Ageing@Work

Smart, Personalized and Adaptive ICT Solutions for Active, Healthy and Productive Ageing with enhanced Workability

Project Acronym:	Ageing@Work
Project Full Name:	Smart, Personalized and Adaptive ICT Solutions for Active, Healthy and Productive Ageing with enhanced Workability
Grant Agreement:	No 826299
Project Duration:	3 years (starting 1 January 2019)

Deliverable 9.4

Data Management Plan – v1

Work Package:	WP9: Project Coordination and Management
Task:	T9.3 Data Management, Ethics, and Standardization
Lead Beneficiary:	SAG
Due Date:	30/06/2019 (M6)
Submission Date:	28/06/2019
Deliverable Status:	Final
Deliverable Style:	ORDP (Open Research Data Pilot)
Dissemination Level:	Public
File Name:	D9.4 Data Management Plan.pdf

Authors

Surname	First Name	Beneficiary
Görlich	Jürgen	SAG
Dubielzig	Markus	SAG
Bautista Montalvá	Juan	UPM
Zacharaki	Evangelia I.	UPAT
Mockałło	Zofia	CIOP-PIB
Balla	Christina	Q-PLAN
Paliokas	Ioannis	CERTH
Wissemann	Ann Kathrin	ASER
Serafin	Patrick	ASER

In case you want any additional information or you want to consult with the authors of this document, please send your inquiries to: info-ageingatwork@iti.gr.

Reviewers

Surname	First Name	Beneficiary
Abril-Jimenez	Patricia	UPM
Moustakas	Konstantinos	UPAT
Nousias	Stavros	UPAT
Paliokas	Ioannis	CERTH

Version History

Version	Date	Modifications made by
0.1	13/05/2019	First draft template created by Jürgen Görlich (SAG)
0.5	21/06/2019	Integration of the descriptions of all foreseen datasets by Jürgen Görlich (SAG)
0.6	24/06/2019	Provision of introductory remarks for chapter 3 by Jürgen Görlich (SAG)
0.7	25/06/2019	Provision/revision of chapter 1 and 2 by Markus Dubielzig (SAG)
0.8	26/06/2019	Revision of Executive Summary and Introduction by Markus Dubielzig (SAG)
0.9	27/06/2019	Integration of the comments from internal Peer Review (SAG + WP9 Partners)
1.0	28/06/2019	Final Version (SAG)

Disclaimer

Executive Summary

This deliverable is the first version of the Data Management Plan (DMP) of the Ageing@Work project in accordance with the EC Open Research Data Pilot (ORDP) regulations of the Horizon 2020 programme (H2020) and will be updated as a living document throughout the project. With regard to the generated project data, it contains preliminary information on the open research data retention and destruction strategy as well as considerations of the limits on the secondary use of the data and its disclosure to third parties.

In order to develop the present deliverable, a "Data Identification Form" template (see Annex A) was first drafted on the basis of the EC guideline document "Template Horizon 2020 Data Management Plan" (see [1]) according to the table structure in Annex I of that EC document. This was distributed to all project partners so as to collect all relevant information concerning the datasets that they are planning to develop during the course of the project. On the basis of the partner feedback received, the preliminary Data Management Plan of the project was then developed, as described in this deliverable.

In this phase of the project, it is foreseeable that a series of data sets will be generated during the lifetime of the project, covering topics ranging from system / user requirements analysis phase (WP2), during the data collection campaigns of WP3, as well as during the pilot trials of WP7.

The research of Ageing@Work includes measuring workers' physical reactions and monitoring their activities at work. The design, implementation and evaluation of these kinds of features will be carried out paying specific attention to privacy and other ethical issues.

Consequently, datasets that will be made publicly available have to be proved by the ethic committee and have to be in line with the GDPR. Since the data collected is mainly personal and sensitive content, it is likely that only statistical data will be published, rather than raw data.

