Project acronym:	WellCo
Project full title:	Wellbeing and Health Virtual Coach
Call identifier:	H2020-SC1-2017-CNECT-1
Type of action:	RIA-Research and Innovation Action
Start date:	1 December 2017
End date:	30 November 2020
Grant agreement no. :	769765

D6.6 – Data Management Plan

WP6	Dissemination and Exploitation
Due Date:	30/12/2020
Submission Date:	12/06/2018
Responsible Partner:	UCPH
Version:	V0.7 June 2018
Status:	Draft
Author(s):	Katarzyna Wac (UCPH)
Reviewer(s):	HIB, CON
Deliverable Type:	R : Report
Dissemination Level:	PU: PUBLIC

VERSION HISTORY

Version	Date	Author	Partner	Description
0.1	28/04/2018	**Katarzyna Wac**	UCPH	Initial Draft Version
0.2-0.5	07/05/2018 to 22/05/2018	**Inmaculada Luengo (PC), Paloma Jimeno**	HIB	Initial Reviewed Versions, creation of new sections: Introduction and WellCo: Data summary and Open access subsections in section 4. Contribution to FAIR guidelines for WP3 and WP5.
0.2	08/05/2018	**Laurens Kemp**	CON	Initial and Final Reviewed Versions
0.4	22/05/2018	**Maj Smerkol**	JSI	Contribution to FAIR guidelines for WP5.
0.5	06/06/2018	**Paloma Jimeno**	HIB	Final Reviewed Versions, Change of template
0.6	11/06/2018	**Laurens Kemp**	CON	Final Reviewed Versions
0.7	12/06/2018	**Katarzyna Wac**	UCPH	Final Draft Version

1 Table of contents

2 List of figures

3 List of tables

4 Glossary

Acronym	Definition
API	Application Programming Interface
DMP	Data Management Plan
GUI	Graphical User Interface
NSB	National Standardization Body
NC	National Committee
CEN	European Committee for Standardisation
DOI	Digital Object Identifier

Executive Summary

This document is the deliverable "D6.6 Data Management Plan" of the European project – "WellCo – Wellbeing and Health Virtual Coach" (hereinafter also referred to as "WellCo, project reference: 769765).

The **Data Management Plan (DMP)** describes the types of data that will be produced, collected and/or processed within the project and how this data will be handled during and after the project, i.e. the standards that will be used, the ways in which data will be exploited and shared (for verification or reuse), and in which way data will be preserved. This DMP has been prepared by taking into account the template of the "Guidelines on Data Management in Horizon 2020" [Version 3.0 of 26 July 2016]. The elaboration of the DMP will allow WellCo partners to address all issues related with data protection, including ethical concerns and security protection strategy. WellCo will take part in the Open Research Data Pilot (ORD pilot); this pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects, such as the data generated by the WellCo platform during its deployment and validation. Moreover, under Horizon 2020 each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results: these publications will be made also available through the public section of the WellCo website. All these aspects have been taken into account in the elaboration of the DMP.

This deliverable 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. So far, in M6 we are at the beginning of the project and we have very little pseudo-anonymized data collected within the WP2 ("Co-design"), stored at each trial site and at the joint repository (Alfesco by HIB). An update will be provided no later than in time for the first review (M12). Other updates will be provided at M24 and M36 detailing which/how the data will be made available to others within the Pilot on Open Research Data (ORD).

Starting from a brief illustration of the WellCo project, and of the ethical concerns that could affect the project and their link with the new General Data Protection Regulation that comes into force this month offering as result some guidelines for data protection and security in WellCo, this report tries to describe the procedures of data collection, storing and processing at M6 of the project.

1 Introduction

The research activities undertaken in the WellCo project have important data protection aspects, in particular due to the sensitive and personal data it collects, processes and stores. This deliverable analyses the **data management implications** of the activities undertaken in the project, and describes the guidelines and procedures put in place in order to ensure compliance with data management requirements.

The structure of the document is as follows:

Initially, section 2 provides a data summary for the WellCo project. In order to link the purpose for the generation and processing of data with the project, **background information of the WellCo project as well as the main objectives for the project** are explained**.**

Then, as many different actors are involved as active participants: elderly, their informal caregivers and professionals, one of the major concerns of the consortium is the protection of their privacy while collecting, analysing and storing sensible data. Thus, **section 3** of this deliverable focuses on the ethics measures that will be taken in each of the countries producing, collecting and/or processing data according to the new European Regulation on Privacy, the General Data Protection Regulation, (GDPR) that has came into force in May 2018. Although **ethic measurements** were already **defined in D2.2 for the countries producing data**, i.e. the **countries where trial sites are performed**, Denmark (DK), Italy (IT) and Spain (ES), this document expands these ethic measurements to cover also the **collecting, storing, processing and re-use of this data** by technical partners during the implementation of the modules envisaged for WellCo. At the end of this section, some **guidelines for data protection and security** are proposed. The aim is to assure maximum privacy for all the personal and sensitive (e.g., ethnicity, health/wellbeing) data within the project as well as after the project end, when this research data will be made as openly accessible as possible.

The final section gathers some FAIR principles with the aim of providing a data management plan that enables to **maximize the access to and re-use of research data**, also ensuring **open access to all peer-reviewed scientific publications and agreed datasets during and after the project.** A **detailed description** of the datasets to be handled in each WP of the project, according to the requirements set out in Annex 1 – Data Management Plan template of the "Guidelines on Data Management in Horizon H2020" [1] is set at the end of this section (Section 5.2). This covers: (a) the handling of research data during and after the project; (b) what data will be generated, collected and processed; (c) what methodology and standards will be applied; (d) whether data will be shared/made open access and how; (e) how data will be curated and preserved.

2 WellCo: Data Summary

This sections aims to make a review of the scope of the project (purpose and objectives) in order to clarify the relation between it and the data generation, collection and processing envisaged in the project.

2.1 The purpose of WellCo

The aim of the WellCo project is to develop and test in a **user-centric**, iterative approach a **"Well-being and Health Virtual Coach for behaviour change"**. WellCo, thereby, seeks to deliver a radical new Information and Communication Technologies (ICT)-based solution in the **provision of personalized advice, guidance and follow-up**

of users for the **adoption of healthier behaviour choices** that help them **to maintain or improve** their **physical cognitive, mental and social well-being** for as long as possible. The whole service is also followed-up and **continuously supported by a multidisciplinary team of experts**, **as well as users' close caregivers** that provide their clinical evidence and knowledge about the user to ensure effectiveness and accuracy of the change interventions.

