06/2021	Maria Pateraki (ICCS) Alexandra Papadaki (ICCS)	Updated first draft
0.2	26/06/2021	Maria Pateraki (ICCS) Alexandra Papadaki (ICCS)	First draft
0.1	06/05/2021	Maria Pateraki (ICCS)	ToC

Table of Contents

List of Tables

List of Abbreviations

ADAPT Asset-Decision-Action-Property-RelaTionship

CC Creative Commons

DMP Data Management Plan

DOI Digital Objective Identifier

DQAP Data Quality Assurance Process

DQC Data Quality Committee

ePD E-Privacy Directive

EPM Ethics and Privacy Manager

FAIR Findability, Accessibility, Interoperability and Reusability

GDPR General Data Protection Regulation

H2020 Horizon 2020 Programme

OA Open Access

ORDP Open Research Data Pilot

QAM Quality Assurance Manager

SVN Subversion

Executive Summary

This deliverable provides the first version of Data Management Plan (DMP) of the FELICE project, structured according to the template suggested by the European Commission[3]. The FELICE project is funded by the European Union's Horizon 2020 research and innovation programme, under the grant agreement No 101017151. It participates in the Open Research Data Pilot (ORD pilot) action, which intends to maximize the use and re-use of open research data across projects. For that reason, FELICE is required to have a DMP, to document and communicate the management aspects related to the project activities.

This initial version of the FELICE DMP will thoroughly analyze the overall data management strategy, the data management procedures specification and the way that FELICE will adhere to the FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles[4]. In FELICE, different types of data are considered, namely scientific publications, other dissemination and communication publications, software and various datasets resulting from development, integration and evaluation efforts. The DMP will provide details about the collection, generation and processing of all data, as well as their storage, accessibility, curation and preservation during and after the completion of the FELICE project.

The DMP is a living document, according to the European Commission guidelines. This means that it will be constantly updated during the project lifetime, to reflect any changes that might arise in any of the datasets, policies and methodologies used.

---

[3] *http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm*
[4] *https://www.force11.org/group/fairgroup/fairprinciples*

1 Introduction

1.1 Purpose of the document

Open Research Data Pilot (ORD Pilot) participating projects – like FELICE – are required to have a DMP[5]. Thus, this document is a formal document that outlines how data will be handled during the project and after the project completion. Likewise, this report describes the data management methodology based on the findable, accessible, interoperable, and re-usable (FAIR) principles. The DMP includes a description on the data to be collected, processed and/or generated; the methodology and standards to make research data FAIR; the data that will be shared/made open access; and how data will be curated and preserved (including past the end of the project) or data sharing policies (e.g. open access, restricted to the consortium, etc.).

1.2 Intended readership

The DMP is of type ORDP and is a public (PU) document. It is aimed at the European Commission, the FELICE Project Officer, the partners involved in the FELICE Consortium, beneficiaries of other European-funded projects, and the general public.

1.3 Relationship with other FELICE deliverables

This deliverable is related to the deliverables of the following Work Packages: WP2, WP4, WP5, WP6, WP8, WP9 and WP10. The technologies developed as part of WP4-WP7 will be integrated and validated as part of WP8 (Integration, pilot implementation and evaluation). Overall, the data collected will be pseudonymized and will only be used for research and system evaluation.

---

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

2 Definitions

In this section, the definitions relevant to the DMP are introduced.

Dataset: An organized collection of data.

Data Management Plan (DMP): A written formal document that outlines how data will be handled during the course of the project and after the project completion.

Data Quality Assurance Process (DQAP): DQAP aims at ensuring the high quality of the data generated/collected during the lifetime of the project.

Digital Object Identifier (DOI): Digital Object Identifier is a unique permanent identifier for a published digital object, standardized by the International Organisation for Standardisation (ISO).

Ethics and Privacy Manager (EPM): The EPM will oversee the FELICE data protection strategy and will advise FELICE partners on data collection and processing procedures, to ensure compliance with GDPR requirements. Dr. Maria Crociani (EUNL), an expert on data protection and privacy, serves as the EPM.

GitHub[6]**:** GitHub is a version-control online repository which supports distributed source code development and management.

FAIR: Research data that is Findable, Accessible, Interoperable and Re-usable. These principles aim to provide a framework to ensure that research data can be accessed and re-used effectively.

Metadata: Metadata is a set of data that provides context or additional information about other data. The main types of metadata include descriptive metadata, structural metadata, and administrative metadata.

Open Access (OA): Open Access refers to the unrestricted access to research results including scientific peer-reviewed publications and research data. There are two complementary mechanisms for achieving open access to research. Green open access means that the authors will publish the accepted manuscript in an online repository. Gold open access means that the publication is directly available free of charge from the publisher and any related costs referred to as Article Processing Charges (APCs) shall be covered by the authors.

Open Research Data: Research data needed to validate the results of the publications that are openly available in digital form for access and re-use by anyone for any purpose.

Open Research Data Pilot (ORDP): The Open Research Data Pilot aims to improve and maximize the accessibility and re-usability of research data generated by H2020 projects, without violating the privacy of sensitive data.

---

[6] *https://github.com*

Quality Assurance Manager (QAM): The QAM will be responsible for the task /deliverable quality. Dr. Andreas Beham (FHOOE) as the FELICE QAM will be liable for (i) developing, implementing, communicating and maintaining the quality plan throughout the lifecycle of the project task; (ii) overall monitoring of the different KPIs; (iii) the identification of the problems during internal audits and initiating corrective actions to eliminate the problem; (iv) ensuring that goals set by the Project Technical Committee (PTC) and GA are fully implemented on a daily basis.