Since this deliverable was drafted at M6 and thus at a relatively early stage of the project, it can naturally only reflect initial intentions of the project partners as to which datasets they will eventually develop within the overall duration of the project. To take account of this development process, three further revisions of this deliverable are planned for the remaining project duration: the first for M18, the second for M30 and the last for M36 of the project, following the H2020 guidelines for Data Management Plans.

Table of Contents

List of Tables

List of Terms and Definitions

Abbreviation	Definition
ANEFA	Asociación Nacional de Empresarios Fabricantes de Áridos
API	Application Programming Interface
AR	Augmented Reality
ASER	Institut für Arbeitsmedizin, Sicherheitstechnik und Ergonomie e.V.
CERIF	Common European Research Information Format
CERTH	Centre for Research and Technology Hellas
CIOP-PIB	Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy
CSV	Comma-Separated Values *(File format usually with a .csv extension)*
DCAT	Data Catalog Vocabulary *(RDF vocabulary)*
DMP	Data Management Plan
DoA	Description of Action
DoW	Description of Work
EDA	Electrodermal Activity
ERP	Ethics Responsible Person
FAIR	Findable, Accessible, Interoperable, and Re-usable
GDPR	General Data Protection Regulation *(EU)*
GSR	Galvanic Skin Response
HR	Human Resource management
HTTP	Hypertext Transfer Protocol *(Application protocol)*
HTTPS	Hypertext Transfer Protocol Secure *(Application protocol)*
IBM	International Business Machines Corporation
ICT	Information and Communications Technology
ID	Identifier *(Symbol which uniquely identifies an person, object or record)*
IoT	Internet of Things
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
KB	Kilobyte
KUL	KU Leuven
Matomo	*A web analytics platform*
MB	Megabyte
MP4	*A digital multimedia container format most commonly used to store video and audio*
MS	Microsoft Corporation
N/A	Not Applicable
OMI	Open Model Initiative
ORDP	Open Research Data Pilot
OSH	Occupational Safety and Health
OWASP	Open Web Application Security Project
PDF	Portable Document Format *(File format developed by Adobe)*

Q-PLAN	Q-PLAN INTERNATIONAL ADVISORS PC
REDCap	Research Electronic Data Capture
RDF	Resource Description Framework *(Family of W3C specifications)*
RDF/XML	*A syntax, defined by the W3C, to express an RDF graph as an XML document*
SAG	Siemens AG
SC	Skin Conductance
SHA-2	Secure Hash Algorithm 2 *(Set of cryptographic hash functions)*
SPSS	Statistical Package for the Social Sciences *(IBM software package)*
SQL	Structured Query Language *(Database language)*
SSL	Secure Sockets Layer
SVN	Apache Subversion *(Software versioning and revision control system)*
SWELL	Smart Reasoning Systems for Well-Being at Work and at Home *(Project)*
UI	User Interface
UPAT	University of Patras
UPM	Universidad Politécnica de Madrid *(Technical University of Madrid)*
VPN	Virtual Private Network
VUM	Virtual User Models
VWM	Virtual Workplace Model
WESAD	Wearable Stress and Affect Detection *(Multimodal dataset)*
WP	Work Package
wVUM	worker Virtual User Model
XML	Extensible Markup Language

Table 1: Terms and Definitions

1. Introduction

Ageing@Work aims to develop a novel ICT-based, personalized system to support ageing workers (aged 50+) into designing fit-for-purpose work environments and managing flexibly their evolving needs. Advanced dynamically adapted virtual models of workers will incorporate specificities in respect to skills, physical, cognitive and behavioral factors, being extended from the work context to personal life aspects interacting with workability, health and well-being.

In order to effectively implement the Ageing@Work project, it is necessary to collect meaningful data (e.g. measuring workers' physical reactions and monitoring their activities at work) in order to gain insights that successfully influence the activities of the project enabling us to deliver evidence-based results to achieve the project objectives.

