Plan Overview

A Data Management Plan created using DMPonline.be

Title: DistriMuSe

Creator: Reda El Hail

Affiliation: KU Leuven (KUL)

Funder: European Commission (Horizon)

Template: Horizon Europe DMP +

Project abstract:

We are surrounded by a variety of more-or-less intelligent technical devices, designed to serve you us or others. Applications onin your mobile phones, wrist-worn health sensors on your wrists, autonomous
vacuum cleaners, robots on the factory floor and increasingly autonomous cars – all pledge to ease your tasks and keep usyou safe and healthy. SThe seamless interplay with these devices gets gainsmore
importane as these devices proliferate and grow in t with the increased autonomy and pervasive presence of the devices. We expect continuously available support fromin the services they provide – yet we want
them to disappear unobtrusively in the background when not needed. In order to provide support in a collaborative environment with human, physical and digital players, the technology needs to be equipped with senses to grasp human presence, their mental and physical state, their activities and their intentions. This is required to ensure human safety, safeguard their health, and allow for natural interaction.
This project intends to improve sensing of human presence, behaviour and health in a collaborative or common environment by means of multi-sensor systems.

ID: 209560

Start date: 01-05-2024

End date: 30-04-2027

Last modified: 21-01-2025

DistriMuSe
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Yes

Processing of personal data is very limited in our data. The personal information such contact information (name, email) will be used to contact the participants.The camera data is only recorded for annotation and data cleaning purposes, while it will not be shared publicly.
Information such as sex and age of the participant willl be shared as public data and constitue no probable high risk on the participant.

DistriMuSe
GDPR record

GDPR record

Have you registered personal data processing activities for this project?

- Yes

The personal data we use corresponds to video, and demographic data of participants.
Identification information (e.g. names, (email) addresses) Personal details (e.g. age, gender)
Audio and video recordings

DistriMuSe
Grant proposal

1. Data summary

1.1 Types of data/other research outputs

During this project radar data will be generated both for UWB radar as for FCMW radar. They differ slightly as for UWB we get Channel Impulse Responses (CIRs), while for the FCMW radar Doppler maps are generated during the different measurement campaigns. The data represent physical changes of the environment and are stored as an array or matrix of complex values. The sizes differ, on the one hand the UWB data is fairly compact with only a few Kilobyte for each second. while the FCMW radar generates 10s of MB each second.

Other research outputs include new Human Activity Recognition (HAR) algorithms. These algorithm will be able to detect activities such as sitting down, standing up, laying down, etc. for multiple people at the same time in some pre-known environments.

2. FAIR principles

2.1 Findability of data/research outputs

Question not answered.

2.2 Accessibility of data/research outputs

Question not answered.

2.3 Interoperability of data/research outputs

Question not answered.

2.4 Reusability of data/research outputs

Question not answered.

3. Resources and responsibilities

3.1 Curation and storage/preservation costs

Question not answered.

3.2 Person/team responsible for data management and quality assurance

Question not answered.

DistriMuSe
Full DMP

Version information

Action number

101139769

Action acronym

DistriMuSe

Action title

Distributed multi-sensor systems for human safety and health

DMP version number

v1.00

Date

Dec 19 2024

1. Data summary

1.1 Will you re-use any existing data and what will you re-use it for?

No, no previously generated data will be incorporated in this project.

1.2 What types and formats of data and other research outputs will the project generate or re-use?

Data content: Numeric sensor values from the UWB, FMCW and WiFi radar sensors will be collected and derived features will be generated from this data.
The UWB sensors will allow for the following data to be collected:

- timestamps of reception
- received power
- channel impulse response used for radar information
- Localization information of the person under test

The FMCW system will be used to collect:

- IQ stream of each receiving antenna
- doppler profile
- Localisation 3D coordinates

The WiFi system will be used to collect:

- timestamps of reception
- IQ stream of each received WiFi packet
- channel state information of each received WiFi packet
- phase difference between two receivers
- doppler frequency of the person under test

