



## **D10.5 Data Management Plan**

Project Acronym: **MARIO**  
Project Title: **Managing active and healthy aging with use of caring service robots**  
Project Number: **643808**  
Call: **H2020-PHC-2014-single-stage**  
Topic: **PHC-19-2014**  
Type of Action: **RIA**

## D10.5

<b>Work Package:</b>	<i>WP10</i>	
<b>Due Date:</b>	M9	
<b>Submission Date:</b>	30/10/2015	
<b>Start Date of Project:</b>	01/02/2015	
<b>Duration of Project:</b>	36 Months	
<b>Organisation Responsible of Deliverable:</b>	Universität Passau (UNI PASSAU)	
<b>Version:</b>	8.1	
<b>Status:</b>	Final	
<b>Author name(s):</b>	Keith Cortis, Adamantios Koumpis <i>UNI PASSAU</i> Dymrna Casey, Kathy Murphy, Oya Beyan, Timur Beyan <i>NUIG</i> Francesco Ricciardi, Daniele Sancarlo, Francesco Giuliani <i>IRCCS</i> Massimiliano Raciti, Diego Reforgiato <i>R2M</i> Ilias Trochidis <i>Ortelio</i> Misael Mongiovì <i>CNR</i> Andy Bleaden <i>Stockport</i> Meftah Ghrissi <i>Robosoft</i> Geoff Pegman <i>RUR</i>	
<b>Quality Reviewer(s):</b>	Thomas Messervey <i>R2M</i> Geoff Pegman <i>RUR</i>	
<b>Nature:</b>	<input type="checkbox"/> R – Report <input type="checkbox"/> P – Prototype <input type="checkbox"/> D – Demonstrator <input checked="" type="checkbox"/> O - Other	
<b>Dissemination level:</b>	<input checked="" type="checkbox"/> PU - Public <input type="checkbox"/> CO - Confidential, only for members of the consortium (including the Commission) <input type="checkbox"/> RE - Restricted to a group specified by the consortium (including the Commission Services)	
Project co-funded by the European Commission within the Horizon 2020 Programme (2014-2020)		

## Revision history

Version	Date	Modified by	Comments
1.0	17/05/2015	Keith Cortis	Initial Table of Contents and First Draft
2.0	17/06/2015	Keith Cortis	Second Draft
2.1	23/06/2015	Andy Bleaden, Dympna Casey	Third Draft edits by Stockport and NUIG
2.2	23/07/2015	Kathy Murphy, Dympna Casey, Francesco Ricciardi	Third Draft edits by NUIG, IRCCS
3.0	28/07/2015	Keith Cortis	Integration of Third Draft edits by Stockport, NUIG, IRCCS
3.1	29/07/2015	Timur Beyan, Oya Beyan, Heike Felzmann	Ethical Issues – Section 2.6.1 by NUIG
3.2	30/08/2015	Keith Cortis, Dympna Casey	Copyright and Intellectual Property Rights – Section 2.6.2 by Uni Passau, NUIG
3.3	02/09/2015	Massimiliano Raciti	Fourth Version Comments and Edits by R2M
4.0	10/09/2015	Keith Cortis	Fourth Version
4.1	16/09/2015	Timur Beyan	Data Confidentiality: Processing Methods – Section 2.8.2 by NUIG
4.2	16/09/2015	Daniele Sancarlo	Addition of Dataset 0 – Section 2 by IRCCS
4.3	17/09/2015	Kathy Murphy	Fourth Version Overall Comments
4.4	18/09/2015	Ilias Trochidis	Selection and Preservation – Section 2.5.1 by Ortelio
4.5	18/09/2015	Misael Mongiovì, Diego Reforgiato, Mary Kagkoura	Edits in Sections 2.3. 2.4, 2.6, 2.8 by CNR and R2M
5.0	30/09/2015	Keith Cortis	Fifth Version
5.1	09/10/2015	Meftah Ghrissi	Storage and backup – Section 2.5.2 by Robosoft
5.2	12/10/2015	Geoff Pegman	Storage and backup – Section 2.5.2 by RURobots
5.3	15/10/2015	Francesco Giuliani, Andy Bleaden	Data Confidentiality: Processing Methods – Section 2.8.2 by IRCCS, Stockport
6.0	16/10/2015	Keith Cortis	Sixth Version
7.0	23/10/2015	Keith Cortis	Seventh Version – Quality Review by Thomas

---

			Messervey
8.0	27/10/2015	Keith Cortis	Eighth Version – Quality Review by Goeff Pegman. Inclusion of Annex 1.
8.1	30/10/2015	Keith Cortis, Timur Beyan, Oya Beyan	Data Confidentiality: Processing Methods – Section 2.8.2.1.3 by NUIG

Copyright © 2015, MARIO Consortium

The MARIO Consortium (<http://www.mario-project.eu/>) grants third parties the right to use and distribute all or parts of this document, provided that the MARIO project and the document are properly referenced.

*THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.*

## Executive Summary

This document describes the Data Management Plan (DMP) for the MARIO EU Project as defined in the Grant Agreement. The plan produced covers “all aspects of data handling, treatment, reporting and access” so that these aspects are clear to the partners. This plan is an evolving document, where new versions of the MARIO DMP will be released during the lifespan of the project that will reflect any changes to the project, such as inclusion of new data sets and changes in the consortium policies or external factors.

## Table of Contents

<b>Executive Summary .....</b>	<b>6</b>
<b>1. Introduction.....</b>	<b>10</b>
1.1. Work Package 10 Objectives .....	10
1.2. Purpose and Target Group of the Deliverable .....	11
1.3. Relations to other Activities in the Project .....	11
1.4. Document Outline.....	11
1.5. About MARIO .....	11
<b>2. Data Management Policy .....</b>	<b>12</b>
2.1. Datasets Reference and Names .....	12
2.2. Datasets Description .....	12
2.3. Standards and Metadata .....	16
2.3.1 Mandatory Metadata .....	17
2.3.2 Existing standards and methodologies to be used.....	17
2.3.3 Files and folders.....	19
2.3.4 Versioning.....	20
2.3.5 Quality assurance processes .....	20
2.4. Data Sharing .....	20
2.5. Archiving and Preservation .....	22
2.5.1 Selection and preservation .....	22
2.5.1.1 Project data specifics.....	22
2.5.1.2 Project data repository/archive .....	22
2.5.1.3 Project data repository/archive costs .....	23
2.5.1.4 Project data time and effort required .....	23
2.5.2 Storage and backup.....	23
2.5.2.1 Patient data .....	23
2.5.2.2 General data.....	24
2.5.2.3 Robot data .....	24
2.5.2.4 Data backups.....	24
2.5.2.5 Data recovery .....	24
2.5.2.6 Data security.....	24
2.6. Ethics and Legal Compliance .....	25
2.6.1 Ethical issues .....	25
2.6.1.1 Consent for data preservation and sharing.....	25
2.6.1.2 Protection of the participants identity .....	25
2.6.1.3 Handling of sensitive data.....	26
2.6.2 Copyright and intellectual property rights .....	26
2.6.2.1 Ownership of data .....	27
2.7. Responsibilities and Resources .....	29
2.7.1 Data Management Activity .....	29
2.7.2 Data Repositories Charges .....	29
2.8. Data Confidentiality .....	29

---

2.8.1 Privacy .....	30
2.8.2 Processing methods .....	30
2.8.2.1 Pilot Sites Regulations.....	31
<b>3. Conclusion.....</b>	<b>36</b>
<b>Annex 1 – Discussions .....</b>	<b>37</b>



## Index of Tables

Table 1: List of existing ontologies under consideration for MARIO .....	18
Table 2: MARIO Dataset sharing details .....	20

## 1. Introduction

This document supports the data management life cycle for all data<sup>1</sup> that will be collected, shared, processed or generated by the MARIO project and reflects the current state of consortium agreements regarding data management.