Repository: A digital repository is an archive for storing and managing digital copies.

Subversion (SVN)[7]**:** Subversion is an open source software versioning and revision control system that tracks changes made to files, folders, and directories. It eases data recovery and provides history of changes.

Zenodo[8]**:** Zenodo is a research data archive/online repository created by OpenAIRE and CERN for sharing research results in a wide variety of formats for all fields of science.

---

[7] *http://subversion.apache.org*
[8] *https://zenodo.org*

3 Data Summary

This section details how research data will be managed along the project, which includes amongst others, the data collection and generation activities, formats, identifications and descriptions, dataset licenses and repositories. In Table 1, the main types of datasets considered in FELICE are summarized. Moreover, FELICE datasets processing will comply with the different data protection regulations that apply to FELICE, as elaborated in FELICE deliverable D1.3 "Ethics and Privacy Manual". The Ethics and Privacy manager's role is to ensure that organization personal data are processed in compliance with the data protection rules.

Table 1: FELICE DATA types

Dataset types	Description
Scientific publications	As part of the dissemination and communication tasks, FELICE will produce scientific publications, such as journal papers, conference papers, white papers, books, workshops, tutorial sessions, etc., that will be made publicly available for a wider audience.
Other dissemination & communication material	As part of the overall dissemination and communication activities (WP9), FELICE will also produce additional promotional materials (e.g., newsletter, flyers, posters, videos), press releases, website news, social media posts (e.g., Facebook, Twitter, LinkedIn).
Software	FELICE partners are committed to contribute to open-source projects and release new software as open-source projects. It is expected that these initiatives will be focused to one partner or a limited group of partners, and with specific licence requirements.
Datasets	The FELICE framework formalization, development, integration and evaluation efforts, are expected to lead to different types of datasets, as further detailed in the following sections.

3.1 Collection and Generation

FELICE is a research and innovation project, so obtaining research data is a critical aspect of its research activities. Within FELICE several types of data are expected to be collected and/or generated (e.g. publications, software and datasets). This will facilitate the realisation of better research and technological evaluation of the FELICE outcomes, in accordance with the project's objectives. However, personal data collection and recording will be avoided whenever possible.

3.2 Format

The formats used for dissemination and communication data should be assessed according to the needs, the long and short-term analysis, the storage and the corresponding sharing purposes. Regarding the other types of data, open access and well documented formats should be used whenever possible, as shown in Table 2. For additional and specific classes of data, different data types might be evaluated and used by the consortium.

Table 2: FELICE Data formats

Type	Recommended Formats
Textual Documents	- Plain text, ASCII (.txt) - XML (.xml) - JSON (.json) - Adobe PDF (.pdf) - Hypertext Mark-up Language (.html) - MS Word (.doc/.docx) - Comma-separated values (.csv) - MS Excel (.xls, .xlsx) - SQL (.sql) - Software-specific formats
Source code	- Python (.py, .npy, .pyc, .pyd, .pyo, .rpy, .whl) - C++ (.h, .hh, .hpp, .hxx, .cpp, .c, .cc, .cxx) - C# (.cs, .csx, .rsp) - Java (.class) - JavaScript / Node.js (js) - Angular (.ts, .html, .scss)
Image	- JPEG (.jpeg, .jpg) - PNG (.png) - TIFF (.tif, .tiff) - BMP (.bmp)
Video	- MPEG-4 (.mp4) - SVO video file (StereoLabsTM) (.svo) - AVI video file (.avi) - BVH human motion file format (.bvh) - OGG video (.ogv, .ogg) - Motion JPEG 2000 (.mj2) - Audio Interchange File (.aif) - WAV (.wav)
Audio	- FLAC (.flac) - MPEG-1 Audio Layer 3 (.mp3) - Audio Interchange File (.aif) - WAV (.wav)
3D information / models	- PLY (.ply) - OBJ (.obj) - STL (.stl) - OFF (.off) 3DS (.3ds) - DXF (.dxf)

3.3 Identification

The data generated and collected within FELICE should be internally uniquely identifiable to univocally identify persistent and actionable data. For data to be publicly available as open access, a DOI will be assigned to achieve easy and clear identification as well as effective and persistent citation. This DOI can also be exploited in all related publications, to link the research results with the corresponding dataset. The FELICE internal identifier should be structured as follows:

FELICE_[type]_[date]_[name]_[version]_[policy]

where:

- **"type"** - describes the type of data (e.g., scientific publication, dissemination and communication material, software, dataset) [REQUIRED].
- **"date"** - is the date in the format "YYYYMMDD" [REQUIRED].
- **"name"** - is a short name for the data [REQUIRED].
- **"version"** - is the version of the dataset in the format v(MAJOR). (MINOR) (e.g. v00.01) [REQUIRED].
- **"policy"** - is the policy to apply to the dataset (e.g., restricted, embargoed, open)[REQUIRED]
- **"_" (underscore**) - is used as a separator between the fields.

3.4 Description

The optimal data management strategies require high-quality data descriptions. To serve this requirement, FELICE will also provide descriptions of the collected and generated data, to facilitate their identifiability. The data descriptions will specify the data name, purpose, type, format, origin, size, utility, privacy principles, metadata, accessibility, storage locations and responsible person. Table 3 provides a reference description template that can be used for the different FELICE data.