Data collection:
The project will collect raw sensor data from custom designed and off-the-shelf devices:

- UWB : Based on Qorvo DW3000 chipset
- FMCW: IWR6843AOPEVM by Texas Instruments
- WiFi: Custom hardware based on the OpenWiFi architecture

Data format: Sampled raw data will be stored as binary .bin or as .csv files. Alongside the raw data, metadata will be stored in JSON files with organized fields and naming conventions that make it easy to link metadata with specific data files. This setup ensures our data is stored efficiently while keeping metadata accessible and well-organized.
Data volume: From 50-500GB

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

The data generated will be used to train Machine Learning (ML) models to recognize activities of persons, such as, standing up, laying down, eating, etc.

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

As mentioned in question 1.2. The expected size is between 50GB to 500GB

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

All the data will be generated during the duration of the project. Measurement campaigns will be set up in order to collect this data.

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

To all research units interested in exploring radar usability to monitor human activities.

2.1 FAIR data: Making data findable, including provisions for metadata

2.1.1 Will data and other research outputs be identified by a persistent identifier?

- Yes: describe below

All research outputs wil be identified by persistent identifiers such as DOI or ORCID.

2.1.2 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.

First, the `README.md` file provides comprehensive information, including details on hardware, software, scenarios, data flows, and storage structure.
Next, a directory is created for each participant, which holds all the experiments conducted for that individual.
Finally, within each experiment folder, there is a UWB `.csv` file for each sensor and an FMCW `.bin` file for the FMCW radar. The `timestamps.csv` file contains the start and end times of each activity.

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

- Yes: describe below

yes, example keywords are: FMCW radar, UWB radar, human activity recognition dataset, Point Cloud data, Range Doppler

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

- Yes: describe below

Yes, a JSON file to store metadata will store metadata,. Harvesters can easily retrieve and parse the JSON file to index.
Here is an example of the structure of the metadata in JSON:

```
{
  "datasetTitle": "Human Activities Dataset Using FMCW and USB Radar Sensors",
  "description": "This dataset contains radar-based measurements of human activities collected using FMCW and USB radar sensors for monitoring and analysis. The activities include walking, running, sitting, and standing, captured in various environmental conditions.",
  "authors": [
    {
      "name": "",
      "affiliation": ""
    }
  ],
  "datePublished": "",
  "keywords": ["FMCW radar", "USB radar", "human activities", "activity recognition", "radar monitoring"],
  "license": "",
  "dataFiles": [
    {
      "fileName": "activity_data.csv",
      "fileFormat": "CSV",
      "fileSize": "50 MB",
      "downloadURL": ""
    },
    {
      "fileName": "readme.txt",
      "fileFormat": "TXT",
      "fileSize": "5 KB",
      "downloadURL": ""
    }
  ],
  "sensorDetails": [
    {
      "sensorType": "FMCW radar",
      "frequencyRange": "",
      "samplingRate": ""
    },
    {
      "sensorType": "USB radar",
      "frequencyRange": "",
      "samplingRate": ""
    }
  ],
  "activitiesMonitored": ["walk", "lay-down", "sit", "stand"],
  "collectionPeriod": {
    "startDate": "2025-01-01",
    "endDate": "2025-03-31"
  },
  "contact": {
    "name": "",
    "email": "",
    "website": ""
  }
}
```

2.2 FAIR data: Making data accessible

2.2.1 Will the data and other research outputs be deposited in a trusted repository?

- Yes: describe below

For preserving the data, the DTAI research group involved in the project will make use of the IT services provided by the DTAI. Ku Leuven ICT admin team offers a set of stable and secure services for storing data and software which can be further extended based upon needs. These services are centrally managed, have ample storage capacity (>200TB) and are automatically backed up.
ManGO servers is an Active Data Management Platform for researchers from KU Leuven. It is based on the open source software iRODS.
ManGO allows researchers to store and manage their data during the active phase of their research projects. This can be done via different clients, such as a command-line interface, a Python API or a web interface.
The KuLeuven GitLab offers all tools needed to support the software lifecycle. These services are offered to the entire KuLeuven researchers and external partners and are guaranteed to provide long-term preservation of the data and code.