The goal of Task 10.5: Knowledge and Data Management, is described in the Description of Work as to produce a data management plan (DMP) where all aspects of data handling, treatment, reporting and access are clear to the partners. This document will describe information about the data (ontologies, schemas, datasets, etc.) used within MARIO and the management of this data (what type of data, how the data will be collected, shared, handled, preserved, any used standard). The data management plan will be hosted within the project website, and a part of it will be made it publicly accessible and spread using the activities of Task 10.4 (Dissemination Activities).

MARIO's DMP is a hybrid of the template provided by the European Commission<sup>2</sup> and the checklist provided by the Digital Curation Centre (DCC)<sup>3</sup> as suggested by the University of Sheffield<sup>4</sup>. The current DMP version reflects the latest status of the datasets descriptions, standards, data sharing and archiving guidelines. It builds upon the first plan that NUIG created for the grant agreement process, which was further extended according to the MARIO indicators (adapted to the EIP AHA indicators).

Given that MARIO's DMP is an evolving document, new versions will be released during the lifespan of the project as required and will reflect any changes to the lifetime of the project, such as inclusion of new data sets and changes in consortium policies or external factors. Such versions will be released as annexes of the project's Periodic Reports, specifically Deliverables 11.2 and 11.4 in Months 18 and 36 respectively.

### 1.1. Work Package 10 Objectives

This WP aims at bringing MARIO results to the widest possible audience for the benefit of people affected by loneliness, isolation and depression, and the MARIO project, its targeted social communities (including people with dementia), market, the scientific community, care givers and medical staff and the robotic community. The specific objectives of WP10 are the following:

1. To design a comprehensive communication strategy
2. To build a stakeholder community that acts as an impact multiplier
3. To carry on creative communication activities in selected domains
4. To conduct effective related dissemination activities
5. To assess both communication and dissemination activities.
6. To manage the knowledge and data gathered.

WP10 will reflect the output of each project WP. It will take into account high impact communication and dissemination channels, networks and instruments of partner

---

<sup>1</sup> The MARIO project consortium has not committed to participate in the Pilot on Open Research Data in Horizon 2020. Nevertheless, the project partners will still aim to provide open access to certain research data as discussed in more detail in Section 2 of this deliverable.

<sup>2</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf) - Annex 1

<sup>3</sup> [http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP\\_Checklist\\_2013.pdf](http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf)

<sup>4</sup> <https://www.sheffield.ac.uk/library/rdm/dmp>

organisations. The output of this WP will be the output dissemination and communication across the defined target audiences. Deliverable 10.5 reports on the last objective listed above.

## **1.2. Purpose and Target Group of the Deliverable**

This deliverable is targeted at all the partners of the MARIO consortium in order to make the data management plan regarding all aspects of data handling, treatment, reporting and access more clear.

## **1.3. Relations to other Activities in the Project**

The present Deliverable appears as result of work carried out under Task 10.5 about Knowledge and Data Management led by partner UNI PASSAU. This is linked to other tasks in the project namely, Task 1.1: Stakeholder workshops and co-producing a needs assessment, led by Stockport, Task 6.3: Development of Marios' adaptive engine, led by CNR; Task 10.4: Dissemination activities, led by Uni Passau and most importantly: Task 1.3: Data Management and Architecture, also led by Uni Passau.

## **1.4. Document Outline**

This deliverable is organised in eight sub-sections (in Section 2) about the Data Management Policy, namely: Datasets Reference and Names, Datasets Description, Standards and Metadata, Data Sharing, Archiving and Preservation, Ethics and Legal Compliance, Responsibilities and Resources and Data Confidentiality. The conclusions of this deliverable are provided in Section 3. One annex has been added to this deliverable; Annex 1 shows a list open issues that were outlined during various discussions about the Data Management Plan.

## **1.5. About MARIO**

MARIO addresses the difficult challenges of loneliness, isolation and dementia in older persons through innovative and multi-faceted inventions delivered by service robots. The effects of these conditions are severe and life-limiting. They burden individuals and societal support systems. Human intervention is costly but the severity of the effects can be prevented and/or mitigated by simple changes in self-perception and brain stimulation mediated by robots.

From this unique combination, clear advances are made in the use of semantic data analytics, personal interaction, and unique applications tailored to better connect older persons to their care providers, community, own social circle and also to their personal interests. Each objective is developed with a focus on loneliness, isolation and dementia. The impact of this work will support progress toward EU scientific and market leadership in service robots and a user driven solution for this major societal challenge. The competitive advantage is the ability to treat tough challenges appropriately. In addition, a clear path has been developed on how to bring MARIO solutions to the end users through market deployment.

## 2. Data Management Policy

### 2.1. Datasets Reference and Names

The following datasets will be produced during the MARIO research project:

- Dataset 0: MARIO survey about the use of companion robots in Persons With Dementia (PWD), caregivers and health professionals
- Dataset 1: MARIO Patient with dementia Profile
- Dataset 2: MARIO Patient Health Status
- Dataset 3: MARIO Social Participation, Connectivity and Life History
- Dataset 4: MARIO – Patient Health History
- Dataset 5: MARIO – Patient Institution Profile
- Dataset 6: MARIO – SME/Industry Sector Profile
- Dataset 7: MARIO Received Care
- Dataset 8: MARIO Human – Robot Interaction Data Set

Uniform Resource Identifiers (URIs) will be used to identify name of resources. URIs will be defined as part of Task 5.1: MARIO Ontology Network which is due after 18 months of the project and is led by CNR.

### 2.2. Datasets Description

The initial description of data sets is given below tables. These descriptions will evolve through the project lifespan and they will be updated in each version of MARIO – DMP.

<b>Dataset Number</b>	0
<b>Dataset Name</b>	MARIO survey about the use of companion robots in PWD, caregivers and health professionals.
<b>Description of Data</b>	Demographic and social data and opinions about the acceptability, functionality and usefulness of companion robots.
<b>Origin of Data</b>	Questionnaires and interviews to the PWD, caregivers and health professionals.
<b>Possible usage</b>	The data could be useful for the project to better understand the more appropriate technological approach to a PWD and his caregivers considering also the opinions of the healthcare workers.
<b>Scientific publications</b>	N/A
<b>Data format</b>	Database records. Interviews/questionnaires.
<b>Data Volume</b>	120 subjects using a questionnaire constituted of about 44 items.

<b>Dataset Number</b>	1
<b>Dataset Name</b>	MARIO Patient with dementia Profile

<b>Description of Data</b>	Patient demographic and social data. Patient social history, such as personal interests, personal photos, place of interests, preferred newspapers, etc.
<b>Origin of Data</b>	Patient documents (to register demographic data). Personal interview for the social data and history. Digital devices (cd, usb stick) for pictures of life history, relatives, etc.
<b>Possible usage</b>	Studies that could include demographic data. Improvement of patient's quality of life with the use of personal historical and social data. For example, MARIO can show to the patient his personal photos to improve his mood or as a way to propose an exercise.
<b>Scientific publications</b>	N/A
<b>Data format</b>	Database records, digital images.
<b>Data Volume</b>	Demographic data will be fixed in size. Patient's photos and preferences will depend on the particular patient and the total amount of patients.

