



**PHC-19-2014: Advancing active and healthy ageing with ICT:  
service robotics within assisted living environments**

*Project Title:*

**Robotic Assistant for MCI Patients at home**



**RAMCIP**

**Grant Agreement No: 643433  
Research and Innovation Action (RIA)**

**Deliverable**

**D9.3. RAMCIP Data Management Plan**

Deliverable No.		<b>D9.3</b>	
Workpackage No.	<b>WP9</b>	Workpackage Title	<b>Dissemination and Exploitation</b>
Task No.	<b>T9.1</b>	Task Title	<b>Dissemination material and publication policy</b>
Lead beneficiary		<b>CERTH</b>	
Dissemination level		<b>PU</b>	
Nature of Deliverable		<b>Report</b>	
Delivery date		<b>1 July 2015</b>	
Status		<b>F: final</b>	
File Name:		<b>RAMCIP Deliverable 9.3.doc</b>	
Project start date, duration		<b>01 January 2015, 36 Months</b>	



This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°643433

## Authors List

<b>Leading Author (Editor)</b>				
	<i>Surname</i>	<i>Initials</i>	<i>Beneficiary Name</i>	<i>Contact email</i>
	Fachantidis	N	CERTH	nfachantid@iti.gr
<b>Co-authors (in alphabetic order)</b>				
#	<i>Surname</i>	<i>Initials</i>	<i>Beneficiary Name</i>	<i>Contact email</i>
1	Rontsis	N	CERTH	nrontsis@iti.gr
2	Tzouvaras	D	CERTH	dimitrios.tzouvaras@iti.gr

## Reviewers List

<b>List of Reviewers (in alphabetic order)</b>				
#	<i>Surname</i>	<i>Initials</i>	<i>Beneficiary Name</i>	<i>Contact email</i>
1	Endo	S	TUM	s.endo@tum.de
2	Partarakis	N	FORTH	partarak@ics.forth.gr

<b>Document history</b>			
Version	Date	Status	Modifications made by
1.0	22/5/2015	First version	CERTH/ALL
2.0	27/5/2015	Second version	ALL
3.0	15/5/2015	Third version	ALL
4.0	30/6/2015	Final version	CERTH

## List of definitions & abbreviations

<b>Abbreviation</b>	<b>Definition</b>
AD	Alzheimer's Disease
ADL	Activities of Daily Living
DMP	Data Management Plan
MCI	Mild Cognitive Impairment
PAKE	Password-authenticated key agreement
SRP	Secure Remote Password
ToF	Time of Flight
VO	Visual Odometry
VUMs	Virtual User Models

## Executive Summary

This deliverable is the first version of the Data Management Plan (DMP) of the RAMCIP project, in accordance to the regulations of the Pilot action on Open Access to Research Data of the Horizon 2020 programme (H2020). It contains preliminary information about the data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

To develop the present deliverable, a "data identification form" template was first drafted on the basis of the H2020 guidelines for the development of projects' Data Management Plans. This was circulated to all project partners so as to collect all relevant information concerning the datasets that are planned to be developed in the course of the project. On the basis of all partners' feedback, the preliminary data management plan of the project has been established, as described in the present deliverable.

As shown from the description of the project datasets provided herein, the project at its present stage is foreseen to develop a series of datasets, related to issues ranging from user requirements analysis, through to evaluation of the algorithms and methods that will enable the target skills of the RAMCIP robot. Specifically, datasets are planned to be collected toward developing and evaluating the RAMCIP robot's object recognition algorithms, home environment modelling and monitoring ones, as well as its human activity, behavior and skills (cognitive and physical) modelling and monitoring methods. Moreover, given the focus of the project on advanced, dexterous manipulations inside the user's home environment, datasets are foreseen to be established concerning the modelling of objects and appliances that should be handled by the foreseen robot through its manipulations, as well as ones related to simulating the robot's manipulator kinematics.

The datasets that will be collected in RAMCIP will help the development and improvement of the skills of the RAMCIP robot, while they can also serve as for e.g. benchmarking datasets to the scientific community of the RAMCIP-related research fields, once made public. Nevertheless, as some of the project's datasets involve data collection from human participants, the respective data collection experiments, as well as the data analysis procedures that will be employed should be carefully handled, under thorough consideration of ethical and privacy issues involved in such datasets. In this line, the present deliverable, in parallel to the deliverable D2.4 "Ethics Protocol", pays due attention to ethical and privacy issues related not only to the above, but also to whether the foreseen datasets can be made public. For all the identified RAMCIP datasets, specific parts that can be made publicly available have been identified in the current first version of the project's DMP. The public datasets of the RAMCIP project will become available through a common repository that will be formulated on the basis of the RAMCIP "data management portal"; this will be a dedicated space of the RAMCIP project website, which will aggregate descriptions of all project public datasets, and provide links to respective dataset download sections to the interested public, as well as centralized data management functionalities to project partners.

Clearly, as the present deliverable has been drafted during the first project stages (at Month 6 of the project), it can only reflect the intentions of the project partners toward developing the overall project's datasets. Two further revisions of the present deliverable are planned during the project's duration; the first for Month 24 and the second for Month 36 of the project, following the H2020 guidelines on Data Management Plans. Therefore, as the project progresses, the current deliverable will be elaborated, so as to reflect how the Data Management Plan of the project has been eventually established, by documenting at its final version (by the end of the project), the detailed descriptions of all the specific

datasets that have been collected, clearly depicting which of them have been made publicly available and under which Data Management framework.

## Table of Contents

<b>Robotic Assistant for MCI Patients at home</b> .....	1
<b>List of definitions &amp; abbreviations</b> .....	4
<b>Executive Summary</b> .....	5
<b>1. Introduction</b> .....	8
1.1 Deliverable structure .....	9
<b>2. General Principles</b> .....	10
2.1 Participation in the Pilot on Open Research Data .....	10
2.2 Security .....	10
2.3 Personal Data Protection .....	10
2.4 The RAMCIP Data Management Portal .....	10
<b>3. Description of the foreseen RAMCIP datasets</b> .....	12
3.1 Dataset "DS1.CERTH-01.ObjectRecognitionDataset" .....	14
3.2 Dataset "DS2.CERTH-02.DomesticSpaceModellingDataset" .....	17
3.3 Dataset "DS3.ACCREA-01.InteractEnvComponentsDataset" .....	21
3.4 Dataset "DS4.CERTH-03.HumanTrackingDataset" .....	24
3.5 Dataset "DS5.SSSA-01.HumanMotionFineDataset" .....	27
3.6 Dataset "DS6.SSSA-02.WalkingSkillsDataset" .....	30
3.7 Dataset "DS7.SSSA-03.CogSkillsDataset" .....	33
3.8 Dataset "DS8.CERTH.04.VirtualUserModelsDataset" .....	36
3.9 Dataset "DS9.TUM-01.LowerBodyKinematicsDataset" .....	39
3.10 Dataset "DS10.ACCREA-02.ManipKinematicsDataset" .....	42
3.11 Dataset "DS11.LUM_ACE-01.UserRequirementsDataset" .....	45
<b>4. Discussion</b> .....	48
<b>5. Conclusions</b> .....	49
<b>Annex I: Dataset Identification Template</b> .....	50

## 1. Introduction

The purpose of the Data Management Plan (DMP) deliverable is to provide relevant information concerning the data that will be collected and used by the partners of the project RAMCIP. These datasets are required for the development and evaluation of the methods that will be researched, developed and used to address the particular research problems of the project. RAMCIP aims at developing a domestic service robot capable of providing proactive and discreet assistance to elderly people with MCI and at early AD stages in their everyday life at home.

Such kind of robot should develop high-level cognitive functions, advanced communication and manipulation skills in order to interact with the patients as well as with its environment. The process of training the robot to achieve such advanced skills require capturing a variety of datasets regarding for instance large and small scale object detection, localization and human tracking, while it is of equal importance to simulate the robot kinematics and the patients' behavior to capture synthetic data, instead of relying exclusively on real patients such as the ones of the primary RAMCIP end user groups.

In this scope, this deliverable extensively describes the RAMCIP consortium plans for each dataset that is planned to be collected throughout the project's duration. It provides preliminary information about the origin and nature of each foreseen dataset, its standards, any similar datasets and corresponding publications, data access and preservation policies.

RAMCIP participates to the Pilot action on Open Access to Research Data which is part of the Horizon 2020 program. Our goal is to provide where possible, accurate and high-quality data to the research community so that the project will contribute to future advancements in the field of assistive robotics. However, since data may contain personal information about human participants, a focus is also given to possible ethical issues and access restrictions regarding personal data so that no regulations on sensitive information are violated.

The DMP, as it is currently formed, is not fixed but will be refined in subsequent revisions of the present deliverables, depending on the needs that may arise in due course. This first version of the RAMCIP DMP mainly depicts the direction of the project regarding the collection of the data.

