This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820767


**CoLLaboratE**

Co-production CeLL performing Human-Robot Collaborative AssEmbly

# D7.8 - Data Management Plan (Preliminary)

Due date: M6

**Abstract:** The present document is a deliverable of the CoLLaboratE project, funded by the European Commission's Directorate-General for Research and Innovation (DG RTD), under its Horizon 2020 Research and innovation programme (H2020). It provides the project reference manual and Quality Plan of the CoLLaboratE consortium. The document consists of a project overview section, where a quick reference of the main project facts is provided. Afterwards, the Project Management Plan is described, whereas in the next section, information relative to control of reporting and documentation is presented. The remaining part is dedicated to the description of the Quality Plan, which sets the quality standards for CoLLaboratE and circulates the message for excellence in the deliverable reports

| Dissemination Level | | |
|---|---|---|
| PU | Public | x |
| PP | Restricted to other programme participants (including the Commission Services) | |
| RE | Restricted to a group specified by the consortium (including the Commission Services) | |
| CO | Confidential, only for members of the consortium (including the Commission Services) | |

## Document Status

| | |
|---|---|
| Document Title | D7.8 - Data Management Plan (Preliminary) |
| Version | 1.0 |
| Work Package | 7 |
| Deliverable | 7.8 |
| Prepared by | CERTH |
| Contributors | ALL |
| Checked by | ALL |
| Approved by | AUTH |
| Due Date | 31/03/2019 |
| Completion Date | 29/03/2019 |
| Confidentiality | Public |

## Document Change Log

Each change or set of changes made to this document will result in an increment to the version number of the document. This change log records the process and identifies for each version number of the document the modification(s) which caused the version number to be incremented.

| **Change Log** | **Version** | **Date** |
|---|---|---|
| First draft | 0.1 | February 1, 2019 |
| Updated | 0.2 | March 12, 2019 |
| Contributing partners' input | 0.7 | March 25, 2019 |
| Internal reviewers' feedback | 0.9 | March 28, 2019 |
| Final | 1.0 | March 29, 2019 |

# EXECUTIVE SUMMARY

The present document is a deliverable of the CoLLaboratE project, funded by the European Commission's Directorate-General for Research and Innovation (DG RTD), under its Horizon 2020 Research and innovation programme (H2020). It presents the first version of the project Data Management Plan (DMP). The current document explains in detail what data will be generated throughout the project's lifecycle, the possible means for the sharing of this data in order to become accessible for verification and reuse, as well as the ways in which it will be curated and preserved. Additionally, it provides the necessary information in order for the Data Management Portal to be afterwards created through this project's activities.

It is strongly emphasized that this is an ongoing document that is being evolved along with the project progress and will be regularly updated in order to reflect up-to-date information.