<b>Dataset Number</b>	2
<b>Dataset Name</b>	MARIO Patient Health Status
<b>Description of Data</b>	The anonymised patient health status data based on comprehensive geriatric assessment (CGI) and Multidimensional Prognostic Index (MPI), including the length of time health professionals (e.g., physicians) spent, Depression scores, cognitive tests.  For the hospital use case this dataset could be enriched with lab and psychological test results.
<b>Origin of Data</b>	MARIO Robot CGA Module. Hospital patient's records (in the hospital use case setting).
<b>Possible usage</b>	Health assessment and quality of life metrics, monitor disease progression.  For the hospital case study: analysis on the impact in terms of health outcomes of MARIO robot made by comparison with the IRCCS existing dataset (IRCCS has a rich dataset on care outcomes for elderly patients in the last three years).
<b>Scientific publications</b>	Sancarlo D, Pilotto A, Panza F, Copetti M, Longo MG, D'Ambrosio P, D'Onofrio G, Ferrucci L, Pilotto A. A Multidimensional Prognostic Index (MPI) based on a comprehensive geriatric assessment predicts short- and long-term all-cause mortality in older hospitalized patients with transient ischemic attack. J Neurol. 2012 Apr;259(4):670-8. doi: 10.1007/s00415-011-6241-4. Epub 2011 Sep 23. PubMed PMID: 21947223.
<b>Data format</b>	Structured data from one or more databases.
<b>Data Volume</b>	To be defined according to the measurement sampling frequency for the parameters that will be tracked.

<b>Dataset Number</b>	3
<b>Dataset Name</b>	MARIO Social Participation, Connectivity and Life History
<b>Description of Data</b>	Social participation of the health participants in the pilots, including any academic and/or patient demographic data e.g., identify any degrees/courses/career training undertaken. Social connectivity and community involvement data of subjects including frequency and type of actions e.g., involvement in social media platforms, such as family Facebook account.
<b>Origin of Data</b>	4-Connect Community Module. 4-Connect My Social Network Module.  Questionnaires to capture quality of life (QoL), social isolation, and other relevant measures. Interviews with support staff, contacts and patients themselves, and review of support staff reports.
<b>Possible usage</b>	Quality of life metrics, monitor disease progression, research on isolation, loneliness and stigmatisation.  Find any correlations (if any) between people that have certain experience e.g., an academic degree and others that do not, and their experience with / acceptance for technological advances that might help them i.e., a robot.
<b>Scientific publications</b>	Linn MW, Sculthorpe WB, Evje M, et al. A social dysfunction rating scale. J Psychiatr Res 1969;6:299–306. (uploaded in basecamp)
<b>Data format</b>	Database records. Interviews/questionnaires.
<b>Data Volume</b>	N/A

<b>Dataset Number</b>	4
<b>Dataset Name</b>	MARIO Patient Health History
<b>Description of Data</b>	Track main causes of death for people with dementia. Profiles of patients across all pilot sites (e.g., hospitals, care homes, etc.). This data will include reviewing of pilot site records, such as cause of entry e.g., falls, other dementia related accidents; duration of stay; procedures carried out e.g., operation, tests, medication, etc.; waiting list duration (if not admitted urgently); stage of dementia, Activities of Daily Living (ADL) dependency scores, FAST (is a functional assessment scale for PWD), Carer Burden (is an instrument used to measure burden of caring).
<b>Origin of Data</b>	Patient personal and medical records. Death medical certifications.
<b>Possible usage</b>	Chronic disease diagnosis, treatment, rehabilitation and possible prevention of certain illnesses for people suffering from the same conditions and/or symptoms.

	Hospital re-admission by diagnosis, monitor disease and/or condition progression, monitoring quality of life.
<b>Scientific publications</b>	N/A
<b>Data format</b>	Database records.
<b>Data Volume</b>	N/A

<b>Dataset Number</b>	5
<b>Dataset Name</b>	MARIO Patient Institution Profile
<b>Description of Data</b>	Profiles of institutions (e.g. nursing home, day care centres, etc.) for patients. This data will include reviewing of institution care records, such as number of days from date of entry, number of patients in same room, meals per day, environmental factors (e.g., single rooms vs. multi-bedded wards), activity (e.g., art classes, reading groups), staffing (e.g., number of caretakers per patient), etc.
<b>Origin of Data</b>	Institution records. Patient personal institution records.
<b>Possible usage</b>	Improvement in level of care and living in institutions, efficiency of health and social care systems, management of health services.  Use environmental profile to enable comparisons e.g., MARIO Kompai robot may be fine for a single room institution but not for multi-bedded wards, MARIO Kompai may help in enabling staff members to focus and spend more time on other important care issues required.
<b>Scientific publications</b>	N/A
<b>Data format</b>	CSV.
<b>Data Volume</b>	N/A

<b>Dataset Number</b>	6
<b>Dataset Name</b>	MARIO SME/Industry Sector Profile
<b>Description of Data</b>	Profiles of SMEs/industry sectors interested in / using MARIO. Such data will include number of SMEs and sectors involved, years of involvement, target sectors, market product/s, customers, platforms used, etc.
<b>Origin of Data</b>	SMEs personal information. Sectors personal information.
<b>Possible usage</b>	Identification of target customers and/or end-users i.e. SMEs/sectors requiring usage of platform.
<b>Scientific publications</b>	N/A
<b>Data format</b>	Database records.
<b>Data Volume</b>	N/A

<b>Dataset Number</b>	7
<b>Dataset Name</b>	MARIO Received Care
<b>Description of Data</b>	Received care in institutional setting and by care givers. Social care plan.
<b>Origin of Data</b>	Interviews and review of health records from GPs, hospital visits, nursing care plan for social history. Costs of care.
<b>Possible usage</b>	Sustainability and efficiency of health and social care systems.
<b>Scientific publications</b>	N/A
<b>Data format</b>	Database records. Interviews.
<b>Data Volume</b>	N/A

<b>Dataset Number</b>	8
<b>Dataset Name</b>	MARIO Human – Robot Interaction Data Set
<b>Description of Data</b>	<p>The abstracted human robot interaction data captured by robots, including human inputs, reaction given by robots, number of users being assisted, tracking data through sensors (e.g., kinect), etc.</p> <p>Data that will be generated at programming time to specify the social capabilities and the knowledge base that the robot has to interact with people (sharing information, acting from linguistic input, learning new discourse patterns, etc.) with focus on the loneliness, isolation and dementia needs (WP3) and the health database (WP4) through the MARIO Ontology Network.</p> <p>Data about the experience of using the Mario Kompai robot from the people with dementia and caregivers will also be collected. This will help identify if the goals set are being reached and what issues are being encountered.</p>
<b>Origin of Data</b>	Robot logs and MARIO ontology network triples Interviews with patients and caregivers.
<b>Possible usage</b>	Human machine interaction research, patient profiling.
<b>Scientific publications</b>	N/A
<b>Data format</b>	RDF for linked data layer (MARIO ontology network). Interviews/questionnaires.
<b>Data Volume</b>	N/A

### 2.3. Standards and Metadata

The MARIO project will use The Resource Description Framework (RDF)<sup>1</sup>, which is a

<sup>1</sup> <https://www.w3.org/RDF/>



standard of World Wide Web Consortium (W3C)<sup>1</sup> specifications as a metadata data model. In particular, most of the system's data will be managed by the RDF/OWL framework and will be accessible through a SPARQL endpoint and a dedicated API. RDF 1.1<sup>2</sup> and OWL 2<sup>3</sup> will be used, where the best practices, ontology design patterns and W3C guidelines will be followed and applied. Existing ontologies will be reused for representing data. As per the data protection policy proposed by the European Commission in January 2012, any data sets held on laptops will be PC password protected and encrypted as per University policy (QA409 2014).

### 2.3.1 Mandatory Metadata

Mandatory metadata provides the key elements for citing and/or searching any project data and makes such data more visible, accessible and usable by any interested party.

