

Handling with AI-enhanced Robotic  
Technologies for flexible ManUfacturing

## D7.2- Data Management Plan



<b>Deliverable ID:</b>	D7.2
Project Acronym:	HARTU
Grant:	101092100
Call:	HORIZON-CL4-2022-TWIN-TRANSITION-01
Project Coordinator:	TEKNIKER
Work Package	WP7
Deliverable Type	R ---Document, report
Responsible Partner:	TEK
Contributors	ALL
Edition date:	30 June 2023
Version:	5
Status:	FINAL
Classification:	PU



This project has received funding from the European Union's Horizon Europe - Research and Innovation program under the grant agreement No 101092100. This report reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

## HARTU Consortium

HARTU “Handling with AI-enhanced Robotic Technologies for flexible manufactUring” (Contract No. 101092100) is a collaborative project within the Horizon Europe – Research and Innovation program (HORIZON-CL4-2022-TWIN-TRANSITION-01-04). The consortium members are:

1	 MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE	FUNDACION TEKNIKER (TEK)	Contact: Iñaki Maurtua inaki.maurtua@tekniker.es
2	 German Research Center for Artificial Intelligence	DEUTSCHES FORSCHUNGSZENTRUM FUER KUENSTLICHE INTELLIGENZ GMBH (DFKI)	Contact: Dennis Mronga dennis.mronga@dfki.de
3	 TECHNOLOGY CENTRE	ASOCIACIÓN DE INVESTIGACIÓN METALÚRGICA DEL NOROESTE (AIMEN)	Contact: Jawad Masood jawad.masood@aimen.es
4	 THE DIGITAL TRANSFORMATION COMPANY	ENGINEERING INGEGNERIA INFORMATICA S.P.A. (ENG)	Contact: Riccardo Zanetti riccardo.zanetti@eng.it
5	 TÜRK OTOMOBİL FABRİKASI A.Ş.	TOFAS TÜRK OTOMOBİL FABRİKASI ANONİM SİRKETİ (TOFAS)	Contact: Nuri Ertekin nuri.ertekin@tofas.com.tr
6		PHILIPS CONSUMER LIFESTYLE BV (PCL)	Contact: Erik Koehorst erik.koehorst@philips.com
7		ULMA MANUTENCION S. COOP. (ULMA)	Contact: Leire Zubia lzubia@ulmahandling.com
8		DEEP BLUE Srl (DBL)	Contact: Erica Vannucci erica.vannucci@dblue.it
9		FMI HTS DRACHTEN B.V. (FMI)	Contact: Floris Goet floris.goet@fmi-improvia.com
10		TECNOALIMENTI S.C.p.A (TCA)	Contact: Marianna Faraldi m.faraldi@tecnoalimenti.com
11	 Politecnico di Bari	POLITECNICO DI BARI (POLIBA)	Contact: Giuseppe Carbone giuseppe.carbone@poliba.it
12		OMNIGRASP S.r.l. (OMNI)	Contact: Vito Cacucciolo vito.cacucciolo@omnigrasp.com
13	 ITRI Industrial Technology Research Institute	INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE INCORPORATED (ITRI)	Contact: Curtis Kuan curtis.kuan@itri.org.tw
14	 INFAR INDUSTRIAL CO., LTD.	INFAR INDUSTRIAL Co., Ltd (INFAR)	Contact: Simon Chen simon@infar.com.tw

## Document history

Date	Version	Status	Author(s)	Description
[01/02/2023]	1	Draft	TEK	Template
[30/05/2023]	2	Draft	All	Contributions of all partners
19/06/2023]	3	Draft	TEK	Integrated version for internal review
26/06/2023	4	Draft	TEK	Ready for internal final review
30/06/2023	5	Final	TEK	Final submitted

## Public Executive Summary

---

HARTU will adhere to FAIR principles:

- Making data findable:
  - All open dataset, publications, and open-source software produced by HARTU will be identifiable and locatable using a persistent Uniform Resource Identifier (URI) and/or Digital Object Identifier (DOI) to ensure the content is easily and uniquely citable.
  - Metadata will be provided to allow discovery.
- Making data accessible
  - Data defined as public will be deposited in a trusted repository such as Zenodo. Additionally, all the links will be posted on the HARTU website and public code will be available on GitHub repository.
- Making data interoperable
  - Partners will observe OpenAIRE guidelines for online interoperability.
- Data re-usability
  - All datasets created by HARTU will provide Readme files that describe the organization and type of data, metadata, variable definitions, units of measurement etc.