# Table of Contents

**Executive Summary** ..... **3**

**Abbreviations and Acronyms** ..... **5**

**1 Introduction** ..... **6**

**1.1 Purpose** ..... **6**

**1.2 General Principles** ..... **6**

1.2.1 Participation in the Pilot on Open Research Data ..... 6

1.2.2 Archiving, Preservation and Re-usability ..... 6

1.2.3 IPR Management & Security ..... 7

1.2.4 Personal Data Protection ..... 7

**2 Teaching from Demonstration Dataset** ..... **7**

**2.1 Dataset reference and name** ..... **7**

**2.2 Dataset Description** ..... **7**

**2.3 Standards and Metadata** ..... **8**

**2.4 Data Sharing** ..... **8**

**2.5 Archiving, Preservation and Re-usability** ..... **8**

**3 Learning of physical human robot cooperation Dataset** ..... **9**

**3.1 Dataset reference and name** ..... **9**

**3.2 Dataset Description** ..... **9**

**4 Gesture Recognition Dataset** ..... **9**

**4.1 Dataset reference and name** ..... **9**

**4.2 Dataset Description** ..... **9**

**5 Ergonomic Performance Dataset** ..... **10**

**5.1 Dataset reference and name** ..... **10**

**5.2 Dataset Description** ..... **10**

**6 Production Planning** ..... **10**

**6.1 Dataset Reference Name** ..... **10**

**6.2 Dataset Description** ..... **10**

**6.3 Archiving, Preservation and Re-usability** ..... **10**

**7 Conclusion** ..... **11**

**8 References** ..... **12**

**ANNEX I** ..... **13**

# ABBREVIATIONS AND ACRONYMS

| Partner's short name | Partner's full name |
|---|---|
| AUTH | ARISTOTLE UNIVERSITY OF THESSALONIKI |
| CERTH | CENTRE OF RESEARCH AND TECHNOLOGY HELLAS |
| ARMINES | ASSOCIATION POUR LA RECHERCHE ET LE DEVELOPMENT DES METHODES ET PROCESSUS INDUSTRIELS |
| JSI | INSTITUT JOZEF STEFAN |
| IDIAP | FONDATION DE L'INSTITUT DE RECHERCHE |
| UNIGE | UNIVERSITA DEGLI STUDI DI GENOVA |
| KU Leuven | KATHOLIEKE UNIVERSITEIT LEUVEN |
| LMS | UNIVERSITY OF PATRAS |
| CRF | CENTRO RICERCHE FIAT SOCIETA CONSORILE PER AZIONI |
| BOR | BLUE OCEAN ROBOTICS |
| ASTI | AUTOMATISMOS Y SISTEMAS DE TRANSPORTE INTERNO SA |
| KOL | KOLEKTOR ORODJARNA NACRTOVANJE IN IZDELAVA ORODIJ TER ORODJARSKE STORITVE D.O.O.S |
| ARCELIK | ARCELIK A.S. |
| ROMAERO | ROMAERO S.A. |

# 1 INTRODUCTION

## 1.1 PURPOSE

The purpose of this deliverable (D7.8 "Data Management Plan (preliminary)") is to deliver an analysis of the main elements of the Data Management Policy that will be used by the consortium with regard to all the datasets generated by the CoLLaboratE project. The DMP is not a fixed document, but will evolve throughout the project's lifecycle. This first version of the DMP includes an overview of the datasets to be produced by the project as well as the specific conditions that are attached to them. The next version of the DMP will be published at M36 through the activities of D7.9 and will describe in more details the data generated as well as the uses identified by the consortium.

## 1.2 GENERAL PRINCIPLES

Through the activities of the CoLLaboratE project [1], pioneer research will be carried out in order to develop and deliver genuine human–robot collaboration for performing assembly tasks in a production cell designed to provide a safe collaborative environment. The project will build upon state-of-the-art methods for teaching the robot assembly tasks using human demonstration, extending them to facilitate human-robot collaboration. To this end, human participants will be involved in the project and data will be collected regarding their assembly's movements and assembly parts in a production line. For the purpose of optimizing the project's development, a process of knowledge management will be implemented. This process will provide the consolidation of the knowledge spiral, enable cooperation and will additionally allow for the creation of new knowledge. All the participants of the project have to cooperate in order to reach the most efficient process of knowledge management. Algorithms that have been implemented to identify objects, grasping and perform gesture recognition, etc. will be used before their adjustment on a production line. Therefore, a database to store data for benchmarking the algorithms developed in the project lifetime and beyond is required. Several experiments will be made using the algorithms both in the lab and/or on a production line and each experiment will derive significant data. Developers will refer to these data with a view to obtain information in order to increase the efficiency of the implemented algorithms. Moreover, a part of the performed experiment's data and of the algorithm's code will be provided to the scientific community as well as to robotics researchers in order to support the optimization of their executive power (e.g. utilizing github repository for open access to code developed in the project lifetime as well as publication to open access journals).