Table 3: FELICE Data description template

Field	Details
Name of the dataset	Short name of the data.
Responsibility	FELICE WP, task, and partners responsible for the data.
Purpose and relation to the objectives of the project	The purpose of collecting / generating the data and the way it serves the objectives of the project.
Data type	Type of collected and / or generated data.

Data format	Data format identification and / or description for complex formats.
Data origin	Identification of the data origin.
Expected size of the data	Estimation of the data size.
Data utility	Foreseen target third-party audience and applications for which the data will be useful.
Privacy principles	Data anonymization methods.
Accessibility policy	One of the following sharing policies: • **Restricted:** The data is only available for project internal use. • **Embargoed:** The end of the embargo period is disclosed. After that date, the data will become public and widely available. • **Open:** The data is public and can be used by other persons or entities outside the project.
Security & storage	The storage location and time period in which the data will be preserved.

3.5 Disclosure control

The disclosure control minimizes the risk of leaking sensitive information (e.g., personal information, business confidential data, etc.) while maximizing the data utility. However, such methods can directly affect the analytical completeness and validity of the data, decreasing its usefulness. For that reason, a trade-off between sensitive data protection and data importance has to be found.

In the context of FELICE, the data collected will be in pseudonymized form and will be used only for research and system evaluation. Pseudonymization is a procedure by which personally identifiable information fields within a data record are replaced by one or more pseudonyms, i.e. artificial identifiers. By doing so, data records become less identifiable while remaining suitable for data analysis and processing.

3.6 Re-usage

FELICE is an ORD Pilot participating project, so the consortium members are encouraged to maximise data re-use, by following the principle of making datasets "as open as possible, as closed as necessary"[9]. In general, it is recommended to use copyright licences to protect the author's rights. However, FELICE partners are recommended to use permissive licences that allow others to freely access, mine, exploit, reproduce and disseminate previously collected and generated data. They are also encouraged to make the data available by default, except for restricted exceptions, and use public repositories. For instance, Creative

---

[9] *https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm*

Commons[10] (e.g., CC-BY and CC0) licences, that maximize the possible re-usage while distributing the data in public repositories, can be used. In any case, whenever datasets are published in a data repository, the license agreement should be specified. Additionally, to optimise data re-usage, FELICE will use search keyword and metadata, to make the data more identifiable, understandable and findable by others. Moreover, the data will remain available and up-to-date for a certain amount of time after the completion of the project on public repositories. Specifically, the Zenodo repository will be used in the case of FELICE.

FELICE will also ensure the high quality of the collected/generated data until the end of the project, by following a *Data Quality Assurance Process (DQAP). DQAP* will ensure for each dataset that at least one internal review, or also additional peer-reviews in case of journal publications, are performed. The evaluation will be performed based on the following criteria:

- **Validity:** Data accurate/valid enough for the intended use.

- **Reliability:** Methods and procedures of data collection and analysis are properly documented and remain always stable.

- **Precision:** The data are detailed enough to provide a proper representation of the examined phenomenon.

- **Integrity:** Data should be maintained and with assured accuracy and consistency.

- **Timeliness:** Data should be regularly collected, up-to-date and available when needed.

- **Completeness:** Data are as complete as possible, minimizing the missing data. That way the corresponding research results reported in scientific publications can be reproducible (c.f. section 3.7).

FELICE DQAP will be coordinated by the FELICE *Quality Assurance Manager (QAM)* Andreas Beham (FHOOE) and the *Data Quality Committee (DQC)* throughout the lifetime of the project and for all datasets. The *DQC* of FELICE includes Alexandra Papadaki (ICCS), Manolis Lourakis (FORTH), Felice Tauro (CRF) and Spyros Vantolas (AEGIS). All DQC members have extensive research experience and have participated in numerous research projects.

3.7 Reproducibility of Results

The reproducibility is the ability of the project results to be obtained by another research team, if using the same methods. In fact, an observation is considered reproducible, if another team can re-observe it after repeating the experiment under the same conditions, using the same data and methods and for different runs, regardless of the location.

---

[10] *https://creativecommons.org/*

Reproducibility is a very valuable characteristic that indicates that the results are generalized enough and are not artefacts of the unique setup. This highlights the significance of the results and minimizes frauds and human errors.

Reproducibility of the results is a critical target for FELICE. To achieve this FELICE will provide detailed results documentation with:

- description of all the methods, equipment, procedures, experimental settings
- description of the data analysis and the data selection
- description of the analytical decisions, for results that depend on statistical inference, and whether the study is exploratory or confirmatory
- discussion of the expected constraints on generality in relation with the effect of different variables and methodological features on the results.
- precision and statistics

3.8 Security

Data security is a key aspect in all stages of the data lifecycle. FELICE will handle this by applying appropriate security principles (e.g., confidentiality, integrity, availability, non-repudiation). All partners will have to ensure data security at collection/generation and disclosure phases, following their internal data management policies, while closely monitored by the FELICE EPM (Dr. Maria Crociani (EUNL)).

The data security depends on the storage location and the time period for which the data is preserved. Within FELICE, the data will be safely stored at certified repositories, where the safe transfer, recovery and storage of sensitive data will be guaranteed (c.f. section 3.9).