HARTU project will use and generate several datasets. HARTU will make them available to stakeholders unless some restrictions apply for commercial or privacy reasons.

The type of datasets include:

- Labelled 2D images and 3D point clouds of objects and environments, some of them generated with simulation tools and the rest by means of real data acquisition.
- Time-series data acquired from robots' sensors. This includes trajectories of robot end effector pose, twist, spatial acceleration, and contact wrenches, as well as position, velocity, and acceleration of the robot joints.
- CAD of the workpieces and the positions of the corresponding contact points.
- Data on electro-adhesives performance (force vs displacement vs voltage) for different objects, data on surface properties of the objects and force-displacement data from the contact/force sensors.

## Table of contents

---

1	Introduction .....	7
2	Data Summary.....	7
2.1	Use of existing data .....	7
2.2	Types and formats of data .....	8
2.3	Purpose of the data generated or re-used.....	9
2.4	Expected size of the data generated or re-used .....	10
2.5	Origin of the data .....	10
2.6	Usefulness of the data outside HARTU .....	10
3	FAIR data .....	11
3.1	Making data findable, including provisions for metadata .....	11
3.2	Making data accessible .....	12
3.3	Making data interoperable.....	14
3.4	Increase data re-use .....	14
4	Other research outputs.....	15
5	Allocation of resources .....	16
6	Data security .....	16
7	Ethics .....	16
7.1	Legal and ethical issues .....	16
7.2	Informed consent .....	20
8	Other issues.....	21

## Acronyms

List of the acronyms	
<b>AI</b>	Artificial Intelligence
<b>DMP</b>	Data Management Plan
<b>DOI</b>	Digital Object Identifier
<b>EDA</b>	Exploratory Data Analysis
<b>FAIR</b>	Findable, accessible, interoperable, Reusable
<b>GDPR</b>	General Data Protection Regulation
<b>URI</b>	Uniform Resource Identifier
<b>URDF</b>	Unified Robot Description Format

# 1 Introduction

---

The Horizon Europe Model Grant Agreement requires that a data management plan ('DMP') is established and regularly updated.

HARTU is using the Template recommended for Horizon Europe beneficiaries (Horizon Europe, Data Management Plan Template, Version 1.0, 05 May 2021) to define this initial version of the DMP.

The DMP is considered 'a living document' and will be regularly updated to reflect changes that may arise or decisions that are implemented.

## 2 Data Summary

---

### 2.1 Use of existing data

Will you re-use any existing data and what will you re-use it for? If yes, describe it. State the reasons if re-use of any existing data has been considered but discarded.

HARTU will re-use existing data:

- Open-source CAD models with different primitive shapes and characteristics will be used to create artificial scenes simulating real picking scenarios to develop a generic object segmentation system.
- It will be also analysed the availability of labelled picking datasets at instance level. A real dataset generated by TEK with different objects labelled at instance level for bin-picking applications from the European project PICKPLACE<sup>1</sup> will be used in the project.
- Publicly available benchmark data sets for assembly operations in industry were considered but discarded. They usually do not contain information about contact forces, something that is considered as crucial for learning assembly skills from demonstration. Also, these data sets are often use-case specific. However, publicly available data sets might be used for initial testing. Also, own programming-by-demonstration data sets from previous projects<sup>2</sup> might be used in this context:
- Existing open/public datasets for object and environmental monitoring model generation.
- Existing open/public datasets of robots, end-effectors, and graspable objects for modelling URDF files based on the Unity and MuJoCo physics engine.
- HARTU will use published data on electro-active soft grippers performance (e.g., electroadhesion grippers) on objects with different materials/geometry for the design of the new electroadhesives.