### 1.2.1 Participation in the Pilot on Open Research Data

CoLLaboratE highly supports the Pilot on Open Research Data launched by the European Commission along with the Horizon2020 programme, and therefore a significant part of research data generated by the project will be made open and will be offered to the Open Research Data Pilot, where CoLLaboratE will participate. To this end, the Data Management Plan provided through the activities of this deliverable, explains in detail what data the project will generate, whether and how it will be exploited or made accessible for verification and reuse, and how it will be curated and preserved.

### 1.2.2 Archiving, Preservation and Re-usability

The datasets, unless otherwise specified, will be preserved in CoLLaboratE internal NextCloud platform. Datasets will remain available online for the duration of the project. The platform is going to be shut down after the project has ended. Then, only datasets qualifying for public use, in

addition to the project database will be also stored in the ZENODO platform [2] which is the open access repository of the OpenAIRE [3].

### 1.2.3 IPR Management & Security

Due to the high innovative nature of the CoLLaboratE project, high level technologies will be developed during the project's lifecycle in order to be afterwards released in the market. Therefore, foreground capable of industrial or commercial application must be protected taking into account legitimate interests. All involved partners have Intellectual Property Rights on the technologies and data developed or collected with their participation. As the partners' economic sustainability highly depends on these technologies and data, CoLLaboratE Consortium will protect all data collected for CoLLaboratE purposes. Additionally, prior notice of dissemination will be given to other participants, whereas any dissemination such as publications and patent applications must indicate the Community financial assistance. Moreover, appropriate measures will be taken for effectively avoiding a leak of data, while all data repositories of this project will be adequately protected.

### 1.2.4 Personal Data Protection

CoLLaboratE involves the carrying out of data collection in order to assess the technology and effectiveness of the proposed solution. This will be carried out in full compliance of any European and national legislation and directives relevant to the country where the data collections are taking place (INTERNATIONAL/EUROPEAN):

i. The Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data;
ii. The General Data Protection Regulation 2016/679 of the European parliament regarding issues with the processing of personal data and the free movement of such data;
iii. The Greek Law 2472/1997: Protection of Individuals with regard to the Processing of Personal Data, and
iv. The Greek Law 3471/2006: Protection of personal data and privacy in the electronic telecommunications sector and amendment of law 2472/1997.

More detailed information regarding data privacy issues can be found in Deliverable 1.2 "Ethics and Safety Manual for CoLLaboratE technology (preliminary)".

# 2 TEACHING FROM DEMONSTRATION DATASET

## 2.1 DATASET REFERENCE AND NAME

DS.01.CERTH.TfDDataset

## 2.2 DATASET DESCRIPTION

The dataset consists of a set of demonstrations performed by human workers for teaching the system collaborative tasks. The tasks involved within the CoLLaboratE application scenarios will be considered.

Regarding the acquisition of the data, an RGBD sensor will be utilised. Thus, two aligned streams will be used, extracted from one depth sensor (640X480) and one RGB camera (640X480). The two sensors operate in a low to medium range area (50cm to 3m). Sampling rate: 10 fps.

This dataset is directly related to "*Objective*2: To develop a framework that enables non-experts teaching human-robot collaborative tasks from demonstration"

## 2.3 Standards and Metadata

While there is no specified standard for Teaching from Demonstration, all safety standards, especially in case of utilizing a collaborative robotic platform during the demonstrations, will be addressed.

| Name | Description |
|---|---|
| ***ISO 10218-1:2011*** | Defines the requirements for the robots and hardware used |
| ***ISO 10218-2:2011*** | Defines the requisites for HRC industrial applications in terms of their system integration |
| ***ISO/TS 15066:2016*** | Describes the collaborative workplace and operations in support to *ISO 10218-2* |

The use of metadata will increase the findability of datasets and publications, while ensuring the acknowledgment of EU funding. It also establishes proper monitoring of the traffic and impact of the H2020 programme. The metadata provision guideline for datasets and publications will strictly follow the one described in the Open Access Infrastructure for Research in Europe, OpenAIRE [4]. A set of specific metadata, tailored to the CoLLaboratE project publications, aiming to promote and acknowledge CoLLaboratE actions, is presented in ANNEX I.