The following is the mandatory metadata of the MARIO project:

European Union  
H2020  
MARIO  
GA643808

### 2.3.2 Existing standards and methodologies to be used

Our analysis identified the following macro-areas of data that will be addressed by the MARIO ontology network:

- personal sphere (personal information, family, friends, care staff, etc.)
- life events and patterns (everyday events monitoring, personal memories, scheduling and calendar, etc.)
- social and multimedia content, e.g. related to life events or potentially interesting things (photos, video, tunes, films, documents, social networks, etc.).
- environment (rooms, doors, furniture, objects, positions, link to RFIDs, etc.)
- health sphere (living patterns, health patterns, vital signs, anything related to CGA and MPI)
- emotional sphere (emotions/sentiments/opinions at a certain time, or related to something or someone)
- open knowledge (speech-derived data, web-extracted data or content: news etc.)
- regulatory sphere: rules, norms, social habits, privacy, etc.
- robot proprioception: robot platform, skills, available actions and applications

The described areas are highly interconnected and have a certain amount of overlap, since it is not possible to make a clear-cut separation of data areas. We will develop an ontology network that covers all areas by connecting specific domain ontologies and reusing existing ones. Table 1 presents a list of existing ontologies that are under analysis for possible reuse within MARIO.

---

<sup>1</sup> <http://www.w3.org/>

<sup>2</sup> <http://www.w3.org/TR/rdf-schema/>

<sup>3</sup> <http://www.w3.org/TR/owl-guide/>

Table 1: List of existing ontologies under consideration for MARIO

Knowledge area	Subarea	Existing ontology
personal sphere (family, friends, care staff)	Relationship between people	<a href="http://vocab.org/relationship/.html">http://vocab.org/relationship/.html</a>
		<a href="http://vocab.org/bio/0.1/.html">http://vocab.org/bio/0.1/.html</a>
	People information	<a href="http://xmlns.com/foaf/spec/">http://xmlns.com/foaf/spec/</a> <a href="http://www.semanticdesktop.org/ontologies/2007/11/01/pimo/">http://www.semanticdesktop.org/ontologies/2007/11/01/pimo/</a>
	Contacts	<a href="http://www.w3.org/TR/vcard-rdf/">http://www.w3.org/TR/vcard-rdf/</a> <a href="http://www.semanticdesktop.org/ontologies/2007/03/22/nco/">http://www.semanticdesktop.org/ontologies/2007/03/22/nco/</a>
	Person visual/audio identification	<a href="http://ns.inria.fr/nicetag/2010/09/09/voc.html">http://ns.inria.fr/nicetag/2010/09/09/voc.html</a> <a href="http://purl.org/spar/datacite">http://purl.org/spar/datacite</a>
	Working hours	
life events and patterns (everyday, scheduling)	Events	<a href="http://motools.sourceforge.net/event/event.html">http://motools.sourceforge.net/event/event.html</a> <a href="http://www.semanticdesktop.org/ontologies/2012/03/06/drmo/">http://www.semanticdesktop.org/ontologies/2012/03/06/drmo/</a>
	Frames	
	Routines	
	Frames	
	Memories	<a href="http://vocab.org/bio/0.1/.html">http://vocab.org/bio/0.1/.html</a>
	Scheduling	
	Temporal ordering	
	Viewpoint	
	Plans	<a href="http://www.loa.istc.cnr.it/ontologies/Plans.owl">http://www.loa.istc.cnr.it/ontologies/Plans.owl</a>
social and multimedia content, e.g. related to life events or potentially interesting (photos, video, tunes, films, documents), etc.	Online social network communities	<a href="http://rdfs.org/sioc/spec/">http://rdfs.org/sioc/spec/</a> <a href="http://www.semanticdesktop.org/ontologies/2011/10/05/dlpo/">http://www.semanticdesktop.org/ontologies/2011/10/05/dlpo/</a>
	Preferences	<a href="http://vocab.ctic.es/reco/reco.owl">http://vocab.ctic.es/reco/reco.owl</a>
	Multimedia content	<a href="http://schema.org">http://schema.org</a>
environment (rooms, doors, furniture, objects, positions, link to RFIDs)	Situational context	<a href="http://www.semanticdesktop.org/ontologies/2011/10/05/dcon/">http://www.semanticdesktop.org/ontologies/2011/10/05/dcon/</a>
	Tempospatial information	<a href="http://observedchange.com/tisc/ns/">http://observedchange.com/tisc/ns/</a>
		Dolce Physical attribute
health sphere (living patterns, health patterns, vital signs, anything related to CGA and	Temporal Relation Inferencing in	<a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041418/">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041418/</a>

MPI)	Clinical Narratives	
emotional sphere (emotions/sentiments/opinions at a certain time, or related to something or someone)	Emotional information about a patient	<a href="http://ns.inria.fr/emoca/">http://ns.inria.fr/emoca/</a>
	MARL	<a href="http://www.gsi.dit.upm.es/ontologies/marll/">http://www.gsi.dit.upm.es/ontologies/marll/</a>
open knowledge (speech-derived data, web-extracted data or content: news etc. Reconciliation and provenance are key)	DBPedia ontology	<a href="http://dbpedia.org/ontology/">http://dbpedia.org/ontology/</a>
regulatory sphere: rules, norms, social habits, privacy, etc.		
Robot proprioception: robot platform, actions, and applications		

The MARIO ontology network will also address the following domain areas that are orthogonal to all areas:

- space
- time
- provenance (of stored knowledge)
- reliability (measures indicating how reliable a certain data is according to certain metrics)
- reconciliation (metadata supporting knowledge reconciliation procedures)

For these aspects we will consider existing ontologies and approaches as well as develop new one based on MARIO application domains and requirements. Possible ontology to reuse include: the time ontology in OWL<sup>1</sup>, the timeline ontology<sup>2</sup> and the PROV ontology<sup>3</sup> for provenance.

### 2.3.3 Files and folders

Each pilot site will have a pilot folder. The overall site folder will hold technical data and contextual data about the site. Each participant in that site will be allocated a code and data will be held for that participant in that file, only the code will appear on the file. The file consist of sub folders related to physiological data, interview data, questionnaire data,, life history data, social care plan, medical data.

<sup>1</sup> <http://www.w3.org/TR/owl-time/>

<sup>2</sup> <http://motools.sourceforge.net/timeline/timeline.html>

<sup>3</sup> <http://www.w3.org/TR/prov-o/>

### 2.3.4 Versioning

Versioning will be handled by numbers and by date. They will be added as metadata of the RDF datasets and OWL files. The version number is composed by a sequence of three numbers separated by dots e.g. 1.0.2, where the first number indicates major revisions, the second one indicates minor revisions that are deployed, and the third number indicates internal revisions. The date is specified in the format yyyyymmdd (e.g. 20150830).

### 2.3.5 Quality assurance processes

Data Quality Assurance methodologies will be applied to ensure that the data entered within the system by people with dementia (through the robot’s touch screen), healthcare staff and MARIO personnel will be accurate and to identify if there are any inconsistencies and missing/incomplete information. With regards to people with dementia, some observational work might be required in order to learn and see what the patients are doing, have certain data check processes (e.g., with every X records, periodically, etc.), to ensure that the recorded information is robust and or the required quality.

## 2.4. Data Sharing

The Ethics and Privacy Supervisory Board, which is part of MARIO Management Structure, will define access procedures and embargo periods during project lifespan with consensus partners and data owners.

Data sharing and re-use policies will comply with Privacy, Ethics and Data Management guidelines of MARIO.

Table 2 provides all the data sharing conditions for the datasets identified in Sections Datasets Reference and Names 2.1 and 2.2 above. This lists the datasets that will be open and restricted, and the shared format that any open datasets will be provided. More specifics for data use, re-use and publishing and restrictions (for consortium or partners only) will be provided in the next versions of the DMP.

