Deliverable Number: D7.3, version: 1.5

Data Management Plan

CAREGIVERSPRO MMD

CAREGIVERSPRO-MMD PROJECT

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

COOSS / Ricerca e formazione

Mobiles Dynamics

FUB UNIVERSITAT A MANRESA

UNIVERSITY OF Hull

CHU Hôpitaux de Rouen

Q-PLAN INTERNATIONAL

CERTH CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 690211"

Document information

Project Number	690211	**Acronym**	CAREGIVERSPRO-MMD
Full title	Self-management interventions and mutual assistance community services, helping patients with dementia and caregivers connect with others for evaluation, support and inspiration to improve the care experience		
Project coordinator	Universitat Politècnica de Catalunya- BarcelonaTech Prof. Ulises Cortés, ia@cs.upc.edu		
Project URL	http://www.caregiversprommd-project.eu		

Deliverable	**Number**	D7.3	**Title**	Data Management Plan - first version
Work package	**Number**	WP7	**Title**	Dissemination, Communication, Exploitation and Business Planning

Date of delivery	**Contractual**	01/05/2016	**Actual**	30/04/2016
Nature	Report ☑ Demonstrator ☐ Other ☐			
Dissemination Level	Public ☑ Consortium ☐			
Keywords				

Authors (Partner)	Atia Cortés (UPC), Cristian Barrué (UPC), Ulises Cortés (UPC)			
Responsible Author	Cristian Barrué		**Email**	cbarrue@cs.upc.edu
	Partner	UPC	**Phone**	+ 34 93 413 40 11

Document Version History

Version	Date	Status	Author	Description
0.1	06-03-2016	Draft	Atia Cortés (UPC)	Start of the document, TOC, review and first structure
1.0	15-03-2016	Draft	Atia Cortés (UPC)	Contribution to different sections
1.1	27-03-2016	Draft	Cristian Barrué (UPC)	Contribution to different sections
1.2	25-03-2016	Draft	Cristian Barrué, Gabriel Verdejo (UPC)	Contribution to section 6
1.3	19-03-2016	Draft	Kevin Paulson (Hull), Atia Cortés (UPC), Dimitrios Daskalakis (QPL), Anastasia Matonaki (QPL)	Review of the document
1.4	20-04-2016	Draft	Ulises Cortés (UPC), Rafa de Bofarull (MDD)	Review of the document
1.5	26-04-2016	Final	Cristian Barrué (UPC)	Integration of review inputs, final contributions

Executive summary

This is a live document that describes the different processes regarding data management, storage and exploitation that have to be agreed and adopted by every member of the CAREGIVERSPRO-MMD Consortium. Over the course of the project this document will be reviewed and updated. Additional information on the data structure or the methodology, a change in responsibility for a task or in the budget, may be included in future versions of the Data Management Plan.

List of Acronyms

Acronym	Title
CERIF	Common European Research Information Format
C-MMD	CAREGIVERSPRO-MMD
DMP	Data Management Plan
DoA	Description of Action
HONCode	Health On the Net Code
QA	Quality Assurance
QC	Quality Control

List of Tables

Table of contents

1 Introduction

This document presents the first version of the Data Management Plan (DMP) for the CAREGIVERSPRO-MMD project. Projects funded by in the Horizon 2020 Open Research Data Pilot are required to develop several versions of a Data Management Plan (DMP), in which they will specify, among other things, what data will be kept for the longer term. In the case of CAREGIVERSPRO-MMD, which is not participating in the Open Research Data Pilot, the DMP is presented on a voluntary basis as a tool that can improve pilot preparation and result analysis. The Consortium will follow the guidelines described in OpenAire[1] platform and the document "*Guidelines on Data Management in Horizon 2020*"[2] . A DMP describes the data management life cycle for all datasets to be collected, processed or generated by a research project. It must cover:

- the handling of research data during & after the project;
- what data will be collected, processed or generated;
- what methodology & standards will be applied;
- whether data will be shared /made open access & how;
- how data will be curated & preserved.