2.2.2 Have you explored appropriate arrangements with the identified repository where your data and other research outputs will be deposited?

- Yes

2.2.3 Does the repository ensure that the data and other research outputs are assigned an identifier? Will the repository resolve the identifier to a digital object?

the repository ensures that all data and research outputs are assigned a persistent identifier, such as a DOI (Digital Object Identifier). The repository will resolve this identifier to the corresponding digital object, ensuring easy and reliable access to the data.

2.2.4 Will all data and other research outputs be made openly available?

- No, certain datasets cannot be shared openly for the following reasons:

No, camera data will be captured the measurement campaigns. The sole purpose of this video data is to verify that the actions of the user match the actions that need to be done. The camera data will be removed immediately after the data capture.

2.2.5 Is an embargo applied to give time to publish or seek protection of the intellectual property (e.g. patents)?

- No

2.2.6 If an embargo is applied (see question 2.2.5), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No embargo is applied

2.2.7 Will the data and other research outputs be accessible through a free and standardized access protocol?

- Yes: describe below

Data will be made available on request. Metadata will be made available using widely accepted standards, and datasets will be provided in open formats to maximize interoperability and usability.

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

Access to restricted data will be managed through a controlled access system. Users will be required to submit a request specifying the intended use of the data, which will be reviewed by the project team or data custodian. Approved users will be granted access under a data-sharing agreement that outlines usage terms, ensuring compliance with legal, ethical, and contractual obligations.

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

The identity of individuals accessing the data will be verified through a secure registration process requiring institutional email addresses and, if necessary, additional credentials such as ORCID iDs. Access will be granted only after identity verification and approval of the data access request.

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

- Yes

2.2.11 Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

- Yes

Yes, the metadata will be made openly available and licensed under a public domain dedication CC0, in compliance with the Grant Agreement. This will ensure the metadata is freely accessible and reusable without restrictions.

2.2.12 Will metadata contain information to enable the user to access the data?

- Yes

Yes, the metadata will include direct links to the data, as well as relevant information such as file formats and access instructions, enabling users to easily locate and retrieve the data

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

The data will remain available and findable for a minimum of 10 years after the project’s completion, in line with Horizon Europe requirements for long-term data preservation.

Yes, the metadata will remain available indefinitely, even if the data is no longer accessible, ensuring that key information about the data is preserved for future reference and citation.

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

No, the data is provided "as is". Other research will need to write their own code to read out the data.

2.3 FAIR data: Making data interoperable

2.3.1
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?

the data will adhere to community-endorsed interoperability best practices. Specifically, we will follow:

- **Data Formats**: Open, machine-readable formats such as CSV, BIN and JSON depending on the data type.

2.3.2 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, if project-specific vocabularies are necessary, we will provide mappings to more widely used and recognized ontologies to ensure compatibility and facilitate integration with existing systems and research.

2.3.3 Will your data and other research outputs include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

- No

2.4 FAIR data: Increase data re-use

2.4.1 How will you provide documentation needed to validate data analysis and facilitate data re-use?

We will provide comprehensive documentation that includes:

- **Data Provenance**: Detailed records of data collection, processing steps, and transformations to ensure reproducibility.
- **Methodological Documentation**: Clear descriptions of the analytical methods, algorithms, and models used, including any assumptions and limitations.
- **Code and Workflow Documentation**: Open source code for data analysis (e.g., Python scripts, Jupyter notebooks) will be fully documented to allow others to replicate and build upon the analysis.
- **Licensing Information**: Clear usage and citation terms to support data re-use and attribution.

This documentation will be made available alongside the data and be openly accessible to facilitate transparent validation, analysis, and reuse of the data by others.

2.4.2
Will your data and other research outputs be made freely available in the public domain to permit the widest re-use possible?
Will your data and other research outputs be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Yes, the data and other research outputs will be made freely available in the public domain to permit the widest re-use possible, in compliance with the FAIR principles and Horizon Europe guidelines.