Table 2: MARIO Dataset sharing details

Dataset	Open / Restricted	Shared Format (if Open Data) e.g. CSV, RDF, etc.	Notes (such as reasons for why dataset cannot be shared)
0: MARIO survey about the use of companion robots in PWD, caregivers and health professionals	Restricted	N/A	Ethical, Rules of Personal Data, IP, Privacy-related
1: MARIO Patient with dementia Profile	Restricted	N/A	Ethical, Rules of Personal Data

2: MARIO Patient Health Status	Restricted	N/A	Ethical, Rules of Personal Data
3: MARIO Social Participation / Social Connectivity	Restricted		Ethical, Rules of Personal Data
4: MARIO Patient Health History	Restricted	N/A	Ethical, Rules of Personal Data
5: MARIO Patient Institution Profile	Open	CSV	N/A
6: MARIO SME/Industry Sector Profile	Restricted	N/A	Ethical, Rules of Personal Data, IP
7: MARIO Received Care	Restricted	N/A	Ethical, Rules of Personal Data
8: MARIO Human – Robot Interaction Data Set	Open data	RDF	N/A

Some of the data collected will be used for testing and demonstration purposes, whereas other for scientific publications. In addition, certain datasets will be made available for independent verification and re-use through the project's website and also through the partners' websites.

The data acquired by the robots and exchanged among them need to be standardised. We propose to publish data in RDF and store it in RDF data storage, such as OpenLink Virtuoso<sup>1</sup>. Data will be maintained and updated in the robot's hardware as well as in central repositories that can be easily shared among the project partners. Updates, inserts, deletes and any operations with the data will be available and permitted. CNR will be responsible for the correct management of the data storage and will provide the other partners with instructions and best practices for its access. Certain data in central repositories will be reachable from the MARIO project website, disseminated using social networks and partners' media channels. External and potential users interested in using the data will have a SPARQL endpoint available. They will be able to query the data according to the MARIO ontologies included in the project website, which can be downloaded and analysed, together with guidelines and documentation. As soon as new data will be collected, they will be loaded in the triple store and will be made available online. Important components of the triple store are the data source adapters, which will allow for the collection of input from robot sensors and other data sources that are used in the pilot cases, and to produce linked data for feeding the triple store.

By definition Open Data “can be freely used, modified, and shared by anyone for any purpose”. In our context, the robots should be able to make use of relevant open data sources for proper actions. The data acquired by robot sensors should be published in the data space with respect to the linked data principles<sup>2</sup>: available on the web, structured, not using a proprietary format, using URIs to denote entities and linked to

<sup>1</sup> <http://virtuoso.openlinksw.com/>

<sup>2</sup> <http://5stardata.info/>

other datasets<sup>1</sup>.

When publishing data to the data space, it has to be linked to other data sets. This linking is very useful for ensuring an optimal data management and integration. It helps enhancing their (re)use and discovering new knowledge from the MARIO domain data put into a wider context. It is important to assess and determine what external datasets are relevant to be linked to MARIO data.

Data from other ontologies will be used to integrate the knowledge acquired by the robots in order to deliver accurate decision analytics. The data space needs to enrich sensor data with relevant information that is required by support services and applications.

## **2.5. Archiving and Preservation**

### **2.5.1 Selection and preservation**

The data collected throughout the project lifetime will be selected and preserved according to the project requirements and specifics. This will also depend on the project data repository/archive chosen and time and effort required. More information about this is provided in the following sub-sections.

#### **2.5.1.1 Project data specifics**

Ethics require us to store data for 5 years. All data gathered during the project lifetime must be kept either in private databases or in public ones.

#### **2.5.1.2 Project data repository/archive**

Two data repositories - Zenodo<sup>2</sup> and OpenLink Virtuoso - are being analysed for the purposes of this project i.e. for storing our data, where the one which best fits the project requirements will be selected.

Zenodo will allow long-term preservation of datasets beyond the project lifetime. It is an open dependable home for the long-tail of science, enabling researchers to share and preserve any research outputs in any size, any format and from any science.

OpenLink Virtuoso - an enterprise grade solution for data access, integration, and relational database management (SQL Tables and/or RDF based Property/Predicate Graphs) will also be considered, especially for storing triples (RDF data). This product is available in Open Source and Commercial editions.

---

<sup>1</sup> T. Berners-Lee, "Linked Data- Design Issues," 2006. [Online]. Available: <http://www.w3.org/DesignIssues/LinkedData.html>. [Accessed: 12-May-2013]

<sup>2</sup> <https://zenodo.org>

### **2.5.1.3 Project data repository/archive costs**

Zenodo is open and free of charge.

OpenLink Virtuoso is open source and free of charge.

### **2.5.1.4 Project data time and effort required**

Each partner is responsible for the curation of data that is gathered and stored on their premises. Preservation will be taken care of through the project's web site, partners' websites and suggested storage locations. Further specifics for data curation and preservation will be provided in next versions of the DMP. The research data from this project will be deposited within a common repository (if no restrictions apply, such as interview data which is confidential and thus can be shared only in synthesis versions), will be made accessible to all project partners, to ensure that the research community has long-term access to the data.

The project team will create a dedicated website to manage and distribute the data; it will be established using a content management system so that data project partners can participate in adding site content over time, making the site self-sustaining. For preservation, we will supply periodic copies of the data to the common Repository. That repository will be the ultimate home for the data. The data generated by this project (e.g., tracking data, etc.) will not pose a disclosure risk because it will be de-identified before being stored in the chosen format.

## **2.5.2 Storage and backup**

Different kinds of data will be collected during the project lifetime, namely i) patient data, ii) general data and iii) robot data. Patient data needs to be secured with regard to its transmission and storage, whereas security for the general data is not mandatory.

### **2.5.2.1 Patient data**

With regards to patient data, storage and backup policies depend on the country rules for the respective pilot sites (i.e. Italy, United Kingdom and Ireland). For example, France is one of the rare countries outside of the United States to create a special legal status for entities that process and store patient health data. A law in 2002 created this status, which was put into effect in 2009. In essence, entities that are not licensed health care establishments must obtain an authorisation from the French Ministry of Health in order to lawfully store patient health data originating from such establishments. The authorisation requires implementation of rigorous measures to ensure security and confidentiality of patient health data at all phases of the data lifecycle; these requirements are generally seen as the most robust in all of Europe and indeed throughout most the world. A bill that is currently pending in the French Parliament would simplify the requirements to obtain authorisation and notably align the security requirements with globally-recognised standards.

For the MARIO project we propose the use of companies that have the necessary authorisation to lawfully store and manage patient health data, such as Domicalis<sup>1</sup>.

### **2.5.2.2 General data**

For storage and backup of general data we propose to use big data cloud services, such as Google Cloud Platform<sup>2</sup> and Amazon Simple Storage Platform<sup>3</sup>. Such services are secure, durable, scalable and allow for better sharing of information with partners.

### **2.5.2.3 Robot data**

Regarding the robot, the only data that we will be collecting is the data logs from the motion sensors (encoders, lasers), the status of the actuators, etc. This data will be used to facilitate diagnosis in case of any malfunction to the robot. Such data can be used locally or remotely by our technicians and engineers. The other data that can be collected by the robot for the use of user functions will be managed by the partners in charge in order to develop such functions (for example, data from the RFID reader).

### **2.5.2.4 Data backups**

All form of data (patient, general and robot) will be backed up within a central repository data locally to a separate hard disk on a daily basis and in an encrypted form to the cloud (in the case of general data) on a weekly basis.

The owner/s of the particular datasets (listed in Section 2.1) will be responsible for the data backups.

### **2.5.2.5 Data recovery**

In the event of an incident the data will be recovered according to the necessary procedures of the data repository owner. The next version of the DMP will contain more details on the exact data recovery procedure that will be adopted in MARIO.

### **2.5.2.6 Data security**

#### **2.5.2.6.1 Risks**

The main data security risks are:

- Data loss prior to daily backup; Mitigation – data will also reside at local partner sites providing the data.
- Catastrophic loss of all data on site prior to weekly backups; Mitigation – Very low probability event. Most data will also reside at local partner sites but maximum exposure is one week of data loss.

---

<sup>1</sup> <http://domicalis.com/>

<sup>2</sup> <https://cloud.google.com/storage/>

<sup>3</sup> <https://aws.amazon.com/s3/>



- Data access by unauthorised 3rd party; Mitigation – Access to data repository would be by partners only and require secure passwords. Cloud data storage will be strongly encrypted with access restricted to the consortium partners or the dataset owners only (depending on the agreements made).