The Data Management Plan will be updated - if appropriate - during the project lifetime (in the form of deliverables D7.7 and D7.8). New versions of the DMP could also be created whenever significant changes arise in the project such as:

- new data sets;
- changes in consortium policies;
- external factors.

2 Project Information

In this section we provide a brief fact sheet of the project details and associated data management requirements

Table 1 Project Fact Sheet

Project Title	CAREGIVERSPRO-MMD
Project Duration	36 months (01/01/16-31/12/18)
Partners	• Universitat Politècnica de Catalunya (UPC, Spain) • Mobile Dynamics (MDD, Spain) • University of Hull (HUL, UK) • Q-PLAN International LTD (QPL, Greece)

[1] https://www.openaire.eu/opendatapilot-dmp

	• Cooperativa Sociale COOSS Marche (COO, Italy) • Fundació-Universitat del Bages (FUB, Spain) • Centre Hopitalier Universitaire de Rouen (CHU, France) • Center for Research and Technology Hellas (CERTH, Greece)
Brief Description	Self-management interventions and mutual assistance community services, helping patients with dementia and caregivers connect with others for evaluation, support and inspiration to improve the care experience
University Requirements for Data Management	UPC is responsible for allocating data in safe environment, maintaining back-ups and processing the data generated
Funding Body	European Commission (Horizon2020 PHC-25-2105)
Grant Number	690211
Budget	4.087.198,75€
Funding Body Requirements for Data Management	For Open Data projects, the ones specified in Guidelines on Data Management in Horizon 2020[2] .

[2]https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

3 Data, Materials, Resources Collection Information

The purpose of this section is to provide a full description of the data that will be generated and stored during this project. The information provided in here might be adapted or updated in further versions of this document.

3.1 Description of the data

All data will be generated through the use of the CAREGIVERSPRO-MMD online platform by several categories of users, *i.e.* health professionals, caregivers and patients. Each category of user will have access to specified content and will be able to generate different types of information according to the permissions granted.

For each user of the platform, different datasets described in this section may be generated. Additional datasets may be generated in the future. The data will be collected before and after the pilot phase of the project.

The platform will also provide means to assess and store data no directly produced by users *i.e.* the interaction among users and the evolution on their activity in the social network, which will also be subject to further analysis.

3.1.1 Personal Dataset

Table 2 Personal Dataset

Data set reference and name
C-MMD-Personal
Data set description
This data set contains all the personal data captured through the registration tools integrated in the C-MMD platform for the dyad (patient and caregiver) and the health professionals. The registration tool collects standard personal information. i.e. as described in EU Data Protection Directive (95/46/EC)[3]: *"Personal data" shall mean any information relating to an identified or identifiable natural person ('Data Subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*. Therefore, the nature of the data corresponds to the values used to represent such concepts (e.g. text, integers). At this moment the registering tool has not been implemented in its final version, further details will be given in future versions.
Standards and metadata
Data will be stored each time a user (be it patient, caregiver or health professional) registers to the platform or modifies their profile. Although at this moment the registering tool and

[3] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

profile management tool have not been defined yet, it is expected that data will be stored in a MySQL database, using noSQL database for complementary purposes. Records will also be related (and identified) with other datasets and the date when the data was recorded. Metadata will include information about the profile creation time, range of possible values, etc. This metadata will be associated to each table and will follow the Common European Research Information Format (CERIF) metadata standard[4].
Data sharing
This dataset will not be shared outside of the Consortium boundaries for ethical and security reasons. Each dataset record belongs to the user and to the Consortium partner responsible for the user. Only the user, people authorised by him/her (*e.g.* caregiver) and authorised personnel of the Consortium partner responsible for the user, can access the record. Data will be available to users and people authorised by them through the C-MMD platform. Authorised personnel of the pilot partner generating the data will be able to access aggregated data in periodic reports and also will be able to access raw data dumped from the database in *csv* files or through a web service. Each access will be identifiable and traceable. Dataset records will be shared among defined Consortium partners anonymised for research purposes in order to be used for the tasks of the project. Anonymisation is the standard procedure followed to preserve confidentiality of participants. Each participant (*e.g.* patient, caregiver, doctor) will sign an informed consent at recruitment phase authorizing access to all his/her data (raw, aggregated, anonymised). Users will agree to the anonymised and aggregated data being used for research and possibly commercial exploitation. The data repository will be in the C-MMD host in the UPC premises (more details are given in section 6).
Archiving and preservation (including storage and backup)
See section 6 and 7.

