Grant No: 688147 • Project Start Date: 2016-03-01 • Duration: 48 months

DELIVERABLE D8.2 REPORT

Initial Data Management Plan

Mary Ellen Foster (GLA), Katie Wright (GLA), Marketta Niemelä (VTT)

Beneficiaries:	**GLA**, VTT
Work Package:	WP8: Dissemination, innovation, and technology transfer

Version:	Final
Nature:	Report (R)
Dissemination level:	Public (PU)
Pages:	9
Date:	2016-09-02

This project has received funding from the European Union's Horizon 2020 research and innovation programme

1 Introduction

1.1 MuMMER overview

The MuMMER project aims to develop a humanoid robot (based on the Pepper platform, a robot created by industry partner Softbank Robotics) that can interact naturally in the challenging environment of a public shopping mall, providing an engaging and entertaining experience to the general public. To interact successfully in the mall, the robot will need to behave in a socially appropriate way. To support this, the project aims to develop and integrate new methods for audiovisual scene processing, social-signal processing, high-level action selection, and human-aware robot navigation. Throughout the project, the robot will be deployed in a large public shopping mall in Finland: initially for short visits to help with development and testing, and later for a long-term field study.

1.2 Deliverable overview

Deliverable 8.2 represents the initial Data Management Plan (DMP) of the MuMMER project. This initial DMP summarises the four main forms of research data that will be collected during the course of MuMMER and also aims to define the process for managing all research data generated over the course of the MuMMER project, including collection practices, distribution, access criteria, ethical practices and preservation.

As the initial DMP, this document reflects the current direction of the project regarding the collection of data. The DMP is a living document which will be amended and updated where required, and changes will be reported to the European Commission via the submission of the final DMP (deliverable 8.4, due in month 48).

2 General Principles

2.1 Open access

As a participant in the Horizon 2020 Open Data Research Pilot, one of the goals of the MuMMER project is to provide – wherever possible – accurate and high-quality data to the research community to enable the project to contribute to the growing field of social robotics. However, since the data that is collected may contain personal information about human participants, we will also ensure that all data is appropriately anonymised before being made available for open access.

2.2 Personal data/ethics[1]

The MuMMER studies will be conducted according to universal ethical principles (for instance, The Charter of Fundamental rights of the EU, 2000) and, in particular, according to Directive 95/46/EC of personal data and data protection and to Directive 2002/58/EC of electronic data and e-privacy. The research will comply with applicable international, EU and national legislations for protection of personal data.

In the MuMMER project, adult, healthy volunteers will participate in the co-design and evaluation activities. Their data and contribution will be gathered through focused discussions (focus groups and workshops), questionnaires, interviews and observations (real time and video). In all cases, it is emphasized that participation is voluntary and it is possible to quit at any time. Whenever applicable, formal ethical approval will be required from the corresponding partners' ethics boards and committees.

During the research, personal data – as it is defined in EU Data Protection Directive (95/46/EC) – will be collected. The data may include age, gender, area of living (no address), education, profession etc. Names or any personal identification data such as social security number are not collected for research purposes.

[1] This section is adapted from the MuMMER grant agreement, Part B, Section 5.1.

However, photographs or videos may be taken. Our principle is to always ask the participating volunteers to sign a written informed consent form in which they give their permission for participating the study, photo/video data collection and use of the data. We will also request separate consent for archiving the data collected to ensure that the participants are aware of and consent to the long term storage of the data.

Some research activities such as evaluation of proof of concepts and the field study will be carried out in the field. Possible arenas are shopping centres (of which at least Ideapark), airports, theatre and cinema lobbies, hotel receptions etc. To understand, design and validate the human-robot interaction in these evaluations, and to reliably provide it to the developers, observation, tracking, photographing and recording video/audio data of the interaction situations is necessary.

In public places, the presence of by-standers and occasional, unplanned interaction events is unavoidable and even desirable. If the place is crowded, it may be impossible to get everyone's written, signed informed consent for recording. Therefore, we will inform all visitors in the entrances (doors, gates) of the place with the following announcement:

- research and recording is carried out in the place (dates and duration given)
- the project name and partners
- the purposes of the project, research and data collection
- how the collected data will be used
- that the collected data will be protected appropriately
- contact information for further questions.

This information is the same as in the informed consent forms used in this project.

We interpret that the visitors, given the chance to get the information, give their informed consent by entering the place. The data will be anonymised before possible publication or sharing outside the MuMMER project.

However, the data needs to include some identifiable information about individuals (faces, recognisable physical properties etc) during the analysis within the project, because they are an essential part of the research and development of the robot behaviour. All identifiable data will be anonymised or deleted after the research goals are met or the project ends.

In all situations, we pay specific attention in anonymising material including children or persons unable to give informed consent. The main research and analysis is based on the data collected from healthy, adult volunteers. However for some research goals, children's opinions and behaviour are of particular importance for the research. When children (of any age) are participants in a study (their opinions are collected or their interaction with the robot is recorded), a specific procedure is applied as follows.