#### **2.5.2.6.2. Access**

In order to keep the data secure, access will be controlled by passwords for each data repository and encryption of long term cloud-based backup. We will ensure that collaborators can access our data securely by providing Web access to them via password control.

## **2.6. Ethics and Legal Compliance**

### **2.6.1 Ethical issues**

#### **2.6.1.1 Consent for data preservation and sharing**

To accomplish its functions, the MARIO robot will need various types of sensitive and personal data. Some of them are patient's data, but much of the reminiscence-related material is about family members, friends and other people. To collect and process these data, informed consent will be provided from participants and other related people before participating in the studies.

In dementia, possible declined cognitive capacities of patients will increase challenges for the realisation of informed consent. The balance between the respect of individual freedom including privacy and dignity and security can be achieved if an adequate informed consent process is realised involving the participants or, if not possible, their representative.

#### **2.6.1.2 Protection of the participants identity**

In MARIO project, names, address, dates directly related to an individual, phone and fax numbers, e-mail addresses, numbers of social security, care record, insurance beneficiary, account, certificate and license, biometric identifiers, full face photographic images and any comparable images etc., are considered as complete or partial identifier data.

To protect the identity of participants;

- All the materials gathered through user studies will be treated as non-identifiable,
- If written data will exist, after transferral to an electronic database, all of it will be destroyed,
- The personal information about the participants will be kept in confidence and will be made anonymous,
- Personal identifiers will be replaced with a numerical code (anonymisation),
- In all electronic systems, there will be no possibility to associate identification code with the name of a participant.

### 2.6.1.3 Handling of sensitive data

To ensure stored and transferred data securely;

- It will not be possible to access sensitive and personal data from outside the MARIO robot itself,
- Only identified caregivers or close family will be authorised by the robot to access such information in joint interactions with the person with dementia,
- Stored data will be specifically protected and never shared with third parties except data which can never be linked to the individual,
- An encrypted password system will be used and assigned new passwords will be defined and changed based on the security policy,
- Sharing with other systems will be possible only over a secure network,
- Shared and/or transferred data will be protected under the same conditions in the corresponding systems.

### 2.6.2 Copyright and intellectual property rights

The nature and scope of intellectual property protection, if any, which pilot trial data should receive in terms of Art. 39 of the TRIPS Agreement have been put back in the spotlight through recent events: First through suggestions by heads of the Dutch, French and United Kingdom (UK) regulatory authorities as well as the European Medicines Agency that such data should not be considered commercially confidential information. Secondly, courts in countries such as Argentina and Brazil have recently decided cases in which they had to balance rights over pilot trial data with competing public health priorities. Both courts decided that public health interests take priority over claims for exclusive rights over pilot trial data. In current literature, little attention has been paid to the fact that data exclusivity may impede efforts by clinical researchers, regulatory authorities and other stakeholders to ensure benefits sharing with clinical research participants.

Moreover, the Directive 2004/48/EU of the European Parliament and of the Council will be followed to ensure enforcement of the Intellectual Property Rights:

(1) The achievement of the internal market entails eliminating restrictions on freedom of movement and distortions of competition, while creating an environment conducive to innovation and investment. In this context, the protection of intellectual property is an essential element for the success of the internal market. The protection of intellectual property is important not only for promoting innovation and creativity, but also for developing employment and improving competitiveness.

(2) The protection of intellectual property should allow the inventor or creator to derive a legitimate profit from his/her invention or creation. It should also allow the widest possible dissemination of works, ideas and new know-how. At the same time, it should not hamper freedom of expression, the free movement of information, or the protection of personal data, including on the Internet.

(3) However, without effective means of enforcing intellectual property rights, innovation and creativity are discouraged and investment diminished. It is therefore necessary to ensure that the substantive law on intellectual property, which is nowadays largely part of the *acquis communautaire*, is applied effectively in the Community. In this respect,

the means of enforcing intellectual property rights are of paramount importance for the success of the internal market.

## **2.6.2.1 Ownership of data**

### **2.6.2.1.1. Ownership of results**

Results are owned by the Party that generates them.

### **2.6.2.1.2. Joint ownership**

Unless otherwise agreed:

- each of the joint owners shall be entitled to use their jointly owned Results for non-commercial research activities on a royalty-free basis, and without requiring the prior consent of the other joint owner(s), and
- each of the joint owners shall be entitled to otherwise Exploit the jointly owned Results and to grant non-exclusive licenses to third parties (without any right to sub-license), if the other joint owners are given:
  - a) at least 45 calendar days advance notice; and
  - b) Fair and Reasonable compensation.

### **2.6.2.1.3. Transfer of results**

1. Each Party may transfer ownership of its own Results following the procedures of the Grant Agreement Article 30.
2. It may identify specific third parties it intends to transfer the ownership of its Results to. The other Parties hereby waive their right to prior notice and their right to object to a transfer to listed third parties according to the project's Grant Agreement Article 30.1.
3. The transferring Party shall, however, at the time of the transfer, inform the other Parties of such transfer and shall ensure that the rights of the other Parties will not be affected by such transfer.
4. The Parties recognise that in the framework of a merger or an acquisition of an important part of its assets, it may be impossible under applicable EU and national laws on mergers and acquisitions for a Party to give the full 45 calendar days prior notice for the transfer as foreseen in the Grant Agreement.
5. The obligations above apply only for as long as other Parties still have - or still may request - Access Rights to the Results.

### **2.6.2.1.4. Dissemination**

#### **2.6.1.4.1. Dissemination of own Results**

During the Project and for a period of 1 year after the end of the Project, the dissemination of own Results by one or several Parties including but not restricted to publications and presentations, shall be governed by the procedure of Article 29.1 of the Grant Agreement subject to the following provisions.

Prior notice of any planned publication shall be given to the other Parties at least 45 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within 30 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is

permitted. The prior notice period may be waived in circumstances where all parties provide their written consent for earlier publication.

An objection is justified if

- a) the protection of the objecting Party's Results or Background would be adversely affected
- b) the objecting Party's legitimate academic or commercial interests in relation to the Results or Background would be significantly harmed.

The objection has to include a precise request for necessary modifications.

If an objection has been raised the involved Parties shall discuss how to overcome the justified grounds for the objection on a timely basis (for example by amendment to the planned publication and/or by protecting information before publication) and the objecting Party shall not unreasonably continue the opposition if appropriate measures are taken following the discussion.

The objecting Party can request a publication delay of not more than 90 calendar days from the time it raises such an objection. After 90 calendar days the publication is permitted, provided that Confidential Information of the objecting Party has been removed from the Publication as indicated by the objecting Party.

#### ***2.6.1.4.2. Dissemination of another party's unpublished results or background***

A Party shall not include in any dissemination activity another Party's Results or Background without obtaining the owning Party's prior written approval, unless they are already published.

#### ***2.6.1.4.3. Cooperation obligations***

The Parties undertake to cooperate to allow the timely submission, examination, publication and defence of any dissertation or thesis for a degree which includes their Results or Background subject to the confidentiality and publication provisions agreed in the project's Consortium Agreement.

#### ***2.6.1.4.4. Use of names, logos or trademarks***

Nothing shall be construed as conferring rights to use in advertising, publicity or otherwise the name of the Parties or any of their logos or trademarks without their prior written approval.

#### ***2.6.2.1.5. Exclusive licenses***

Where a Party wishes to grant an exclusive licence to its Results and seeks the written waiver of the other Parties pursuant to Grant Agreement Article 30.2, the other Parties shall respond to the requesting Party within 45 calendar days of the request. Any Party's failure to respond (whether in the negative or the positive) to the request within such 45 calendar days shall be deemed to constitute written approval of the waiver by the non-responding Party.