---

<sup>1</sup> <https://pick-place.eu/>

<sup>2</sup> TransFIT - Flexible Interaction for infrastructures establishment, FKZ 50RA1701, <https://robotik.dfki-bremen.de/en/research/projects/transfit/>

- In addition, we will use existing published data on soft grippers design and performance with different classes of objects.

## 2.2 Types and formats of data

### What types and formats of data will the project generate or re-use?

2D images will be generated and stored in PNG, JPEG or TIFF format for generic object segmentation and 3D image data type in the point cloud data format for the object and environment monitoring. As for the instance-level labels associated with the objects in the scenes, the labels will be stored in COCO JSON format.

In the tasks involving the estimation of grasping points (T3.1, T3.2), CAD models of the objects will be used (valid formats are STL, STEP and OBJ). The association of the CAD model with its respective grasping points will be done using XML or JSON format files.

In the context of programming-by-demonstration, HARTU will mostly use time-series data acquired from robot sensors. This includes trajectories of robot end effector pose, twist, spatial acceleration, and contact wrenches, as well as position, velocity, and acceleration of the robot joints. Furthermore the use of motion capture data, e.g., using VR devices of tracking gloves will be considered. The data will be stored as binary log-files (e.g., rosbag), as well as in text-file format, e.g., csv.

For the simulation infrastructure for learning, the robots and end-effectors will be defined in the URDF data format.

In addition, we will also generate data from the tool-part contact quality monitoring and real prototype evaluation. During the tool-part monitoring (testing data) the force and torque data will be time series e.g., csv format. The evaluation of the prototype data will be of several types ranging from objective data type (performance comparison with benchmarks) and subjective data type (questionnaires, surveys, etc.). More concrete information will be provided once the evaluation campaigns will be designed.

We will collect data on electro-adhesives performance (force vs displacement vs voltage) for different objects (.csv), (2) generate and store data on surface properties of the objects (.csv) and (3) collect force-displacement data from the contact/force sensors.

We will collect grasping data for the pick and place of different objects in the form of (1) videos .mp4, (2) dataset .csv with an evaluation of success/failure of each grasp, (3) dataset of 3D shapes (e.g., point clouds .ply) with selected contact point positions and grasp outcome (success/failure).

With respect to type of data gathered in WP1, related to human factors, it will be mainly qualitative data deriving from interviews and observations, some quantitative indicators will be derived from online questionnaires and from the analysis of Human-AI robots interactions. The format of the data will vary from 2D images in png format, .docx and .mov for the interviews, as well as .xls and .pdfs. for the data analysis.

In WP6 the type of data gathered will be mainly related to dissemination materials from the partners involved in the project, and the format of the data is expected to be mainly .pdfs when



considering presentations, flyers and standard video formats for promotional videos, i.e., .mov, .avi, .mp4, etc.

## 2.3 Purpose of the data generated or re-used

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The purpose of generating a dataset of 2D images of different picking scenes using CAD models is to generate a generic object segmentation model, in order to help to better understand the scene and help to better detect the grasping points. Generated object segmentation model is related with the “R3: Object segmentation module” key exploitable result.

The 3D scenes datasets will be used to perform the self-supervised grasping point identification and grasping policies that later will be used to decide in a real scene which is the best candidate among the possible grasping points.

The 2D and 3D data will also be used for training the neural network models for object and environment monitoring.

The purpose of time-series data acquired from robot sensors, including trajectories of robot’s end effector pose, twist, spatial acceleration, and contact wrenches, as well as position, velocity, and acceleration of the robot joints, is to train machine learning models. They will allow the robot to adapt to changes in the assembly process, e.g., different variants of the same part, modifying the trajectory or control policy. The goal is to make production lines more flexible, and easier to reconfigure. Also, ML models are trained to let the robot adapt its control parameters with respect to the environment, and the status of the assembly process.