## 2.4 Data Sharing

Due to the sensitive nature of the data, access and sharing policy must comply with the General Data Protection Regulation 2016/679 of the European parliament regarding issues with the processing of personal data and the free movement of such data, along with the national laws of each consortium member. Data must also be anonymous to protect the identities of the involved persons.

During the lifetime of the project, the datasets are going to be stored in a database. Members of the consortium will be able to view and download the dataset at any time. Access to third parties will be provided upon request, given it has been examined and agreed to.

After careful consideration, datasets qualifying for public use, in addition to the project database will be also stored in ZENODO [5] which is the open access repository of the OpenAIRE [3].

## 2.5 Archiving, Preservation and Re-usability

For the proper creation, usefulness and re-usability of the dataset, compliance must be ensured with the standards and guidelines specified. Therefore, quality assurance and quality assessment comments will be requested from the dataset providers.

The datasets will be preserved in CoLLaboratE internal NextCloud platform. Datasets will remain available online for the duration of the project. The platform is going to be shut down after the project has ended. Then, only the qualifying as public components of the dataset, stored in ZENODO are going to be publicly available.

# 3 LEARNING OF PHYSICAL HUMAN ROBOT COOPERATION DATASET

## 3.1 DATASET REFERENCE AND NAME

DS.02.AUTH.LphrcDataset

## 3.2 DATASET DESCRIPTION

This dataset consists of physical collaboration between a human with a robotic manipulator aiming to investigate dynamic load sharing, stiffness adaptation, the human's intentions and online estimation of the collaborative movement's goal and duration.

The data that will be collected only include measurements from the robot's internal sensors such as the robot's position and the external forces applied by the human to the robot.

This dataset is related to "Objective 1: To equip the robotic agents with basic collaboration skills easily adaptable to specific tasks"

For this dataset, the same principles with DS.01 apply for standards, metadata, data sharing, archiving, preservation and re-usability.

# 4 GESTURE RECOGNITION DATASET

## 4.1 DATASET REFERENCE AND NAME

DS.02.ARMINES.GRDataset

## 4.2 DATASET DESCRIPTION

The dataset consists of a set of gestures performed by human workers for enabling the system to recognize professional gestures performed during collaborative tasks. The tasks involved within the collaborate application scenarios will be considered.

Regarding the acquisition of the data, an RGB-D sensor will be utilized. Thus synchronized streams will be used, extracted from one depth sensor (640x480) and one RGB sensor (640x480). The two sensors operate in a low to medium range (50cm to 3m). Sampling rate: 29fps.

For this dataset, the same principles with DS.01 apply for standards, metadata, data sharing, archiving, preservation and re-usability.

# 5 ERGONOMIC PERFORMANCE DATASET

## 5.1 DATASET REFERENCE AND NAME

DS.03.ARMINES.EPDataset

## 5.2 DATASET DESCRIPTION

The dataset consists of a set of professionally-captured human motions of industrial workers for recognition of dangerous work postures and other ergonomic risks factors. The tasks involved within the CoLLaboratE application scenarios will be considered as well as others which are known to have high ergonomic risk.

The data corresponds to Biovision Hierarchy (BVH) motion capture files recorded by using the BioMed bundle system from NANSENSE Inc. [6]. The system consists in 54 inertial measurement units placed throughout the body and hands (suit and gloves). Sampling rate: 1000 Hz.

For this dataset, the same principles with DS.01 apply for standards, metadata, data sharing, archiving, preservation and re-usability.

# 6 PRODUCTION PLANNING

## 6.1 DATASET REFERENCE NAME

DS.01.LMS.PlanDataSet

## 6.2 DATASET DESCRIPTION

This dataset is a precompiled dataset that contains data required for the demonstration of the production planning system that will interact with humans and robots.