The original data will be safely stored in the secure, local, private and non-publicly-accessible repository of the responsible partner. All partners will comply with the GDPR (c.f. section 7.1), to personally ensure secure storage and data transfer.

The datasets that will be shared only within the consortium will be safely stored on the centralized password-protected SVN repository[11] of the project, which is accessible over secure network connections. Larger files will be shared in the Seafile file store[12] maintained by the project. For their safety, regular procedures for backup strategy and data recovery will be followed.

After the end of the project, open data will be archived and preserved in the Zenodo repository. Regarding the confidential data that will be

---

[11] *http://subversion.apache.org*
[12] *https://storage.felice-project.eu*

deposited to the SVN or the Seafile, the data will be preserved for a minimum of one year, though decisions regarding their long-term preservation and curation have not yet been made.

FELICE will apply data anonymization techniques to guarantee the preservation of anonymity of the users' sensitive and personal data. Anonymization or pseudonymization will be applied to the data before their transmission to partners of the consortium and dissemination. At the end of the data retention period, full anonymization and data sanitization methods shall be applied.

3.9 Repositories

Four different repositories are used by FELICE for communication, organization/management and data storage. These repositories are the project website, the project repository, the cloud service and the Zenodo[13] platform. Additional internal communication tools that might be used by the consortium partners are independent and are not considered as repositories of the project. In the following section the repositories used by the project are further analysed.

3.9.1   FELICE Website

The FELICE website (https://www.felice-project.eu/) will be a channel for communication and for releasing publicly datasets of the project. It will contain information about the project, including public project details, news, events, dissemination and communication materials (e.g., flyers, posters, papers) as well as public deliverables. The website will also provide access to several public project datasets or refer visitors to the Zenodo repository.

The FELICE website is hosted by UNISA on an Ubuntu 20.04 LTS virtual server. The service is ensured by Nginx over HTTPs and the database is hosted on the same server. A backup of the database and the pages is performed daily to remote server at UNISA, hosted on a different physical server. The website is developed and maintained in compliance with the most modern and strict security requirements and guidelines. It will also be designed in accordance with the provisions regarding respect for user privacy deriving from the applicable EU regulations (as enlisted in the FELICE deliverable D1.3 "Ethics and Privacy Manual"). FELICE website will be made available during the project and for at least five years after the project end.

3.9.2   FELICE SVN Repository[14]

The FELICE SVN repository will be the repository that will accumulate all the project data, including open, embargoed and restricted ones, while

---

[13]*https://zenodo.org*
[14]*http://subversion.apache.org*

not making them publicly available. This repository is a safe, password-protected repository that will only have internal use between the authenticated and authorized FELICE members. Therefore, the repository can be used by all partners to store and manage data. The FELICE repository is managed by ICCS and will be available during and five years after the completion of the project.

3.9.3   FELICE Large Seafile File Store[15]

The Large Seafile file store is a cloud service that will share large files within the FELICE consortium. It is a safe, password-protected service, which will be accessible only by the authenticated and authorized FELICE members. It will accommodate all the large files (like videos, images, multimedia files etc.) that are cumbersome to be stored in the SVN repository. Each partner will have their own folder for sharing their data and will be personally responsible for their management.

3.9.4   Zenodo

FELICE will also provide publicly available datasets, using an online platform. In fact, the platform that will be used for this repository will be the Zenodo platform, which is a general-purpose open-access repository operated by CERN[16]. It stores files and metadata, provides version control and assigns DOIs to all uploaded elements. Zenodo is an open and accessible repository that enables access to data without restrictions and retains datasets for the lifetime of the repository, which is at least 20 years[17]. It uses a JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core[18], MARCXML[19], BibTeX[20], CSL[21], DataCite[22] and Mendeley[23]. The data record metadata will make usage of the vocabularies applied by Zenodo and reference to any external metadata will be done with a resolvable URL.

---

[15] _https://storage.felice-project.eu_
[16] _https://about.Zenodo.org/infrastructure/_
[17] _https://about.Zenodo.org/policies/_
[18] _https://dublincore.org/_
[19] _https://www.loc.gov/standards/marcxml/_
[20] _http://www.bibtex.org/_
[21] _https://citationstyles.org/_
[22] _https://datacite.org/_
[23] _https://www.mendceley.om/_

4 FAIR Data

FELICE, as a Horizon 2020 funded project that participates in the ORD pilot, will release its publicly available data in such a manner to comply with the FAIR data principles[24], *i.e. Findable, Accessible, Interoperable, and Reusable.*

Findability is related to how discoverable the data are from other researcher, what metadata accompany them and according to what standards they are created. Findability is also related to the extent to which the data are identifiable, well documented, with proper naming conventions, clear versioning and search keywords. Accessibility is related to the availability and access of all types of data and the software documentation. Interoperability depends on the licensing strategy that affects the data reusability. It also specifies the vocabulary and mapping to commonly used ontologies. Finally, reusability defines the ability of the data to be reused by third parties. The duration of the reusability, the restrictions and the quality assurance process depends on the corresponding project.

Apart from the publicly available data, other data will also be collected and generated that are confidential due to internal regulations and/or legal reasons. These data will not be publicly available. Such data will be either shared within the consortium or become accessible in-house after a proper agreement is signed. In this manner FELICE will ensure a well-structured data management for both public and confidential data.

4.1 Making data findable, including provisions for metadata

The data generated or collected within FELICE should be easily discoverable by researchers. This can be achieved by providing rich metadata and a unique and persistent identifier.