The activities related to data management along the project RAMCIP are planned as follows:

- M6: Preliminary analysis and production of the first version of the Data Management Plan (contained in this document).
- M16: Writing of the specifications for the project's data management portal, where information over the project's datasets and links to download locations shall be provided where applicable (e.g. where a publicly available version of a dataset exists).
- M17-M19: Development of the data management portal (to be carried out by CERTH), as a dedicated part of the RAMCIP website.
- M20: The data management portal is operational.
- M24: Second version of the Data Management Plan, describing actual, proven procedures implemented by the project during its data collection efforts, and preparing the sustainability of the data storage after the end of the project.
- M36: Final Data Management Plan, reflecting on the lessons learnt through the project, and describing the plans implemented by RAMCIP for sustainable storage and accessibility of the data.

In the next section, we provide a list of the foreseen datasets of the project RAMCIP, with detailed descriptions on the aforementioned specifications.

## **1.1 Deliverable structure**

In the rest of the deliverable, Chapter 2 first summarizes key general principles that are involved in the Data Management Plan of the RAMCIP project, such as ones related to data security and personal data protection, whereas it also provides a description of the project's plans toward the development of the data management portal. Chapter 3 serves as the core chapter of the present deliverable, as it describes, in the detail level possible at this early project stage, the datasets that are planned to be collected during the RAMCIP project. Finally, Chapters 4 and 5 provide a discussion on this 1<sup>st</sup> version of the RAMCIP Data Management Plan and draw the conclusions of the present deliverable.

## **2. General Principles**

### **2.1 Participation in the Pilot on Open Research Data**

RAMCIP participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon2020 programme. The consortium believes firmly in the concepts of open science, and the large potential benefits the European innovation and economy can draw from allowing reusing data at a larger scale. Therefore, all data produced by the project may be published with open access – though this objective will obviously need to be balanced with the other principles described below.

### **2.2 Security**

The datasets foreseen to be collected through RAMCIP are of high value and may contain sensitive personal data. Special care should be taken to prevent such datasets to leak or become hacked. This is another key aspect of RAMCIP data management, and all data repositories used by the project will include effective protection.

A holistic security approach will be followed, in order to protect the pillars of information security (confidentiality, integrity, availability). The security approach will consist of a methodical assessment of security risks followed by their impact analysis. This analysis will be performed on the personal information and data processed by the proposed system, their flows and any risk associated to their processing.

Security measures will include the implementation of PAKE protocols, such as the SRP protocol, and protection about bots such as captcha technologies. Moreover, the pilot sites shall apply monitored and controlled procedures related to the data collection, their integrity and protection. The data protection and privacy of personal information will include protective measures against infiltration as well as physical protection of core parts of the systems and access control measures.

### **2.3 Personal Data Protection**

RAMCIP activities will involve human participants for various human activity and behaviour analysis –related data collection purposes. Therefore, it is clear that in some cases personal data will have to be collected. Such data will be protected in accordance with the EU's Data Protection Directive 95/46/EC "on the protection of individuals with regard to the processing of personal data and on the free movement of such data". Further information on how personal data collection and handling should be approached in the RAMCIP project are provided in the deliverable D2.4 "Ethics Protocol" of the project.

All personal data collection efforts of the project partners will be established after giving subjects full details on the experiments to be conducted, and obtaining from them a signed informed consent form, following the respective guidelines set in the D2.4 deliverable.

### **2.4 The RAMCIP Data Management Portal**

RAMCIP will develop a data management portal as part of its website. This portal will provide to the public, for each dataset that will become publicly available, a description of the dataset along with a link to a download section. The portal will be updated each time a new dataset has been collected and is ready of public distribution. The portal will however not contain any datasets that should not become publicly available.

The initial version of the portal will become available during the 2<sup>nd</sup> year of the project, in parallel to the establishment of the first versions of project datasets that can be made publicly available. The RAMCIP data management portal will enable project partners to manage and distribute their public datasets through a common infrastructure.

### 3. Description of the foreseen RAMCIP datasets

In this chapter we provide detailed information about the datasets that are planned to be captured by the partners of the RAMCIP project. In order to meet the requirements of the DMP according to the Pilot of Open Access of the Horizon 2020, each partner provided the description of their datasets using the template given in Annex I, which was formed by following the EC guidelines of the dataset aspects that should be reported in DMPs of the H2020 projects<sup>1</sup>.

#### Datasets Naming Conventions

Concerning the convention followed for naming the RAMCIP datasets, it should be noted that the name of each dataset comprises: (a) a prefix 'DS' indicating a dataset, along with its unique identification number, e.g. "DS1", (b) the name(s) of the partner(s) responsible to collect it, e.g. CERTH, along with an identifier denoting the internal numbering of the dataset concerning the specific partner, e.g. -01, and (c) a short title of the dataset summarizing its content and purpose, e.g. Object Recognition Dataset.

#### Summary of the foreseen RAMCIP datasets

In the following, Table 1 provides a short description of each dataset, whereas the detailed description of each dataset, in accordance to the H2020 DMOP template is provided in the following sections. There are in total 11 datasets foreseen in the project, covering a series of research dimensions on the skills the RAMCIP robot should develop.

**Table 1. Summary of datasets planned to be collected during the course of the RAMCIP project**

No	Name	Description
2.1	DS1.CERTH-01. Object Recognition Dataset	A large scale dataset of images and associated annotations will be collected aiming at benchmarking object recognition and grasping algorithms in a domestic environment.
2.2	DS2.CERTH-02. Domestic Space Modeling Dataset	A collection of RGB-D data with great spatial coherence using the Kinect2 sensor of multiple places concerning indoor scenarios both for large and small scale circumstances.
2.3	DS3.ACCREA-01. Interactive Environmental Components Dataset	A collection of CAD data containing models of usable/interactive elements of RAMCIP user's surroundings, like light switches, water taps, cooker knobs, door handles etc.
2.4	DS4.CERTH-03. Human Tracking Dataset	Dataset for human identification, pose and gestures tracking, facial expressions monitoring and activity tracking along obtained with Kinect2 sensor mounted on a mobile robotic base (e.g. Turtlebot).

<sup>1</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)

2.5	DS5.SSSA-01. Human Motion for Fine Biomechanical Analysis Dataset	Dataset for the training and evaluation of the Fine-grained Body Motion Tracking Task by SSSA.
2.6	DS6.SSSA-02. Human Walking Dataset	Dataset for characterizing the walking behavior of subjects and identification of changes in the motion patterns, based on RGB-D cameras.
2.7	DS7.SSSA-03. Human Cognitive Skills Dataset	Characterization of the variations of a cognitive skill such as knob turning, or shirt button closing. A set of actors will be asked to perform the selected task and RGB-D will be collected using Kinect 2.
2.8	DS8.CERTH.04. Virtual User Models Dataset	Virtual User Models (VUMs) of robot users (e.g. MCI patients), encoding their cognitive and motor skills, behavioral aspects, as well as human-robot interaction and communication preferences.
2.9	DS9.TUM.01. Lower-body kinematic Dataset	Dataset consisting of kinematics of lower-body interaction by pairs of human participants.
2.10	DS10.ACCREA.02 Manipulator kinematics chains Dataset	A set of Simulink/CAD/Gazebo models for simulation, optimization and development purposes.
2.11	DS11.LUM_ACE.01 User Requirements Dataset	Dataset with the pictures and videos taken during the workshops with stakeholders in Lublin and Barcelona as well as anonymized questionnaires, which were filled in by the different stakeholders groups

### 3.1 Dataset "DS1.CERTH-01.ObjectRecognitionDataset"

#### General Description

A large scale dataset of images and associated annotations will be collected aiming at benchmarking object recognition and grasping algorithms in a domestic environment.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS1.CERTH-01. Object Recognition Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>The dataset will include a collection of RGB-D images of household objects captured from various viewpoints using an Xtion and/or Kinect2 sensor. Small sized objects will be captured by placing them on a turntable. Larger objects will be captured by moving the sensor around the object and capturing as much views as possible. Fiducial markers will be used to obtain an accurate estimation of the camera pose for each view point.</p> <p><b>Nature and scale of data</b></p> <p>The data will consist of 3D models of objects that will be created either with CAD software, 3D scanner or by merging RGB-D point clouds as well as test images depicting realistic scenarios for evaluation. The target is that the dataset will contain at least 100 objects. A set of tools for adding more objects in the dataset will also be available.</p> <p><i>Data Format:</i> Training: PLY, OBJ for 3D models, Testing: PNG, JPG for images, TXT for annotations</p> <p><b>To whom could the dataset be useful</b></p> <p>The dataset will be valuable for benchmarking algorithms for object recognition, robotics navigation and grasping.</p> <p><b>Related scientific publication(s)</b></p> <p>The dataset will accompany our research results in the field of object recognition and grasping.</p> <p><b>Indicative existing similar data sets</b></p> <p>There are several public datasets containing RGB-D images of objects aimed at object recognition.</p> <p>The UW dataset (<a href="http://www.cs.washington.edu/rgb-d-dataset/">http://www.cs.washington.edu/rgb-d-dataset/</a>)</p> <p>The Berkley's B3DO dataset (<a href="http://kinectdata.com/">http://kinectdata.com/</a>)</p> <p>The Berkley's BigBird dataset (<a href="http://rll.berkeley.edu/bigbird/">http://rll.berkeley.edu/bigbird/</a>).</p> <p><i>Part of our dataset will be considered for integration in the B3DO dataset that is designed to be extensible.</i></p>	
<b>3.</b>	<b>Standards and metadata</b>
Indicative metadata include a) foreground-background masks for training images,	

b) camera calibration information, c) camera pose matrix for each viewpoint, d) object identifier, description category label and 3D pose annotation, e) 3D object model in CAD format. The metadata will be in a format that may be easily parsed with open source software.