## 2.7. Responsibilities and Resources

### 2.7.1 Data Management Activity

The University of Passau team is responsible for implementing the DMP and ensuring that it is reviewed and revised.

Each of the 9 datasets listed in the sections above will be the responsibility of a specific partner, who will take care of the data management activity in relation to the respective dataset.

### 2.7.2 Data Repositories Charges

No charges will be applied by the data repositories given that we use the open source and/or open dependable technologies as mentioned in the previous sections (e.g. OpenLink Virtuoso, Zenodo, etc.).

## 2.8. Data Confidentiality

Within one year after the conclusion of the pilot trial, a summary of the results should be made publicly available. This summary should include information on the primary outcome, of any secondary outcome and statistical analyses. All data should be recorded according to standards procedures.

MARIO users participate voluntarily in pilot activities with the assurance of help to improve scientific knowledge about their disease and related therapeutic innovations. Therefore, with the publication of the results the researchers meet the ethical obligation to users enrolled, as defined by the Declaration of Helsinki (more information in Section 2.8.2). The publication of all results from the pilot activities and reduction in reporting bias will help produce more reliable systematic reviews on the efficacy and safety of health interventions.

The summary of the results of a pilot activity should contain at least the items provided by the results page [clinicaltrials.gov](https://clinicaltrials.gov): summaries of the participants, the original protocol, and any corrections, summary of the results for the primary endpoint and secondary default, details of any adverse events and statistical analysis.

As already mentioned, data will be stored in triple stores and certain datasets will be made publicly available. However, personal data will be kept anonymous using the necessary encryption methods. To do that, named entities (such as proper nouns or organisations) will be detected and encrypted. Access to personal data will be allowed only to members of the consortium and according to the existing Consortium agreement. More information is provided in the sub-sections below.

### 2.8.1 Privacy

The health and social personnel working and following the study will identify the patients with a code: the data, relating to a patient collected during the study, with the exception of his/her name, will be recorded, processed and stored along with the code, the date of birth, the sex, weight and height (all the above variables to be determined under the specifications of the study). Only the health and social personnel and authorised parties can connect the code to the patient's name.

### 2.8.2 Processing methods

The data, processed using electronic tools also will be disclosed only in a totally anonymous way, such as scientific publications, statistical and scientific meetings. The participation in the study implies that in accordance with the pilot activity rules, the ethics committee, health authorities and foreign companies will be able to know the data concerning a person, even in content original medical records, in ways that ensure the confidentiality of patient's identity.

MARIO partners should act in accordance to the following documents for data confidentiality:

- International regulations, such as the Nuremberg Code of 1949, the Declaration of Helsinki, 1980 OECD Recommendations of the Council Concerning guidelines Governing the Protection of Privacy and Trans-Border Flows of Personal Data etc.
- The legal documents by the European Union such as the Directive 95/46/EC of the European Parliament and of the Council (1995), Directive 2002/58/EC on Privacy and Electronic Communications (ePrivacy Directive), Directive 2006/24/EC on Data Retention (Data Retention Directive) the Handbook on European Data Protection Law (2013) etc. and,
- The relevant regulations and guidelines of all participating countries (e.g. Data Protection Guidelines on research in the Health Sector of Data Protection Commissioner of Ireland).

In the MARIO project, there is no "open data" policy (namely, we do not aim to share all the data publicly). In principle, certain data will not be disclosed for the general public other than partners. The project partners still aim to provide open access to certain research data and results (as specified in Section 2.4). This will be exposed publicly under the following circumstances:

- The patients' personal information will be strictly anonymised and kept in confidence. Hence, personal identifiers will be replaced with numerical codes (pseudonymisation) which cannot be associated with the identity information of participants.
- Substituting personal identification data with numerical codes does not assure total confidentiality all the time. Therefore, individuals and organisations using confidential information have to take responsibility for deciding what is justified and acceptable also on a case by case basis.

- All personal information needs to be encrypted and placed in a secure place with the needs of research, as well as of information processing. Collected data should be protected from intrusive access and only authorised people will have access them.
- Collected personal data will be used for only research purposes by authorised people. Commercial use of project data is strictly forbidden.
- Participants' names and unanonymised identifiers will not be given in any publication or presentation.
- Important data for researches such as age and gender can be publically made available in publications and presentations. Sometimes it is possible to disclose patients' identity without using personal identifiers (de-anonymisation). All stakeholders should consider this risk before they publish anonymised research data.
- In visual materials which will be used for publications or presentations, any identification data of participants will be hidden to reveal sensitive health information.
- To improve individual responsibility about data protection, managers of the study sites will inform members of researches against possible threats, risks, unauthorised breaches and precautions.
- To provide sensitive data un-disclosed, all collected data should be protected. Privacy protection technologies will make a barrier against unauthorised exposure of private information.
- Database managers and registry administrators should keep ahead the confidentiality issues as an obligation.

### **2.8.2.1 Pilot Sites Regulations**

Certain pilot sites that are within EU member states have specific regulations pertaining to their country. These are specified in more detail in the following sub-sections.

#### **2.8.2.1.1. Italy - IRCCS**

The Italian central authority follows the general EU regulations in these subjects<sup>1</sup>. To be compliant with the current rules, IRCCS as a hospital have to:

- Inform the patients, with a proper form to be signed by them, about what the hospital is doing with their data.
- Be very careful in the transmission of sensitive data outside the hospital information system. In this case, IRCCS are asked to operate a physical and

---

<sup>1</sup> Please note that the EU will be producing a "final" shared regulation that will hold for all EU Countries by the end of the year

logical division between the tables storing data that can be used to identify the patient from the tables containing measurements (i.e. made with the robot in the case of IRCSS). The former are usually stored in the hospital information system and the latter can be stored outside the hospital, usually marked with a code referring to a specific patient but through which it is essentially impossible (except for the hospital) to “discover” the name of the patient these data refer to.

- Ask permission to the external provider to store the data concerning the hospital’s patients. This is usually made through a document that states that the hospital expects the provider to treat with the proper care (backups, anti-intrusion measures, etc.) the patients’ data.

### **2.8.2.1.2. United Kingdom – Stockport Council**

In the UK the Data Protection Act 1998 (DPA) applies to personal data. Information which relates to a living, identifiable individual i.e. information which is about individuals and identifies them is their personal data.

Organisations such as Stockport Metropolitan Borough Council are ‘data controllers’ for the purposes of the DPA because they process personal data. ‘Processing’ means anything that is done to the data including just holding them in a file or computer system.

The DPA gives individuals (known as data subjects) a number of rights in relation to their personal data and sets out rules that must be followed by data controllers when they process personal data. As a data controller Stockport Council must ensure it complies with the DPA.

The DPA gives individuals the right to ask any organisation which processes personal data about them for a copy of that information.

They are entitled to see personal data that are held electronically as well as those held in most manual records. These types of requests are called subject access requests (SARs).

The DPA requires organisations such as Stockport Council to treat individuals’ personal data fairly.

To help meet this requirement, individuals should be told why their personal data is being held, what the organisation will do with it and the types of third parties, if any, their personal data may be disclosed to.