Children participate only if they volunteer (by verbal response, clear nodding or other clear gesture or expression) AND their guardian/legal representative volunteers. The guardian/legal representative is asked to sign a fully informed consent for the child's participation and possible video or other recording for research purposes. Specific attention is paid that the child is not in a risk of any harm during the study. The guardian/legal representative is invited to monitor and supervise the study situation with the researcher. Both the child and the guardian/legal representative are told that the child can withdraw from the study any time. Also the guardian/legal representative can withdraw the child from the study any time. The collected data is always anonymised unless necessary for the research purpose (e.g. facial expression analysis) and this is made clear in the informed consent. The data is deleted after the research goals are met or the project ends.

Research data, which may include personal data as described above, is planned to be transferred between MuMMER project members. The consortium includes Switzerland, which has been considered to be a third country offering an adequate level of protection (on the Commission list of countries offering adequate protection http://ec.europa.eu/justice/data-protection/document/international-

transfers/adequacy/index_en.htm) in accordance with Article 25 of the Directive 95/46/EC and so the transfer may take place.

We will ask for ethical opinion from relevant institutional ethical committees, e.g. the Ethical committee of social sciences of Tampere area (Tampereen alueen ihmistieteiden eettinen toimikunta, http://www.uta.fi/tutkimus/etiikka/arviointitmk/kokoonpano.html, in Finnish only) to ethically evaluate the planned research or relevant parts of it. VTT is represented in the Ethical committee.

2.3 Storage and access provision

All data created and collected during MuMMER will be stored internally on a private gitlab server hosted at Softbank Robotics, to which only project partners will have access. All data in gitlab will be private by default; where appropriately anonymized and processed data is to be made openly available, we will make use of the University of Glasgow's Enlighten Research Data repository at http://researchdata.gla.ac.uk/.

2.4 File formats

Wherever possible, we will use open and/or archival formats for all MuMMER data. This includes TIFF for images, FLAC for audio files, MPEG-4 for video, RTF for text, and CSV for spreadsheet data.

3 Description of MuMMER Datasets

3.1 Dataset naming Conventions

With regards to the naming convention MuMMER datasets, each name will be created as follows:

a. Dataset number containing prefix “DS” followed by a unique identifier number, e.g. DS1
b. Partner short name who will be managing the dataset, e.g. VTT
c. Short title of the dataset summarising the data contained within the dataset, e.g. Consumer Research

E.g. “DS1.VTT.Consumer Research”

3.2 Summary of Datasets

The following table provides the name and a short description of each dataset. In total we envisage three high-level datasets to be collated during the MuMMER project, encompassing a wide variety of data files.

Name	Description
DS1.VTT. Consumer Research	All data collected as part of the co-design activities carried out by VTT throughout the project.
DS2.Partner. Interaction with Robot	All data collected during interactions with the MuMMER robot, at all partner sites.
DS3.Partner. Software Development	All data collected for the purposes of developing and training the software models included in the MuMMER system.

In the remainder of this document, we provide an initial DMP for each of these datasets. Note that the canonical version of the DMPs are stored in DMPonline at https://dmponline.dcc.ac.uk/; please contact the MuMMER administrator to be given access to the online DMPs.

4 MuMMER Datasets

4.1 Consumer Research DMP

The canonical version of this DMP is stored in DMPonline (https://dmponline.dcc.ac.uk/). Please contact the MuMMER administrator if you require access to the online DMP.

4.1.1 Data set description

WP1: Use scenarios, acceptance and success metrics will be led by VTT. The objective of this WP is to ensure that the MuMMER robot and its implemented and foresighted future applications will be user-driven in terms of (Human-Robot Interaction) HRI, socially and ethically accepted, and interesting to commercial end users. This goal will be achieved through an intensively applied co-design approach that engages consumer users and other relevant stakeholders in the technical development throughout the design. Consumers will be engaged using several forms of market research including demonstrations, workshops, discussion events, interviews and surveys. The co-design activities will produce use scenarios to guide the development, increase user acceptance toward robotic applications in consumer markets and develop success metrics for human-interactive mobile robots. This DMP aims to address the storage of MuMMER consumer research data in line with H2020 guidelines.

Origin of data

Data collected by VTT at the following events:

- Demonstrations
- Workshops with consumers
- Discussion events
- Interviews

In addition VTT will collect data from the following:

- Surveys/questionnaires
- Consent Forms

Nature and scale of data

- Demonstrations – photos and video recordings in Finnish. Summary of results in English to be shared with Partners
- Workshops with consumers - video recordings in local language. Summary of results in English to be shared with Partners
- Interviews - video/audio recording in Finnish. Summary of results in English to be shared with Partners
- Surveys and questionnaires – Internet questionnaire as well as paper questionnaires in word format or via digital feedback screens
- Consent Forms – paper questionnaires in RTF format

To whom the dataset could be useful

Raw data in local language will be assessed and by VTT a summary of results (in English) will be shared with the consortium.