In the case of the simulation infrastructure for learning, the main objective will be to model as accurately and realistically as possible the robot, the end-effectors and the graspable objects.

Data corresponding to the tool-part quality monitoring will be two-fold. First, it will be used for properly calibrating the fiber bragg grating (FBG) based sensors output and its representation as force/torque values. Second, it will be used for the verification of the performance of the quality monitoring.

Data from literature and new one generated in the experimentation with electro-active grippers will help us shaping the design of the research to develop new electro-active fingertips and soft grippers.

Data related to human factors can be classified according to the sources from which they originate:

1. **Industrial use cases** will provide the opportunity to gather private data coming from end-users and different working scenarios. These data will be used to understand the social impacts that AI and cobots could have in current manufacturing settings in terms of awareness, trust, ethics and liability, but also the current operators’ skills set and competences. The qualitative analysis will provide feedback from humans about their current experience as well as imagine future

possible collaborations with AI. The individual data gathered will be only shared anonymously within the project and, more broadly, with the scientific community.

2. **Human-AI interactions** will be analysed to understand where teaming and collaboration could best benefit the end users as well as the pilots' goals. A journey map will be generated to create a technical log of the human interaction with the system. The scope and results of this data collection will be of high importance for the scientific community studying how to better design AI systems with a human-centric trustful approach. The data collected will be anonymised.

Such data will be used with the objective to define how to better design AI systems (both in terms of the tools and interfaces) considering the end-user needs. The data will be analysed from a human factors perspective, including its ethical and legal aspects, and will guide the development of trustworthy AI-based robotic handling.

## 2.4 Expected size of the data generated or re-used

What is the expected size of the data that you intend to generate or re-use?

Each dataset may size several Gigabytes, e.g., it is estimated that for 2D segmentation it could be around 250 Gbs (50.000 scenes x 5 Mb); a few Gbs for contact classification and imitation learning per robot task; several Gbs corresponding to human factors data, or some tenths of Gbs for electro-adhesion performance data.

## 2.5 Origin of the data

What is the origin/provenance of the data, either generated or re-used?

There are many different origins of the data due to their different character, as explained in previous sections:

- Publicly available 2D and 3D images datasets, datasets generated in PICKPLACE project and others generated in HARTU, either through simulation and real experiments.
- Time-series data acquired from the robot's sensors in experiments developed in HARTU.
- Data generated by the FBG interrogator.
- The data collected on human factors will come from the project's industrial use cases.
- For electroactive fingers and soft grippers the origin of data will be the academic literature, publicly available datasets and those generated through experiments performed in HARTU.

## 2.6 Usefulness of the data outside HARTU

To whom might your data be useful ('data utility'), outside your project?

The main purpose of the data is, of course, to develop the HARTU's objectives. However, the data generated in the project may also be of interest to other audiences:

- Researchers and companies working in robotics (manipulation) and 2D/3D perception.
- Researchers in tactile sensing.
- Scientific community from the fields of Human Machine Interaction, Human Factors and Human Centred Design.
- Researchers working on electrically-enhanced soft robotics and soft grippers

## 3 FAIR data

---

### 3.1 Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

All open dataset, publications, and open-source software produced by HARTU will be identifiable and locatable using a persistent Uniform Resource Identifier (URI) and/or Digital Object Identifier (DOI) to ensure the content is easily and uniquely citable.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata will be provided to allow discovery. It will include the most relevant information:

- Digital identifier.
- Date and time.
- Description and other necessary details depending on data type.
- Reference to the project and grant.
- (Open source) software required to read data.
- Responsible person and/or company and contact details.
- Type of data.
- Version.
- Keywords.
- Experimental setup, including hardware used (e.g., robot version), description of the robot task, etc.
- Link to open-source software used in the experiments.
- Access and licensing info.
- When relevant:
  - Bibliographic information.
  - Related publications and reports.
  - Language.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Will metadata be offered in such a way that it can be harvested and indexed?