Most of the data contained in this dataset will be stored in a Relational Database Management System (RDBMS) that we be used during the demonstration of the Production Planning module. Moreover, data will be collected while the module is being developed and be revised according to the requirements of the demonstration. Data will fit the developed model along with additional supportive files such as 3D files, configuration files etc.

## 6.3 ARCHIVING, PRESERVATION AND RE-USABILITY

For the proper creation, usefulness and re-usability of the dataset, compliance must be ensured with the standards and guidelines specified. Therefore, quality assurance and quality assessment comments will be requested from the dataset providers.

The datasets will be preserved in dedicated database server on LMS premises. Datasets will remain confidential for the duration of the project due to certain intellectual property rights of the end-users. After the project has ended, the server is going to be shut down and only the qualifying as public components of the dataset, stored in ZENODO are going to be publicly available.

# 7 CONCLUSION

This deliverable constitutes a first draft analysis of the procedures and infrastructures that will be implemented by CoLLaboratE in order to effectively manage the data produced through the project's activities. One of the key elements of the Data Management Plan constitutes the Data Management Portal, which will handle and manage the large amount of datasets collected from the devices used for the CoLLaboratE purposes.

Special care will be given in order for the Data Management Portal to allow specific access to all partners participating in the process of data production. Additionally, editing and access rights will be managed in an appropriate way. Moreover, special attention will be given by the CoLLaboratE data management plan to the appropriate collection and publication of metadata. All necessary information will be stored in order to facilitate the optimal use as well as the re-use of these datasets. Each data producer will be responsible for managing the respective data and metadata, whereas all data and metadata will be integrated in the Data Management Portal. Specific flexibility levels are required by the Data Management Portal regarding the public datasets, as well as attention towards the IPR rights of every partner and the European and National regulations and directives regarding personal data privacy and protection.

In conclusion, the current document presents a first overview of the datasets expected to be used and the kind of data gathered for the CoLLaboratE purposes, as well as for the specific challenges that need to be considered for their effective management. It is emphasized that this constitutes an ongoing document and will therefore be updated throughout the project's lifecycle. The final, updated version of this document will be delivered in M36 through the activities of D7.8, and will provide a more detailed Data Management Plan, whereas the Data Management Portal will be at its final stage by that time.

# 8 REFERENCES

[1] CoLLaboratE Grant Agreement Annex I – "Description of Action" (DoA)

[2] ZENODO: https://zenodo.org/

[3] OpenAIRE Horizon 2020 Project: https://www.openaire.eu

[4] OpenAIRE Horizon 2020 Project Guidelines: https://guidelines.openaire.eu/en/latest/

[5] ZENODO: https://zenodo.org/

# ANNEX I

The metadata provision policy is clearly described in OPENAIRE GUIDELINES. However, not all types of metadata contribute to increasing discoverability of publications and acknowledgement of EU funding. Hence, an effort has been made, to create a project-specific template of metadata provision, ensuring the inclusion of the most critical metadata. The metadata presented in the template must accompany every deposited publication, in addition to basic bibliographic information.

- EU funding acknowledgement:
  - Contributor: "European Union (EU)" & "Horizon 2020"
- Peer Reviewed type (e.g. accepted manuscript; published version).
- Project Information:
  - Grant number: “820767”
  - Name of the action: “Research and Innovation action”
  - Project Name: “Co-production CeLL performing Human-Robot Collaborative AssEmbly”
  - Project Acronym: “CoLLaboratE”
- Publication Date.
- Authors and Contributors.
- License.

Whenever possible, a Scientific Publication, as soon as possible and at the latest at the time of the publication, will be deposited in a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication, in a repository for scientific publications, considering OpenAIRE Repositories such as Zenodo, European PubMed Central, arXiv.org, Online Research Database in Technology, etc. Moreover, CoLLaboratE consortium will aim at depositing at the same time the research data needed to validate the results presented in the aforementioned scientific publications.