This information is usually contained within a Privacy Notice within the constraints of the following 4 areas:

- National Fraud Initiative<sup>1</sup>
- Adult Social Care<sup>2</sup>
- Schools and Children Social Care<sup>1</sup>

---

<sup>1</sup><http://www.stockport.gov.uk/services/councildemocracy/yourcouncil/freedomofinformationdataprotection/dataprotection/privacynotices/nationalfraudinitiative>

<sup>2</sup><http://www.stockport.gov.uk/services/councildemocracy/yourcouncil/freedomofinformationdataprotection/dataprotection/privacynotices/privacynoticeasc/>



- Schools Workforce Census<sup>2</sup>

Stockport Council also ensures it complies with the Caldicott principles which were laid down by the Government. They must be followed by employees of Stockport Council working in social care settings with 'person- identifiable information'; however it is good practice for all service areas to follow these principles. They are outlined below:

- Principle 1: Justify the purpose. Every proposed use or transfer of person-identifiable information within or from an organisation should be clearly defined and scrutinised, with continuing uses regularly reviewed by an appropriate guardian.
- Principle 2: Don't use person-identifiable information unless it is absolutely necessary. Person- identifiable information should not be used unless there is no alternative.
- Principle 3: Use the minimum necessary person-identifiable information. Where use of person -identifiable information is considered to be essential, each individual item of information should be justified with the aim of reducing identifiably.
- Principle 4: Access to person- identifiable information should be on a strict need-to-know basis. Only those individuals who need access to person-identifiable information should have access to it and they should only have access to the information they need to see.
- Principle 5: Everyone should be aware of their responsibilities. Action should be taken to ensure that those handling person - identifiable information are aware of their responsibilities and obligations to respect individual confidentiality.
- Principle 6: Understand and comply with the law. Every use of person-identifiable information must be lawful. Someone in each organisation should be responsible for ensuring that the organisation complies with legal requirements.

More information relating to this area is available here within Stockport's Policy on Managing Personal Information Document<sup>3</sup>.

### **2.8.2.1.3. Ireland – NUIG**

The main Irish law dealing with data protection is the Data Protection Act 1988 which was amended by the Data Protection (Amendment) Act 2003 (hereafter briefly mentioned as the Act 2003). This Act 2003 brought national law into line with the EU Data Protection Directive 95/46/EC.

In this Act 2003, **personal data** defined as data relating to a living individual who is or can be identified either from the data or from the data in conjunction with other information that is in, or is likely to come into, the possession of the **data controller**.

---

<sup>1</sup><http://www.stockport.gov.uk/services/councildemocracy/yourcouncil/freedomofinformationdataprotection/dataprotection/privacynotices/privacynoticeschoolschildsocialcare>

<sup>2</sup><http://www.stockport.gov.uk/services/councildemocracy/yourcouncil/freedomofinformationdataprotection/dataprotection/privacynotices/privacynoticeschoolsworkforcecensus>

<sup>3</sup><http://www.stockport.gov.uk/2013/2986/48046/managingpersonalinfopolicyfeb2012>

The definition of personal data includes any information that relates to an identifiable, living individual. When people give their personal details to an organisation, this organisation has a duty to keep these details private and safe (means data protection).

A **data controller** is an individual or a legal person who controls and is responsible for the keeping and use of personal information on a computer or in structured manual files. In Ireland for the MARIO project, NUIG is a “data controller” for the purposes of the Data Protection Act because they process personal data. “Processing” means performing any operation or set of operations on data, including:

- Obtaining, recording or keeping data,
- Collecting, organising, storing, altering or adapting the data,
- Retrieving, consulting or using the data,
- Disclosing the information or data by transmitting, disseminating or otherwise making it available,
- Aligning, combining, blocking, erasing or destroying the data.

Under the Act 2003, people have rights regarding the use of their personal details when their data held on a computer or on paper or another manual form as part of a filing system, and made up of photographs or video recordings of their image or recordings of their voice. The owners of the personal data have rights to:

- Have your details used in line with data protection regulations
- Information about your personal details
- Access your personal details
- Know if your personal details are being held
- Change or remove your details
- Prevent use of your personal details
- Remove your details from a direct marketing list
- Object
- Freedom from automated decision-making
- Refuse direct marketing calls or mail.

These rights help data owners to make sure that the information stored about them is factually correct; only available to those who should have it and only used for stated purposes. As a data controller NUIG must ensure it complies with the Act 2003 and other relevant national regulations.

In this project, if people do not feel comfortable about processing their data, they can contact NUIG for further clarification. If they believe that the organisation is still not respecting their data protection rights, they can contact the Office of the Data Protection Commissioner to ask for help.

Data controllers have certain responsibilities in how they handle personal information. These are summarised in terms of eight fundamental rules which must be followed:

- Obtain and process information fairly
- Keep it only for one or more specified, explicit and lawful purposes

- Use and disclose it only in ways compatible with these purposes
- Keep it safe and secure
- Keep it accurate, complete and up-to-date
- Ensure that it is adequate, relevant and not excessive
- Retain it for no longer than is necessary for the purpose or purposes
- Give a copy of his/her personal data to an individual, on request

A basic data protection checklist in terms of national regulation which NUIG will consider as a data controller is provided below:

- Are the individuals whose data you collect aware of your identity?
- Have you told the data subject what use you make of his/her data?
- Are the disclosures you make of that data legitimate ones?
- Do you have appropriate security measures in place both internally and externally to ensure all access to data is appropriate?
- Do you have appropriate procedures in place to ensure that each data item is kept up-to-date?
- Do you have a defined policy on retention periods for all items of personal data?
- Do you have a data protection policy in place?
- Do you have procedures for handling access requests from individuals?
- Are you clear on whether or not you should be registered?
- Are your staff appropriately trained in data protection?
- Do you regularly review and audit the data which you hold and the manner in which they are processed?

### 3. Conclusion

This document provides the basis and structure for the Data Management Plan of the MARIO project. It covers all aspects on how the partners plan to collect, handle, share, archive and preserve the data throughout the project lifetime in accordance to the respective ethical and legal requirements. The Data Management Plan will be updated as part of the project's periodic reports. The lead beneficiary of this deliverable (University of Passau), together with the project coordinator, technical director, pilot manager and innovation manager, serve as references for any questions in relation to data management activities in MARIO.

## Annex 1 – Discussions

This annex provides some of the important discussions that took place within communication involved towards the compilation of the final version of the Data Management Plan (DMP). This list provides a number of issues that are still open in nature and their latest update at the time of deliverable submission. These issues will be further discussed and updated within the next version of the DMP in Month 18 (as an annex to the project's Periodic Reports i.e. Deliverables 11.2).

Open Issues	Latest Update
Dataset 4 on patient health history has now changed and is still changing.	It's not a major problem if we have to change / include more / exclude some datasets in the future, so what we have in the DMP is a good idea of where we're at in M9. That's why in the DMP we propose that this will be revised at least twice i.e. in Month 18 and Month 36 (as annexes to the project's Periodic Reports i.e. Deliverables 11.2 and 11.4). The best idea is that we only update the information about the datasets when we're in a better position to indicate what data we will be actually collecting or not, i.e. by M18.
At the moment we are planning on asking residents to complete surveys on quality of life, depression, cognitive status, social isolation levels, we will also be conducting interviews with them to capture their experience of having MARIO as a companion. Many of the KPI's we had initially submitted are not realistic and cannot be measured, thus are being revised and updated.	The revised indicators will be included in the next version of the DMP. Several changes are being made which have yet to be agreed so it will no doubt change, but it might provide a better picture of what has been removed and what we are likely to be collecting, mainly survey or interview data pre and post the introduction of MARIO. We may want to access people with dementia's medical or nursing notes to see how often the CGA was completed previously and how long it took and to compare the results taken by MARIO versus the nurse, but not sure.
Basecamp not mentioned in DMP.	This project management tool is more aimed to be used amongst us to discuss and share resources, and manage the project's tasks, rather than to share datasets with members outside of the project consortium, A few additional portals are mentioned in the DMP (we might also need to revise this in the future), but the target in this case is different for each.
Association of data management and social media.	Social media is more aimed as a method to promote any open datasets that we will

	be publishing throughout the project lifetime, in order to share this knowledge with any parties that might be interested in using such data. Further discussions are still required on the subject in this respect with R2M and the pilot sites.
Can be confusing to keep separate project data vs. robot data vs. pilot data.	This is a very delicate topic, since certain data cannot be associated with others due to privacy and ethics concerns, especially for the pilot data.
Restricted datasets listed in Section 2.4.	So far we have determined which ones are open / restricted (this might also change later on). We have to further determine which ones will be restricted to the consortium or partners only.