2.2 Objectives of WellCo

As gathered at proposal stage, the main objectives of the WellCo project and those that explain the purpose of data collection/generation in the scope of the project are listed below:

- **Objective 1 (O1). Develop novel ICT based concepts and approaches for useful and effective personalised recommendations and follow up in terms of preserving physical, cognitive, mental and social well-being for as long as possible.**

  WellCo provides an innovative technology framework, based on **last mile AI technologies**, that establishes a solid ground for a highly personalised environment where WellCo will be incorporated in a seamless way in the user's daily activities by means of **dynamic profiles** that take into consideration all the **context around the user** (from **user reported outcomes**, to **profile information**, **Life Plan** or **data derived from the monitoring of the user**). This personalization will allow the platform to provide adapted goals and recommendations to users with the aim of leading to a behavioural change on a healthy lifestyle. This change process will be followed-up and continuously supported by a multidisciplinary team of professionals and users' relatives or informal caregivers as main supporters.

- **Objective 2 (O2). Validate non-obtrusive technologies for physical, cognitive, social and mental wellbeing.**

  WellCo aims to fuse **data that can from multiple sources: static data** such as Profile, life goals (defined along e.g., Life Plan method), etc. and **dynamic** data **derived from the monitoring of the user:** data from **wearable bracelets**, **smartphone sensor data** and the implementation of **deep learning techniques to extract sentiment features of the user** based on his/her speech and body gestures. The aim is to infer not only the individual behaviour but also the social, cognitive and environmental context surrounding him/her in order to provide highly adapted and personalised guidelines and recommendations that could be adapted to individual's' daily routine.

  WellCo as a non-obtrusive solution will result in a higher amount of data and quality since users will be more likely to engage longer with our solution. The "observer effect" will be minimized resulting in data quality that will closely match the natural behaviour of the subjects.

- **Objective 3 (O3). Evidence of user-centred design and innovation, new intuitive ways of human-computer interaction and user acceptance.**

  WellCo key activities to optimize engagement and adoption are focused on the personalisation and affective awareness; so the **solution is strictly**

aligned with the user Life Plan. WellCo addresses behavioural aspects including hesitation, engagement and discouragement in the adaptation of the interactive interface. Furthermore, WellCo **includes user's emotional state into the adaptation** of the interactive interface, which is essential in considering the user needs for engagement, thereby furthering adaptive user interface knowledge. User centred design is specifically addressed in T3.3 of the project with the **personalization of the interactive user interfaces** to the needs and preferences of individual users **based on context-of-use** using user profiles, context models and heuristics context aware models, e.g. rules or decision trees. In order to provide an intuitive user interaction with the application, WellCo provides **speech interaction by means of an affective aware virtual coach** that is always active in the device (that could be de-activated on the settings) and **Natural Language technologies**, so WellCo will be able to understand user's daily-life conversation in different languages and guide the user through advice and recommendations (de-activation is always possible, and instead normal interaction through touch screen could be used).

Regarding user acceptance, to ensure the usability and personalisation of the platform, **WellCo design will be developed jointly with technical, business and end user partners through all the project life** (starting from the needs identification prior to the proposal phase). On tasks T2.4 WellCo **Co-design will be developed, and mock-ups** are expected **to be shared and designed together with the set of users, involved also in the requirements phase.**

- **Objective 4 (O4). Cost-effective analysis to maximize the quality and length of life** in terms of activity and independence for people in need of guidance and care due to age related conditions because of self-care, lifestyle and care management.

  Evidence suggests that **self-management**, especially for people with long-term conditions**, can be effective through behavioural change and self-efficacy** (for example for diabetes patients) and may reduce drug and treatment costs and hospital utilisation, which is translated on savings for the National Health Systems. **WellCo** will aim to support this evidence by **sharing project results and ensuring open access to all peer-reviewed scientific publications** as well as **research data supporting them**, as long as it respects the **balance among openness and protection of scientific information, commercialization and IPR, privacy concerns, security and data management and preservation questions**.

2.3 Types of data generated, collected and processed in WellCo

As extracted from the previous sections of WellCo, different types of data coming from multiple data sources will be available in WellCo. Mainly this data will consist of:

- **Static Data (O1&O3)** needed to perform a static modelling of the user, i.e.:
  - **User profile information**
  - **Life Plan information** - different areas of a user's life like health, work, community involvement, relationship with friends and families, etc.
- **Dynamic Data (O2&O3)** needed to dynamically model the user and adapt the recommendations to the social, cognitive and environmental context surrounding him/her.
  - **Wearable Bracelet**

        - **TicWatch S**[1] – heart rate, steps, distance, calories, sleep quality, GPS and accelerometer.
        - **Nokia Steel HR**[2] - heart rate, steps, distance, calories, sleep quality.
    - **Personal Smartphone/ Tablet Device**
        - Record visible WiFi access point;
        - Localisation via GPS;
        - Counting of number of SMS and phone calls sent / received / missed (no actual content of SMS or phone calls will be stored);
        - Patterns of use of specific app categories (e.g. social media, browsing, email, photography, etc.). WellCo will never track individual apps to ensure preservation of privacy;
        - Lastly screen on / off events that could provide interesting input in assessing mental state (e.g. anxiety, stress, sleep quality);
        - Record ambient sound (extract features in real time, no storage) – Affective Computing;
        - Record video (extract features in real time, no storage) – Affective Computing.
    - **Patient Report Outcomes;**
        - Self-reported nutrition, physical activity, sleep, stress etc.
    - **Expert and Informal Caregivers reported Outcomes**

These **data will be originated by the target users involved in each of the trial sites defined in WellCo**, Denmark (DK), Italy (IT) and Spain (ES) on the part of SDU (DK), FBK (IT) and GSS (ES). For more information about the sample size and enrolment procedures of these users, please see D2.1 User Involvement Plan.

The data previously originated in trial sites will be **collected, processed and stored according to three phases** that define the core of WellCo – co-design, implementation and validation. These phases suppose an enlargement of the initial phases described in D2.2 Ethics, Gender and Data Protection Compliance Protocol and that only covered the collecting, processing and storing of data from beneficiaries in charge of trial sites, i.e. FBK, GSS and SDU. A new phase has been included that aims to cover the management of data by technical partners in order to handle data as part of the work they perform for the implementation of algorithms and technologies that ensure the provision of effective personalized recommendations.

#	Phase	Description	Partners involved
1	Co-Design	The first phase, consists of requirements gathering and concept development of WellCo. The data from participants will be captured, stored and processed by the personal involved in trial sites according to the ethics measures defined in D2.2.	FBK, GSS, SDU
2	Implementation	The second phase will consist on the collection and processing of the data derived from the profile and	FBK, JSI, UCPH,

[1] www.mobvoi.com/eu
[2] health.nokia.com/es/en/steel-hr

	monitoring of each of the users involved in trial sites in Spain, Italy and Denmark. This processing will allow the development of the modules described in WP4 and WP5 (see figure 1)	MONSENSO , HIB, CON
3 Validation	The third phase will be the validation of each of the three prototypes envisaged in WellCo in the different trial sites defined in the project. The data from this validation will be captured, stored and processed by personnel involved in these trials and provide it to technical user with the aim of improving the coming prototype.	FBK, GSS, SDU

Table 1.- Data Collecting, Processing and Storage phases.

In order to clarify phase 2 of the table above and with the aim of determining the relation between the processing of the data and the achievement of the different goals expected in WellCo, the initial architecture design is included in the picture below with the aim of offering a clearer view of how the data available in WellCo will feed each of the modules composing the architecture.

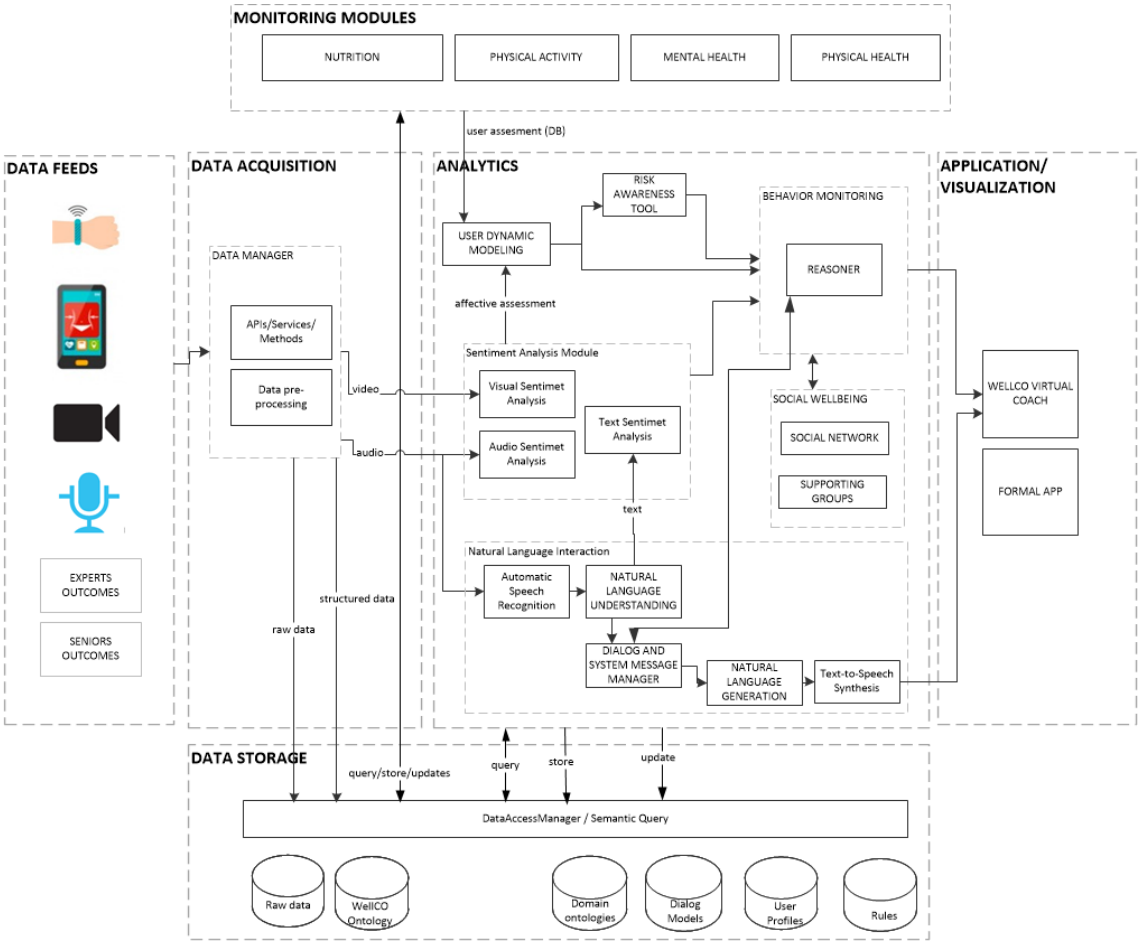

Figure 1.- WellCo Platform conceptual architecture and main components.

As already mentioned, this is a living document so it is expected that the figure above changes along the project lifetime since the first version of this document has been delivered in M6, i.e. when WP3 WellCo Prototyping and Architecture has just started.

3 WellCo: Ethical Issues

As part of the engagement on ethics, the WellCo consortium has been committed to ensure that ethical principles and legislation are applied in the scope of the activities performed in the project from the beginning to the end. For this reason, the consortium has identified relevant ethical concerns already during the preparation of the project proposal and, then, during the preparation of the Grant Agreement. During this phase, ethics issues have been already covered as part of D2.2 Ethics, Gender and Data Protection Compliance Protocol and later, in D7.1 POPD Requirement No.2 and D7.2 H – Requirement No.3.

In the context of this deliverable, it can be determined that the ethical issue that could have more impact on data handling and sharing during and after the project is that regarding privacy and data protection issues, especially relevant because of the entry into force of the new General Data Protection Regulation this month that establishes a common framework for data protection in Europe.

The following section aims to describe how the founding principles of the new European Regulation on Privacy, the General Data Protection Regulation, (GDPR), will be followed in the WellCo consortium. Then, these principles will be used in the coming section to set out specific guidelines for accurate and compliant use of personal data within the boundaries of the GDPR. It is important to mention that this deliverable is a living document and as far as GDPR-related developments are clearer, further details will be included in it. Additionally, it is important to note that some of the details of the data management implementation are also mentioned within deliverable D2.2 “Ethics, Gender and Data Protection Compliance Protocol”.

3.1 Alignment with the GDPR

This deliverable will describe how the data will be handled during and after the project. 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 accurate and compliant use of personal data. On the other hand, it means that in some cases, tools and partner specific guidelines are not yet available. This deliverable is a living document and as far as GDPR-related developments are more clearer, further details will be included in it. Additionally, it is important to note that some of the details of the data management implementation are also mentioned within deliverable D2.2 “Ethics, Gender and Data Protection Compliance Protocol”.

In this chapter we will describe how the founding principles of the GDPR will be followed in the WellCo consortium. Then we will set out specific guidelines for accurate and compliant use of personal data within the boundaries of the GDPR.

3.2 Lawfulness, fairness and transparency

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

All data gathering from individuals will require informed consent of the test subjects, participants, or other individuals who are engaged in the project. Informed consent requests will consist of an information letter and a consent form (generic template in an appendix of D2.2 “Ethics, Gender and Data Protection Compliance Protocol”). This will state the specific causes for the experiment (or other activity), how the data will be handled, safely stored, and if/how 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.

Along the project, we will try to anonymize the personal data as far as possible, however we foresee this will not be possible for all instances; some data will be pseudo-anonymized where the identity of the participants will not be known to researchers, but based on the data content collected one may get back and discover this identity.

A specific consent will be acquired to use the cumulative data for open research purposes; including presentations at conferences, publications in journals as well as, once accurately anonymized, depositing a bulk data set in an open repository at the end of the project. This clause is included in the informed consent form.

The consortium is going to be as transparent as possible in the 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 if/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 their data to be removed from the cumulative results.

3.3 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 WellCo project will not collect any data that is outside the scope of the project. Each researcher will only collect data necessary within their specific work package and task activity (see Section 4.2).

3.3.1 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 state assessment and coaching activities will be collected. However since participants are free in their answers, both when using the WellCo coaching 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 participants in their answer possibilities; we will rather limit the scope of the data being processed to the minimum one necessary for coaching to work.

3.3.2 Accuracy

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

All the collected data will be checked for consistency and will be stored with the metadata for which timeframe that data applies; for example "age" could be stored as "age in 2018" and once captured, would be automatically updated. However since some of the dataset register self- reporting data from the participants, we cannot check this data for accuracy.

3.3.3 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 project purposes will be deleted as soon as possible. All personal and sensitive data will be made anonymous as soon as possible. At the end of the project, if the data has been accurately anonymized, then it will be stored in an open repository. If data cannot be anonymized, the pseud-

anonymized datasets will be stored for a maximum of the partner's archiving rules within the institution. For example, a complete data set will be archived at the UCPH for 10 years, according to its data policy. Each partner has its individual data policy.

3.3.4 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:

- Along phase 1 and 3 of the project, data sets with personal data will be stored at dedicated servers at the trial sites (DK, IT and ES) complying with all GDPR regulations. Decisions with respect to data storage for the project's phase 2 (and beyond) will be made accordingly.
- Access to these servers will be managed by the project controller and will be given only to authorized individuals who need to access data for accomplishing the tasks within WellCo. Access can be retracted if necessary.
- In some cases pseudo-anonymized data sets can further be shared through the WellCo Alfresco platform and code repository by HIB, only if the datasets are sufficiently encrypted. The key to the encryption will be handed out by the collaborating parties and will be changed when access needs to be revoked.
- All WellCo collaborators with access to the identifiable, non-anonymized personal data will need to sign a confidentiality agreement, i.e., the "Contract for Data Controller" as defined in D2.2.
- None of the WellCo datasets can be copied outside of the secure servers, 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 accurate and compliant 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.

3.3.5 Accountability

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

There is no one responsible for the data management in the project; we assume a role of separate Data Controller at each of the trial sites, controlling the same data types across the trials sites (D2.2). Furthermore, at project level, the project management is responsible for the accurate data management within the project. In the next section, guidelines will be described for each partner to follow in case of datasets with personal and sensitive data. The project management will regularly check whether the partners follow these guidelines. For each data set, a responsible data Controller has to be appointed at the partner level. This person is held accountable for this specific data set.

4 Guidelines for Data Protection and Security

As part of the above principle, a guideline for data protection and security has been established in this section with the aim of ensuring that all researchers keep up the

principles of lawful and ethical data management along the whole project duration and after. The guidelines established in this DMP are embraced within the consortium and the project management will ensure these principles will be followed.

It is important to highlight that, because of the fact that the first version of this DMP is published at M6, when there are still many uncertainties about the data collected in the project, additional details are going to be inserted in here along the project progress, as well as given within D2.2.

4.1 Purpose limitation and data minimisation

Researchers will apply the principles of purpose limitation and data minimisation to the different types of data defined in section 2.3. Each researcher will take care not to collect any data that is outside the scope of his/her research and will not collect additional information not directly related to the goal of his/her research.

4.2 Personal information

As soon as the parameters in the data-sets defined in section 2.3 are identified, the researchers need to indicate whether the data set will contain personal information.

In cases where the parameters themselves contain 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 is co-called pseudo-anonymized and 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 trial site server, or encrypted on Alfresco and/or in a common code repository.
- 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) to which only WellCo authorized staff has access. This applies to informed consent collected in paper form or documents generated along the user requirements phase (e.g., results of the brainstorm with the users). Once the data has been digitised, the physical copies shall be securely destroyed.

4.3 Anonymisation and pseudo-anonymisation

The data controller will make sure the personal data is anonymized as quickly as possible after its collection. When the data cannot be anonymized completely, it will be pseudo-anonymized as much as possible – the personal identifier must be stored separately. The authorized personnel, data controllers, will store the key between the pseudo-anonymized files and the list of participants. The key will be stored in a separate physical location from the original files. We keep in mind that the research subjects should be able to withdraw their data completely from the WellCo at any point in time, hence the key must be stored securely but be feasible to be accessed.

Part of the WellCo platform relies on 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 (M6) it is undecided: we are looking into the different possibilities of hosting a server at each trial site (DK, IT, ES), as well as assuring that each technical partner (MONSENSO, UCPH, FBK, HIB, CON, JSI) has its own GDPR-compliant server. As far as for the client side technology, we are looking into the possibilities of pseudo-anonymizing the client side, e.g., the tablet or a smartphone on which the app runs, may be serving as a random, yet unique identifier for the project. However, the implications of the privacy-by-design provisions in the GDPR cannot be settled up front

and will be contributed to this document along research and development in WP3 and WP4.

4.4 Informed consent

When collecting personal information, researchers are required to get informed consent from the study participants. In D2.2 we provided a standardized EU informed consent template, which can always be supplemented with additional consent requests, depending on the project stage, time involvement, as well as risks and benefits of the involvement.

Consent should cover all processing activities carried out for the same purpose or purposes. When the proposing has multiple purposes, consent should be given for all of them.

4.5 End users' rights

The user can submit a request to see which information about him/her is being kept on our files through the contact person on the consent form. He/she can request to delete his information up to 48 hours after the experiment has taken place. Furthermore he/she can request that no additional data collection will take place starting immediately from the time of request.

4.6 Storage and researchers' access to data

Personal and sensitive user data will be stored safely and in a secure environment; potentially at each trial site. Backups are an important aspect of the server management and shall also be GDPR compliant. For example backup of secure servers at UCPH are made every 24 hours by the system itself. A common security protocol will be established once the project reaches the maturity level, for all the partners storing personal data (defining authentication, authorization and encryption; protection against unauthorized access, internal threats and human errors, etc.).

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 contributing to WellCo activities. All users that are granted access to the datasets will need to sign a Data Protection Contract (see Appendixes of D2.2). Access can be restricted or revoked, when researchers are not complying with the guidelines or when their contract is terminated.

4.7 Encryption

When researchers want to share personal data files through Alfresco and/or the code repository, 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 to participating parties, however as secure as needed. Possibilities for encryption as build in Word and Excel or can leverage PGP keys (more advanced option).

If a scientist keeps data files with personally identifiable data on own personal computer or on a separated hard drive for data analysis purposes, he/she has to use BitLocker of FileVault for the encryption of the hard drive.

4.8 Open data and FAIR principles.

Within the WellCo project, we endorse the European Commission's motto: "*to make the data as open as possible, but as closed as necessary*". We are committed to protect the privacy of the participants 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 (having the research data findable, accessible, interoperable and reusable) 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, also in this document, once more information on the data sets comes to light. So far we discuss the FAIR principles along each WP activities and tasks (c.f., Section 3).

4.9 Privacy statements

WellCo will actively communicate the privacy and security measures it takes through all media channels (from consent forms to websites) with a privacy statement. We will adjust the statement to fit the target group, purpose, and level of privacy.

4.10 Update of the DMP

The DMP deliverable 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 WellCo Data Management Plan Details

Within this section the work package leaders 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 European Commission 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.

It is important to notice that, as long as it is possible from a privacy point of view, it is our intention to make all the below-mentioned written data openly available in order to validate the data presented in scientific publications and on a voluntary basis. Only 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 will be excluded.

5.1 Provisional FAIR Guidelines for WellCo Data Sets

5.1.1 Findable

Digital Object Identifier (**DOI**) is a unique alphanumeric string assigned by a registration agency (the International **DOI** Foundation) to identify content and provide a persistent link to its location on the Internet. The publisher assigns a **DOI** when an article is published and made available electronically.

As already specified within the GA Article 29.2, with respect to the open access for the peer reviewed publications, the bibliographic metadata must be in a standard format and must include all of the following: the terms “European Union (EU)” and “Horizon 2020”; the name of the action, acronym and grant number; the publication date, and length of embargo period if applicable, and a persistent identifier, e.g., a DOI.

Each dataset within the WellCo project will get a unique Digital Object Identifier (DOI). If/when the data set will be stored in a trusted repository the name might be adapted in order to make it more findable. To construct a DOI, we may assign it a name containing three elements along the pattern *UserModel.WellCo-data-set.datasetID.version.WellCo_controller*, where *UserModel* is the logical name that is associated with the user state assessment component (e.g., physical, mental health), *WellCo-data-set* is the data set name, and *datasetID* and *version* are assigned by the *WellCo_controller* (i.e., a specific project partner).

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

5.1.2 Accessible

As described before, our intention is to open up as many WellCo data as possible. However, if we cannot guarantee the privacy of the participants by accurate anonymization of the data or the IPR of the owner beneficiary are under risk, the data set might be opened up under a very restricted license or it will remain completely closed. This document will be updated along the project development with which data will be made accessible and which not as well as the reasons for opting out.

For those project results to be made openly available, WellCo will adhere to the pilot for open access to research data (ORD pilot) adopting an open access policy of all projects results, guidelines and reports, providing on-line access to scientific information that is free of charge to the reader. Open access will be provided in two categories: **scientific publications** (e.g. peer-reviewed scientific research articles, primarily published in academic journals) and **research data** (Subsections below).

Open access to scientific publications

According to the European Commission, "under Horizon 2020, each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results" (see also Article 29.2 of the GA). The WellCo Consortium adheres to the EU open access to publications policy, choosing as most appropriate route towards open access **self-archiving** (also known as "**Green Open Access**"), namely "a published article or the final peer-reviewed manuscript is archived (deposited) in an online repository before, alongside or after its publication. Repository software usually allows authors to delay access to the article ("embargo period"). The Consortium will ensure open access to the publication within a maximum of six months.

The dissemination of WellCo results will occur by mean of activities identified in the initial plan for exploitation and dissemination of results (PEDR), such as creation of the web page for the project, public workshops, press releases, participation in international events, etc. In compliance with the Grant Agreement, **free-online access will be privileged for scientific publication**, following the above-mentioned rules of "green" open access. All relevant information and the platform textual material (papers, leaflets, public deliverables, etc.) will be **also freely available on the project website.** In order to guarantee security, this textual material will be available in **protected PDF** files. In specific cases and according to the rules of open access, the dissemination of research results will be managed by **adopting precautionary IPR protection protocols,** not to obstacle the possibility of protecting the achieved foreground with preventive disclosures.

Open access to scientific publications (Open Research Data Pilot)

According to the European Commission, "research data is information (particularly facts or numbers) collected to be examined and considered, and to serve as basis for reasoning, discussion, calculation". Open access to research data is **the right to access**

and reuse digital research data under the terms and conditions set out in the Grant Agreement.

Regarding the digital research data generated in the action, according to the Article 29.3 of the GA, the WellCo Consortium will:

Deposit in a research data repository *and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate – free of charge for any user – the following:*

i. The data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

ii. Other data, including associated metadata, as specified and within the deadlines laid down in this data management plan;

iii. Provide information – via the repository- about tools and instruments at the disposal of the beneficiaries and necessary for validating the results.

WellCo Consortium will make a great effort, **whenever possible**, to make this research data available **as open data or through open services**. It is important to note that because of the low maturity of this document and some existing uncertainties about the data collected in the project, additional details are going to be inserted in here as the project progresses.

5.1.3 Interoperable

We are considering generating project specific ontologies in order to normalize and make data from different sources interoperable. Additionally we consider suitable metadata standards, for example: DataCite[3]. Depending on the scientific field where the data set will originate from, additional meta-data standards might be used.

5.1.4 Reusable

When 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 or the scientific publication. A period of embargo will only be necessary if a data set contains specific IPR or other exploitable results will justify an embargo. The length of embargo will be negotiated on an individual basis.

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 (see section 5.1.1).

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 practices that we foresee will be replaced soon, these set will not be stored for eternity. Furthermore data sets collected leveraging specific technological solutions, might become out-dated, which will also limit their time of reusability.

5.2 DMP within WellCo Work Packages

The following section represents a work in progress where a FAIR approach, allocation of resources, data security, ethical aspects and other issues will be detailed along each work package tasks and activities.

[3] https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf

WP2: Co-design (GSS, M1-M36)

DMP Component	Issues to be addressed
1 Data Summary	• 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

Answers

In T2.2 and T2.3: datasets on user requirements to serve a proper development of WellCo (technical and functional requirements), and additionally to elaborate on WellCo personas (thus, lifestyle and lifestyle patterns, main concerns in well-being, and personal goals) and to describe and validate scenarios, wireframes and user journeys.

In T2.4.: datasets on validation and feedback of users regarding a clickable mock-up, prototype 1, prototype 2 and prototype 3 (final version of WellCo) and feedback of users in order to measure the success of the project.

Specific datasets are:

- Notes and minutes of brainstorming, workshops, focus groups (.DOCX)
- Recordings and notes from interviews with stakeholders (.DOCX)
- Cultural probes: data form the user´s filled diaries, WhatsApp messages sent, personal interviews about users' feelings in the cultural probes process.
- Reports after individual interviews on a questionnaire for technical and functional requirements.
- Reports after individual interviews to offer feedback on wireframes and user journeys
- Reports for personal feedback on the clickable mock-up, prototype number 1. 2 and 3
- Reports on monitoring through wearable devices
- Reports for personal feedback on success of the project
- Ex-ante and Ex-post evaluations referred to the participants in the test trials

Files are pseudo-anonymized and stored in for example in .DOCX, .PDF, .XLSX formats

Size: ±100MB so far

2.1 FAIR: Findable	• 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

Answers

The metadata associated with each dataset:

- Organization name, contact person
- Type of activity where data was collected, date

Further metadata might be added at the end of the project in line with meta data conventions. No deviations from the intended FAIR principles are foreseen at this point.	
2.2 FAIR: 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
Answers No data is going to be publically available at this point. No deviations from the intended FAIR principles are foreseen at this point.	
2.3 FAIR: 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?
Answers Data is stored in interoperable format (DOCX) that can be opened by anyone authorized to do so. No deviations from the intended FAIR principles are foreseen at this point.	
2.4 FAIR: Reusable	• 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
Answers N/A at this stage	
3. Allocation of resources	• 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
Answers The work to be done in making the data FAIR will be covered by the assigned budget for producing the deliverables.	

4. Data Security	• Address data recovery as well as secure storage and transfer of sensitive data

Answers

The original data is stored in a dedicated trial site server. Namely, handwritten notes will be stored under lock in the offices of the trial site owner (FBK, GSS and SDU) in a physical storage space separate from the participant lists of workshops and interviewees. The pseudo-anonymized data (interview summaries, co-design reports) are shared on Alfresco (managed by HIB).

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 WellCo project.

Machine-written notes (i.e. data files in .DOCX and .XLSX format) will be stored in Alfesco space provided by HIB. Access is granted in line with the project's procedures.

All the data will be destroyed, once the research has end, thus the project has end. Once destroyed, the data processor must certify their destruction in writing and must deliver the certificate to the data controller.

Additionally, GSS has to follow the procedure described in the document *Report On The Security Measures To Be Adopted For The File "UNIQUE RECORD OF USERS OF THE SOCIAL RESPONSIBILITY SYSTEM*". This document states, i.e. "Personal data will be cancelled when they are no longer necessary for the purpose for which they were collected or registered. However, they may be kept for as long as any type of responsibility can be demanded, but in any case it must be determined".

5. Ethical Aspects	• 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

Answers

Ethical consent has been acquired from the participants so far.

Ethical approval for the studies, is under evaluation (UCPH to date). GSS does not require the ethical approval.

Additionally, HIB, as leader of WP7, will guarantee the compliance of the Ethical deliverables from within WP2.

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

Answers

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

WP3: Prototyping And Architecture (HIB, M6-M30)

DMP Component	Issues to be addressed
1 Data Summary	• 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

Answers

This WP will collect, pre-process and store all the research data derived from the monitoring of the user in a centralized server where it will be anonymized as long as possible. The collection of this data will allow an initial pre-processing of it. The idea behind this pre-processing is to enable the normalization of this data in order to be interoperable with the rest of modules of WP4 and WP5 where a complete processing will be performed. This WP will also:

- Generate and store deliverables D3.1, D3.2, D3.3, D3.4 and D3.5 in the common repository in Alfresco. D3.2 is a public document so it will be also available in the project webpage;
- Code for prototypes as well as system logs will be shared in the common code repository;
- Intermediate documents for requirements and architecture design will be shared through the common repository in Alfresco.

The previous collection/generation of research data will allow the re-use of this normalized data to: on the one hand, help to develop WellCo as a novel ICT based platform for useful and effective personalised recommendations and follow-up in terms of preserving or improving wellbeing (O1, as in Section 2.2) and, on the other hand, contribute to the validation of non-obtrusive technologies for physical, cognitive, social and mental wellbeing (O3)

Deliverables will be in .DOCX and .PDF. The format for the research data has not been decided yet.

Data in this module will be re-used by modules in WP4 and WP5. Also, after being normalized and anonymized, and as soon as it does not affect data protection or IPR, this data will be made open in ORD.

This research data will be originated in the smartphone/tablet and wearable devices worn by the users participating in trials in Spain, Denmark and Italy. Deliverables and code will be originated by the beneficiaries participating in this WP, i.e. HIB, FBK, UCPH, JSI, CON, MONSENSO.

State the expected size of the data is not known yet.

As already explained, this data, once normalized, will serve as input for the modules implemented in WP4 and WP5.

2.1 FAIR: Findable	• 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

Answers

All WellCo datasets will use a standard format for metadata according to the described in section

5.1.1. Further metadata might be added at the end of the project in line with meta data conventions. Each dataset within the WellCo project will get a unique Digital Object Identifier (DOI). If/when the data set will be stored in a trusted repository the name might be adapted in order to make it more findable

Identifiability of data is already explained above.

The naming conventions for deliverables are described in the project handbook for the project. The naming conventions for datasets are explained in section 5.1.1.

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.

The version will be included as part of the naming conventions.

2.2 FAIR: 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

Answers

Due to the initial stage of the project, there is still some uncertainty on the specific data to be handled. Deliverables will be shared around consortium partners in Alfresco repository and those, which are public, will be available in the project webpage. Regarding datasets, as gathered in the GA and along the whole document, they will be made open as soon as they do not represent a risk for IPR and data protection.

For those project results to be made openly available, WellCo will adhere to the pilot for open access to research data (ORD pilot).

Methods or software tools needed to access the data are not known yet.

The consortium will decide, and specify where the data and associated metadata, documentation and code are deposited

The consortium will decide, and specify how access will be provided in case there are any restrictions.

2.3 FAIR: 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?

Answers

Deliverables will be delivered in .PDF format in order to ensure that the format is always kept. Regarding datasets, as part of this WP, ontology will be designed with the aim of performing an initial pre-processing that enables the normalization of this research data.

Use of standard vocabulary for all data types present in our data set, to allow inter-disciplinary interoperability is mentioned above.

2.4 FAIR: Reusable	• 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
Answers Whenever possible, the datasets will be licensed under an Open Access license The datasets will be made available for re-use twelve-months later to project end, or on partner-by partner basis, as agreed with all the project partners. In the case of deliverables in WellCo for this WP, they will be stored in Alfresco and published in the project web page as soon as they are delivered in the EC Portal (without any embargo period). As explained in section 4, the 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. HIB as responsible for this WP will perform a quality control of the datasets processed in this work package by editing and moderation, cleaning, pre-processing, adding metadata, transforming to a more convenient format or providing easier access. The datasets will be available for reuse till the quality assurance tasks performed over each data determines that these datasets are out-dated	
3. Allocation of resources	• 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
Answers The work to be done in making the data FAIR will be covered by the assigned budget for producing the deliverables.	
4. Data Security	• Address data recovery as well as secure storage and transfer of sensitive data
Answers HTTPS will be used as application protocol for WellCo. HTTPS is an extension of HTTP for secure communication over a computer network; Transport layer Security (TLS) or Secure Sockets Layer (SSL) encrypts it. Moreover the system also includes: • Authorization and authentication processes; • Periodic backups of the databases and the code; • Firewall inspection trough White Lists; • Intrusion detection and prevention mechanisms.	
5. Ethical Aspects	• 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
Answers The guidelines for data protection and security defined in Section 3 will be followed for the data available in this WP. Some of the aspects that will be covered are: data minimization, protection of personal information through anonymisation and pseudo-anonymisation, rights for the user to give his/her consent and to ask for access to his/her data, rectification of data, removal and portability.	
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Answers

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

WP4: Physical, Cognitive And Mental User Assessment (UCPH, M1-M21)

DMP Component	Issues to be addressed
1 Data Summary	• 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

Answers

To design, implement and evaluate the WellCo user assessment services we will collect the following data

- User state assessment model specification (.DOCX, .XLSX, .PDF) and implementation
  - Self-assessed variables or data collected via a wearables dataset (Heart rate, Steps, Distance, Calories, Sleep quality, Accelerometer, Gyroscope and Magnetometer); estimated size: 10 MB/day
  - Potentially Smartphone datasets (WiFi patterns usage, applications usage, GPS, ON-OFF and ambient sound measurements); estimated size: 10MB/day
  - Behavioural features, derived from the above sources like "step counts", “hours of sleep”; estimated size: few kB-1MB/day
- API designs for wearables and smartphones dataset (.DOCX, .PPT)
  - 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).
- System logs (performance, debugging, benchmarking of service quality)
  - Stored on device: SQLite format
- Source code (Java, Python, PHP, .APK, etc.)

2.1 FAIR: Findable	• 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

Answers

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

2.2 FAIR: 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
Answers To access the data we will likely leverage MySQL technology. Some examples of open source options are: DBeaver, SQLelectron or SequelPro. Accessibility of the data for others will only be provided if we assure that the data is anonymized and based on it will not be possible to identify or retrace a person (for instance through location tracking).No deviations from the intended FAIR principles are foreseen at this point.	
2.3 FAIR: 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?
Answers We use standard models for encoding the data (e.g., JSON, CSV). No uses of specific ontologies are planned so far. No deviations from the intended FAIR principles are foreseen at this point.	
2.4 FAIR: Reusable	• 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
Depending on the content of the data set and whether it contains personal information, re-use by third parties could be possible. No deviations from the intended FAIR principles are foreseen at this point.	
3. Allocation of resources	• 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.	
4. Data Security	• Address data recovery as well as secure storage and transfer of sensitive data
Answers An anonymized universal unique identifier will be used to identify the data collected from each user; it will not be possible to reveal the identity of the user solely based on this ID. However there might be a combination of data possible, with which you can identify a person, for example 24hour location tracking.	

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. In all cases data will be stored according to the project's guidelines on personal data.	
5. Ethical Aspects	• 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
Answers The end user will receive an information leaflet and will sign a consent form. This way we ensure the individual 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	• 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.	

WP5: Behaviour Modelling And Lifestyle Coach (JSI, M8-M29)

DMP Component	Issues to be addressed
1 Data Summary	• 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

Answers

To design, implement and evaluate the WellCo Behaviour Modelling and Lifestyle Coach, the following data will be collected:

- Features extracted from an initial pre-processing of video in real time. Video is never stored. These data will be only shared in case of need to support peer-reviewed scientific reports
- Data for speech emotion analysis (affective computing) – recorded sound– saved or process in real-time
- Data from WP4 (physical activity, nutrition specifications, mental assessment, behavioural features) and sentiment analysis will be used for dynamic user modelling-
- Data directly gathered from the wearable bracelet.
- Sensors and data embedded in the smartphone or tablet device of the user.
- Static data of the user such as Profile Information, Life Plan and Reported Outcomes and Expert/Informal caregiver reports. These data will be shared after anonymisation.
- Above mentioned will be used to provide personalized recommendations to the user through the virtual coach in order to ensure the adoption and maintenance of healthier behaviour change habits as gathered in Objective 1 (O1, Section 2.1.) of WellCo.

Although the format and synchronization of these data have still to be decided, we are considering the possibility of having specific ontologies in order to normalize data formats and make them interoperable among the different modules of WP5.

Regarding the re-use of the data in this WP, we plan to make them as open as possible. Because of this, as the project reaches maturity and we have more certainty about the data, we will define some measures to ensure that IPR and data privacy is taken into consideration by design as well as which data is feasible to be made open without prejudice to the foregoing. The uncertainty about this data makes difficult to determine the expected size of this data as well as to define to whom will it be useful.

2.1 FAIR: Findable	• 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

Answers

In case we decide to publish the anonymized dataset for speech sentiment analysis, data will be provided as audio recordings or files, containing the extracted features (CSV or ARFF file formats). Annotations will be provided as CSV files. That coincides with standard practice regarding to publication of recorded speech datasets.

However, due to the uncertainty about the data to be shared in this module, there is not yet a final decision about how we plan to make these data findable. For sure, we will use a standard format for metadata and naming as is already described in section 4. Further metadata might be added at the end of the project in line with these meta data conventions.

2.2 FAIR: 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
Answers Following the ideas described along the project, since the initial stage of the project, there is still some uncertainty on the specific data to be handled. Datasets will be made open as long as they serve as support to scientific publications in the project and also under anonymized basis, considering that neither IPR or data privacy of users from which this data was originated are at risk.	
2.3 FAIR: 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?
Answers Data will be stored in standard formats, such as WAV files for recorded audio, CSV and ARFF for metadata and annotations, some data may be in a database, such as MySQL. The interoperability of data will be made possible thanks to the use of ontologies that will ensure that data is converted to common formats that enable interoperability both among the different modules in this WP and the scientific community when making them open.	
2.4 FAIR: Reusable	• 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
Answers As already mentioned, whenever possible, the datasets will be licensed under an Open Access license. Once we decide which data in this WP is reused we will establish quality assurance measures to ensure that all datasets in this WP are cleared of bad records, with clear naming conventions, and with appropriate meta- data conventions applied as well as the responsible for this.	
3. Allocation of resources	• 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
Answers	

The work to be done in making the data FAIR will be covered by the assigned budget for producing the different modules collecting and processing these data.	
4. Data Security	• Address data recovery as well as secure storage and transfer of sensitive data
Answers N/A at this moment.	
5. Ethical Aspects	• 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
Answers We will ensure transparency by making data subjects aware of the type of data collected and processed in WellCo as well as which of these datasets will be shared, always after a complete anonymisation process, in Open Research repositories. Informed consent will be always required before performing any of these actions. These features are quite interesting for the case of recording speech data of the interaction of the user with the virtual coach in early prototypes– in the user's normal environment, not in a laboratory, the collected data may include personal information.	
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
Answers No other procedures need to be put in place for project management data.	

WP6: Dissemination and Exploitation (CON, M2-M36)

DMP Component	Issues to be addressed
1 Data Summary	• 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
Answers The following data is going to be considered: • Conference/journal publications (.PDF) • Exploitation plan (.DOCX, .PDF) • Standardization activities (.DOCX, .XLSX) • Dissemination materials (.PPT, .PDF, .JPG, videos) including website (.html) with embedded content, as well as connected to Google Analytics to evaluate its reach	
2.1 FAIR: Findable	• 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
Answers Data related to dissemination and exploitation will be findable –to the best of the consortiums' capacity- utilizing digital communications best practices, e.g. hashtag, metadata, keywords. In social media, WellCo posts will be findable and discoverable by the name, while for posts to different media (e.g. 3rd party blogs), the posts will refer to the project website. At this moment we foresee no separate datasets to be posted in repositories at the end of the project.	
2.2 FAIR: 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
Answers Most of this data will be made public, although there might be made an exception when it comes to data concerning the project exploitation. We foresee most data will be published online, just not in online repositories, since it does not contain specific research data.	

2.3 FAIR: 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?
Answers This is not applicable for data related to dissemination and exploitation.	
2.4 FAIR: Reusable	• 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
Answers The data related to dissemination and exploitation will be reusable. The reference to original materials will be kept.	
3. Allocation of resources	• 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
Answers The work to be done in making the data FAIR will be covered by the assigned budget for producing the deliverables.	
4. Data Security	• Address data recovery as well as secure storage and transfer of sensitive data
Answers This is not applicable for data related to dissemination – containing only the cumulative, anonymized data representation. For the WellCo website visitors, privacy statement will be provided.	
5. Ethical Aspects	• 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
Answers All participants in the consortium have agreed with posting their pictures online for dissemination items and project updates.	
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Answers

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

6 Conclusive Remarks

This deliverable provides a description of the data management strategies taken in account during the project. It describes and outlines the existing regulations to which WellCo must comply, and defines how data will be collected, stored, shared and most important protected. Important measures are mentioned about the protection of the data, which should be taken into account during the project. This is a “living document” and an update will be provided no later than in time for the first review (M12). Other updates will be provided at M24 and M36 detailing which/how the data will be made available to others within the Pilot on Open Research Data (ORD).