As explained, the metadata will include search keywords.

### 3.2 Making data accessible

#### **Repository:**

Will the data be deposited in a trusted repository?

Data defined as public will be deposited in a trusted repository such as Zenodo. Additionally, all the links will be posted on the HARTU website and public code will be available on GitHub repository.

Articles published as open access on journals will be made available through trusted repositories, and links included in the project website.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

It is not required for ZENODO. HARTU will create and manage its own community.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes, it does by assigning a persistent Uniform Resource Locator (URI) and DOIs

#### **Data:**

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Certain types of research development may require a level of confidentiality, and making the dataset and labels public could compromise the integrity of the research or the privacy of the participants involved. In such cases, it may be necessary to restrict access to the dataset, limit the amount of data shared or limit its use to non-commercial purposes only, in order to protect the interests of all parties involved.

Due to privacy and security concerns and possibly due to the commercial interests of the industrial partners, the users' research data gathered will not be made openly accessible as primary data but will be made accessible in a processed form, after their anonymisation and analysis.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Within consortium without any restriction but outside consortium. At least until the publication date of the scientific article or until the patent application is deposited.

Will the data be accessible through a free and standardized access protocol?

HARTU project will make selected datasets available to the public through either free, standard protocols by providing free downloads from Zenodo, Gitlab or from the project website.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

No restriction on the use of the data will exist by default, apart from giving the appropriate credit to the data creator and follow selected license. Any scientific article that uses the dataset generated in the project will have to cite the project and the different articles of the project that are related to it. The datasets that are defined as restricted, i.e., those that are important due to the commercial interests of the industrial partners will be required to fill out a form and sign it agreeing to comply with the terms and conditions set out in this document.

How will the identity of the person accessing the data be ascertained?

For public datasets, the identity of the person accessing the data is not required.

For the datasets classified as restricted, some personal data (name, surname, email, country and company) will be asked in the form to grant permission to download the data, ensuring that all privacy guidelines are respected.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

There is no need for a data access committee.

#### **Metadata:**

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

The metadata will be publicly available and includes sufficient information, such as direct links, to allow users to download the data.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Data published in trusted repositories (Zenodo, GitHub, project website) is expected to remain available according to the rules of the repositories.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g., in open source code)?

It will be necessary to provide documentation, references and links to the data reading/processing software, which will also be included, for example MessagePack, <https://msgpack.org/>.

### 3.3 Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

No universal standard exists for data exchange and reuse in robotics; we will utilize best practice whenever achievable. Partners will observe OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories, OpenAIRE Guidelines for Data Archives, OpenAIRE Guidelines for CRIS Managers based on CERIF-XML. These guidelines can be found at: <https://guidelines.openaire.eu/en/latest/>.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Yes

Will your data include qualified references<sup>3</sup> to other data (e.g. other data from your project, or datasets from previous research)?

Published datasets will contain proper citations to related scientific papers that reference data and vice versa.

### 3.4 Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

All datasets created by HARTU will be required to provide Readme files that describe the organization and type of data, metadata, variable definitions, units of measurement, etc.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

The datasets that are defined as restricted, i.e., those that are confidential or have commercial interests for the industrial partners, will be required to fill out a form and sign it agreeing to comply with the terms and conditions established. This dataset will be available only for non-commercial use.

Regarding non-critical datasets, HARTU project requires that all published datasets should be openly accessible and licensed under a MIT license or a comparable license.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

---

<sup>3</sup> A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

Yes, if correctly referenced and with the specific conditions for restricted datasets.

Will the provenance of the data be thoroughly documented using the appropriate standards?

YES

Describe all relevant data quality assurance processes.