As described in section 3.3, FELICE will use a standardized naming convention for all the project data and will facilitate citation by assigning DOIs to the uploaded online datasets. Search keywords for publications and datasets will also be used as a measure of increasing their re-usability. Furthermore, the data will be accompanied with rich metadata, providing context and additional information about the datasets. These provisions will make the data more findable and will increase the re-usability and impact of the project (c.f. section 3.4).

4.2 Making data openly accessible

As per Article 29.2 of the Model Grant Agreement under Horizon 2020, the FELICE consortium will ensure open access to all peer-reviewed

---

[24] *https://www.force11.org/group/fairgroup/fairprinciples*

scientific publications related to its results. The consortium will strive for gold open access for its publications. Additionally, as per Article 29.3, the FELICE consortium will also ensure open access to research data. The decision on whether to publish through open access will have to account for the potential necessity of protecting sensitive data.

All public datasets, scientific publications, software and deliverables will be made openly available, free of charge on Zenodo (c.f. section 3.9.4). The final version of DMP will include possible exceptions that might arise in cases of privacy concerns or protection for commercial exploitation. Confidential data will be available internally on the SVN repository[25] or Large Seafile File Store of the project (c.f. section 3.9.2, 3.9.3).

4.3 Making data interoperable

FELICE will make data interoperable as described in section 3.7, by using commonly used metadata vocabularies and the shared metadata standards. Table 4 presents the general overview of the FELICE metadata structure, which however might slightly be updated throughout the project.

Table 4: FELICE Metadata

Title	Name of the data
Creator / Responsible Partner	Name of the partner responsible for the data created
Dataset Identifier	Data's internal reference number
DOI	(if applicable)
Dataset Description	A brief description of the data
Work Package/Deliverable	Associated work package and deliverable/task
Source	How the data have been generated
Processing	How the data have been processed
Repository	The repository where the data will be uploaded
Language	All languages used in the data
Code list	Explanation of codes or abbreviations used
Type	Types of the data
Format	Formats of the data
Expected Size	An approximation of the size of the data
Keywords	Keywords describing the content of the data
Version	Unique identifier for each version of the data
Date of Repository Submission	Release date (preferred format yyyy-mm-dd)
Necessary software	Necessary software needed to create, view or analyse data
Rights	Any rights information on the use of the data
Access Information	Where and how your data can be accessed by other researchers

---

[25] *http://subversion.apache.org*

4.4 Increase data re-use (through clarifying licenses)

FELICE will increase data re-use by encouraging third parties to access, process, reproduce and disseminate the project's publicly available data free of charge (c.f. section 3.6).

5 Allocation of resources

FELICE data will follow the FAIR principles, as also mentioned in section 4. To achieve this, publications, software and datasets will be made publicly accessible on different free-of-charge, self-archived and open access repositories (c.f. section 3.9), enabling the findability, accessibility, interoperability and reusability by the research community.

The cost of the necessary actions for making the data FAIR, are mainly related to personnel costs and include preparation of data by each project partner for publishing, updating and maintaining the data, data hosting and backup, data sharing, and security. These costs will be covered by the project funds since according to Article 6 and Article 6.2.D.3 of the Grant Agreement (GA), costs related to open access to research data in Horizon 2020 are eligible for reimbursement during the duration of the project.

The FELICE EPM, QAM and DQC will be responsible for the data management within FELICE. They will ensure the drafting and updating of the DMP, and will additionally control the appropriate storage, management, and sharing. However, every individual partner is responsible for implementing and respecting the policies of the FELICE Data Management Plan.

After the completion of the project, the data will be preserved for a long-term period. The related costs though are hard to estimate at this first version of the Data Management Plan. The data will be preserved for up to 3 years after the end of the project. However, preserving datasets on the Zenodo repository where a single dataset file does not exceed 50 GB, is free of charge. Moreover, internal datasets of the project will be stored and preserved in the SVN repository[26] hosted by the coordinator. These costs are estimated to be zero. Nevertheless, the final decision regarding the costs of preserving datasets has not been made yet.

---

[26] *http://subversion.apache.org*

6 Data security

As analysed more extensively in section 3.8, FELICE will ensure the data security by applying appropriate security principles through the lifecycle of the data, while complying with the GDPR and being monitored by the FELICE EPM (Dr. Maria Crociani (EUNL)). The data will be stored and preserved through and after the completion of the FELICE project in certified repositories, as discussed in section 3.9.

7 Ethics aspects

Partners within FELICE will comply with the highest of ethical standards to ensure the ideal balance between research objectives and the means by which the project partners go about achieving these. Research activities will comply with ethical principles, and applicable international, EU and national law, satisfying by design all relevant compliance requirements for each specific activity. They will ensure respect for people and for human dignity, fair distribution of burden and research benefits, while at the same time they will protect the value, rights and interests of all research stakeholders.

FELICE deliverable D1.3 "Ethics and privacy manual" analyses very extensively the FELICE ethics and privacy aspects. It specifies all the considerations taken into account within FELICE regarding the research with human participants during the pilot and evaluation execution. It also specifies the security measures, guidelines and rules that will be followed by FELICE, to ensure the users' privacy and safety, respecting their rights as volunteer test subjects. In the following paragraphs, a short overview of the application of GDPR and ePrivacy Directive to the FELICE project is provided.