Some pictures and videos may be used for dissemination purposes

Related scientific publications

The summary data derived from these studies will be used as the basis for scientific publications. Pictures may also be used as part of scientific publication, where appropriate consent has been obtained.

4.1.2 Standards and metadata

The following metadata will be recorded regarding the consumer research data:

- Demonstrations - metadata required: demonstration host, group size, location, date and time, duration
- Workshops with consumers - metadata required: workshop host, workshop attendees, location, date and time, duration
- Interviews - metadata required: interviewer, interviewee, location, date and time, duration
- Surveys/questionnaires – metadata required: location, date and time
- Photos - saved in TIFF format with location, description and date in the title

4.1.3 Data sharing

Data will be shared internally with project partners using the MuMMER gitlab server. Where appropriate, reusable data such as log files or (where permission has been obtained) video data will be shared using the University of Glasgow's Enlighten Research Data server.

4.1.4 Archiving and preservation (including storage and backup)

During the course of the MuMMER project, all data will be stored on the MuMMER project gitlab server, which is backed up regularly. Public data will be released through the University of Glasgow's Enlighten Research Data server, where it will persist after the end of the project.

4.2 Human-Robot Interaction DMP

The canonical version of this DMP is stored in DMPonline (https://dmponline.dcc.ac.uk/). Please contact the MuMMER administrator if you require access to the online DMP.

4.2.1 Data set description

This data set will encompass all system logs, video recordings, and user questionnaire responses arising from users interacting with the MuMMER robot, both in lab settings and in the public deployment locations (including but not necessarily limited to Ideapark). It will also include any after-the-fact annotations of the video data that is gathered.

Within the project, this data will be useful for analysing the success of all robot deployments throughout the project: this information will then be used to inform the development and refinement of the robot scenarios, as well as to help the developers to enhance the system for future deployments.

The data will also be used as the basis for scientific publications describing the robot deployments.

Outside of the MuMMER project, this data could also be useful to guide the work of other developers of similar public-space robot systems.

4.2.2 Standards and metadata

System logs:

- Timestamped entries for every system event, internal state update, and message exchanged among the system components
- Every log includes a synchronisation signal to allow it to be matched up later with video data

Videos:

- High-quality HD videos, compressed with a useful codec such as H.264
- If multiple videos are included of a session, they will be synchronised with each other using a synchronisation signal

Questionnaires:

- Every questionnaire annotated to indicate the date, time, and location of the session it relates to, in order to allow questionnaire responses to be correlated with videos and/or logs

Annotations:

- Any annotated data will be stored alongside the raw data.

4.2.3 Data sharing

Data will be shared internally with project partners using the MuMMER gitlab server. Where appropriate, reusable data such as log files or (where permission has been obtained) video data will be shared using the University of Glasgow's Enlighten Research Data server.

4.2.4 Archiving and preservation (including storage and backup)

During the course of the MuMMER project, all data will be stored on the MuMMER project gitlab server, which is backed up regularly. Public data will be released through the University of Glasgow's Enlighten Research Data server, where it will persist after the end of the project.

4.3 Software Development Data DMP

The canonical version of this DMP is stored in DMPonline (https://dmponline.dcc.ac.uk/). Please contact the MuMMER administrator if you require access to the online DMP.

4.3.1 Data set description

This data set will encompass all audiovisual recordings specifically designed to help in the development of the technical components of MuMMER. Note that this dataset is distinct from the Human-Robot Interaction dataset: while the data in that set is related to interactions with the deployed robot system, the data in this set is specifically gathered and designed to help in developing and training the software components of the system.

Within the project, this data will be useful for ensuring that the components of the robot system function appropriately in the target environment: for example, this will include recordings made with the robot's own sensors of users interacting in the target deployment locations within Ideapark.

The data will also be used as the basis for scientific publications describing the technical development.

Outside of the MuMMER project, this data could also be useful to guide the work of other developers of similar technical components, both as training and as test data.

4.3.2 Standards and metadata

Videos:

- High-quality HD videos, compressed with a useful codec such as H.264
- If multiple videos are included of a session, they will be synchronised with each other using a synchronisation signal

4.3.3 Data sharing

Data will be shared internally with project partners using the MuMMER gitlab server. Where appropriate, reusable data such as log files or (where permission has been obtained) video data will be shared using the University of Glasgow's Enlighten Research Data server.

4.3.4 Archiving and preservation (including storage and backup)

During the course of the MuMMER project, all data will be stored on the MuMMER project gitlab server, which is backed up regularly. Public data will be released through the University of Glasgow's Enlighten Research Data server, where it will persist after the end of the project.