#### 4. Data sharing

##### **Access type**

Widely open.

##### **Access Procedures**

A web page will be created on the RAMCIP data management portal (hosted at the RAMCIP web site) that should provide a description of the dataset and links to a download section.

##### **Embargo periods**

Some datasets maybe available only after the corresponding paper is accepted and published.

##### **Technical mechanisms for dissemination**

A link to the dataset from the RAMCIP web site (RAMCIP data management portal). The link will be provided in all relevant RAMCIP publications. A technical publication describing the dataset and acquisition procedure will be published.

##### **Necessary S/W and other tools for enabling re-use**

The dataset will be designed to allow easy reuse with commonly available tools and software libraries.

##### **Repository where data will be stored**

The dataset will be accommodated at the data management portal of the project website.

#### 5. Archiving and preservation

##### **Data preservation period**

The dataset will be preserved online for as long as there are regular downloads. After that it would be made accessible by request.

##### **Approximated end volume of data**

The dataset is expected to be several Gigabytes.

##### **Indicative associated costs for data archiving and preservation**

Probably a dedicated hard disk drive will be allocated for the dataset. No costs are currently foreseen regarding its preservation.

##### **Indicative plan for covering the above costs**

Small, one-time costs covered by RAMCIP.

#### 6. Partners activities and responsibilities

***Partner Owner / Data Collector***

CERTH

***Partner in charge of the data analysis***

CERTH

***Partner in charge of the data storage***

CERTH

***WPs and Tasks***

The data are going to be collected within activities of WP3 in Task 3.1 and will mainly be used for analysis in the scope of WP3, WP5 and WP6 tasks

## 3.2 Dataset "DS2.CERTH-02.DomesticSpaceModellingDataset"

### General Description

The space modeling dataset comprises the collection of RGB-D data with great spatial coherence using the Kinect2 ToF (Time of Flight) sensor. It is expected that multiple places will be recorded concerning indoor scenarios both for large and small scale circumstances. The collected dataset will contain fully registered color (RGB) images with their respective depth maps. The collection area concerns domestic real home or home-like environment.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS2.CERTH-02. Domestic Space Modeling Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>During the acquisition procedure the sensor motion should be as smooth as possible, which combined with the high frame rate of the sensor will ensure great overlap among the captured scenes. Therefore, the collected data will be suitable for mapping and navigation experimentation.</p> <p><b>Nature and scale of data</b></p> <p>The Domestic-Space-Modelling dataset will be split into two parts:  In PART I the recording will be carried out in static environment, providing thus the required data to assess the developed solutions (mapping, navigation) in their basis.  In PART II the recording will be carried out in a dynamic environment including also human activity. Thus the acquired data will be used for the assessment of the developed algorithms in their higher level (map recalling, planning, and re-planning) and their real performance in human inhabited environments.  Moreover, the acquired dataset will be accompanied with accurate ground-truth measurements (of the robot location and pose, as well as of the modelled space) for the evaluation of the mapping and localization algorithms.  <i>Data Format:</i> PNG, JPG image format</p> <p><b>To whom could the dataset be useful</b></p> <p>The dataset could be useful for the benchmarking of mapping and robotic navigation solutions.</p> <p><b>Related scientific publication(s)</b></p> <p>The results of the developed algorithms along with the Domestic-Space-Modelling dataset can be disseminated in International Conferences and Journals of the robotics field.</p> <p><b>Indicative existing similar datasets</b></p> <p>Similar datasets have already been collected in the past such as:  (<a href="http://vision.in.tum.de/data/datasets/rqbd-dataset">http://vision.in.tum.de/data/datasets/rqbd-dataset</a>) provided by the Technische</p>	

Universität München.  
<http://robotics.pme.duth.gr/kostavelis/Dataset.html>) provided by Laboratory of Robotics and Automation, DUTH.  
 Contrary to the aforementioned cases where the data have been collected with the RGB-D sensor Kinect1, our Domestic-Space-Modelling dataset will be captured with a Kinect2 sensor which is more accurate and retains greater resolution.  
 Since the publicly available datasets are recorded with a Kinect1 sensor, a direct integration with the Domestic-Space-Modelling dataset is problematic mainly due to the fact that a) the data are collected in different environments and b) the resolution is different between the acquired data.

### 3. Standards and metadata

The Domestic-Space-Modelling should be accompanied by accurate ground-truth data ensuring the validity of the developed algorithms as well as their reuse in future research activities.

The possible metadata to be produced can be summarized as follows:

- The point clouds (textured/pseudo-colored) of each instance transformed in real world coordinates (x, y, z).
- The produced 3D/2D map as a result of the developing procedure within the RAMCIP project, providing a benchmark
- The consecutive Visual Odometry (VO) transformations reproducing the trajectory of the robot, also for benchmarking.

### 4. Data sharing

#### **Access type**

Only portions of the PART I of the dataset that contain a static environment could be publicly released. These portions concern data collected from the LUM apartment, simulating a home-like environment, without human presence.

For the rest parts of this dataset, including e.g. data collected from real apartments and those dynamically updated through human activities, only private access is given. The access is granted to RAMCIP partners whose research and development activities have a direct dependency (e.g. map recalling, planning and re-planning), on the basis that a respective informed consent has been taken from the human subjects participated in the data collection.

The latter parts of the dataset, including models of real apartments and in some cases dynamically updated through human activity, cannot become publicly available. Regardless of the informed consent for publication, such data could lead to a recognition of participant's identity and details of his/her home environment. Thus, it raises significant privacy and ethical concerns and publication of such a dataset should be prevented, as further explained in the project's ethics protocol (deliverable D2.4).

On the contrary, as the LUM home-like environment concerns a public space, the respective home environment modelling and monitoring dataset, without human presence, would not be subject to such privacy and ethical issues.

The dataset will be accompanied with a specific technical report describing the calibration, the acquisition procedure as well as technical details of the architecture of the robot.

#### **Access Procedures**

For the public part of this dataset a web page will be created on the RAMCIP data management portal (hosted at the RAMCIP web site) that should provide a description of the dataset and links to a download section.

The private part of this dataset will be stored at a specifically designated private space of CERTH, in dedicated hard disk drives, on which only members of the CERTH research team whose work directly relates to these data will have access. For the other RAMCIP partners to obtain access to these data, they should provide a formal request to the CERTH's primarily responsible for the data storage, including a justification over the need to have access to these data. Once deemed necessary, CERTH will provide the respective data portions to the partner.

***Embargo periods (if any)***

None

***Technical mechanisms for dissemination***

A link to the public part of this dataset from the RAMCIP web site (data management portal). The link will be provided in all relevant RAMCIP publications. A technical publication describing the dataset and acquisition procedure will be published.

***Necessary S/W and other tools for enabling re-use***

The dataset will be designed to allow easy reuse with commonly available tools and software libraries.

***Repository where data will be stored***

The public part of this dataset will be accommodated at the data management portal of RAMCIP.

**5. Archiving and preservation**

***Data preservation period***

The public part of the dataset will be preserved online for as long as there are regular downloads. After that, it would be made accessible by request.

The private part of the dataset will be preserved by CERTH at least until the end of the project.

***Approximated end volume of data***

The dataset is expected to be several Gigabytes.

***Indicative associated costs for data archiving and preservation***

Probably two dedicated hard disk drives will be allocated for the dataset; one dedicated to the public part and one to the private. No costs are currently foreseen regarding its preservation.

***Indicative plan for covering the above costs***

Small one-time costs covered by RAMCIP.