Data quality assurance processes may involve the following steps:

- Data acquisition: Data will be captured using appropriate equipment and techniques, and the data will be properly calibrated to eliminate any distortions or errors. Data from different sensor sources are synchronized in time by suitable measures.
- Data pre-processing: Cleaning and pre-processing the data to remove any noise or artifacts that could affect the dataset.
- Data validation: Verifying the completeness of the data by performing an exploratory data analysis (EDA) and comparing it against reference datasets.
- Quality control: Monitoring the data acquisition and processing procedures to ensure that the data meets the desired quality.
- Documentation: Properly documenting the data collection and pre-processing procedures and providing detailed metadata to help understand the data and its limitations for future use.
- Feedback: Asking for feedback from users to identify any issues or errors in the data.

## 4 Other research outputs

---

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

The software algorithms for using and processing the data will be available on GitHub. Reusability is ensured by developing generic, hardware-independent software that can be configured for other hardware and use cases.

Research results will be published in open-access journals and presented in scientific conferences, workshops and tutorials.

## 5 Allocation of resources

---

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?

The planned repositories (Zenodo, GitHub) do not charge for their use.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

If needed, it will be covered by EU grant and corresponding funding scheme in Taiwan.

Who will be responsible for data management in your project?

TEK is the responsible and will receive assistance from all others.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Data will be published on platforms like Zenodo, which ensure long-term storage.

## 6 Data security

---

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

In addition to secure local server storage with backup on the cloud, data will be stored on public servers that ensure long-term and secure storage.

Whenever sensitive information will be included, data will be anonymized and comply with GDPR regulations.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes

## 7 Ethics

---

### 7.1 Legal and ethical issues

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Although the objectives include the study of the human-machine interaction and the treatment of datasets, the project falls out of the scope of the personal data and there is no risk of privacy or fundamental rights affection. In order to fulfil some of the objectives, i.e., the delivery of a Skills



Transformation Map and guidance about the interaction human-machine, HARTU foresees to work with humans but no vulnerable groups will be targeted and, in all cases, proper, complete and transparent information and voluntary consent compilation will be ensured. The data processing foreseen by HARTU might involve the collection of sensitive personal data, based on tracking and observation of participants on-site. In such a context, the results of the research have zero risk and burden to the participants and will benefit the individual or the group represented by the participant. The individuals are recognized as the starting point and key players of HARTU's value chain in the design of the technological solutions. They will gain benefits as employees (e.g., through insights provided to create trustworthy AI/robotic systems). Furthermore, they will even profit from HARTU as members of a group with an educated workforce, which will take advantage of the improved design of interactions with robotic systems, thanks to their insights.

The international scientific cooperation is also an objective of HARTU project. Therefore, HARTU consortium includes the scientific cooperation with Turkey and Taiwan. But no potential risks are foreseen in this sense: activities carried out outside the EU are similar to those planned inside although with different technological challenges, therefore, no eventual infraction of any EU law is at risk.

Furthermore, the AI developments within the project will be designed accordingly with the trustworthiness guidelines for all the AI-based techniques /systems made in Europe. All the developments performance will be properly quantified and tracked ensuring the possible detection of deviations from expected behavior. Considering the scope of this project, only strictly necessary data will be captured and processed, and datasets will be done fully accessible and provided as open datasets unless there are industrial restrictions, ensuring the multidisciplinary, participatory and collaborative nature of the work. The work proposed here does not go into conflict with any fundamental right of individuals, safety or discrimination. However, we would like to point out that the developments will be designed as explainable solutions to users, ensuring the transparency. Special attention will be provided to alert all potential users that the assessment of the AI model is not definite and independent test on the suggested solution regarding its compliance with regulations need to be done to validate the results.