3.1.2 Screening Dataset

Table 3 Screening Dataset

Data set reference and name
C-MMD-Screening
Data set description
This data set contains all the clinical and social data captured through the screening tools integrated in the C-MMD platform for the dyad (patient and caregiver). The screening tools

[4] http://www.eurocris.org/cerif/main-features-cerif

implement standard evaluation scales for different conditions (physical, psychosocial, neurological, functional, *etc*.). Therefore, the nature of the data corresponds to the values used to evaluate such scales. At this moment the screening tool has not been implemented yet, further details will be given in future versions.
Standards and metadata
The data will be stored following the standard numeric scales defined by each screening tool each time that a user (be it patient, caregiver or health professional) uses one of the screening tools. Although at this moment the screening tool has not been defined it is expected that data will be stored in a MySQL database, using noSQL database for complementary purposes. Records will also be related (and identified) with the user to which the recorded data belong and the date when the data was recorded. Metadata will include information about the scale recorded, range of possible values, etc. This metadata will be associated to each table and will follow the Common European Research Information Format (CERIF) metadata standard[5].
Data sharing
This dataset will not be shared outside of the Consortium boundaries for ethical and security reasons. Each dataset record belongs to the user and to the Consortium partner responsible for the user. Only the user, people authorised by Him/her (*i.e.* caregiver) and authorised personnel of the Consortium partner responsible for the user can access the record. Data will be available to users and people authorised by them through the C-MMD platform. Authorised personnel of the pilot partner generating the data will be able to access aggregated data in periodic reports and also will be able to access raw data dumped from the database in csv files or through a web service. Each access will be identifiable and traceable. Dataset records will be shared among the Consortium partners anonymised for research purposes in order to be used in the tasks of the project. Anonymisation is the standard procedure followed to preserve confidentiality of participants. Each participant will sign an informed consent at recruitment phase authorizing access to all his/her data (raw, aggregated, anonymised). Users will agree to the anonymised and aggregated data being used for research and possibly commercial exploitation. The data repository will be allocated in the C-MMD host in the UPC premises (more details in section 6).
Archiving and preservation (including storage and backup)
See section 6 and 7.

[5] http://www.eurocris.org/cerif/main-features-cerif

3.1.3 Treatment Dataset

Table 4 Treatment Dataset

Data set reference and name
C-MMD-Treatment
Data set description
This dataset contains all the treatment information for each dyad. The treatment information will come from: (1) a specific toolset integrated in the platform for that purpose, (2) through the API to connect with national healthcare systems where possible. The nature of the data corresponds to medication descriptions, doses, schedules and follow-up of the adherence. At this moment the data-capturing tool has not been implemented, further details will be given in future versions.
Standards and metadata
The data will be stored following the numeric/text standards each time that a user (be it patient, caregiver or health professional) uses the treatment management interface to introduce or modify information about the pharmacological treatment being followed and the adherence regime to the treatment. Although at this moment the treatment management tool has not been defined it is expected that data will be stored in a MySQL database, using noSQL database for complementary purposes. Records will also be related (and identified) with the user to which the recorded data belong and the date when the data was recorded. Metadata will include information about the data recorded, range of possible values, etc. This metadata will be associated to each table and will follow the Common European Research Information Format (CERIF) metadata standard[6].
Data sharing
This dataset will not be shared outside of the Consortium boundaries for ethical and security reasons. Each dataset record belongs to the user and to the Consortium partner responsible for the user. Only the user, people authorised by him/her (*i.e.* the caregiver) and authorised personnel of the Consortium partner responsible for the user, can access the record. Data will be available to users and people authorised by them through the C-MMD platform. Authorised personnel of the pilot partner generating the data will be able to access aggregated data in periodic reports and also will be able to access raw data dumped from the database in *csv* files or through a web service. Each access will be identifiable and traceable. Dataset records will be shared among the Consortium partners, anonymised for research purposes, in order to achieve with the tasks of WP6. Anonymisation is the standard procedure followed to preserve confidentiality of participants. All described accesses to data (raw, aggregated, anonymised) will be authorised though an informed consent signed by the participant at the recruitment phase. Users will agree to the