**6. Partners activities and responsibilities**

***Partner Owner / Data Collector***

CERTH

***Partner in charge of the data analysis***

CERTH

***Partner in charge of the data storage***

CERTH

***WPs and Tasks***

The data are going to be collected within activities of WP3 in Task 3.1 and are planned to be mainly used in the research efforts of same task, as well as in the context of WP5 activities.

### 3.3 Dataset "DS3.ACCREA-01.InteractEnvComponentsDataset"

#### General Description

Interactive environmental components dataset comprises the collection of CAD data. The prepared dataset will contain models of usable/interactive elements of RAMCIP user's surroundings, such as light switches, water taps, cooker knobs, door handles etc.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS3.ACCREA-01. Interactive Environmental Components Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b> CAD models will be created using a 3D-scanning device and regular calliper/ruler methods.</p> <p><b>Nature and scale of data</b> Dataset will be available in a form of SolidWorks library files package. <i>Data Format:</i> SolidWorks format</p> <p><b>To whom could it be useful</b> The collected data will be used for simulations and development of the RAMCIP manipulator, mobile platform, elevation mechanism and dexterous hand kinematic chains. Models could be imported into the Gazebo environment simulation, which will be used for testing components and system integration by most of technical RAMCIP partners.</p> <p><b>Related scientific publication(s)</b> N/A</p> <p><b>Indicative existing similar data sets</b> Several websites provide free bases of everyday objects, although not all of them are applicable for RAMCIP uses because of their artistic purposes instead of mechanical/simulation ones.</p>	
<b>3.</b>	<b>Standards and metadata</b>
<p>The possible metadata to be produced can be summarized as follows:</p> <ul style="list-style-type: none"> <li>the parsing routines used to read and absorb the data for developing purposes,</li> <li>the 3D CAD maps of selected user environments with modelled objects placed on specified world coordinates</li> </ul>	
<b>4.</b>	<b>Data sharing</b>
<p><b>Access type</b> The entire dataset or part of it can be publicly available. Its public part will be</p>	

accessible through the data management portal of the RAMCIP project. The dataset will be accompanied with photographs and datasheets of chosen, more complex models.

### ***Access Procedures***

For the public part of this dataset a web page will be created on the RAMCIP data management portal (hosted at the RAMCIP web site) that should provide a description of the dataset and links to a download section.

The private part of this dataset will be stored at a specifically designated private space of ACCREA, in dedicated hard disk drives, on which only members of the ACCREA research team whose work directly relates to these data will have access. For the other RAMCIP partners to obtain access to these data, they should provide a formal request to the ACCREA's primarily responsible for the data storage, including a justification over the need to have access to these data. Once deemed necessary, ACCREA will provide the respective data portions to the partner

### ***Embargo periods***

None

### ***Technical mechanisms for dissemination***

A link to the dataset from the Data management portal. The link will be provided in all relevant RAMCIP publications.

### ***Necessary S/W and other tools for enabling re-use***

The dataset will be designed to allow easy reuse with commonly available tools and software libraries.

### ***Repository where data will be stored***

The public part of this dataset will be accommodated at the data management portal of the project website.

## **5. Archiving and preservation**

### ***Data preservation period***

The dataset will be preserved online for as long as there are regular downloads. After that it would be made accessible by request.

### ***Approximated end volume of data***

The data are expected to be several hundred of Megabytes.

### ***Indicative associated costs for data archiving and preservation***

Probably a dedicated hard disk drive will be allocated for the dataset. No costs are currently foreseen regarding its preservation.

### ***Indicative plan for covering the above costs***

The cost will be covered at the local hosting institute in the context of RAMCIP.

## **6. Partners activities and responsibilities**

### ***Partner Owner / Data Collector***

ACCREA

***Partner in charge of the data analysis***

ACCREA, TUM, CERTH, SHADOW

***Partner in charge of the data storage***

ACCREA, CERTH

***WPs and Tasks***

The data are going to be collected within activities of WP5 in Task 5.1 and 5.4, to serve the respective project tasks' research efforts.

### 3.4 Dataset "DS4.CERTH-03.HumanTrackingDataset"

#### General Description

Dataset for human identification, pose and gestures tracking experiments, along with high-level activities monitoring (e.g. Activities of Daily Living – ADLs, such as cooking or eating), obtained with Kinect2 sensor mounted on a mobile robotic base (e.g. Turtlebot).

The dataset is also planned to include facial expressions monitoring and activity tracking during different affective states, to be used for WP4 affect-related analyses.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS4.CERTH-03. Human Tracking Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>The dataset will be collected using a Kinect2 sensor mounted on a mobile robotic base (e.g. Turtlebot robotic platform). During the acquisition procedure the robot motion should be as smooth as possible.</p> <p><b>Nature and scale of data</b></p> <p>It is expected that the collection experiment will be carried out in two phases, that will split the dataset into three parts:</p> <p>In PART I the recording will focus on monitoring of low-level human activities, such as pose, gestures and actions.</p> <p>In PART II the recording will deal with monitoring of high-level domestic activities, such as cooking and eating.</p> <p>In PART III the recording will focus on facial expressions and activity monitoring during different affective states of the user.</p> <p><i>Data Format:</i> PNG, JPG for images, XML or TXT for annotations</p> <p><b>To whom could the dataset be useful</b></p> <p>The collected data will be used for the development and evaluation of the human activity monitoring and the affect recognition methods of the RAMCIP project. The different parts of the dataset could be useful in the benchmarking of a series of human tracking methods, focusing either on human identification, on pose and gesture analysis and tracking, on high-level activity recognition and on affect-related human activity analysis.</p> <p><b>Related scientific publication(s)</b></p> <p>The dataset will accompany our research results in the field of human activity monitoring and affect recognition.</p> <p><b>Indicative existing similar datasets</b></p> <p>HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion, IJCV 2010.</p> <p>Cornell Activity Datasets: CAD-60 &amp; CAD-120 (<a href="http://pr.cs.cornell.edu/humanactivities/data.php">http://pr.cs.cornell.edu/humanactivities/data.php</a>)</p>	

It should be noted that although several RGB-D datasets dealing with human activity analysis are publicly available at present (e.g. the MSRDailyActivity3D dataset - <http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc>), to the best of our knowledge, no domestic human activity tracking datasets, focusing on low-level actions, high-level activities and affect, recorded through the Kinect2 sensor mounted on a mobile robot base currently exist.

### **3. Standards and metadata**

The dataset will be accompanied with detailed documentation of its contents. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset and (c) annotated pose, action, activity and affective state of the monitored person per time interval.

### **4. Data sharing**

#### ***Access type***

Due to ethical reasons, only the data captured in the LUM premises by normal healthy control subjects could become publicly available, while the rest of them will be private to serve the RAMCIP R&D objectives.

The inclusion of a (normal healthy control) subject's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication (see deliverable D2.4).

#### ***Access Procedures***

For the portions of the dataset that will be made publicly available, a respective web page will be created on the data management portal (hosted at the RAMCIP web site) that will provide a description of the dataset and links to a download section.

The private part of this dataset will be stored at a specifically designated private space of CERTH, in dedicated hard disk drives, on which only members of the CERTH research team whose work directly relates to these data will have access. For further RAMCIP partners to obtain access to these data, they should provide a proper request to the CERTH primarily responsible, including a justification over the need to have access to these data. Once deemed necessary, CERTH will provide the respective data portions to the partner.

#### ***Embargo periods***

None

#### ***Technical mechanisms for dissemination***

For the public part of the dataset, a link to this will be provided from the Data management portal. The link will be provided in all relevant RAMCIP publications. A technical publication describing the dataset and acquisition procedure will be published.

#### ***Necessary S/W and other tools for enabling re-use***

The dataset will be designed to allow easy reuse with commonly available tools and software libraries.

#### ***Repository where data will be stored***

The public part of this dataset will be accommodated at the data management portal of the project website.

## **5. Archiving and preservation**

### ***Data preservation period***

The public part of this dataset will be preserved online for as long as there are regular downloads. After that it would be made accessible by request.

The private part of the dataset will be preserved by CERTH at least until the end of the project.

### ***Approximated end volume of data***

The dataset is expected to be several Gigabytes.

### ***Indicative associated costs for data archiving and preservation***

Probably two dedicated hard disk drives will be allocated for the dataset; one for the public part and one for the private. There are no costs associated with its preservation.

### ***Indicative plan for covering the above costs***

Small one-time costs covered by RAMCIP.

## **6. Partners activities and responsibilities**

### ***Partner Owner / Data Collector***

CERTH

### ***Partner in charge of the data analysis***

CERTH, SSSA, LUM

### ***Partner in charge of the data storage***

CERTH

### ***WPs and Tasks***

The data are going to be collected within activities of WP3 and WP4, to mainly serve the research efforts of T3.2, T3.4 and T4.2.