The HARTU Ethical Policy relies on the following actions:

- Ethical Issues will be addressed at a first level within the consortium, for potential ethics issue arising from the activities to be performed during the project.
- Organization of open discussion and consultation with stakeholders on the privacy and ethics issues arising from project research. These participatory events will also address ethical and legal concerns and expectations.
- All the subjects involved in the demonstration activities will be strictly volunteers. The voluntary participants will be recruited from the pilot organizations following recruitment procedures and inclusion/exclusion criteria previously set, including also details on the sampling and recruitment process for each of the demonstrators. The HARTU (scientific and technological) objectives and project advancement, as well as use cases' refinement and pilots' planning and operation, will drive the identification of voluntary subjects. The

demonstrators will serve as a basis for identifying the number of volunteers (i.e., the sample size), as well as relevant statistical parameters of the sample. All necessary steps will be taken to eliminate bias within the selection process in order to avoid discrimination based on physical and cognitive aptitude and political, social, religious and cultural, gender orientation. Under no circumstances will vulnerable subjects be selected as HARTU voluntary participants; this includes persons under the age of 18 and any other person unable to give the informed consent.

- Informed consent procedures for the participation of humans will be defined and implemented. HARTU will maintain transparency (i.e., it will never intentionally deceive, mislead or withhold information from participants as to the purpose and nature of the investigation). All participants in the demonstrators will be informed of all aspects of the activities that might reasonably be expected to influence willingness to participate. The foundation of ethical research is the principle of informed consent. The voluntary participants, who agree to take part in a specific exercise, after the described information session and delivery of an information sheet, will be asked to fill and sign a consent form. The informed consent form provided to, and signed by, the volunteers involved in the use cases will explain all details of the use case or validation activity according to specific National and EU legislation. The project researchers will also discuss before and after each experiment with participants to maintain on-going consent. Informed consent procedures for data collection, storage and processing will be also defined and implemented.

Each potential voluntary participant will be adequately informed, before being involved in HARTU, and on the project and ethics and regulatory frameworks. They will be provided with an information sheet describing the project, its objectives and status, an explanation of the particular research activity related to the pilot and the test, the information to be collected and how that information will be used. Then they will be invited to sign a consent form. The Consent Form will clearly state that the project will preserve the user's right to privacy and anonymity and that the data arising from his/her participation will be only used for scientific research and system validation purposes. This form will detail the objectives, all elements (e.g., purpose, duration, procedures), and all the steps of the trial or activity, as well as data subject's rights. Participants will have the right to withdraw from the trial at any time, and to withdraw retrospectively any consent given and to demand that their personal data be destroyed. The Consent Form will be tailored specifically to each different test/technology and demonstrator's features. In all cases volunteers can withdraw at any time during any interview, observation, workshop and/or test session.

- If applicable, the ethics approvals for the involvement of humans or any other relevant ethics authorisation/approval in each demonstrator sites will be obtained and kept of file.

As regards data collection, processing, storing, and sharing, the HARTU Ethical Policy will outline adequate procedures and safeguards ensuring efficient data and privacy protection, based on the following:

- After the end of the project, all collected data that can relate to individuals will be put on hold and users will be noticed to extend their presence from the platform or get a copy of their data and remove it completely from the platform.
- All the test subjects will be informed and given the opportunity to provide their consent to any monitoring and data acquisition process.
- No personal or sensitive data will be centrally stored without consent. In addition, data will be scrambled where possible and abstracted in a way that will not affect the final project outcome.
- No data collected will be sold or used for any purposes other than the current project.
- A data minimisation policy will be adopted at all levels of the project. This will ensure that no data which is not strictly necessary to the completion of the current study will be collected.
- Anonymisation of the users' data right at their source for sensitive information used within the HARTU demonstrators, by removing the direct identifiers (e.g., name, address etc.),
- The data to be stored in the platform will be held securely using state-of-the-art encryption methods and the other techniques described in the project.
- Any ancillary personal data obtained during the course of the research will be immediately cancelled. However, the plan is to minimize this kind ancillary data as much as possible.
- Any incidental findings will be kept strictly confidential and erased from files under request from the enrolled subject.

In order to carry this process forward, the Consortium will work from the very early stages of the project together with the demonstrator users towards understanding the complete procedures needed to obtain consent for researching such data. Apart from the already research information, all other possible information sources will be analysed and included in the process of identifying:

- What data and information are relevant to the HARTU concept that will benefit the whole project.
- Whether consent to access this kind of data is necessary (e.g., open data publicly available vs. personal data protected by various laws).

All of the above will be discussed internally within the consortium and will be checked against national laws and GDPR and other EU legislations towards understanding the differences in each country and identifying the process to be followed for each pilot. A continuous interaction will be established and maintained between the Coordinator and the WP Leaders and the Ethics Manager in order to ensure the efficient management of ethical, privacy and data protection issues. This will allow us to regularly seek input from all the WP Leaders, prepare and regularly update the

Ethical Policy, as well as provide guidance and recommend workflows that comply with current European and national legislation and ethical standards and are well accepted by society. This will:

- Raise awareness by leaders of all relevant work packages of salient aspects of the project's impact on society, to couple scientific excellence with social awareness and responsibility and to align innovative outcomes to the values, needs and expectations of European society;
- Ensure that leaders of all relevant work packages include in their workflows the processes needed to make their project solutions and activities compliant with existing EU and national legislations and HE ethical guidelines;
- Ensure that any complaints or adverse incidents will be referred to the Coordinator and the ethics experts at DBL, for independent review and appropriate action, if necessary.

## 7.2 Informed consent

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Informed consent represents a strategy to preserve confidentiality and ensure anonymization that is strongly related to privacy and data protection rights (European Commission, 2010). Informed consent implies the consent of subjects to voluntarily participate in the research and to release any personal data. In this sense, prior to taking part in to the research activity, participants should have obtained all the information needed to make an informed decision to participate.

All participants taking part in research activities within HARTU, must give explicit consent to participate. For this purpose, they will receive an information letter and a consent form containing information on:

- What data will be collected;
- Means of collection (interview, observation, etc.);
- The purpose of the project and the expected results;
- Information on anonymization of data and confidentiality;
- Information on secure data storage, retention period and procedures for de-identification of any personal data;
- Information on access rights and authorised personnel;
- The rights of participants (i.e., participation is voluntary and can be withdrawn at any time without explanations and repercussions);
- Contact information for requests.

Moreover, consent to use data in aggregated form for publication purposes will be acquired. Participants should provide explicit consent in written form before participating. The informed consent form has to include the information sheet and a certificate of consent.

The research activities during HARTU involve the participation of humans and personal data collection and handling, to the extent that the project will interact with individuals in the demonstrators and gather, process and share data on their work environment. The purpose of this

research fully respects the human dignity and ethical values are full to the side of the individual who will be involved on a voluntary basis: the project will contribute to improving the general quality of work of business experts and will serve them with value-added information. The free and aware choice of the data subjects is assured by the provision to them of the risk assessment and privacy exposure report regarding their shared data. In this way, both in the validation phase and in the post-project uptake of HARTU developed solutions, it will be concretely ensured that the individual's consent is freely given, specific, informed and unambiguous, as requested by GDPR art. 7 and Recitals 32. Likewise, situations, where individuals feel pressurized to share, are avoided, while safeguarding human dignity at the maximum extent.

## 8 Other issues

---

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

Gender aspects will be considered to investigate how the introduction of new AI technologies can impact the gender balance and composition of the workforce, as well as the roles covered. Based on OECD PIACC data, 26% of women existing jobs is at risk due to automation, in middle term. Therefore, we will safeguard that the future career opportunities for HARTU's operators will be taking well into consideration gender particularities.

HARTU will include all findings and guidelines related to the gender dimension in the set of SSH public guidelines that it will produce (R16). In the initial stage of requirements capture, gender sensitive SSH methodologies will be selected. This will be reflected, for instance, in the design ethnography observation studies that will be conducted in the five use cases. The observation studies will involve users representing different genders, and the DBL researchers from HARTU will observe and enquire about any specific aspects that might be gender sensitive. This will include tailoring interview scripts with questions that may help elucidate the presence of aspects sensitive from a gender perspective.