7.1 GDPR

The General Data Protection Regulation (GDPR) is the European Regulation on privacy since May 2018. Within FELICE, all sensitive data shall be processed lawfully, fairly and transparently, adhering to the GDPR. A more extensive analysis regarding the compliance of FELICE with GDPR is provided in the FELICE deliverable D1.3 "Ethics and privacy manual".

7.2 ePrivacy Directive (ePD)

The 2002 ePrivacy Directive 2002/58/EC[27] (amended in 2009[28]) is an important legal instrument for privacy in the digital age, and more specifically the confidentiality of communications and the rules regarding tracking and monitoring. The entry into force of the GDPR requires the EU legislator to update this text and the European Commission published a proposal on 10 January 2017. This new text will have to tackle the rapidly evolving technological landscape, with issues such as confidentiality of machine-to-machine communication (Internet of Things) or the confidentiality of individuals' communication on publicly accessible networks (such as public Wi-Fi). These European privacy laws regulate how data is allowed to be collected, processed and stored.

---

[27] *http://data.europa.eu/eli/dir/2002/58/oj*
[28] *http://data.europa.eu/eli/dir/2009/136/oj*

As a Directive, it is transposed into EU national laws rather than being imposed in a unified way as a Regulation is. FELICE project activities will design, develop, evaluate, and showcase solutions which highly rely on electronic communication. From these, new or improved products will emerge, as well as research outputs will be published and shared open and freely, satisfying also the EC's Open Access policy. The project will fully comply with Article 5, which directs the 27 EU member states to ensure that the storage of information or the gaining of access to information already stored on users' devices, is only allowed on the condition that the user has given their consent, having been provided with clear and comprehensive information about the purposes of the processing.

8 Other issues

There are no other national/funder/sectorial/departmental procedures that project partners must adhere to.

9 Data

This chapter will present the first version of the FELICE collected and generated data descriptions.

9.1 Scientific publications

FELICE research activities will lead to scientific publications created by the project partners. The authors/editors are responsible to make their publications publicly available, except if certain restrictions apply. The scientific publications should be provided free of charge in the Zenodo FELICE repository.

Table 5: FELICE scientific publications description

Field	Details
Name of the dataset	FELICE scientific publication
Responsibility	All partners of the consortium, editor and task leader
Purpose and relation to the objectives of the project	This dataset contains project journals and conferences papers/ posters, white papers, or other scientific publications.
Data type	NA
Data format	pdf
Data origin	Project research output
Expected size of the data	100Mb
Data utility	Researchers and scientists
Privacy principles	NA
Accessibility policy	Open except if certain restrictions apply
Security & storage	Zenodo repository

9.2 Other dissemination and communication publications

FELICE activities will lead to dissemination and communication (non-scientific) publications, like white papers, magazine articles, web pages, promotional materials (brochures, flyers, newsletters, posters, etc.), press releases, website news, posts (e.g. Facebook, Twiter, LinkedIn) and videos produced during the FELICE project. These publications will be publicly available free of charge.

Table 6: FELICE Dissemination and communication publications description

Field	Details
Name of the dataset	FELICE dissemination and communication publication
Responsibility	Dissemination manager, all partners of the consortium
Purpose and relation to the objectives of the project	Dissemination and communication of the FELICE results to increase its impact.

Data type	This dataset contains the research results in terms of publications and the data collection on the website and social networks. It involves different dissemination documents and videos.
Data format	Pdf, json, mp4
Data origin	Project dissemination and communication activities
Expected size of the data	10Gb
Data utility	The dissemination and communication publications will be useful for other researchers and groups in industry that are interested in increasing the impact of a project.
Privacy principles	NA
Accessibility policy	Open
Security & storage	Zenodo repository

9.3 Software

FELICE research activities will lead to the development of software serving the needs of the software modules. The different modules will require implementations in different programming languages and will provide several source code files as described below. The developed software will be available on Github[29] and will be subject to the quality assurance processes, as described in FELICE deliverable D1.1 "Quality assurance plan".

Table 7: FELICE Software source code description

Field	Details
Name of the dataset	Software source code
Responsibility	PRO, FORTH, ICCS, FHOOE, IML, AEGIS, UNISA, CAL-TEK, TUD, ACC
Purpose and relation to the objectives of the project	The source files serve the needs for all the software modules of the FELICE project.
Data type	Text files
Data format	py, npy, pyc, pyd, pyo, rpy, whl, h, hh, hpp, hxx cpp, c, cc, cxx, cs, csx, rsp, class, js, ts, html, scss, txt, json, csv, xml
Data origin	NA
Expected size of the data	To be defined
Data utility	The developed software will be useful to other research groups, tech companies and other research projects outside FELICE, working on related research topics or commercial areas.
Privacy principles	Github privacy principles
Accessibility policy	Github policy

---

[29] *https://github.com*

Security & storage	Github

9.4 Datasets

Apart from the publications and software, several research datasets will be collected and/or generated within FELICE, to serve the planned experiments and software modules. Descriptions of the currently foreseen datasets are provided next.

9.4.1   Object detection and localization

The object detection and localization dataset will serve the needs of the Task 4.1, to facilitate the human-robot collaboration. It will depict tools and components commonly met in industrial environment, focusing more on objects with challenging geometric and surface characteristics.