All personal data collected/generated to this end will be controlled and processed on the basis of informed consent, in full compliance with the General Data Protection Regulation (GDPR) (see [2]) as well as other relevant applicable EU and national regulations, protecting the data subjects' rights and freedoms in relation to the processing of their personal data. Further details on ethical aspects, privacy and protection of personal data are provided in deliverable D2.4 "Ethics Protocol" chapter 2 ff as well as in the deliverables D1.1, D1.2 and D1.3.

With this background and in line with the "Guidelines on FAIR Data Management in Horizon 2020", Ageing@Work incorporates a sound data management across the entire life cycle and beyond to ensure that conflicts between privacy, data protection and open access to research data are properly addressed, especially when it comes to making project data Findable, Accessible, Interoperable and Reusable (FAIR).

This is the 1st version of the DMP therefore chapter 3 does only describe foreseen datasets.

However the DMP is a living document and will evolve during the projects lifetime (M18, M30 and M36), covering the entire data management life cycle (even after the end of the project) addressing the following aspects: (i) Data summary; (ii) FAIR data (iii) Allocation of resources; (iv) Data security; (v) Ethical aspects; and (vi) other issues. Along these lines, data collected/generated will be centrally stored, scrambled where possible and abstracted in a way that will not affect the outcomes of the project.

2. General Principles

2.1 Participation in the Pilot on Open Research Data

Ageing@Work participates in the Pilot on Open Research Data launched by the EC along with the H2020 programme. The consortium believes in the concepts of open science, and in the benefits that the European innovation ecosystem and economy can draw from allowing reusing data at a larger scale however, as the collected data include e.g. measuring workers' physical reactions and monitoring their activities at work and home (for data protection, security reasons, but also for measuring privacy aspects), qualitative research raw data may not be made openly accessible, but only in processed form.

2.2 Security

The datasets that will be collected in the Ageing@Work project includes measuring workers' physical reactions and monitoring their activities at work - therefore sensitive and personal data which definitely have an influence on the professional development of the participants, if the data fall into the wrong hands.

To protect these data and control unauthorized access to the Ageing@Work data repositories, only authenticated personnel will have access to pilot-specific data. During the proposed system lifecycle, a holistic security approach will be followed, in order to protect the pillars of information security (confidentiality, integrity availability) from a misuse perspective.

Towards the protection of personal data of volunteer pilot participants, the following issues will be considered:

- No personal data will be collected without definite **permission** of the human participants;
- All data associated with a recognizable person will be **held private**.
- Individual data on subjects will be used in strictly confidential terms and will only be published as statistics (**anonymously**).
- Every personal data collected throughout the project will be treated with respect to the **protection of fundamental human rights** (e.g. separating general and personal data, handling encrypted personal data and identities, erasing irrelevant personal data)
- All data collected is anonymous, but the data is referred to as pseudonymized because personal data is securely stored and only one person has access to this data but not to the pilot data. The contact details are stored if participants wish to participate in a later pilot phase. If the participants do not wish to participate in any further pilot phase, the data will be completely anonymized.
- Any data or information about a person will be kept confidential, regardless of how this data was collected: Therefore, all data collected within the Ageing@Work project will be treated confidentially. Accidental acquisition does not replace the mandatory procedure, whereby researchers require explicit consent to obtain, store and use information about participants

- The acquired data will under no circumstances be used for commercial purposes.

2.3 Personal Data Protection

In order to achieve the goals of Ageing@Work personal data of ageing workers will be collected (questionnaires and observations) and processed, including in some cases sensitive personal information. The Ageing@Work Consortium is fully aware of this and the related challenges in terms of ethics, privacy and data protection. In this scope, it is self-evident that all personal data collected/generated to this end will be controlled and processed on the basis of informed consent, in full compliance with the General Data Protection Regulation (EU 2016/679) [2] as well as other relevant applicable national regulations, protecting the data subjects' rights and freedoms in relation to the processing of their personal data, including rectification and erasure. The Ageing@Work project will protect the gathered personal data in a series of ways, respecting the principles of data minimization, anonymity, informed consent, confidentiality, rights to personal data access, including rectification and erasure. The names of employees participating in the questionnaire-based surveys or in the pilots will not be made public in any document nor will their participation be communicated to other pilot participants or colleagues. As mentioned above, all personal data will immediately at least be pseudonymized (if the participant wants to be involved again, or irreversibly anonymized if not.