### 3.5 Dataset "DS5.SSSA-01.HumanMotionFineDataset"

#### General Description

This dataset is aimed at the training and evaluation of the Fine-grained Body Motion Tracking Task by SSSA (T3.3) in the activities not covered by other datasets from RAMCIP.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS5.SSSA-01. Human Motion for Fine Biomechanical Analysis Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>The dataset will be collected by asking people to perform motions compatible with the project aims, recorded with Kinect2 RGB-D camera placed on a fixed structure with the same point of view of the robot at short range distance, together with Vicon Motion Capture markers.</p> <p><b>Nature and Scale of Data</b></p> <p>The dataset will comprise two sets of data: Hand motions and Arm motions, computed with 5 people each.</p> <p>The size of the dataset will be on the order of 2-3 GB.</p> <p><i>Data Format:</i> PNG, JPG image format</p> <p><b>To whom could the dataset be useful</b></p> <p>The dataset will be helpful for research because it will combine a marker-based tracking with Kinect 2 recordings in specific short ranges shots. The biomechanical measure provided by Vicon will provide a ground truth to the Kinect data.</p> <p><b>Related scientific publication(s)</b></p> <p>A scientific publication will be created for analyzing data and proposing a new mechanism for a body tracking.</p> <p><b>Indicative existing similar data sets (including possibilities for integration and reuse)</b></p> <p><a href="http://pr.cs.cornell.edu/humanactivities/data.php">http://pr.cs.cornell.edu/humanactivities/data.php</a> (no Vicon)</p> <p><a href="http://vision.imar.ro/human3.6m/description.php">http://vision.imar.ro/human3.6m/description.php</a></p> <p><a href="http://pr.cs.cornell.edu/humanactivities/">http://pr.cs.cornell.edu/humanactivities/</a></p>	
<b>3.</b>	<b>Standards and metadata</b>
<p>The dataset will be accompanied with detailed documentation of its contents. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset and (c) annotated pose, action, activity and affective state of the monitored person per time interval.</p>	

<b>4.</b>	<b>Data sharing</b>
<p><b>Access type</b></p> <p>Only data from normal healthy control subjects will be acquired, and such data will not contain personal information. This will allow the release of data, after anonymization, to the public. The collection of data from these subject will require the adoption a consent form that will follow the guidelines of deliverable D2.4 (Ethics Protocol).</p> <p><b>Access Procedures</b></p> <p>For the portions of the dataset that will be made publicly available, a respective web page will be created on CERTH's RAMCIP portal that will provide a description of the dataset and links to a download section.</p> <p>The private part of this dataset will be stored at a specifically designated private space of SSSA, in dedicated hard disk drives, on which only members of the SSSA research team whose work directly relates to these data will have access. For further RAMCIP partners to obtain access to these data, they should provide a proper request to the SSSA's primarily responsible, including a justification over the need to have access to these data. Once deemed necessary, SSSA will provide the respective data portions to the partner.</p> <p><b>Embargo periods</b></p> <p>None</p> <p><b>Technical mechanisms for dissemination</b></p> <p>For the public part of the dataset, a link to this will be provided from the Data management portal. The link will be provided in all relevant RAMCIP publications. A technical publication describing the dataset and acquisition procedure will be published.</p> <p><b>Necessary S/W and other tools for enabling re-use</b></p> <p>The data will be published as ROS bag and in a form easily loadable by MATLAB. The ROS solution is quite good for existing tools but it is not good on the long term due to the complexity of the representation and the associated dependencies.</p> <p><b>Repository where data will be stored</b></p> <p>The dataset will be made available over a dedicated website under the domain of SSSA. The data management portal will provide links to the dataset's download section.</p>	
<b>5.</b>	<b>Archiving and preservation</b>
<p><b>Data preservation period</b></p> <p>The data will be available on the PERCRO SSSA website with an expected lifetime of 10 years given the history of PERCRO and the backup procedures of SSSA. The digital signature of the whole dataset, or the storage of the dataset in a git repository could provide support for the correct duplication and preservation.</p> <p><b>Approximated end volume of data</b></p> <p>Few GBs.</p>	

**Indicative associated costs for data archiving and preservation**

None if kept on SSSA server.

**6. Partners activities and responsibilities****Partner Owner / Data Collector**

SSSA

**Partner in charge of the data analysis**

SSSA, CERTH

**Partner in charge of the data storage**

SSSA

**WPs and Tasks**

The data are going to be collected within activities of WP3 in Task 3.3

### 3.6 Dataset "DS6.SSSA-02.WalkingSkillsDataset"

#### General Description

This dataset will be created for the purpose of characterizing the walking behavior of subjects based on RGB-D cameras. This characterization is part of the motor based skill assessment of the subject for the identification of changes in the motion patterns.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS6.SSSA-02. Human Walking Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data (e.g. indicative collection procedure, devices used etc.)</b></p> <p>Actors will be asked to perform different walking behaviors introducing purposely created walking errors, such as missed steps, and foot sliding over the terrain: the specific types of walking patterns will be obtained from known and structured walking problems taken from literature and experts. This recording will be performed with a Kinect 2 sensor and Vicon Motion capture setup as ground truth. The data will be labelled by the task.</p> <p><b>Nature and scale of data</b></p> <p>For this dataset, 5 subject will be recorded with 1 minute for each type of action The dataset will be on the order of GB. <i>Data Format:</i> PNG, JPG for images, XML or TXT for annotations</p> <p><b>To whom could the dataset be useful</b></p> <p>This dataset will be very valuable for research due to the validation with Vicon and the use of Kinect 2. The collected data will be used for the development and evaluation of the human tracking and motor skills estimation. The different parts of the dataset could be useful in the benchmarking of a series of human tracking methods.</p> <p><b>Related scientific publication(s)</b></p> <p>Such dataset is not existent from literature, and it will be used for characterizing skill level in a new publication.</p> <p><b>Indicative existing similar datasets</b></p> <p>Various activity datasets do exist, but none deals with variability in walking patterns. In addition this dataset will provide Vicon measures together with the Kinect2.</p>	
<b>3.</b>	<b>Standards and metadata</b>
<p>The dataset will be accompanied with detailed documentation of its contents. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset and (c) annotated pose, action, activity and affective state of the monitored person per time interval.</p>	

<b>4.</b>	<b>Data sharing</b>
<p><b>Access type</b></p> <p>Only data from normal healthy control subjects will be acquired, and such data will not contain personal information. This will allow the release of data, after anonymization, to the public. The collection of data from these subject will require the adoption of a consent form that will follow the guidelines of deliverable D2.4 (Ethics Protocol).</p> <p><b>Access Procedures</b></p> <p>For the portions of the dataset that will be made publicly available, a respective web page will be created on the data management portal that will provide a description of the dataset and links to a download section.</p> <p>The private part of this dataset will be stored at a specifically designated private space of SSSA, in dedicated hard disk drives, on which only members of the SSSA research team whose work directly relates to these data will have access. For further RAMCIP partners to obtain access to these data, they should provide a proper request to the SSSA's primarily responsible, including a justification over the need to have access to these data. Once deemed necessary, SSSA will provide the respective data portions to the partner.</p> <p><b>Embargo periods</b></p> <p>None</p> <p><b>Technical mechanisms for dissemination</b></p> <p>Publishing and RAMCIP project advertising. Eventually robotics mailing list advertising.</p> <p><b>Necessary S/W and other tools for enabling re-use</b></p> <p>The data will be published as ROS bag and in a form easily loadable by MATLAB. The ROS solution is quite good for existing tools but it is not good on the long term due to the complexity of the representation and the associated dependencies.</p> <p><b>Repository where data will be stored</b></p> <p>The public part of the dataset will be made available over a dedicated website under the domain of SSSA. The RAMCIP data management portal will provide links to the respective dataset's download section.</p>	
<b>5.</b>	<b>Archiving and preservation</b>
<p><b>Data preservation period</b></p> <p>The data will be available on the PERCRO SSSA website with an expected lifetime of 10 years given the history of PERCRO and the backup procedures of SSSA. The digital signature of the whole dataset, or the storage of the dataset in a git repository could provide support for the correct duplication and preservation.</p> <p><b>Approximated end volume of data</b></p> <p>Few GBs</p>	

**Indicative associated costs for data archiving and preservation**

None if kept on SSSA server.

**6. Partners activities and responsibilities****Partner Owner / Data Collector**

SSSA

**Partner in charge of the data analysis**

SSSA, CERTH

**Partner in charge of the data storage**

SSSA

**WPs and Tasks**

The data are going to be collected within activities of WP3 in Task 3.3 and Task 3.5.

### 3.7 Dataset "DS7.SSSA-03.CogSkillsDataset"

#### General Description

The aim of this dataset is to characterize the variations of a cognitive skill such as procedural memory and spatial memory (knob turning, shirt button closing). A set of actors will be asked to perform the selected task with variations and different levels of quality.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS7.SSSA-03. Human Cognitive Skills Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data (e.g. indicative collection procedure, devices used etc.)</b> The dataset will be collected by using a RGB-D camera compatible with the RAMCIP system (Kinect 2).</p> <p><b>Nature and scale of data</b> 4 subjects will be asked to perform a series of task for up to 5 minutes covering the cognitive skills mentioned. The dataset will be on the order of 2-3 GB. <i>Data Format:</i> PNG, JPG for images, XML or TXT for annotations</p> <p><b>To whom could the dataset be useful</b> The dataset will be useful for the cognitive skills recognition and estimation module of RAMCIP and also for external experts for training machine learning based recognizers.</p> <p><b>Related scientific publication(s)</b> Such dataset is not existent from literature, and it will be used for characterizing skill level in a new publication.</p> <p><b>Indicative existing similar data sets (including possibilities for integration and reuse)</b> There are no datasets of this kind, in particular related to a specific task as RAMCIP proposed.</p>	
<b>3.</b>	<b>Standards and metadata</b>
The dataset will be accompanied with detailed a documentation of its contents. Indicative metadata include: (a) description of the experimental setup and procedure that led to the generation of the dataset, (b) documentation of the variables recorded in the dataset and (c) annotated pose, action, activity and affective state of the monitored person per time interval.	
<b>4.</b>	<b>Data sharing</b>
<p><b>Access type</b> Only data from normal healthy control subjects will be acquired, and such data will not contain personal information. This will allow the release of data, after anonymization, to the public. The collection of data from these subject will require</p>	

the adoption a consent form that will follow the guidelines of deliverable D2.4 (Ethics Protocol).

### ***Access Procedures***

For the portions of the dataset that will be made publicly available, a respective web page will be created on the data management portal that will provide a description of the dataset and links to a download section.

The private part of this dataset will be stored at a specifically designated private space of SSSA, in dedicated hard disk drives, on which only members of the SSSA research team whose work directly relates to these data will have access. For further RAMCIP partners to obtain access to these data, they should provide a proper request to the SSSA primarily responsible, including a justification over the need to have access to these data. Once deemed necessary, SSSA will provide the respective data portions to the partner.

### ***Embargo periods***

None

### ***Technical mechanisms for dissemination***

For the public part of the dataset, a link to this will be provided from the data management portal. The link will be provided in all relevant RAMCIP publications. A technical publication describing the dataset and acquisition procedure will be published.

### ***Necessary S/W and other tools for enabling re-use***

Data will be published as ROS bag and in a form easily loadable by MATLAB. The ROS solution is quite good for existing tools but it is not good on the long term due to the complexity of the representation and the associated dependencies.

### ***Repository where data will be stored***

The public part of the dataset will become available over a dedicated website under the domain of SSSA. The data management portal will provide links to the respective dataset's download section.

## **5. Archiving and preservation**

### ***Data preservation period***

The data will be available on the PERCRO SSSA website with an expected lifetime of 10 years given the history of PERCRO and the backup procedures of SSSA. The digital signature of the whole dataset, or the storage of the dataset in a git repository could provide support for the correct duplication and preservation.

### ***Approximated end volume of data***

Few GBs.

### ***Indicative associated costs for data archiving and preservation***

None if kept on SSSA server.

## **6. Partners activities and responsibilities**

### ***Partner Owner / Data Collector***

SSSA

***Partner in charge of the data analysis***

SSSA, CERTH, TUM, FORTH, LUM, ACE

***Partner in charge of the data storage***

SSSA

***WPs and Tasks***

The data are going to be collected within activities of WP3, to serve the needs of the project research efforts in Task 3.5 and Task 3.6

### 3.8 Dataset "DS8.CERTH.04.VirtualUserModelsDataset"

#### General Description

This dataset concerns the RAMCIP VUMs; these will be Virtual User Models (VUMs) of robot users (e.g. MCI patients), encoding their cognitive and motor skills, behavioral aspects, as well as human-robot interaction and communication preferences. The models will be an XML-based specification of parameters that are taken into account in the context of the RAMCIP user modeling methodology. The dataset is thus planned to include for each indicative robot user case, a semantic representation of a series of parameters related to the above.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS8.CERTH.04. Virtual User Models Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>This dataset will be derived by analyzing the datasets of Human Tracking (2.4), Walking Skills (2.6) and Human Cognitive Skills (2.7) described above, toward modeling behavioral aspects as well as cognitive and motor skills of the participants of the respective data collection experiments, into VUM representations.</p> <p><b>Nature and scale of data</b></p> <p>The dataset will be in the form of XML-based representations of the parameters involved in the RAMCIP VUMs.  <i>Data Format:</i> XML file format</p> <p><b>To whom could the dataset be useful</b></p> <p>This dataset will be used in the development of the RAMCIP user modelling methodology of WP3. The dataset could also be useful for researchers investigating behavioral traits, as well as cognitive and motor skills correlates to MCI.</p> <p><b>Related scientific publication(s)</b></p> <p>The developed VUMs dataset can be disseminated in International Conferences and Journals of robotics and health (e.g. MCI-related) domains.</p> <p><b>Indicative existing similar datasets</b></p> <p>Virtual Human Models encoding anthropometric and kinematic parameters of the human body, focusing on the elderly and disabled have derived from the VERITAS FP7 project. Knowledge derived from the VERITAS VUMs could be integrated into the RAMCIP VUMs which, however, will also focus on the cognitive and behavioral traits of elderly with MCI.</p>	
<b>3.</b>	<b>Standards and metadata</b>
The dataset will be accompanied with detailed documentation of its contents; detailed documentation of the variables involved in the RAMCIP VUMs will be	

provided. Guidelines for Virtual Human Modelling derived from the VUMS cluster ([http://vums.iti.gr/index8091.html?page\\_id=64](http://vums.iti.gr/index8091.html?page_id=64)) can be followed, as well as related XSD and XML specifications will be followed. The relevance of following also usiXML-based paradigms to develop respective (e.g. Human Robot Communication -related) parts of the RAMCIP VUMs will be investigated.

#### 4. Data sharing

##### **Access type**

Anonymized versions of the RAMCIP VUMs can formulate open models encoding behavioral traits and human robot communication preferences, cognitive and motor skills of MCI patients.

##### **Access Procedures**

A web page will be created on the RAMCIP data management portal that should provide a description of the dataset and links to a download section.

##### **Embargo periods (if any)**

None

##### **Technical mechanisms for dissemination**

A link to the anonymized dataset from the Data management portal. The link will be provided in all relevant RAMCIP publications. A technical publication describing the dataset and acquisition procedure could be published.

##### **Necessary S/W and other tools for enabling re-use**

The dataset will be designed to allow easy reuse with commonly available XML editing tools and software libraries.

##### **Repository where data will be stored**

The public part of the dataset will be accommodated at the data management portal of RAMCIP.

#### 5. Archiving and preservation

##### **Data preservation time**

The dataset will be preserved online for as long as there are regular downloads. After that it would be made accessible by request.

##### **Approximated end volume of data**

The dataset's end volume is expected to be at the level of Megabytes.

##### **Indicative associated costs for data archiving and preservation**

Probably a dedicated hard disk drive will be allocated for the dataset. There are no costs associated with its preservation.

##### **Indicative plan for covering the above costs**

The cost will be covered at the local hosting institute as a part of the standard network system maintenance.

6.	<b>Partners activities and responsibilities</b>
	<p><b><i>Partner Owner / Data Collector</i></b> CERTH</p>
	<p><b><i>Partner in charge of the data analysis</i></b> CERTH, SSSA, FORTH, LUM, ACE</p>
	<p><b><i>Partner in charge of the data storage</i></b> CERTH</p>
	<p><b><i>WPs and Tasks</i></b> The data are going to be collected within activities of WP3 in Task 3.4, to serve the project's research efforts within Task 3.4, Task 3.5 and Task 3.6.</p>

### 3.9 Dataset "DS9.TUM-01.LowerBodyKinematicsDataset"

#### General Description

This dataset will contain kinematics of lower-body interaction by pairs of human participants in which one human participant assists wearing a shoe of another seated participant in line with a scenario description of the RAMCIP project. The dataset will be used to train the predictive controller of the RAMCIP system in R&D activities under T6.3. Furthermore, the dataset will be used to provide ergonomic guidance for designing the control of the RAMCIP system.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS9.TUM.01. Lower-body kinematic Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>The data will be collected from volunteers of healthy participants using a Qualisys passive-marker motion tracking system. The small light-weight markers will be placed on the foot, tibia, and femur of both legs to capture the position and orientation of the lower-limb segments. Furthermore, the pose of the torso will be captured with markers placed on the sternum. A separate set of markers will track the positions of the shoe and the hand of the assisting person. The dataset will be obtained in accordance with the local ethics requirements at TUM, Germany, for human subject testing.</p> <p><b>Nature and scale of data</b></p> <p>The raw data are images of the reflective markers taken by each motion tracking camera at a pre-set frequency. The cameras use reflections of infrared light on the special markers to visualize their positions, thus the raw data do not record any personal information. The centroid of each marker image is then triangulated from multiple cameras to estimate its position in the Cartesian coordinate. The position data are completely anonymous and will be used as a dissemination material.</p> <p>The dataset will consist of repetitions of the same motions from multiple pairs of participants. Each pair will perform approximately 10 repetitions of a given movement scenario. Data will be collected from 10 – 20 pairs of participants.</p> <p><i>Data Format:</i> PNG, JPG for images, XML or TXT for annotations</p> <p><b>To whom could the dataset be useful</b></p> <p>Roboticians, biomechanists, ergonomic designers.</p> <p><b>Related scientific publication(s)</b></p> <p><i>Not Available</i></p> <p><b>Indicative existing similar datasets</b></p> <p>CMU Graphics Lab Motion Capture Database Multisensor-Fusion for 3D Full-Body Human Motion Capture</p>	

<b>3.</b>	<b>Standards and metadata</b>
<p>The marker position data obtained from the recording of human participants will be processed in Matlab and then converted into a c3d file format (<a href="http://www.c3d.org">www.c3d.org</a>). The c3d format is a public domain, binary file supported by most of major motion capture system and animation software. The anonymized files will become available with general information about the file including participant's gender, age group, and a short description of movements being performed.</p>	
<b>4.</b>	<b>Data sharing</b>
<p><b>Access type</b></p> <p>In accordance with the ethical requirement regarding data obtained from human participants, anonymized dataset will be available to a restricted group. Personal information regarding the participants will be kept strictly private.</p> <p>The description of data may be publically disseminated in a form of publication. Published data including articles, book chapters, and conference proceedings are available in print or electronically from publishers, subject to subscription or printing charges. The source codes will be retained at the local site, open to access by a restricted group (e.g. consortium), subject to privacy, confidentiality, and intellectual property right policy of the developer(s) with respect to the local national registrations.</p> <p><b>Access Procedures</b></p> <p>The request form of the raw data may be submitted to the principal investigator of the developing site, and upon approval, the data will be electronically transferred.</p> <p>Published materials may be accessed from the publishers, subject to subscription or printing charges.</p> <p><b>Embargo periods</b></p> <p>None</p> <p><b>Technical mechanisms for dissemination</b></p> <p>A standard publication procedure is taken for dissemination.</p> <p><b>Necessary S/W and other tools for enabling re-use</b></p> <p>The dataset will be stored as MATLAB, c3d, and QTM (Qualisys Tracking Manager) files.</p> <p><b>Repository where data will be stored (institutional, etc., if already existing and identified)</b></p> <p>The dataset will be accommodated at the data management portal of RAMCIP. The network repository will also be used to host all relevant materials at the local institutes where all data are periodically backed up.</p> <p>The published materials will be hosted by the publishers.</p>	
<b>5.</b>	<b>Archiving and preservation</b>
<p><b>Data preservation period</b></p> <p>The datasets for publication will be stored up to 3 years at the local site following</p>	

publication for published material.

***Approximated end volume of data***

The dataset's end volume is expected to be 100-500 megabytes

***Indicative associated costs for data archiving and preservation***

A dedicated hard drive will be used to preserve the dataset. It is estimated to be around 100 euros.

***Indicative plan for covering the above costs***

The cost will be covered by the local hosting institute

**6. Partners activities and responsibilities**

***Partner Owner / Data Collector***

TUM

***Partner in charge of the data analysis***

TUM, CERTH, LUM

***Partner in charge of the data storage***

TUM

***WPs and Tasks***

The data are going to be collected within activities of WP6 in Task 6.3

### **3.10 Dataset "DS10.ACCREA-02.ManipKinematicsDataset"**

#### **General Description**

Set of Simulink/CAD/Gazebo models for simulation, optimization and development purposes. The prepared dataset will contain models of selected manipulator kinematics, which will allow RAMCIP partners to choose the best solution for user requirements and dexterous manipulation tasks.

<b>1.</b>	<b>Data set reference and name</b>
<b>DS10.ACCREA.02 Manipulator kinematics chains Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>Anthropomorphic kinematics of previously developed by ACCREA and commercial manipulators will be considered, as well as new concept ones. Our goal is to select solutions meeting user and safety requirements and also being capable of performing RAMCIP manipulation tasks.</p> <p><b>Nature and scale of data</b></p> <p>Dataset will be available in a form of Simulink/Solid Works/Gazebo/URDF files package.  <i>Data Format:</i> SolidWorks / URDF file format</p> <p><b>To whom could it be useful</b></p> <p>The collected data will be used for simulations and development of the RAMCIP manipulator. Models could be imported into the Gazebo environment simulation, which will be used for testing components and system integration by most of technical RAMCIP partners.</p> <p><b>Related scientific publication(s)</b></p> <p>None</p> <p><b>Indicative existing similar data sets</b></p> <p>Several commercial manipulators' kinematics will be considered in the design of the most suitable one for the RAMCIP project.</p>	
<b>3.</b>	<b>Standards and metadata</b>
<p>The possible metadata to be produced can be summarized as follows:</p> <ul style="list-style-type: none"> <li>• the parsing routines used to read and absorb the data for developing purposes</li> <li>• selection of different object grasping/manipulating scenarios based on RAMCIP requirements along with results of simulations.</li> </ul>	
<b>4.</b>	<b>Data sharing</b>

<p><b>Access type</b></p> <p>The entire dataset or part of it can be publicly available. It could be uploaded to the main site of the RAMCIP project.</p> <p><b>Access Procedures</b></p> <p>A web page will be created on the project's data management portal that should provide a description of the dataset and link to a download section.</p> <p><b>Embargo periods</b></p> <p>None</p> <p><b>Technical mechanisms for dissemination</b></p> <p>A link to the dataset from the data management portal. The link will be provided in all relevant RAMCIP publications.</p> <p><b>Necessary S/W and other tools for enabling re-use</b></p> <p>The dataset will be designed to allow easy reuse with commonly available tools and software libraries.</p> <p><b>Repository where data will be stored</b></p> <p>The dataset will be accommodated at the data management portal of RAMCIP, being accessible through the RAMCIP website.</p>	
<b>5.</b>	<b>Archiving and preservation</b>
<p><b>Data preservation period</b></p> <p>The dataset will be preserved online for as long as there are regular downloads. After that it would be made accessible by request.</p> <p><b>Approximated end volume of data</b></p> <p>The data is expected to be several hundred of Megabytes.</p> <p><b>Indicative associated costs for data archiving and preservation</b></p> <p>Probably a dedicated hard disk drive will be allocated for the dataset. There are no costs associated with its preservation.</p> <p><b>Indicative plan for covering the above costs</b></p> <p>The cost will be covered at the local hosting institute in the context of RAMCIP.</p>	
<b>6.</b>	<b>Partners activities and responsibilities</b>
<p><b>Partner Owner / Data Collector</b></p> <p>ACCREA</p> <p><b>Partner in charge of the data analysis</b></p> <p>ACCREA, SHADOW</p> <p><b>Partner in charge of the data storage</b></p> <p>ACCREA</p>	

***WPs and Tasks***

The data are going to be collected within activities of WP7

### **3.11 Dataset "DS11.LUM\_ACE-01.UserRequirementsDataset"**

#### **General Description**

The user requirement dataset comprises the pictures and videos taken during the workshops with stakeholders in Lublin and Barcelona as well as anonymized questionnaires, which were filled in by the different stakeholders groups. Since the raw data are collected in the local languages, the videos and summary of the collected data have to be translated into English

<b>1.</b>	<b>Data set reference and name</b>
	<i>Identifier for the data set to be produced</i>
<b>DS11.LUM_ACE.01 User Requirements Dataset</b>	
<b>2.</b>	<b>Data set description</b>
<p><b>Origin of Data</b></p> <p>Materials collected for and during workshops conducted at LUM and ACE with medical personnel and caregivers. The surveys were conducted by LUM and ACE teams.</p> <p><b>Nature and scale of data</b></p> <ul style="list-style-type: none"> <li>• Videos which were taken during the workshops with medical personnel and caregivers – in local languages.</li> <li>• Pictures taken during the workshops</li> <li>• Transcripts of videos and summary in English</li> <li>• Completed questionnaires – paper versions and scans – in local languages</li> <li>• Informed consents of the workshop participants in local languages – paper versions and scans.</li> <li>• Excel sheets and summary of the survey results</li> </ul> <p><i>Data Format:</i> MPG, AVI format for videos, JPG for images, DOC/PDF for transcripts, questionnaires and papers, XLS for survey results.</p> <p><b>To whom could it be useful</b></p> <p>Raw data – videos and questionnaires in local languages can be assessed and summarized by local LUM and ACE teams in the scope of user requirements analysis and definition of the RAMCIP use cases. These data should also be available for the local Ethics Committees on their requests.</p> <p>Some videos and pictures may be used for publications and presentations.</p> <p>The transcripts, tables and summaries can be used by the entire RAMCIP consortium for a preparation of the functional and technical specifications.</p> <p><b>Related scientific publication(s)</b></p> <p>The summaries of the dataset can to be published as user requirements analysis - related publication. Some pictures can be also part of the scientific publications.</p> <p><b>Indicative existing similar data sets</b></p>	

No similar data sets are available for public.

### 3. Standards and metadata

The dataset will be accompanied by the metadata describing the demographics of the samples from which the questionnaires were collected and the data collection process will be described analytically.

The results of the workshops are going to be described and categorized. The results of the questionnaires will be shown in an Excel data sheets with the statistical analysis will be conducted.

### 4. Data sharing

#### ***Access type (widely open, restricted to specific groups, private)***

Based on the ethical rules and legal requirements, the data which contain personal data such as images of people and their opinions, cannot be available for public. Summaries of the data can be published as public deliverables and scientific publication.

#### ***Access Procedures***

The datasets with personal data of the workshop participants (videos, pictures, informed consents) will be stored at the special locked cabinet (paper) or servers (videos, pictures and scans) at LUM and ACE and only the members of the RAMCIP team will have access to them.

#### ***Embargo periods (if any)***

None

#### ***Technical mechanisms for dissemination***

The summaries of the data will be published as user requirements in the appropriate deliverables and scientific publications. Some videos and pictures can be part of the scientific publications and presentations, but for dissemination of the videos and pictures, the written confirmation of LUM or ACE (depend on where the data has been recorded) is needed to ensure that the publication does not violate personal rights of the participants of the workshops.

#### ***Necessary S/W and other tools for enabling re-use***

N/A

#### ***Repository where data will be stored (institutional, etc., if already existing and identified)***

LUM's and ACE's internal servers for electronic data and locked cabinets at LUM and ACE for paper documents.

### 5. Archiving and preservation (including storage and backup)

#### ***For how long should the data be preserved?***

In Poland the videos and pictures will be kept for 5 years after the end of the project and the paper documentations have to be kept for 20 years after the end of the project as required by the local regulations.

In Spain there is no time limits for how long data should be kept. Therefore all

source data will be kept as long as possible.

***Approximated end volume of data***

Videos and pictures – 8 GB.

Informed consents – 18 pages

Questionnaires – 789 pages

***Indicative associated costs for data archiving and preservation***

No additional costs if kept on LUM and ACE servers and spaces.

**6. Partners activities and responsibilities**

***Partner Owner / Data Collector***

LUM, ACE

***Partner in charge of the data analysis***

LUM, ACE

***Partner in charge of the data storage***

LUM, ACE

***WPs and Tasks***

The collection of this dataset and its analysis is part of WP2 activities, concerning the research efforts of Task T2.1 in the scope of user requirements analysis.

## 4. Discussion

One of the main concerns of our project is to make the datasets or the portions of them that can become public, easily discoverable and accessible. In most cases, a published scientific paper will introduce a new dataset so that the community can learn about it, and later evaluate and refer to it. In this case the dataset will be also related with a certain DOI. Publishing a scientific paper along with a new dataset not only helps in making the dataset known to a wider community, but also the peer review process ensures about its reliability and quality of the context.

All the metadata and various file formats used will adhere to commonly used practices as much as possible, including commonly used software, while the description will ensure clarity and ease of use by third parties.

The datasets will be announced on the project's website with extensive description and with a download link; the RAMCIP "data management portal" of the project's website will be responsible to enable such a centralized repository. However, when a dataset cannot be publicly available, it will be accessible to the members of the consortium only via internal servers.

As it can be seen from the analysis above, the total space required would be on the order of several TB. In order to balance the period of the datasets availability and the preservation costs, we will keep the public datasets available on dedicated servers (being accessible through a the Data management portal), for as long as there is sufficient demand for them, under specific licensing schemas which will be defined at subsequent project phases, when the datasets will be established. In a later time, they will be distributed only by request.

Lastly, we will secure any personal data of healthy controls and patients involved in the data acquisition from being publicly leaked, while anonymity will be exercised in all cases.

## 5. Conclusions

In this first version of the DMP we have performed a detailed analysis of the datasets that the partners of the RAMCIP project plan to collect and use, toward developing the various skills of the RAMCIP robot. Foreseen datasets contain ones from captured models of objects and domestic environments through to human tracking, and behavioral modeling, as well as questionnaires related to the analysis of the RAMCIP user requirements. Each dataset was separately analyzed, with emphasis given on the nature of the data, the accessibility and its possible access type, as well as any ethical issues that may arise from manipulating sensitive personal information. This deliverable will serve as a guide to build the infrastructure for efficiently managing, storing and distributing the amount of data collected, especially concerning the portions of the RAMCIP datasets that will be made publicly available.

It should be finally noted that as this deliverable has been drafted during the first period (first six months) of the project, its content concerning the description of the RAMCIP datasets depicts the plans of the RAMCIP consortium at this early stage, whereas further, more detailed information over the RAMCIP datasets will be provided in the subsequent versions of the RAMCIP DMP deliverable. The RAMCIP DMP will be a living document, continuously updated till end of the project, as part of the RAMCIP WP9 activities.

## **Annex I: Dataset Identification Template**

Below is the template document that was used by each partner of the project to describe the datasets that will be created during the research period of the project RAMCIP.



***PHC-19-2014: Advancing active and healthy ageing with ICT:  
service robotics within assisted living environments***

*Project Title:*

**Robotic Assistant for MCI Patients at home**



**RAMCIP**

**Grant Agreement No: 643433**

**Research and Innovation Action (RIA)**

**Data Identification Template**

***Internal Report***

<b>Title</b>	<b>Data Identification Template (to be filled in by all partners)</b>
<b>Authors</b>	<b>ALL PARTNERS</b>
<b>Status:</b>	<b>F</b>
<b>Distribution:</b>	<b>All Partners</b>
<b>Document ID</b>	<b>RAMCIP_DataIdentificationTemplate_PARTNER-NAME.doc</b>



This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°643433

## Executive Summary

This “data identification form” template of RAMCIP has been build on the basis of the EC document on “Guidelines on Data Management on Horizon 2020”<sup>1</sup> The structure of the data identification form follows the template provided in Annex I of the Horizon 2020 Data Management Guidelines.

The purpose of this template is to facilitate each partner of the RAMCIP project in identifying and properly describing the dataset(s) that are planned to be generated during the project, toward the development of the project’s overall Data Management Plan.

---

<sup>1</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)

**The table(s) below should be filled in by each partner. At the title of each section, guidelines for the field are provided in italics. One table instance should be filled in for each expected data set.**

## Dataset "PARTNER-DX.DatasetName"

<b>1.</b>	<p><b>Data set reference and name</b></p> <p><i>Identifier for the data set to be produced</i></p>
<b>2.</b>	<p><b>Data set description</b></p> <p><i>Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.</i></p> <p><b>General Description</b></p> <p><b><i>Origin of Data (e.g. indicative collection procedure, devices used etc.)</i></b></p> <p><b><i>Nature and scale of data</i></b></p> <p><b><i>To whom could the dataset be useful</i></b></p> <p><b><i>Related scientific publication(s)</i></b></p> <p><b><i>Indicative existing similar data sets (including possibilities for integration and reuse)</i></b></p>
<b>3.</b>	<p><b>Standards and metadata</b></p> <p><i>Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.</i></p>
<b>4.</b>	<p><b>Data sharing</b></p> <p><i>Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).</i></p> <p><i>In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).</i></p>

<p><b><i>Access type (widely open, restricted to specific groups, private)</i></b></p> <p><b><i>Access Procedures</i></b></p> <p><b><i>Embargo periods (if any)</i></b></p> <p><b><i>Technical mechanisms for dissemination</i></b></p> <p><b><i>Necessary S/W and other tools for enabling re-use</i></b></p> <p><b><i>Repository where data will be stored (institutional, etc., if already existing and identified)</i></b></p>	
<p><b>5.</b></p>	<p><b>Archiving and preservation (including storage and backup)</b></p> <p><i>Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.</i></p>
<p><b><i>For how long should the data be preserved?</i></b></p> <p><b><i>Approximated end volume of data</i></b></p> <p><b><i>Indicative associated costs for data archiving and preservation</i></b></p> <p><b><i>Indicative plan for covering the above costs</i></b></p>	