Table 8: FELICE Object detection and localization dataset description

Field	Details
Name of the dataset	Object detection and localization dataset
Responsibility	ICCS, WP4, T4.1
Purpose and relation to the objectives of the project	The dataset aims to support research in object detection and localization focusing on specific objects with challenging characteristics, like weak texture, reflective surfaces, symmetries and occlusions, commonly met in industrial (car manufacturing) environments.
Data type	Benchmark dataset of colour (RGB) and depth (D) images (RGB-D data) will be collected demonstrating specific assembly tools and components. The 3D information/models of the known objects will be available in the form of ply files. The camera calibration will be provided in text or structured file. The annotations (ground truth data) for the poses of the known objects for each image will be provided in text or structured files. Documentation of data structure, source code to parse the data etc in text files.
Data format	png, ply, xls, csv, doc, pdf, txt, html, xml, json Png image files, ply files containing the 3D information/models of the known objects. Text, csv or structured json or xml files containing the camera calibration and the object pose labels. Documentation will be provided in txt and pdf files.
Data origin	Visual sensors (cameras) installed at the CRF working environment or on-board the robotic platform, part of the annotation (ground truth) data will be provided by experts.
Expected size of the data	To be defined
Data utility	The dataset will be useful for other research groups, private technological companies and other research projects outside FELICE, working on related research topics or commercial areas.

Privacy principles	The image data will be post-processed in order to remove the identity of potential users depicted in the scene with masking/blurring face characteristics.
Accessibility policy	Open
Security & storage	Zenodo repository

9.4.2   Robot localization and mapping

The robot localization and mapping dataset will serve the needs of the Task 4.2, to facilitate the robot navigation. It will depict an industrial environment, captured under realistic conditions and during common manufacturing activities.

Table 9: FELICE Robot localization and mapping dataset description

Field	Details
Name of the dataset	Robot localization and mapping dataset
Responsibility	FORTH, WP4, T4.2
Purpose and relation to the objectives of the project	The dataset aims to support research and development in robot localization and mapping (SLAM) in a realistic industrial environment (specifically, a car assembly line). The dataset will be captured during normal manufacturing activities. Despite the abundance of public SLAM datasets, they generally do not adequately represent the particularities exhibited by FELICE's environment (in particular, frequent dynamic scene motions).
Data type	Video sequences and stills of (RGB) colour and (D) depth data (RGB-D data) will be recorded by a moving robot as it navigates in the assembly line. Additionally, camera calibration information and possibly some sort of ground truth localization data will be collected. The data will be accompanied by relevant documentation.
Data format	avi, mp4, svo, jpg, png, xml, json, xls, csv, doc, pdf, txt, html AVI, MPEG-4 video files, JPG or PNG images, proprietary video files (e.g. SVO by StereoLabsTM), text or CSV files for calibration and ground truth data. Documentation will be provided in text and PDF files.
Data origin	Visual sensors (cameras) installed on-board the robotic platform.
Expected size of the data	Several Gbs
Data utility	The dataset will be useful to researchers and practitioners working on related research topics or commercial areas.
Privacy principles	The video data will be post-processed in order to remove the identity of humans by masking/blurring their facial characteristics and only after obtaining approval from those involved through the informed consent forms.
Accessibility policy	Restricted. The data will visually depict indoor spaces owned by a private company (CRF), hence appropriate

	clearances are needed for sharing them. Access will be provided through password protection mechanism.
Security & storage	FELICE Seafile cloud service

9.4.3   User actions and postures

User actions and postures dataset will serve the needs of the Task 4.3, to facilitate human body posture estimation. It will capture videos of workers' actions and postures under realistic conditions and during common manufacturing activities.

Table 10: FELICE User actions and postures dataset description

Field	Details
Name of the dataset	User actions and postures dataset
Responsibility	ICCS/FORTH, WP4, T4.3
Purpose and relation to the objectives of the project	The dataset aims to support research in human body pose estimation, tracking and action/posture recognition in the context of symbiotic human-robot collaboration in a realistic industrial (car manufacturing) environment. Despite the reasonably accurate performance of state-of-the-art algorithms in controlled or semi-controlled scenarios, there are limitations in coping with complex, realistic scenarios. FELICE aims to capture video data with workers' actions and postures during manufacturing activities.
Data type	Both qualitative and quantitative data will be generated. Video files and image sequences of (RGB) colour images and (D) depth images (RGB-D data) will be collected demonstrating the assembly actions and postures of workers during assembly tasks in the shop floor. Additionally, for each video the annotations (ground truth data) corresponding to semantic/linguistic labels/types of actions, postures and user ids will be generated in structured text files. Information on the 2D/3D human body pose estimated and tracked across time (coordinated and angles of body joints estimated) based on a skeletal model will be provided as numeric data. Documentation of data structure, source code to parse the data etc.
Data format	avi, mpeg-4, svo, bvh, jpg, png, xml, json, xls, csv, doc, pdf, txt, html AVI, MPEG-4 video files, jpg or png images, svo video files (proprietary file format of StereoLabsTM containing stereo colour and depth frames), text or csv files of linguistic/semantic action/posture labels, text or structured json or xml files of numeric values corresponding to 2D/3D positions and angles of human body joints estimated, .bvh file format of human motion data. Documentation will be provided in txt and pdf files.
Data origin	Visual sensors (cameras) installed at the CRF working environment, part of the annotation (ground truth) data

	will be provided by experts.
Expected size of the data	To be defined
Data utility	The dataset will be useful for other research groups, tech companies and other research projects outside FELICE, working on related research topics or commercial areas.
Privacy principles	The video data will be post-processed in order to remove the identity of users with masking/blurring face characteristics only after obtaining approval from the data subjects involved through the informed consent forms. Video data sequences will be made available in uncompressed format and compressed format based on the ITU-T H.264 codec.
Accessibility policy	Open
Security & storage	Zenodo repository

9.4.4   Speech command

The speech command dataset will serve the needs of the Task 4.4, to facilitate human-robot vocal and textual communication.