[6] http://www.eurocris.org/cerif/main-features-cerif

anonymised and aggregated data being used for research and possibly commercial exploitation. The data repository will be allocated in the C-MMD host in the UPC premises (more details in section 6).
Archiving and preservation (including storage and backup)
See section 6 and 7.

3.1.4 Intervention Dataset

Table 5 Intervention Dataset

Data set reference and name
C-MMD-Intervention
Data set description
This data set contains all the intervention contents created by the consortium members during the lifetime of the project. These intervention contents include posts, articles, tips, multimedia, tutorials, webinars and any kind of educational content produced to support the caregiving process and the healthy ageing lifestyle. These intervention contents will be introduced in the platform through specific tools designed for that purpose (*e.g*. the ones available in Wordpress to edit blog posts). Standards in multimedia and text posts storage will be followed. At this moment the editor tools have not been implemented, further details will be given in future versions.
Standards and metadata
The data will be stored following the standard text/media formats following best practices for data management (see section 6). Although at this moment the editing tool has not been defined, it is expected that data will be stored in a MySQL database, using noSQL database for complementary purposes. Records will also be related (and identified) with the user authoring the contents and the date when the data was recorded. As explained in section 5.1 of DoA and later in this document in section 4, all contents created will follow the HONCode. Metadata will include information about the intervention recorded and a list of tags or keywords that relate the content with specific symptoms, conditions or problems that the content refers to (*e.g.* a video about Alzheimer could have the tags *Alzheimer*, *dementia*, *cognitive decline*, etc.) This metadata will be associated to each table and will follow the Common European Research Information Format (CERIF) metadata standard[7].
Data sharing

[7] http://www.eurocris.org/cerif/main-features-cerif

Each dataset record belongs to the Consortium partner responsible for creating it. All the Consortium and suitable users[8] are authorised to access the recorded contents. Data will be available to users and people authorised by them through the C-MMD platform. Aggregated data about the amount of contents generated and specific metadata (*e.g.* tags) will be available as well as access to raw data dumped from the database in files to selected Consortium members. Dataset records, particularly aggregated data, will be shared among the Consortium partners for research purposes in order to be used in the tasks of the project. Users will agree to the anonymised and aggregated data being used for research and possibly commercial exploitation. The data repository will be allocated in the C-MMD host in the UPC premises (more details in section 6).
Archiving and preservation (including storage and backup)
See section 6 and 7.

3.1.5 Dissemination Dataset

Table 6 Dissemination Dataset

Data set reference and name
C-MMD-Dissemination
Data set description
This data set contains all the dissemination contents created by the consortium members during the lifetime of the project. These dissemination contents include scientific papers, newsletters, multimedia, press articles, conferences and any kind of dissemination content produced to support the communication activities of the project and dissemination of results. These contents created from different sources will be stored in a database/filesystem.
Standards and metadata

[8] In the case of patients or caregivers, contents should be available depending on their specific needs

The data will be stored following the standard text/media formats following best practices for data management (see section 6). Records will also be related (and identified) with the user authoring the contents and the date when the data was recorded. Metadata will include information about the dissemination data recorded, the target audience, identifier (*i.e.* DOI, URI), authors, title of the publication, time of publication, related event (*e.g.* conference, forum, *etc.*) and a list of tags or keywords that relate the content with specific topics or results. This metadata will be associated to each table and will follow the Common European Research Information Format (CERIF) metadata standard[9].
Data sharing
Each dataset record belongs to the Consortium partner/s responsible for creating it. These contents are open for access. The data repository will be allocated in the C-MMD host in the UPC premises (more details in section 6).
Archiving and preservation (including storage and backup)
See section 6 and 7.

3.1.6 Dataset Summary

Table 7 Dataset Summary

Dataset	Who	Ownership	Access
Personal Dataset	User	Yes	Yes , full
	Partner (recruiting)	Yes	Yes, full to authorised personnel
	Rest of Consortium	No	Yes, only anonymised and aggregated data
	World	No	No
Screening Dataset	User	Yes	Yes , full
	Partner (recruiting)	Yes	Yes, full to authorised personnel
	Rest of Consortium	No	Yes, only anonymised and aggregated data

[9] http://www.eurocris.org/cerif/main-features-cerif

	World	No	No
Treatment Dataset	User	Yes	Yes , full
	Partner (recruiting)	Yes	Yes, full to authorised personnel
	Rest of Consortium	No	Yes, only anonymised and aggregated data
	World	No	No
Intervention Dataset	User	No	Yes, depending on their needs
	Partner (authoring)	Yes	Yes, full to authorised personnel
	Rest of Consortium	No	Yes, full to authorised personnel
	World	No	Limited and depending on project needs and exploitation policies
Dissemination Dataset	User	No	Yes
	Partner (authoring)	Yes	Yes
	Rest of Consortium	No	Yes
	World	No	Yes

3.2 Quality Assurance Process

Every data gathering process is susceptible to contamination in the absence of adequate preventive measures. Data contamination results from a process or phenomenon, other than the one of interest, which can affect the variable values. Data contamination results in erroneous values in the data set. In general, there are two types of errors that can occur in a data set. Firstly, errors of commission, which are the result of incorrect or inaccurate data being included in the data set. This may happen because of a malfunctioning instrument that produces faulty results, data that are mistyped during entry, or other problems.

Errors of omission are the second type of errors. These result from data or metadata being omitted. Situations that result in omission errors occur when data are inadequately documented, when there are human errors during data collection or entry, or when there are anomalies in the field that affect the data.

Quality assurance/quality control (QA/QC) activities should be an integral part of any inventory development processes as they improve transparency, consistency, comparability, completeness and accuracy.

Quality control (QC) is defined as a system of checks to assess and maintain the quality of the data inventory being compiled. Quality control procedures are designed to provide routine technical checks to measure and control the data consistency, integrity, correctness and completeness; and to identify and address errors and omissions. Quality control checks should cover everything from data acquisition and handling, application of approved procedures and methods, and documentation. Examples of general quality control checks include:

- checking for transcription errors in data input;
- checking that scale measures are within the range of acceptable values;
- checking that proper conversion factors are used;

In future versions of this document we will provide more details on the QC protocols to be adopted during the project lifetime.

Quality assurance (QA) is a planned system of review procedures conducted outside the actual inventory compilation by personnel not directly involved in the inventory development process. It is a non-biased, independent review of methods and/or data summaries that ensures that the inventory continues to incorporate correctly the scientific knowledge and data generated. Quality assurance procedures may include expert peer reviews of data summaries and audits to assess the quality of the inventory and to identify where improvements could be made. If deemed necessary, selected members of the Advisory Board may perform this task in the course of the project lifecycle.

4 Ethics, Intellectual Property, Citation

4.1 Ethics

The lack of ethical principles standardization at international level may potentially lead to the abuse of data collection, use and storage by exploiting differences between societies with regard to established ethical standards. Ethics of data collection, and data use and storage in medical applications, is of growing importance since the quality and quantity of medical data usage is growing quickly both in Europe and worldwide. Great concerns are raised about data protection and privacy issues in the area of biometric and health applications with growing markets that might be affected by insufficiently protected sensitive information.

The healthcare providers that are involved in the project follow strict ethical codes. All ethical, legal and regulatory issues will be studied in detail in T8.6 and presented in the incremental versions of D8.3. The most relevant findings will be included in the final version of this document.

4.2 Intellectual Property

With regard to property and ownership of medical data and records, there are two distinct views. From the standpoint of practitioners (i.e., healthcare providers, hospitals), patient medical records are their property because they are the ones who write, compile and produce the records (data producers). At the same time, patients tend to believe that medical records belong to them as they provide the relevant information.

Nevertheless, the project will produce data assets that do not correspond to medical records. For instance:

- Intervention contents and guidelines;
- Gamification reports;
- Treatment adherence reports;
- Aggregated medical data reports; and
- Reports and statistics of platform usage.

The ownership and IPR of these assets will be detailed in future versions of this document. The resulting agreements will be compliant with corresponding legislation (i.e. Data Protection Act, Copyright, Freedom of Information Act, *etc*.).

4.3 Citation

An article, paper or presentation that refers to, or draws, information from a data set should cite the data set, just as it would cite other sources such as books and articles. A citation gives appropriate credit to the data set creator(s), and allows interested readers to find the data set so they can confirm the data is being correctly represented, or can use it in their own work. There is no universal standard for formatting a data set citation.

There are many different styles for formatting citations, such as APA and Chicago Manual of Style. In addition, most scientific publications have their own style, either unique to themselves or based on an existing style. A few of these styles, such as APA 6th edition, specify how to cite data sets. However, most citation style manuals do not currently cover citing data sets. Consequently, adaptation of the styles' general format can be applied to the needs of data sets.

At this early stage, the information used to cite C-MMD data sets could be:

- Author(s) (*the principal investigator can be used as the "author" of a data set*)
- Title
- Year of Publication
- Publisher (*partner producing the dataset*)
- Version
- Access information (*doi or url*)

5 Access and Use of Information

One of the objectives of the CAREGIVERSPRO-MMD project is to develop the solution into a commercial product. This is the main reason why the Consortium has decided that potentially publishable data will not be available for open access until the end of the project, once the exploitation paths have been defined.

However, results of the pilot execution and platform evaluation will be made publicly available through the deliverables **D6.1 – Mid-Pilot preliminary analysis report, D6.2 – Final Pilot analysis report** and **D6.3 – User feedback and usability report.**

More details on specific dataset access regimes are defined in section 3.1.

6 Storage and Backup of Data

In order to safeguard the appropriate preservation of the data, portion of the budget has been allocated in the data storage and backups during the lifespan of the project and at least for the following two years.

The data will be stored in databases installed on the same server that holds the CAREGIVERSPRO-MMD platform. These Databases are only accessible locally (i.e. only available to the server itself) in order to prevent any connection from outside. The system and server configuration have been arranged in order to support local data encryption to avoid physical access to the hard disk drive. This measure would prevent access to the data if the physical storage was stolen or accessed directly.

The server has a local firewall that only allows secure web connections to the Internet and verified IP addresses for development/updates of the C-MMD application. A local log file records every access to the server.

The server is located in the UPC campus Data Center. This data center is a dedicated 250m^2 facility with controlled access, personal ID cards for authorized staff and video surveillance 24x7. The server has dedicated bandwidth and backup power system in order to guarantee availability.

A daily backup procedure has been designed in order to ensure data integrity and recovery. This backup has two main subsystems:

1. File system backup: A daily copy of every file in the file system is stored in compressed format.
2. Database backup: A daily dump of every database/table is stored in a single file.
3. Daily encryption and compression of log files.

Optionally, this backup can be physically moved to a safe location outside the UPC Data Center if the personal data requires this level of protection. A specific budget is reserved for this task.

A 30-day window backup system has been programmed and enough disk space has been reserved for a monthly operation.

6.1 Best Practices for File Formats

The file formats used have a direct impact on the ability to open those files at a later date and on the ability of other people to access those data.

6.1.1 Proprietary vs Open Formats

Data should be saved in a non-proprietary (open) file format when possible. If conversion to an open data format will result in some data loss from the files, it should be considered saving the data in both the proprietary format and an open format. Having at least some of the information available in the future is better than having none.

When it is necessary to save files in a proprietary format, it will be included a readme.txt file that documents the name and version of the software used to generate the file, as well as the company who made the software.

6.1.2 Guidelines for Choosing Formats

When selecting file formats for archiving, the formats should ideally be:

- Non-proprietary;
- Unencrypted;[10]
- Uncompressed;
- In common usage by the research community;
- Adherent to an open, documented standard:
    - Interoperable among diverse platforms and applications
    - Fully published and available royalty-free
    - Fully and independently implementable by multiple software providers on multiple platforms without any intellectual property restrictions for necessary technology
    - Developed and maintained by an open standards organization with a well-defined inclusive process for evolution of the standard

6.1.3 Some Preferred File Formats[11][12]

- Containers: TAR, GZIP, ZIP
- Databases: XML, CSV
- Geospatial: SHP, DBF, GeoTIFF, NetCDF
- Moving images: MOV, MPEG, AVI, MXF
- Sounds: WAVE, AIFF, MP3, MXF
- Statistics: ASCII, DTA, POR, SAS, SAV
- Still images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP
- Tabular data: CSV
- Text: XML, PDF/A, HTML, ASCII, UTF-8
- Web archive: WARC

[10] Data will be encrypted in the UPC server for security reasons

[11] http://www.digitalpreservation.gov/formats/

[12] http://www.loc.gov/preservation/resources/rfs/data.html

7 Archiving and Future Proofing of Information

The national legislation (European compliant) of the server site (Spain) compels UPC to preserve all data and access records for **two years** after the project completion. The server will remain in the same safe location in order to preserve physical and logical access. Consequently, the data will be kept in the server and will be accessible under the same terms that will be agreed among partners during the project lifespan.

All public project deliverables will be available at least for **five years** after the project completion at the project portal.

Selected datasets, databases, standalone documents, and even software may be made public or open for exploitation at the end of the project. These resources may prove useless without explanatory notes (metadata) accompanying them. Metadata will be clearly linked to the materials so that they can adequately inform any future user about the material. For example, a published dataset will typically be accompanied by a metadata document that explains the various fields, their usefulness and summarises the purpose of the dataset in general. These documents will be stored along with the dataset and made accessible in the same manner as the dataset (*e.g.* online, or download). Contact information will be provided accordingly in case that the future user needs further clarification.

8 Resourcing of Data Management

This section outlines the staffing and financial details of the data management within the CAREGIVERSPRO-MMD project. The former aspect provides information about the role and responsibilities of the partners that generate the data and those who control it. The latter aspect describes the financing process for data management and data storage.

8.1 Roles in Data Management

Each pilot partner (HUL, COO, FUB, CHU) is responsible for the data generated in their own pilots by the different stakeholders of the platform as **data producers**. Each pilot partner will assign a responsible person from his or her institution for this task to be designed for the next version of this document.

The UPC is responsible for all the aspects related with data storage and backup as **data processor**.

MDD and CERTH as the main developers of the C-MMD platform will be responsible as **data processor** and **service provider** of all the aspects related with data gathering, data integrity, access logging, *etc*.

As specified in section 5.1.3 of DoA, specific agreements will be signed among partners in order to grant access to the different datasets for the different uses (data storage, data processing, service provision).

8.2 Financial Data Management Process

As mentioned before, the Consortium has reserved a portion of the project budget for data hosting and backup.

9 Review of Data Management Process

The follow-up of this plan will be reported in future versions of this document, where detailed protocols and measures will be described to ensure the compliance with the plan along with preliminary results on the observed evolution. UPC as main contributor to this plan, supported by the roles described in section 8.1, will perform the follow-up.

External reviewers of the Consortium as well as selected members of the Advisory Board will support the peer-review process.

10 Statements and Personnel Details

10.1 Statement of Agreement

The Consortium agree to the specific elements of the plan as outlined.[13]

Project Coordinator

Title	
Designation	
Name	
Date	
Signature	

Project Manager

Title	
Designation	
Name	
Date	
Signature	

[13] To be signed in the next version of this document

Management Board (one table for each member)

Title	
Designation	
Name	
Date	
Signature	