Yes, the data and research outputs will be licensed using standard reuse licenses, such as CC BY (Creative Commons Attribution) or CC0 (Public Domain Dedication), ensuring compliance with the Grant Agreement and facilitating maximum reuse.

2.4.3 Will the data and other research output produced in the project be useable by third parties, in particular after the end of the project?

- Yes

The data and other research outputs will be made openly accessible, properly documented, and licensed for reuse, ensuring they remain usable by third parties even after the end of the project. For instance, this dataset might be a benchmark for machine learning researchers community to compare their models and publish their best modelling results.

2.4.4 Will the provenance of the data and other research outputs be thoroughly documented using the appropriate standards?

- Yes

Detailed records of data collection, processing steps, and transformations will be documented to insure similar pre-processing of the data among users.

2.4.5 Describe all relevant data quality assurance processes.

1. **Data Collection Standards**: Standard operating procedures will be followed during data collection to ensure consistency, accuracy, and reliability of the data.
2. **Data Validation**: Automated and manual validation checks will be applied to verify the accuracy, completeness, and integrity of the data. This includes cross-checking sensor data against known anomalies that can occur such as interference.
3. **Data Cleaning**: Outliers, missing values, and inconsistencies will be identified and addressed using appropriate data cleaning techniques, ensuring data integrity.
4. **Version Control**: All datasets, code, and analysis workflows will be version-controlled to track changes and ensure reproducibility of analyses.
5. **Metadata Quality**: Metadata will be standardized and thoroughly documented to accurately describe the dataset, ensuring ease of use and interoperability.
6. **Documentation**: Clear and comprehensive documentation will accompany all datasets.

3. Other research outputs

3.1 Do you have any additional information, that was not addressed in the previous sections, which you wish to provide regarding other research outputs that are generated or re-used throughout the project?

NO

4. Allocation of resources

4.1 What will the costs be for making data and other research outputs FAIR in your project?

- **Data Storage and Hosting**: Costs associated with depositing data in trusted repositories, ensuring long-term storage, and maintaining access to data.
- **Metadata Creation**: Time and resources required for developing standardized, high-quality metadata that meets FAIR principles.
- **Licensing and Compliance**: Costs associated with acquiring appropriate licenses for open access and ensuring compliance

with Horizon Europe requirements.
- **Personnel**: Staff time for data management, quality assurance, and ensuring the application of FAIR principles to all outputs.

4.2 How will these be covered?

The cost are part of the Distrimuse project overhead.

4.3 Who will be responsible for data management in your project?

Researcher will be responsible for the technical aspects regarding documentation, pseudonymization and metadata annotation with standards of the collected data sets during the data collection campaigns.

4.4 How will long term preservation be ensured?

All data will be stored long term, and the responsibility for data preservation will transition from the Project lead to department heads to ensure continuity. The data can be preserved for longer than 10 years on ManGo server. Long term data preservation is already available and part of the daily operation of the partners and is accounted for in the overhead.

5. Data security

5.1 What provisions are or will be in place for data security?

Our pseudonymization of this personal information forms a first security measure that is taken to ensure that the data cannot be linked back to the individuals who participated. The IDLab and KU Leuven Cloud storage system used has been designed to accommodate all types of research data, including sensitive and non-sensitive information. It foresees 2-factor authentication for anyone that logs into the platform. As such, access is only provisioned to persons involved in the research. Additionally, the data will be stored in an encrypted format using AES. Finally, this private cloud is managed by an experienced team with decades of experience in security and data privacy.

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

- Yes

We plan to use Mango servers to store the collect data along with its metadata and ensure its storage after the project for at least 10 years. However, personal data such as identification information will be stored on a paper stored in the closet of the researcher.

6. Ethics

6.1 Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

- Yes

The shared data should be completely anonymised and pseudonymized so it would not inherit any personal aspects of the participants.

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

- Yes

Informed consent forms contain everything about data sharing and preservation in long term.

7. Other issues

7.1 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)?

- No