Table 11: FELICE Speech command dataset description

Field	Details
Name of the dataset	Speech command dataset
Responsibility	UNISA, WP4, T4.4
Purpose and relation to the objectives of the project	Recording the speech commands that the robot must recognize, in terms of voice samples and text.
Data type	Voice samples and text of the speech command
Data format	Wav, json Voice samples in WAV, ground truth (path, text, duration) in JSON
Data origin	To be defined
Expected size of the data	To be defined
Data utility	Research groups and private sector
Privacy principles	NA
Accessibility policy	Open
Security & storage	Zenodo repository

9.4.5   Gesture command

The gesture command dataset will serve the needs of the Task 4.4, to facilitate human-robot gestures.

Table 12: FELICE Gesture command dataset description

Field	Details
Name of the dataset	Gesture command dataset
Responsibility	UNISA, WP4, T4.4
Purpose and relation to the objectives of the project	Recording video samples of the gesture commands that the robot must recognize.
Data type	Video samples and metadata for the gesture commands
Data format	mp4, json Video samples in MP4, ground truth (path, type, duration) in JSON
Data origin	To be defined
Expected size of the data	To be defined
Data utility	Research groups and private sector
Privacy principles	NA
Accessibility policy	Open
Security & storage	Zenodo repository

9.4.6   ADAPT workflow/meta-model

The Asset-Decision-Action-Property-RelaTionship (ADAPT) workflow/ meta-model dataset will serve the needs of the WP7, to facilitate the workflow execution.

Table 13: FELICE ADAPT workflow / meta-model dataset description

Field	Details
Name of the dataset	ADAPT workflow / meta-model
Responsibility	FHOOE, WP7, T7.3
Purpose and relation to the objectives of the project	Data format for the ADAPT Workflow (meta) models. These are generated/modified based on data from the process optimization in a previous step in the toolchain. Required for the workflow execution by the workflow runtime.
Data type	Structured text data (ADAPT-workflow elements such as actions, decisions, assets and their relationship to each other)
Data format	xml
Data origin	Data is generated at runtime. Outcome depends on input data provided by the process optimizer.
Expected size of the data	< 5MB
Data utility	Research groups (workflow performance / execution times)
Privacy principles	NA
Accessibility policy	Restricted. Data is generated (dynamic / changing). Data must be shared with other elements of the toolchain but

	might contain sensitive data such as execution time.
Security & storage	FELICE SVN repository

9.4.7   Adaptive workstation operation dataset

The adaptive workstation operation dataset will serve the needs of the WP5, to develop an adaptive workstation for industrial environments and to facilitate the human-workstation interaction.

Table 14: FELICE Adaptive workstation operation dataset description

Field	Details
Name of the dataset	Adaptive workstation operation dataset
Responsibility	TUD, WP5, T5.2
Purpose and relation to the objectives of the project	Anonymised data of the person (in advance): - to allow physical adaptation of the workstation based on anthropometric and other data - to allow cognitive adaptation of information - to allow adaptation of illumination Some anonymised data related to stress-strain concept of the test persons (optional, in concept): - to give recommendations on how to reduce stress
Data type	Structured text data of anonymised test persons: age, sex, some anthropometric data (e.g.) body height, work experience. Some anonymised physiological data (optional, in concept): e.g. heart rate or heart rate variability (Excel format)
Data format	xlsx, csv, doc, pdf
Data origin	Some anonymised personal data: Surveys filled out by the test persons or measured before experimentation (using scales, measuring tape). Anonymised data from: Measurement systems e.g. polar 10 belt or smartwatch.
Expected size of the data	Some MBs
Data utility	The data might be useful for ergonomists and research groups, working on adaptive systems or industrial groups who want to improve working conditions and reduce risks related to work-related injuries.
Privacy principles	Use a worker-ID for anonymization when collecting data.
Accessibility policy	Restricted. Data is only to be shared after anonymization and signed agreement from the data owners that allows to share the data with third partners/the public.
Security & storage	The datasets will be stored or are already stored on the FELICE project server and TUD's data is stored on the IAD institutes server.

10 Conclusions

The FELICE project is expected to provide the consortium partners with the opportunity to collect, generate and process different types of research data as part of their research and development work. The current deliverable presented the Data Management Plan according to which all the identified FELICE types of data (scientific publications, dissemination and communication publications, software and datasets) will be handled during the implementation and after the completion of FELICE. The presented DMP was created in order to follow the FAIR principles[30] and the H2020 ORD principle, according to which the project data should be made "as open as possible, as close as necessary"[31].

As expected, for this first DMP version, it was not yet feasible to define all the details of the data to be collected or generated within FELICE. In the future and during the project, the DMP will be constantly updated, to incorporate the changes that might arise regarding the datasets, the methodologies and policies.

---

[30] *https://www.force11.org/group/fairgroup/fairprinciples*
[31] *https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm*