



D1.2: Data Management Plan

Dissemination level: Public

Document type: Report

Version: 1.0.1

Date: February 28, 2018 (original)

March 5, 2019 (this version)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains.

Document Details

Project Number	769553
Project title	Council of Coaches
Title of deliverable	Data Management Plan
Due date of deliverable	February 28 th , 2018
Work package	WP1
Author(s)	Jorien van Loon (CMC), Rasmus Nielsen (DBT), Oresti Banos (CMC), Tessa Beinema (RRD), Randy Klaassen (CMC), Alison Pease (UDun), Dennis Reidsma (CMC), Catherine Pelachaud (UPMC), Sofoklis Kyriazakos (iSPRINT), Alvaro Fides Valero (UPV), Maria Jesus Arnal Parada (UPV).
Reviewer(s)	Rasmus Nielsen (DBT)
Approved by	Coordinator
Dissemination level	Public
Document type	Report
Total number of pages	48

Partners

- University of Twente – Centre for Monitoring and Coaching (CMC)
- Roessingh Research and Development (RRD)
- Danish Board of Technology Foundation (DBT)
- Sorbonne University (SU)
- University of Dundee (UDun)
- Universitat Politècnica de València, Grupa SABIEN (UPV)
- Innovation Sprint (iSPRINT)

Abstract

This Data Management Plan will describe how the projects handle data. Council of Coaches intends to comply with the General Data Protection Regulation (GDPR). Therefore, it has set out guidelines for the consortium how to safely and securely handle sensitive data. Each Work Package will describe the datasets that are being collected and/or processed. The Project Management has developed provisional guidelines on how to implement the FAIR principles to the datasets.



Corrections

- v1.0.1
 - Correctly applied EU logo on header page.
 - Changed UPMC to Sorbonne University (SU).
 - Numerous small textual fixes.

Table of Contents

1	Introduction	6
2	Objectives	7
3	Data Management and the GDPR.....	8
3.1	Lawfulness, fairness and transparency	8
3.2	Purpose limitation	8
3.3	Data minimisation	9
3.4	Accuracy	9
3.5	Storage limitation.....	9
3.6	Integrity and confidentiality	9
3.7	Accountability	10
4	Guidelines for data management on project level.....	11
4.1	Guidelines	11
4.1.1	Purpose limitation and data minimisation	11
4.1.2	Personal information.....	11
4.1.3	Anonymization and pseudonymisation.....	11
4.1.4	Informed consent	12
4.1.5	End users' access to data.....	12
4.1.6	Storage and researchers' access to data	12
4.1.7	Encryption.....	13
4.1.8	Open data and FAIR principles.	13
4.1.9	Privacy statements	13
4.1.10	Update DMP.....	13
5	Data management per WP.....	14
5.1	Provisional Guidelines.....	14
5.1.1	Overall	14
5.1.2	Findable.....	14
5.1.3	Accessible	14
5.1.4	Interoperable.....	15
5.1.5	Re-Use.....	15
6	WP1: Data for sound, effective and efficient project management	16
7	WP2: Data for Responsible Research and Innovation	19
8	WP3: Data for providing personalised coaching.....	23
9	WP4: Data for the development of the Holistic Behaviour Analysis Framework.....	26
10	WP5: Data for Dialogue and Argumentation Framework.....	30
11	WP6: Data for Human Computer Interfaces	33

12	WP7: Data for Continuous integration and Demonstration	36
13	WP8: Data for Dissemination and Exploitation	41
	Annex 1: European Informed Consent Form.....	44



Symbols, abbreviations, and acronyms

AAL	Ambient Assisted Living
BSS	Biomedical Signals and Systems
CMC	Centre for Monitoring and Coaching
D	Deliverable
DANS	Data Archiving and Networked Services
DBT	Danish Board of Technology Foundation
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
DSA	Data Seal of Approval
EC	European Commission
ELAN	EUDICO Linguistic Annotator
EU	European Union
EULA	End User License Agreement
FAIR	Findable, Accessible, Interoperable, Re-use
GB	Gigabyte
GDPR	General Data Protection Regulation
GPS	Global Positioning System
HMI	Human Media Interaction
ICT	Information and Communication Technology
IPR	Intellectual Property Right
ISO	International Organization for Standardization
ISPRINT	Innovation Sprint
M	Month
MS	Milestone
RRD	Roessingh Research and Development
RRI	Responsible Research and Innovation
STS	Science and Technology Studies
SU	Sorbonne University
TA	Technology Assessment
UDun	University of Dundee
UPV	Universitat Politècnica de València
UT	University of Twente
WDS	World Data System
WP	Work Package

1 Introduction

This Data Management Plan (DMP) is the second deliverable in the first Work Package (WP) of the Council of Coaches project. It will describe how the data will be handled during and after the project. The focus during the project will be on providing privacy for the personalised and medical data that will be used. After the project has ended the focus will shift towards making the data as openly accessible as possible. By means of anonymising the personalised data, the consortium will try to open up as much data as possible. However, since privacy of the patients comes first, we foresee this will not be possible for all datasets.

In May 2018, the new European Regulation on Privacy, the General Data Protection Regulation, (GDPR) will come into effect. In this DMP we will describe the measures to protect the privacy of all subjects in the light of the GDPR. We will not describe any measures on a national level before May 2018, since we the little personal data we have before May 2018, will be stored in the Netherlands at a location that is already GDPR compliant.

This DMP is a living document. At this stage in the research a lot of questions concerning the data are still open for discussion. Questions concerning opening up the data or answers to questions related to the Findable, Accessible, Interoperable, Re-use (FAIR) principles will only have a provisional answer in this DMP. We will add relevant information to the DMP as soon as it is available. An update will be provided before the end of the first period, in time for the first review. Another update will be provided before the end of the project in order to describe how the data will be made open access.

In this document, firstly objectives of the document will be described. Secondly, the relationship between the GDPR and the implementation of data management within the Council of Coaches project will be discussed. In the chapter 4 guidelines will be provided for the partners within the consortium on working with datasets that contain personal information. In the last chapters each WP will describe the dataset that will be collected or generated within the WP. Furthermore, the project management will describe the intentions we have with the datasets on following the FAIR principles and opening the datasets.

Version 1.0.0 has been submitted to the PO through the portal. In the latest version, V1.0.1 updates have been provided on the text to specify further the guidelines for the consortium, based on input from the Privacy Officer of the UT.

2 Objectives

In this deliverable the data management of the Council of Coaches Project will be described.

Discussions on Privacy by Design and Open data will be ongoing during the project.

Our aim is for the DMP to not only be a description of the datasets within the Council of Coaches project, but for the DMP to serve two additional goals:

On the one hand, it will provide guidelines for data management to all partners in the project.

On the other hand, it can serve as a tool to create awareness on the different topics of Open Access, Privacy and Personal Data, and the FAIR principles. This will help the participants to make as much of the research data Openly Accessible as possible, while staying within the boundaries of the Privacy regulations.

3 Data Management and the GDPR

As of May 2018, the GDPR will come into play. This means all partners within the consortium will have to follow the same new rules and principles. On the one hand it makes it easier for the project management to set up guidelines for the correct use of personal data. On the other hand, it means that in some cases, tools and partner specific guidelines are not yet available.

In this chapter we will describe how the founding principles of the GDPR will be followed in the Council of Coaches project. In the next chapter we will set out specific guidelines for proper use of personal data within the boundaries of the GDPR.

3.1 Lawfulness, fairness and transparency

Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject.

The council of coaches describes all handling of personal data in this DMP. Some of the answers requested can at the moment of writing not be provided for. Therefore, we chose to let the DMP be a living document. As soon as information about data sets become available, this will be updated in the DMP. Furthermore, additional updates will be provided to the Project Officer, before both reviews.

All data gathering from individuals will require informed consent of the test subjects, patients, or other individuals who are engaged in the project. Informed consent requests will consist of an information letter and a consent form. This will state the specific causes for the experiment (or other activity), how the data will be handled, safely stored, and shared. The request will also inform individuals of their rights to have data updated or removed, and the project's policies on how these rights are managed (see below).

We will try to anonymise the personal data as far as possible, however we foresee this won't be possible for all instances. Therefore, further consent will be asked to use the data for open research purposes, this includes presentations at conferences, publications in journals as well as depositing a data set in an open repository at the end of the project.

The consortium tries to be as transparent as possible in their collection of personal data. This means when collecting the data information leaflet and consent form will describe the kind of information, the manner in which it will be collected and processed, if, how, and for which purpose it will be disseminated and if and how it will be made open access. Furthermore, the subjects will have the possibility to request what kind of information has been stored about them and they can request up to a reasonable limit to be removed from the results.

3.2 Purpose limitation

Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes

The Council of Coaches project won't collect any data that is outside the scope of the project. Each researcher will only collect data necessary within their specific work package.

3.3 Data minimisation

Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed

Only data that is relevant for the project's research questions and the required coaching strategies will be collected. However, since patients are free in their answers, both when working with the Council of Coaches technology or in answering open ended research questions, this could result in them sharing personal information that has not been asked for by the project. This is normal in any coaching relationship and we therefore chose not to limit the patients in their answer possibilities. Since this data can be highly personal, it will be treated according to all guidelines on special categories of personal data and won't be shared without anonymisation or explicit consent of the patient.

3.4 Accuracy

Personal data shall be accurate and, where necessary, kept up to date.

All data collected will be checked for consistency. However, since some of the dataset register self-reporting data from the patients, we cannot check this data for accuracy. Since all data is gathered within a specific timeframe, we chose not to keep the data up to date, since it would hinder our research. However, we will try to capture the data as accurately as possible, for example "age" could be stored as "age in 2018". This will remove the necessity of keeping this information up to date.

3.5 Storage limitation

Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed

All personal data that will no longer be used for research purposes will be deleted as soon as possible. All personal data will be made anonymous as soon as possible. At the end of the project, if the data has been anonymised, the data set will be stored in an open repository. If data cannot be made anonymous, it will be pseudonymised as much as possible and stored for a maximum of the partner's archiving rules within the institution. A complete data set will be stored at the UT for project archiving for 10 years, according to UT's data policy.

3.6 Integrity and confidentiality

Personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures

All personal data will be handled with appropriate security measures applied. This means:

- Data sets with personal data will be stored at a SharePoint server at the UT that complies with all GDPR regulations and is ISO 27001 certified.
- Access to this SharePoint will be managed by the project management and will be given only to people who need to access the data. Access can be retracted if necessary.
- Data sets with personal information could further be shared through the Council of Coaches Dropbox folder, only if the datasets are sufficiently encrypted. The key to the encryption will be handed out by the project management and will be changed when access needs to be revoked.

- All people with access to the personal data files will need to sign a confidentiality agreement.
- These data files cannot be copied, unless stored encrypted on a password protected storage device. In case of theft or loss, these files will be protected by the encryption.
- These copies must be deleted as soon as possible and cannot be shared with anyone outside the consortium or within the consortium without the proper authorization.

In exceptional cases where the dataset is too large, or it cannot be transferred securely, each partner can share their own datasets through channels that comply with the GDPR.

3.7 Accountability

The controller shall be responsible for, and be able to demonstrate compliance with the GDPR.

At project level, the project management is responsible for the correct data management within the project. In the next chapter guidelines will be described for each partner to follow in case of datasets with personal data. Whether the partners follow these guidelines will be regularly checked by the project management. For each data set, a responsible person has been appointed at partner level, who will be held accountable for this specific data set. Each researcher will need to make a mention of a dataset with personal information to their Data Protection Officer, in line with the GDPR regulations.

4 Guidelines for data management on project level.

Data management is an ongoing process. The first version of the DMP will be published at month 6, when there are still many uncertainties about the data collected in the project. For many tasks, due to the nature of the research, we do not know what kind of data we will collect specifically. This results in not having a good overview of whether this data will contain personal data (or data that can be combined into personal data) or confidential parts of for example software. Therefore, we have established guidelines for data management to ensure that all researchers will keep up the principles of lawful and ethical data management and end users will be able to trust the system with their personal data.

The guidelines established in this DMP are embraced within the consortium and the project management will ensure these principles will be followed.

4.1 Guidelines

The sections below describe the ten basic guidelines that will be adopted in the Data Management process in Council of Coaches.

4.1.1 Purpose limitation and data minimisation

As soon as a researcher has identified which information to collect, the principles of purpose limitation and data minimisation come into play. Each researcher will take care not to collect any data that is outside the scope of his or her research and will not collect additional information not directly related to the goal of his research.

4.1.2 Personal information

As soon as the parameters in the data set are identified, the researchers need to indicate whether the data set will contain personal information.

In cases where the parameters themselves contains no personal information, but the various parameters can be merged to show a distinct pattern that can be linked to a specific person, the data set will be classified as containing personal information as well.

When the dataset contains personal information or otherwise information that needs to be kept confidential, the following privacy principles should be taken into account:

Sensitive data should be stored at either the dedicated SharePoint server at the UT, or encrypted on Dropbox. Preferably the SharePoint, since access management will be better implemented. However, for short terms with a limit amount of users, the Dropbox option is a viable one.

In the case of personal data collected in physical form (e.g. on paper), it shall be stored in a restricted-access area (e.g. locked drawer) where only relevant staff has access to. Once the data has been digitised, the physical copies shall be removed.

Personal data should be deleted as soon as possible.

4.1.3 Anonymization and pseudonymisation

The researcher will make sure the personal data is anonymised as quickly as possible. When the data cannot be anonymised completely, it will be pseudonymised as much as possible. The key between the pseudonymised file and the list of participants will be stored by the project management on a separate physical location from the original files. Do keep in mind that the research subjects should be able to withdraw their data within 48 hours of the experiment.

Part of the technology to be developed is a client server technology. Both the client and the server should incorporate the privacy rules as set out in the GDPR as of May 2018. At the moment we are looking into the different possibilities of hosting a server at either RRD, UT, or UPV to be sure we will abide by the rules. As far as client-side technology, we are looking into the possibilities of anonymizing the client side, so the phone on which the app runs, won't be serving as a unique identifier for the project.

However, the implications of the privacy-by-design provisions in the GDPR cannot be settled up front. As part of the Responsible Research Innovation work in WP 2, we will incorporate therefore work to include privacy by design techniques in all relevant development phases, ensuring that privacy will be maintained on all levels for our patients.

4.1.4 Informed consent

When collecting personal information, researchers are required to get informed consent from the patients. The standardized EU informed consent form template is used see Annex 1. However, this can always be supplemented with additional consent requests. For the online collection of data for the final prototype, we are looking into a privacy by design solution where the consent is an integral part of the technical development of the system.

Dissemination of this (personal) information for scientific conferences, journal papers or other dissemination items may only occur with direct consent of the patients or if the data is completely anonymised.

4.1.5 End users' access to data

The user can submit a request to see which information about him is being kept on our files through the contact person on the consent form. He can request to delete his information up to 48 hours after the experiment has taken place. Furthermore, he can request that no additional data collection of the patient will take place starting immediately from the time of request. The right to change the personal data can only be used when the data has not been processed for research purposes. This in order to keep body of the research data intact and to not change the research results. The end users have the right to know who the data will be shared with and this right will be described in the consent form.

4.1.6 Storage and researchers' access to data

Personal data will need to be stored safely and in a secure environment. This can either be the SharePoint server hosted by the UT, or a partner's own solution that complies with the GDPR. For short term transfer of data, an encrypted file might be put on Dropbox. The researcher is responsible for the correct encryption and the access management of the encryption key.

Access to this secure environment can be granted or revoked by either the researchers responsible for the data, or the project management on a case to case basis and will not be given out by default to all researchers. All users that are granted access to the SharePoint will need to sign a confidentiality agreement about the data on the server. Access can be restricted or revoked, when researchers are not complying with the guidelines or when their contract is terminated.

Backups of Dropbox and the UT's SharePoint are made every 24 hours by the system itself.

The UT's SharePoint server is a secure environment that is ISO 27001 certified.

4.1.7 Encryption

When you want to share personal data files through Dropbox, the data files will need to be encrypted. Each researcher is free to use their own preferred encryption tools, to make the process as easily available as possible. Possibilities for encryption as build in Word and Excel encryption or PGP keys.

If you keep data files with personal data on your personal computer or on a separated hard drive for data analysis purposes, you can use BitLocker or FileVault for the encryption of your hard drive.

4.1.8 Open data and FAIR principles.

Within the Council of Coaches project, we endorse the EC's motto: to make the data as open as possible, but as closed as necessary. We are committed to protect the privacy of the people involved, and the confidentiality of specific results or agreements. In these cases, the data will not be made available for public use.

In all other cases we will try our best to make the research data as broadly available as possible. This means the FAIR principles will be held, but at the moment it is not possible for us to give definitive answers on how these will be held. We intent to discuss those in more detail, once more information on the data sets comes to light.

4.1.9 Privacy statements

Actively communicate the privacy and security measures you take through all media channels (from consent forms to websites) with a privacy statement. You can adjust the statement to fit the target group, purpose, and level of privacy.

4.1.10 Update DMP

The DMP is a living document. The fact that at the moment there are still many uncertainties about the data does not release us of the obligation to ethically and lawfully collect, process, and store this data. All researchers have the responsibility to keep the DMP up to date, so the DMP will reflect the latest developments in data collection.

5 Data management per WP

The work package leaders have been asked to describe the different data sets that will be used within their WP as well as possible. For the description of the work packages the standard EC template for a data management plan has been used. However, many questions concerning the FAIR principles cannot be answered at this moment. Therefore, we have specified provisional guidelines concerning these principles below. If not otherwise specified in the Work Package description, these provisional guidelines will for now apply to the data set. Description in the Work Packages that deviate from these intentions will be mentioned in the description of the work packages.

5.1 Provisional Guidelines

5.1.1 Overall

Since the consortium has no extensive detailed knowledge of Open Data and the best practices of opening up the data set, we intend to collaborate with the information specialists of the University of Twente. This department offers specialised support for all aspects of data management including Open Access, Open Data, Archiving and Data Management.

For secure storage we closely work together with the ICT department to keep up with the latest technologies and rules and regulations regarding the GDPR.

If possible, from a privacy point of view, it is our intention to make all the above-mentioned written data openly available except those parts of the data that pertain to practices and technologies covered by any secrecy clauses in the consortium agreement or in the exploitation agreements reached within the consortium or between the consortium and external parties.

5.1.2 Findable

Each dataset will get a unique Digital Object Identifier (DOI).

Deliverable 1.1 on Quality, Risk, and IPR management has dictated the naming conventions and versioning guidelines that will be used within the project. When the data set will be stored in a trusted repository the name might be adapted in order to make it better findable.

Keywords will be added in line with the content of the datasets and with terminology used in the specific scientific fields to make the datasets findable for different researchers.

5.1.3 Accessible

As described before, our intention is to open up as many data sets as possible. However, if we cannot guarantee the privacy of the participants, the data set might be opened up under a very restricted license or it will remain completely closed.

All open data set will be stored in a trusted repository. At the moment we are looking into DANS and 4TU Centre for research data.

DANS is a Data Seal of Approval (DSA) and World Data System (WDS) trusted repository, based in the Netherlands. ¹ In DANS access can be restricted, ranging from Open Access, Open Access for registered users to restricted access. For the final dataset the appropriate level of access will be chosen.

¹ https://dans.knaw.nl/en/front-page?set_language=en

4TU Centre for Research Data is a DSA Trusted repository, also based in the Netherlands². The centre provides knowledge, experience and the tools to archive research data in a standardized, secure and well-documented manner. The data will be accessible here for the (restricted) public.

Furthermore, datasets might be stored in partners' repositories, and other (inter)national trusted repositories with a Data Seal of Approval. The definitive list will be added to the final version of the DMP.

If a dataset will be stored in a trusted repository with a limited access license, a Data Access Committee will be set up at the end of the project. They will decide on a case-to-case basis if access will be granted and for how long.

5.1.4 Interoperable

We are looking into suitable metadata standards, for example: DataCite³ and Dublin Core⁴.

Depending on the scientific field where the data set will originate from, additional meta-data standards might be used.

5.1.5 Re-Use

If possible, the data set will be licensed under an Open Access license. However, this will depend on the level of privacy, and the Intellectual Property Right (IPR) involved in the data set.

A period of embargo will only be necessary if a data set contains specific IPR or other exploitable results that will justify an embargo.

Our intention is to make as much data as possible re-useable for third parties. Restriction will only apply when privacy, IPR, or other exploitations ground are in play.

All data sets will be cleared of bad records, with clear naming conventions, and with appropriate meta-data conventions applied.

The length of time, the data sets will be stored will depend on the content of the data set. For example, if the data set contains medical practices that we foresee will be replaced soon, these set won't be stored for eternity. Furthermore, data sets with specific technological aspects, might become outdated and we will apply an appropriate time for storage.

² <http://researchdata.4tu.nl/en/home/>

³ https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadadataKernel_v4.1.pdf

⁴ https://github.com/dcmi/repository/blob/master/mediawiki/wiki/User_Guide.md

6 WP1: Data for sound, effective and efficient project management

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>The purpose of the data collection in this WP is sound, effective, and efficient project management to ensure the project is delivered on time, within budget and with outstanding quality.</p> <p>The following datasets will be used for project management purposes. These data will be kept confidential within the project, so we have chosen the formats that were easiest for the partner to work with.</p> <ol style="list-style-type: none"> a. Agenda and minutes of all meeting held. (.doc format) b. Financial information of the partners, as provided to the EC for the two periodic reports. (.xls/.doc/email format) c. Templates of deliverables, presentations, posters etc. (.doc/.ppt format) d. Logo's and other corporate identity items. (several graphical formats like vector files, jpg, png etc.) e. All final deliverables and reports will be uploaded in the EC portal in a pdf format. <p>No data is being re-used.</p> <p>The data will be collected/generated by the project manager before during, or after project meetings as well as during periodical and financial reporting periods.</p> <p>Graphical elements will be generated by iSPRINT, in charge of dissemination and exploitation activities.</p> <p>The data will probably not exceed 1 Gigabyte (GB).</p> <p>The data will be useful for project managers, project partners and where it concerns the deliverables and financial information generated during periodical reports, to the EC.</p>	
<p>FAIR Data 2.1. Making data findable, including provisions for metadata</p>	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision) ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? ▪ Outline naming conventions used ▪ Outline the approach towards search keyword ▪ Outline the approach for clear versioning ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
<p>The DMP and periodic reporting and financial information will be shared with the EC through the portal. This should make it accessible for the committee as well as the consortium. No further</p>	

keywords will be provided and no further measures will be taken to improve the discoverability of the data.	
2.2 Making data openly accessible	<ul style="list-style-type: none"> Specify which data will be made openly available? If some data is kept closed provide rationale for doing so Specify how the data will be made available Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Specify where the data and associated metadata, documentation and code are deposited Specify how access will be provided in case there are any restrictions
<p>Most of this data will not be made public, with exception to this Data Management Plan, which will be publicised through the EC's portal. Furthermore, the Periodic reports and financial information will be provided to the EC, but are still confidential.</p> <p>Therefore, we believe it is not necessary to make this data findable, openly accessible or otherwise future proof.</p>	
2.3. Making data interoperable	<ul style="list-style-type: none"> Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
No project management data will be made interoperable.	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> Specify how the data will be licenced to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
No data re-use will be made possible for project management data.	
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project

	<ul style="list-style-type: none"> Describe costs and potential value of long-term preservation
No additional costs will be made for the project management data.	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>Most PM data will be stored on the Dropbox folder, since the dataset contains no personal data. Dropbox access is granted by the project management and can be revoked at any time. Dropbox makes its own backups at least every 24 hours. Any data that is sensitive will be stored on either the SharePoint server or encrypted on Dropbox, in line with the project's guidelines on personal data.</p> <p>At the end of the project most project management data will be archived through JOIN, the archiving system of the UT. It will be stored for 10 years.</p>	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of Description of Action (DoA) and ethics deliverables. Include references and related technical aspects if not covered by the former
None of the project management data is subject to ethical considerations, outside of privacy. This aspect has been covered in the data security questions.	
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
In line with the UT policy on archiving data, Project Management data will be archived at the end of the project for a 10-year period in JOIN.	

7 WP2: Data for Responsible Research and Innovation

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>The purpose of the data collection in this WP is understanding opinions on societal responsibility and user needs in order to achieve the objective of ensuring the research and innovation process in the project follows the principles of Responsible Research and Innovation and to develop new tools and coaching methods.</p> <p>The following datasets are being collected:</p> <ol style="list-style-type: none"> a. Notes and minutes of brainstorm and workshops (.doc format) b. Recordings and notes from interviews with stakeholders (.mps, .doc format) c. Transcribed notes/recordings or otherwise 'cleaned up' or categorised data. (.doc, .xls format) <p>The files are initially stored as word and excel files. If it is possible to anonymise them and can be used for open access, these files will be stored in the equivalent Open Office format or as pdf.</p> <p>No data is being re-used. The data will be collected/generated by DBT and RRD before during, or after project meetings and through interviews with stakeholders. All data gathering will take place within the EU and by/from EU citizens.</p> <p>The data will probably not exceed 2 GB, where the main part of the storage will be taken up by the recordings.</p> <p>The data will be useful for other project partners and in the future for other research and innovation groups or organizations developing virtual coaching applications and researchers in the field of Responsible Research and Innovation (RRI), Science and Technology Studies (STS), and Technology Assessment (TA).</p>	
<p>FAIR Data 2.1. Making data findable, including provisions for metadata</p>	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision) ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? ▪ Outline naming conventions used ▪ Outline the approach towards search keyword ▪ Outline the approach for clear versioning ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
<p>At the moment, the following metadata will be created for the data files:</p> <ul style="list-style-type: none"> ▪ Author ▪ Institutional affiliation ▪ Contact e-mail 	

<ul style="list-style-type: none"> ▪ Alternative contact in the organizations ▪ Date of production ▪ Occasion of production <p>Further metadata might be added at the end of the project in line with meta data conventions.</p> <p>All data files will be named so as to reflect clearly their point of origin in the Council of Coaches structure as well as their content. For instance, brainstorming data from the RRI workshop in task 2.1 will be named "T2.1 – RRI Workshop – Brainstorm results".</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	<p>2.2 Making data openly accessible</p> <ul style="list-style-type: none"> ▪ Specify which data will be made openly available? If some data is kept closed provide rationale for doing so ▪ Specify how the data will be made available ▪ Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? ▪ Specify where the data and associated metadata, documentation and code are deposited ▪ Specify how access will be provided in case there are any restrictions
<p>Depending on the answers the subjects will give, the dataset might contain personal information. The answers might be connected to the subject's age, gender, professional position, etc. There may be a conflict here with the need for pseudonymisation, because with such a small group of respondents it will be very easy to connect specific responses to specific respondents by triangulating with e.g. social media profiles.</p> <p>If it turns out the dataset does contain personal information, then it will be treated in line with the project's guidelines.</p> <p>Open Office should be sufficient to open the document and spreadsheet files.</p> <p>We foresee no restrictions to the dataset, if and when completely anonymised.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
<p>2.3. Making data interoperable</p>	<ul style="list-style-type: none"> ▪ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. ▪ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
<p>By storing the data in Open Office format, these data files can be read by commercial administrative tools, like Microsoft Office as well. In case MP3 files will be recorded, these are universal and can be played through multiple software tools.</p>	

<p>The collected data will be ordered so as to make clear the relationship between questions being asked and answers being given. It will also be clear to which category the different respondents belong (consortium members, external stakeholder).</p> <p>The data will use common social science data collection practice. One potential deviation in terms of privacy has to do with whether the answers will contain personal information.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> Specify how the data will be licenced to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
<p>The data will probably be stored in a trusted repository with an Open Access license. At the moment, there is no intention for patenting the information.</p> <p>By posting the data in an open repository with the Data Seal of Approval, we will ensure that the data will be made available for re-use. Only the final data set will be submitted in the repository.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project Describe costs and potential value of long-term preservation
<p>The work to be done in making the data FAIR will be covered by the ordinary working budget for producing the deliverables. DBT's project manager Rasmus Øjvind Nielsen will be responsible for the data management for this purpose.</p>	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>Workshop and interview data will be gathered in the form of notes and audio recordings.</p> <p>Audio recordings and handwritten notes will be stored under lock in the offices of the DBT in a physical storage space separate from the participant lists of workshops and interviewees.</p> <p>Audio recordings and handwritten notes (e.g. Post-its) will be destroyed once they have been added to the machine-written notes from the workshops or interviews. In cases where audio recordings or handwritten notes are never added to the machine-written notes, they will be destroyed in any case no later than the end of the Council of Coaches project.</p> <p>Machine-written notes (i.e. data files in word or excel format) will be stored in a SharePoint space provided by Twente University. Access will be granted in line with the project's procedures.</p>	

5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
<p>All workshop participants and interviewees will be asked to sign a consent form giving consent to use of the data in the Council of Coaches project's analyses and for the sharing of the data with others through FAIR measures. Consent for the two different uses will be specific. Consenting to the use of collected data for the purposes of the Council of Coaches project's analyses will be a mandatory requirement for workshop participation and the conduct of interviews. In cases of non-consent to FAIR use of the data, the statements produced by the person in question will be marked with a non-personal marker and eliminated from the dataset before publication.</p> <p>When during data collection, it turns out that the responses are possible ways to identify individuals, the data will be treated as personal data and will be stored in line with the project's guidelines.</p>	
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
No other procedures need to be put in place for project management data.	

8 WP3: Data for providing personalised coaching.

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>Data is collected for providing personalised coaching, in terms of coaching strategy, manner of coaching, and timing in order to generate the content of the conversations the individual coaches can have with the end user.</p> <p>At the moment we do not have a clear picture on what kind of information will be necessary to provide these conversations. It will be personal information on different topics of coaching (e.g. Dietary, Activity).</p> <p>In order to determine the correct coaching strategy datasets will be gathered through 3 sources:</p> <ul style="list-style-type: none"> ▪ Through wearable sensors (e.g. steps and heart rate monitors); ▪ Through surveys the user will fill out during the process; ▪ Through conversations the user will have with the different coaches. <p>At the moment, since we are still working out what kind of data will be gathered, there has been no decision taken on the format of the data files nor can we say anything about the size of the data.</p> <p>No data is being re-used.</p> <p>The data will be useful for other project partners and in the future for other research and innovation groups or organizations developing virtual coaching applications and researchers in the field of monitoring and coaching, as well as medical professionals and health and life style specialists.</p>	
<p>FAIR Data 2.1. Making data findable, including provisions for metadata</p>	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision) ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? ▪ Outline naming conventions used ▪ Outline the approach towards search keyword ▪ Outline the approach for clear versioning ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
<p>No deviations from the intended FAIR principles are foreseen at this point.</p>	
<p>2.2 Making data openly accessible</p>	<ul style="list-style-type: none"> ▪ Specify which data will be made openly available? If some data is kept closed provide rationale for doing so ▪ Specify how the data will be made available ▪ Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it

	<p>possible to include the relevant software (e.g. in open source code)?</p> <ul style="list-style-type: none"> ▪ Specify where the data and associated metadata, documentation and code are deposited ▪ Specify how access will be provided in case there are any restrictions
<p>It is unclear at this time whether responses will be contain any personal information ("today is my birthday", together with a time stamp) So, at the moment it is unclear the amount of privacy the datasets will need. It is our intention to make the datasets as open as possible, but if this turns out to be violating privacy regulations, we choose to keep the datasets closed or with very restricted access.</p> <p>Since we do not know what the data set will look like, we do not have any specific methods or software in mind to access the data. It is our intention to make the data as accessible as possible this includes storing the data in a broadly used file format.</p> <p>No deviations from the intended FAIR principles are foreseen at this point.</p>	
2.3. Making data interoperable	<ul style="list-style-type: none"> ▪ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. ▪ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
<p>We will try to make the data as interoperable as possible, depending on how the data set will look like.</p> <p>No deviations from the intended FAIR principles are foreseen at this point.</p>	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> ▪ Specify how the data will be licenced to permit the widest reuse possible ▪ Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed ▪ Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why ▪ Describe data quality assurance processes ▪ Specify the length of time for which the data will remain re-usable
<p>Depending on the content of the data set and whether it contains personal information, re-use by third parties could be possible.</p> <p>No deviations from the intended FAIR principles are foreseen at this point.</p>	
3. Allocation of resources	<ul style="list-style-type: none"> ▪ Estimate the costs for making your data FAIR. Describe how you intend to cover these costs ▪ Clearly identify responsibilities for data management in your project

	<ul style="list-style-type: none"> Describe costs and potential value of long-term preservation
<p>The work to be done in making the data FAIR will be covered by the regular working budget for producing the deliverables. From RRD Harm op den Akker will be responsible for the data management for this purpose.</p>	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>Non-sensitive data is stored on a Dropbox folder for the entire consortium to access. Sensitive data, in terms of personal data and privacy is stored on a SharePoint portal, hosted at the University of Twente. Backups are made through Dropbox and the UT ICT systems every 24 hours.</p> <p>After the project, the data files will be anonymised before the will be posted in an open repository.</p> <p>Collection of the questionnaires will be done through the Qualtrics system. At the moment there is an agreement between SurfNet and Qualtrics, which would make it sufficient for use within the Dutch university system. However, it is unclear if the necessary security measures are indeed in place. However, we've looked into different questionnaire systems and Qualtrics, despite the limitations, still seems to be the best (and most secure) option available.</p>	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
<p>When during data collection we figure that the responses are possible ways to identify individuals, the data will be treated as personal data and will be stored in line with the project's guidelines.</p> <p>The end user will receive an information leaflet and will sign a consent form. This way we ensure the patient is fully informed about the nature of the research and the data collection that takes place and they give their (full) consent for the research.</p> <p>Furthermore, in case of accidental finding through the nature of the responses we will contact the end user immediately.</p>	
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
<p>Any research data that cannot be made publically available will be stored according to the archiving guidelines of RRD.</p>	

9 WP4: Data for the development of the Holistic Behaviour Analysis Framework

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>For the development of the Holistic Behaviour Analysis Framework, it is necessary to:</p> <ol style="list-style-type: none"> 1. Develop and validate new models to recognise users' behaviours both in the interaction with the system and during their everyday affairs. 2. Extract features quantifying the user behaviour (physical, social, emotional, cognitive) <p>At this point is not possible to list the specific sensor data that will be collected; therefore, we provide examples that reflect to a large extent what we will eventually use for measuring user's behaviours.</p> <p>We plan to collect diverse digital traces from the user explicit and implicit interaction with their smartphones, smartwatches and ambient sensors. Smartphone and smartwatch data logs will be stored temporarily on the devices in a relational database (e.g., SQLite). The temporal data will be transmitted over HTTPs in the form of data objects (e.g., JSON) to a secure server where it is persisted in another relational database management system (e.g., MySQL).</p> <p>The raw sensory data will be used to generate different levels of behavioural data. Relevant features will be computed through mathematical and statistical modelling. Examples of this features are "step counts" (calculated from the raw acceleration data), "minutes spent in a given location" (calculated from the raw Global Positioning System (GPS) data) or "heart rate variability" (calculated from the raw photo-plethysmography data). These features will be computed automatically and populated in the knowledge base. These features will be also used as inputs to the machine learning models that will render a decision on the current user behaviour (e.g., "resting" from steps counts and minutes spent in a given location).</p> <p>The collected data will be used to train and validate new machine learning models aimed to recognise the behaviour of the users. At this point it is not clear yet whether the data will be used for each user separately or in combination as to create personalised or general models (most possibly both).</p> <p>The size of the data sets is difficult to estimate since the choice of sensors and sampling rates can fairly affect the amount of generated data. The size can range from a few KB of data per user and day if sensors with low data generation rate are considered (e.g., GPS) to tens of MB if highly data productive sensors are considered (e.g., accelerometers). It is safe to say that the collected raw sensor data will be around 10 MB per user and day. The processed data will represent a fairly compressed version of the raw sensory data, thus in the order of KBs.</p> <p>The processed data will be used to populate the knowledge base, which will in turn become available to different components of the Council-of-Coaches system.</p> <p>Moreover, the processed data could be of much relevance for other groups conducting research in social and behavioural computing. Likewise, the raw sensory datasets collected in this project, after full anonymization, could be used in the benchmarking of new machine learning and artificial intelligence models.</p>	

FAIR Data 2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision) ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? ▪ Outline naming conventions used ▪ Outline the approach towards search keyword ▪ Outline the approach for clear versioning ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
No deviations from the intended FAIR principles are foreseen at this point.	
2.2 Making data openly accessible	<ul style="list-style-type: none"> ▪ Specify which data will be made openly available? If some data is kept closed provide rationale for doing so ▪ Specify how the data will be made available ▪ Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? ▪ Specify where the data and associated metadata, documentation and code are deposited ▪ Specify how access will be provided in case there are any restrictions
<p>It is unclear at this time from the data it will be possible to retrace a person. (for instance, through location tracking). At the moment it is unclear the amount of privacy the datasets will need. It is our intention to make the datasets as open as possible, but if this turns out to be violating privacy regulations, we choose to keep the datasets closed.</p> <p>Since we do not know what the data set will look like, we do not have any specific methods or software in mind to access the data. The methods/software may change depending on how the raw sensor data & behavioural data is integrated in/with the knowledge bases.</p> <p>Any software that allows access to relational databases (i.e., MySQL) will do. Some examples of open source options are: DBeaver, SQLelectron or SequelPro.</p> <p>It is our intention to make the data as accessible as possible this includes storing the data in a broadly used file format.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
2.3. Making data interoperable	<ul style="list-style-type: none"> ▪ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. ▪ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

<p>We use standard models for encoding the data (e.g., JSON, CSV).</p> <p>At this moment we do not plan to use any specific ontologies. We would have no problem in either case to implement the necessary changes at the point of need.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> Specify how the data will be licenced to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
<p>Depending on the content of the data set and whether it contains personal information, re-use by third parties could be possible.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project Describe costs and potential value of long-term preservation
<p>The work to be done in making the data FAIR will be covered by the regular working budget for producing the deliverables. From CMC-BSS Oresti Banos will be responsible for the data management for this purpose.</p>	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>An anonymised universal unique identifier will be used to identify the data collected from each user. This identifier will in no way allow to reveal the identity of the user.</p> <p>However, there might be a combination of data possible, with which you can identify a person, for example 24hour location tracking.</p> <p>The raw sensor data will be transmitted over HTTPs in the form of data objects (e.g., JSON) to a secure server where it is persisted in another relational database management system (e.g., MySQL). Any further information on the server it at the moment of writing not available yet.</p> <p>In all cases data will be stored according to the project's guidelines on personal data.</p>	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

When during data collection we figure that the responses are possible ways to identify individuals, the data will be treated as personal data and will be stored as such.

The end user will receive an information leaflet and will sign a consent form. This way we ensure the patient is fully informed about the nature of the research and the data collection that takes place and they give their (full) consent for the research.

Furthermore, in case of accidental finding through the nature of the responses we will contact the end user immediately.

6. Other

- Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

No other procedures need to be put in place for project management data.

10 WP5: Data for Dialogue and Argumentation Framework

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>Data will be collected through videotaping sessions where experts pretend-play a coaching session with an actor. This way data can be collected to study the verbal and nonverbal behaviours of the participants such as body language, interaction between the different experts, and use of language and dialogue in order to make the virtual coaching session as realistic as possible. These sessions will be audio-recorded and videotaped from multiple angles.</p> <p>The sessions will be annotated to generate annotation files. The analysis will be conducted on different levels: at the behaviours level, we will annotate with automatic tools (e.g. facial expression analysis, body movement analysis, prosody analysis, text transcription, etc.) and manually whenever necessary. We will also annotate higher level information such as level of engagement, attitude the participants have toward each other, emotion, etc. Other annotations may be conducted such as dialog strategies, turn taking dynamism, and interruption types.</p> <p>These annotations will be further analysed and used to build virtual coaches with their own specific style that ought to be defined at different levels (cf WP6): behaviour, emotional sensibility, attitude, interactional sensibility.</p> <p>The format of the data will be audio and video stream using format such as MP3, MPEG or AVI.</p> <p>At the moment we are choosing between different annotation tool such as Elan and NOVA (developed by Elisabeth André's group at University of Augsburg). The analysis of the data will be used to develop a computational model of the virtual coach's nonverbal behaviours. It will correspond to software code and a library of behaviour defined in a lexicon (see d2.1 and D6.1).</p> <p>Considering Re-use, at the moment we are looking at existing databases such as the NoXi database available at https://noxi.aria-agent.eu/, after signing an End User License Agreement (EULA). We will refer to existing databases to learn about specific phenomena (e.g. interruption). These databases are either publicly available or have been gathered by ourselves (and are publicly available after signing a EULA).</p> <p>The re-enactment scenarios are provided by RRD and UDun.</p> <p>The expected size of the data is going to be multiple GB's depending on the size of the videos. It should be large enough to allow for a machine learning approach.</p> <p>The data will be useful to UPMC to design different virtual coaches. The data could also be used to study communicative behaviours, interactive behaviours, engagement level, etc. Researchers from social sciences, computational linguistic, affective computing and social signal processing may find the databases very interesting, especially since it gathers data of group conversation that have been barely available so far.</p>	
<p>FAIR Data 2.1. Making data findable, including provisions for metadata</p>	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision) ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use

	<p>of persistent and unique identifiers such as Digital Object Identifiers?</p> <ul style="list-style-type: none"> ▪ Outline naming conventions used ▪ Outline the approach towards search keyword ▪ Outline the approach for clear versioning ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
No further deviations from the intended FAIR principles are foreseen at this point.	
2.2 Making data openly accessible	<ul style="list-style-type: none"> ▪ Specify which data will be made openly available? If some data is kept closed provide rationale for doing so ▪ Specify how the data will be made available ▪ Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? ▪ Specify where the data and associated metadata, documentation and code are deposited ▪ Specify how access will be provided in case there are any restrictions
No further deviations from the intended FAIR principles are foreseen at this point.	
2.3. Making data interoperable	<ul style="list-style-type: none"> ▪ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. ▪ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
No further deviations from the intended FAIR principles are foreseen at this point.	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> ▪ Specify how the data will be licenced to permit the widest reuse possible ▪ Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed ▪ Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why ▪ Describe data quality assurance processes ▪ Specify the length of time for which the data will remain re-usable
No further deviations from the intended FAIR principles are foreseen at this point.	
3. Allocation of resources	<ul style="list-style-type: none"> ▪ Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

	<ul style="list-style-type: none"> Clearly identify responsibilities for data management in your project Describe costs and potential value of long-term preservation
The work to be done in making the data FAIR will be covered by the regular working budget for producing the deliverables. From UDUN Alison Pease will be responsible for the data management for this purpose. From UPV Catherine Pelachaud will be responsible for the transcribed files.	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
The videos as captured during the interview with the experts will be stored and shared through the University of Dundee's own box.com solution. This storage solution fully complies with the GDPR. Access to the videos will be granted only by request through Mark Snaith.	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
The experts will receive an information leaflet and will sign a consent form. This way we ensure the person is fully informed about the nature of the research and the data collection that takes place and they have options in giving their consent for the use of data for the research.	
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
No other procedures need to be put in place for project management data.	

11 WP6: Data for Human Computer Interfaces

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>In order to design, implement and evaluate user interactions with the coaches, the following kinds of data will be collected.</p> <p>Corpora (video and audio) of representative conversational behaviour by humans, to be used for defining and generating Council of Coaches coaching behaviour:</p> <p>Video and audio; might be published as corpus. For this data to be used and stored consent from the human participant will need to be required. If no consent can be acquired, we will need to anonymise the video footage. It will be stored in MP3 and MP4 or avi format.</p> <p>Annotations of video/audio/etc from experiments and corpora:</p> <p>The video and audio files will be annotated, in order to make the data more accessible. The annotations are made in some kind of common annotation tool (often ELAN); generally, XML storage format will be used for the codes, and word (or Open Office) documents for the coding scheme.</p> <p>Data from local study sessions (experimental and explorative):</p> <p>Local study session will be conducted to analyse the impact of the system, the behaviour of users with system, and the perception of system behaviour by user. The following kinds of data will be collected:</p> <ul style="list-style-type: none"> ▪ System logs (as described in chapter 12); ▪ Data collected using the sensor system (as described in chapter 9); ▪ Video and audio recordings for analysis of the user in interaction with the system. This will be stored in mp3, mp4 or avi format. ▪ Dialogue logs (speech and/or transcription: what has been said during the dialogue) in Elan, XML and doc format. ▪ Questionnaires and interviews with user about the experience. The format of these are still up for discussion. <p>Data from demonstrator sessions (so not collected for a study, but to showcase our system):</p> <p>When the system is showcases, the system will automatically collect the same kind of data as when it is in experimental mode. This data will be discarded as soon as possible after the demo. However, we might opt to use some data as PR material. End Users will always be asked to sign a consent form in that case.</p> <p>Processed data from studies fit for publication:</p> <p>We may want to publicly release some data collected in studies for journal publication or presentations at conferences. This requires anonymization, as well as the right kind of consent.</p> <p>The data originates from experiments with humans and virtual demonstrators as conducted on the platform as developed within the project.</p> <p>The size of the data set will probably be several GB's, depending on the length and quality of the video footage.</p>	

Moreover, the raw sensory data as well as the processed data could be of much relevance for other groups conducting research in social and behavioural computing.	
FAIR Data 2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision) ▪ Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? ▪ Outline naming conventions used ▪ Outline the approach towards search keyword ▪ Outline the approach for clear versioning ▪ Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
No further deviations from the intended FAIR principles are foreseen at this point.	
2.2 Making data openly accessible	<ul style="list-style-type: none"> ▪ Specify which data will be made openly available? If some data is kept closed provide rationale for doing so ▪ Specify how the data will be made available ▪ Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? ▪ Specify where the data and associated metadata, documentation and code are deposited ▪ Specify how access will be provided in case there are any restrictions
No further deviations from the intended FAIR principles are foreseen at this point.	
2.3. Making data interoperable	<ul style="list-style-type: none"> ▪ Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. ▪ Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
No further deviations from the intended FAIR principles are foreseen at this point.	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> ▪ Specify how the data will be licenced to permit the widest reuse possible ▪ Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed ▪ Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

	<ul style="list-style-type: none"> Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
No further deviations from the intended FAIR principles are foreseen at this point.	
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project Describe costs and potential value of long-term preservation
The work to be done in making the data FAIR will be covered by the regular working budget for producing the deliverables. From CMC-HMI Randy Klaassen will be responsible for the data management for this purpose.	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>Personal data that is collected goes on an encrypted hard disk that we can carry around from location to location; a backup will then be made (encrypted as well) and stored at HMI in a safe place, and password to encryption stored behind lock and key.</p> <p>The consent forms used for the video, audio and experimental data collection will be stored in locked cupboard at HMI. In case of pseudonymisation, the key to the consent / ptcp number is stored at a physically different location.</p> <p>Any information used for publications or presentations will be made fully anonymous, unless there is the right consent from the end user.</p> <p>Data to be opened at the end of the project will be in line with the project's guidelines as well as the HMI group's own data policy. (Since this policy is still under development, no specifics can be given on what it will contain. However, at the end of the project, the data policy should be in place and should be used as an additional guideline on data management.)</p>	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
The end user will receive an information leaflet and will sign a consent form. This way we ensure the end user is fully informed about the nature of the research and the data collection that takes place and they give their (full) consent for the research.	
6. Other	<ul style="list-style-type: none"> Refer to other national / funder / sectorial / departmental procedures for data management that you are using (if any)
Data files that contain personal data that cannot be made openly available will be stored according the UT-HMI data policy for 10 years in a secure environment.	

12 WP7: Data for Continuous integration and Demonstration

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>To facilitate the integration of the various software components developed in the project as well as maintaining two working demonstrators, several datasets will be collected:</p> <p>Knowledge base:</p> <p>The information in the knowledge base is about the user and the environment and is the one used by the system itself to provide its features. The knowledge base is going to be implemented in cooperation with WP3.</p> <p>The format is still to be decided but it is safe to assume it will be some kind of database, whether relational, Non-SQL or semantic.</p> <p>The data in the knowledge base will be collected from different sources. It will be generated based on the raw sensor data coming from wearable sensors, surveys with the end users, and through conversation the end user will have with the system. The data will be updated during the trials, by means of machine learning through the user input of the users of the system.</p> <p>For the size of the data it is difficult to estimate the exact amount of data generated until the system is running as expected, especially if the format is not decided yet. As a reference, a real-life deployment of universAAL that gathers environment information generates around 4GB each two months. It is safe to assume that the project's technology can generate knowledge information in the range of gigabytes per month.</p> <p>The knowledge base will be used by developers and technicians in the project for troubleshooting and debugging. Externally the knowledge base can be interesting for researchers in the field of coaching, human behaviour and other social sciences.</p> <p>System logs:</p> <p>To monitor the adequate operation of the entire system, analyse its performance and issues, and troubleshoot any system errors that may happen. The system logs will be useful for developers and technicians in the project for troubleshooting and debugging.</p> <p>User-related logs:</p> <p>To monitor the interaction of the user with the system, to analyse how the system is used, how many times, in which conditions, at which hours... This can provide interesting info on how the system can be used, and to troubleshoot any errors: it gives information on how to reproduce errors and find out if it was being improperly used. The user logs will be useful for developers and technicians in the project for troubleshooting and debugging. The logs are useful for behaviour analysis as well, to improve usability and to determine whether the system is successful in its goals. Depending on what kind of information the logs will contain, it might be of interest to social scientists focussing on online behaviour of patients.</p>	

Probably all of the logs will be text files following commonly used logger formats depending on the technology used. For instance, Java-based software can use Log4j logs, which are commonly used. The information recorded in these logs represent information generated by the programs themselves during their execution. Therefore, its content depends on what the developers decided to log at each point in their program. For software developed by the project it can be determined by ourselves if need be. For external libraries not developed by the project, the information recorded in their produced logs is detailed in their respective documentation.

All data will be collected during the trials of the project and by testing the prototypes. For the size of the data it is not possible to estimate in advance an accurate measurement of the log files. These can be configured to record more or less information based on what we finally need. It is also difficult to estimate the exact amount of data generated until the system is running as expected. As a reference: A recent pilot based on universAAL platform generated around 10MB of logs per day. It is safe to assume that the project can generate logs in the range of tens to hundreds of MB per day.

FAIR Data 2.1. Making data findable, including provisions for metadata

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

At the moment we foresee the system logs only to be used during the project. They are of no interest to keep for after the project, since they are used for troubleshooting and analyse performance issues. The user logs could be of interest to social scientist, but this will need to be determined after we've decided on what kind of data will be collected.

For log files it is common practice to include the date of the log in the filing name as well, so we will keep with this practice.

No further deviations from the intended FAIR principles are foreseen at this point.

2.2 Making data openly accessible

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

System logs:

These can be, in principle, made public. However, they are not expected to offer much information of interest to others outside the project. There are two possible restrictions to consider if this data would be published:

- 1) User-identifiable data: There may be log entries that could be used to infer user-related information. These should not be there in the first place if logs are going to be shared.
- 2) "Confidential" program logic: There could be logs produced by software that is considered confidential by some developer, and could be used for reverse-engineering.

Almost all logging libraries and utilities in any program can be configured to follow a format that can be read by many log-analysis tools. These allow easy reading of logs and can produce statistics out of them. Documentation about logs will be provided with the log files.

The logs are regular files (text files) and can be made public in any online method for file sharing.

Knowledge base:

This can be made public and could be one of the most interesting data to be analysed. There is one possible restriction however: User-identifiable data: There may be log entries that could be used to infer user information and affect privacy restrictions (e.g. inferring private data by analysing behaviour or when 24-hour location tracking is performed, home location and therefore an identifiable person might be retraced).

There are some technologies that could facilitate making the database accessible. For instance, if the Knowledge base was implemented or connected to FIWARE, it is possible to use CKAN Generic Enabler to publish its stored data as publicly available CKAN repository. We will keep this in mind while developing the database.

All databases (or similar data-storage technologies) provide some method or language to access the data. For instance, there is SQL for relational databases or SPARQL for semantic stores. These can be used to read data in bulk or performing complex queries for statistics. Public access to the databases would require that these are exposed online with a query endpoint.

Public access to the databases would require that these are exposed online with a query endpoint. Some engines already include query endpoints, but this requires that the project maintains a server for accessing the hosted database.

Source Code:

It is our intention to include the relevant software in open source, but it will need to be discussed which part will be made accessible. Open source code published by the project will be available in GitLab: <https://gitlab.com/CouncilOfCoaches>. This can be done free of charge. At the moment this is an invitation only group, to be used for the project members who have a Gitlab account. Parts of the software can be made available at a later stage, while other parts can remain private.

No further deviations from the intended FAIR principles are foreseen at this point.

2.3. Making data interoperable

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

<p>System/User logs: These can follow established formats for logging that are widely used and known by developers and technicians.</p> <p>Knowledge base: Almost all possible options of technologies to be used in the knowledge base follow a well-known format or query language.</p> <p>The use of standards will depend on the underlying technology, but standards will be used whenever possible. However, it is highly unlikely that there are existing standards in this domain that cover all the needs of the project, so some degree of custom modelling is expected.</p> <p>No further deviations from the intended FAIR principles are foreseen at this point.</p>	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> Specify how the data will be licenced to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
No further deviations from the intended FAIR principles are foreseen at this point.	
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project Describe costs and potential value of long-term preservation
<p>The work to be done in making the data FAIR will be covered by the regular working budget for producing the deliverables. UPV will be responsible for the data management for this purpose. At the moment we are looking for the right person to do this.</p>	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>At the moment we are looking into setting up a secure server to be able to integrate the different parts of the software. This server will comply with all privacy regulations. Communication with external sources will go through an https connection and regular backups will be made.</p> <p>Further decisions on the server have not been made.</p>	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
All personal data stored and processed will have the consent of the end users.	
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

No other procedures need to be put in place for project management data.

13 WP8: Data for Dissemination and Exploitation

<p>DMP component Issues to be addressed</p> <p>1. Data summary</p>	<ul style="list-style-type: none"> ▪ State the purpose of the data collection/generation ▪ Explain the relation to the objectives of the project ▪ Specify the types and formats of data generated/collected ▪ Specify if existing data is being re-used (if any) ▪ Specify the origin of the data ▪ State the expected size of the data (if known) ▪ Outline the data utility: to whom will it be useful
<p>For the dissemination of the project results, the communication to raise awareness of the project in terms of training, ecosystem building. And the pre-marketing for exploitation of results the following dataset will be collected.</p> <ol style="list-style-type: none"> 1) Publications in journals and on conferences describing project results. These publications will be published through green or gold open access publishing as much as possible. These will be available in pdf format. All partners will be responsible for disseminating the research data through papers. These results will be useful for researchers in the same or adjoining fields. The dataset will probably be limited to 1 GB of data. 2) Exploitation plans will be written in a .doc format. However, depending on the level of detail described in the plans, these may be kept confidential within the consortium. The partners will be the author of the plans, but separate exploitation agreements might be set up with third parties. Both partners and third parties are expected to directly benefit from the exploitation of the project results. The dataset will probably be limited to 1 GB of data. 3) Ecosystem building for the Open Agent Platform. Our assumption at the moment is mainly that building an ecosystem will be an online activity, where the consortium will have no specific ownership of for example forum discussions. So, the data will be created by the ecosystem and available and useful for everyone who's interested. 4) Standardization activities. At the moment we assume that these will contain mainly .doc files or limited size. These could be of interest to the standardization community while developed within the consortium. The size of the files will be several MB's. 5) Training activities. For now, we assume these training activities will consist of videos or presentations. These can be posted online for everyone to see. The video's will be put in mp4 or similar format and be posted on an online platform to be expected to best reach our target groups. Presentations will be made in a PowerPoint of similar Open Office format and are expected to be used in the face to face training of different stakeholders. Our expectation at time of writing is that all materials will be posted online, being available for everyone. The size of the data set will be depending on the size of the video footage. But we assume it won't exceed 10 GB. <p>Other communication items like websites, press releases, interviews and other dissemination items will be made to create awareness of the project and to create a community of interested subjects. All this information will be made available at least through our own website. The website will be continued for at least 2 years after the project has finished. At the moment we do not know the size of the dataset, since it will depend on the content that will be posted on there. Different target groups will be reached through several media as described in the communication plan. (D8.2). The website will make use of Google Analytics, in order to further improve the use of the website. Google Analytics completely anonymises the data of the visitors of the website and stores this information their own servers.</p>	
<p>FAIR Data 2.1. Making data findable, including provisions for metadata</p>	<ul style="list-style-type: none"> ▪ Outline the discoverability of data (metadata provision)

	<ul style="list-style-type: none"> Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? Outline naming conventions used Outline the approach towards search keyword Outline the approach for clear versioning Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
<p>Data related to communication, dissemination and pre-marketing will be findable –to the best of the consortiums’ capacity- utilizing digital communications best practices, e.g. hashtag, metadata, keywords. In social media, Council of Coaches posts will be findable and discoverable by the name, e.g. @Council_Coaches (Twitter handle), while for posts to different media (e.g. 3rd party blogs), the posts will refer to the project website, i.e. council-of-coaches.eu.</p> <p>At this moment we foresee no separate datasets to be posted in repositories at the end of the project.</p>	
2.2 Making data openly accessible	<ul style="list-style-type: none"> Specify which data will be made openly available? If some data is kept closed provide rationale for doing so Specify how the data will be made available Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Specify where the data and associated metadata, documentation and code are deposited Specify how access will be provided in case there are any restrictions
<p>Most of this data will be made public, although there might be made an exception when it comes to data concerning the exploitation of results. We foresee most data will be published online, just not in online repositories, since it does not contain specific research data.</p>	
2 2.3. Making data interoperable	<ul style="list-style-type: none"> Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
<p>This is not applicable for data related to communication, dissemination and pre-marketing.</p>	
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> Specify how the data will be licenced to permit the widest reuse possible Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

	<ul style="list-style-type: none"> Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why Describe data quality assurance processes Specify the length of time for which the data will remain re-usable
Data related to communication, dissemination and pre-marketing will be allowed for reuse, following standard digital practices, i.e. naming the source of the information (e.g. http://council-of-coaches.eu/news/xxx). In addition, photos from the consortium may be released under specific Creative Commons licenses.	
3. Allocation of resources	<ul style="list-style-type: none"> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs Clearly identify responsibilities for data management in your project Describe costs and potential value of long-term preservation
There are no costs related to making data FAIR. Costs related to the support and maintenance of the digital infrastructures are not considered, as they will occur in either way.	
4. Data security	<ul style="list-style-type: none"> Address data recovery as well as secure storage and transfer of sensitive data
<p>Information posted on the website will be posted through WordPress. They will provide backups of the website as well. For other purposes: non-sensitive data is stored on a Dropbox folder for the entire consortium to access. Sensitive data, in terms of personal data and privacy is stored on a SharePoint portal, in line with the project's guidelines.</p> <p>Privacy statements will be provided on the website and the newsletter.</p> <p>The newsletter is sent through MailChimp, which will provide their own backups. All registered users can opt out of the newsletter at any time.</p>	
5. Ethical aspects	<ul style="list-style-type: none"> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
All participants in the consortium have agreed with posting their pictures online for dissemination items and project updates.	
6. Other	<ul style="list-style-type: none"> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
No other procedures need to be put in place for project management data.	

Annex 1: European Informed Consent Form

This template is intended for creating the Informed Consent Form for your research. For further support, please contact Marlon Domingus at domingus@ubib.eur.nl

INFORMED CONSENT FORM

Project Title	[COMPLETION BY PRINCIPAL INVESTIGATOR]
Purpose of the Study	This research is being conducted [COMPLETION BY PRINCIPAL INVESTIGATOR]. I am inviting you to participate in this research project about [COMPLETION BY PRINCIPAL INVESTIGATOR]. The purpose of this research project is [COMPLETION BY PRINCIPAL INVESTIGATOR].
Procedures	<p>You will participate in an interview lasting approximately [COMPLETION BY PRINCIPAL INVESTIGATOR]. You will be asked questions about [COMPLETION BY PRINCIPAL INVESTIGATOR]. Sample questions include: "[COMPLETION BY PRINCIPAL INVESTIGATOR]".</p> <p>You must be at least 18 years old [ADDITIONALLY, WHERE APPROPRIATE OTHER CONDITIONS, COMPLETION BY PRINCIPAL INVESTIGATOR].</p>
Potential Risks and Discomforts	There are no obvious physical, legal or economic risks associated with participating in this study. You do not have to answer any questions you do not wish to answer. Your participation is voluntary and you are free to discontinue your participation at any time.
Potential Benefits	<p>Participation in this study does not guarantee any beneficial results to you. As a result of participating you may better understand [Additions from researcher if necessary].</p> <p>The broader goal of this research is to [COMPLETION BY PRINCIPAL INVESTIGATOR].</p>

Confidentiality	<p>Your privacy will be protected to the maximum extent allowable by law. No personally identifiable information will be reported in any research product. Moreover, only trained research staff will have access to your responses. Within these restrictions, results of this study will be made available to you upon request.</p> <p>As indicated above, this research project involves making audio recordings of interviews with you. Transcribed segments from the audio recordings may be used in published forms (e.g., journal articles and book chapters). In the case of publication, pseudonyms will be used. The audio recordings, forms, and other documents created or collected as part of this study will be stored in a secure location in the researchers' offices or on the researchers password-protected computers and will be destroyed within ten years of the initiation of the study.</p>
Compensation	[COMPLETION BY PRINCIPAL INVESTIGATOR]
Right to Withdraw and Questions	<p>Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalised or lose any benefits to which you otherwise qualify. The data you provided before you stopped participating however will be processed in this research; no new data will be collected or used.</p> <p>If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the primary investigator:</p> <p>[COMPLETION BY PRINCIPAL INVESTIGATOR]</p>

<p>Statement of Consent</p>	<p>Your signature indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree that you will participate in this research study. You will receive a copy of this signed consent form.</p> <p>I agree to participate in a research project led by [COMPLETION BY PRINCIPAL INVESTIGATOR]. The purpose of this document is to specify the terms of my participation in the project through being interviewed.</p> <p>1. I have been given sufficient information about this research project. The purpose of my participation as an interviewee in this project has been explained to me and is clear.</p> <p>2. My participation as an interviewee in this project is voluntary. There is no explicit or implicit coercion whatsoever to participate.</p> <p>3. Participation involves being interviewed by (a) researcher(s) from the [COMPLETION BY PRINCIPAL INVESTIGATOR]. The interview will last approximately [COMPLETION BY PRINCIPAL INVESTIGATOR] minutes. I allow the researcher(s) to take written notes during the interview. I also may allow the recording (by audio/video tape) of the interview. It is clear to me that in case I do not want the interview to be taped I am at any point of time fully entitled to withdraw from participation.</p>
------------------------------------	---

	<p>4. I have the right not to answer any of the questions. If I feel uncomfortable in any way during the interview session, I have the right to withdraw from the interview.</p> <p>5. I have been given the explicit guarantees that, if I wish so, the researcher will not identify me by name or function in any reports using information obtained from this interview, and that my confidentiality as a participant in this study will remain secure. In all cases subsequent uses of records and data will be subject to standard data use policies at the EU (Data Protection Policy).</p> <p>6. I have been given the guarantee that this research project has been reviewed and approved by [COMPLETION BY PRINCIPAL INVESTIGATOR] and by the EU Ethics Committee. For research problems or any other question regarding the research project, the EU Ethics Committee may be contacted through [information of the contact person at the Ethics Committee at EU [COMPLETION BY PRINCIPAL INVESTIGATOR]].</p> <p>7. I have read and understood the points and statements of this form. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.</p> <p>8. I have been given a copy of this consent form co-signed by the interviewer.</p>	
Signature and Date	NAME PARTICIPANT	NAME PRINCIPAL INVESTIGATOR
	SIGNATURE	SIGNATURE
	DATE	DATE