3. Description of the foreseen Ageing@Work Datasets

This chapter provides detailed information about the datasets that are planned to be captured by the partners of the Ageing@Work project. In order to meet the requirements of the DMP set by the EC Open Research Data Pilot (ORDP) regulations of the Horizon 2020 programme (H2020), each partner provided the description of their datasets using the template given in Annex A, which is based on the EC guideline document "Template Horizon 2020 Data Management Plan" (see [1]).

In this first attempt to collect all the needs regarding the data collections consortium partners have identified the kinds of data to be collected and the way they will be used to achieve project objectives. The proposed datasets describe several characteristics of the data structures and properties to be used by various tasks throughout project's lifecycle. Those datasets correspond to the contents of the database which will be developed as a backend of the Ageing@Work platform plus some more datasets to be used for offline data collection – when needed. With respect to the need of involved partners to have ownership and control over the data they will collect at this first stage of the project, a certain degree of overlaps between datasets is considered as highly expected. After the development of the digital knowledge-base, as part of the Ageing@Work platform, any overlaps will be resolved and datasets will be minimized accordingly. Changes will be reported in the public D9.5 deliverable on M18.

Dataset Naming Conventions

The following notation is used to name the Ageing@Work datasets: Each name of a dataset consists of: (a) a prefix "DS" indicating a dataset together with its unique identification number, e.g. "DS01", (b) a hyphenated combination of the name of the partner responsible to collect it, e.g. "CERTH" and an identifier denoting the internal numbering of the dataset concerning this respective partner, e.g. "CERTH-01", and (c) a short title of the dataset summarizing its content and purpose given in upper camel case notation, e.g. "PersonalDataAndBasicDemographics".

Summary of the foreseen Ageing@Work Datasets

The following *Table 2* below provides a brief summary of all data sets currently foreseen in the Ageing@Work project. In the subsequent sections 3.1 through 3.11, the detailed descriptions of these datasets are then provided according to the structure indicated in the template from Annex A, as mentioned above.

No	Dataset Name	Short Description
DS01	DS01.CERTH-01.PersonalDataAndBasicDemographics	A dataset of personal information, identification and basic demographics (online user profile).
DS02	DS02.CERTH-02.InteractionHistory	A collection of data related to the personal navigation routes and the user generated data on the online platform.
DS03	DS03.CERTH-03.AgeingWorkerActivitiesMonitoring	A data collection related to the worker activities in working and home environments (physical places).
DS04	DS04.CERTH-04.WorkerVirtualUserModel.	A dataset used to describe the physiological, psychometric, behavioral, and social parameters of the users (including disabilities and functional limitations) in order to perform studies in the 3D simulation and in workplace ergonomics optimization.
DS05	DS05.CIOP-PIB-01.EmployeeDataset	This dataset aims at defining the list of factors related to employees' work ability, quality of life and productivity.
DS06	DS06.Q-PLAN-01.NewsletterSubscription	This dataset will be used in the context of online subscriptions to the project's newsletter for the collection of the necessary personal data.
DS07	DS07.Q-PLAN-02. DataForDisseminationAndCommunicationReporting	This dataset will be used for the periodic monitoring of the miscellaneous dissemination activities of the project.
DS08	DS08.UPAT-01.StressDetection	This dataset serves the aim of monitoring worker affective traits in realistic workplace environments.
DS09	DS09.UPAT-02.VirtualWorkplace	This dataset concerns the establishment of a virtual workplace model.
DS10	DS10.UPM-01.BackendData	This dataset allows the support of the end-user website and is used for the visualization of certain statistics and the algorithmic processing in order to fulfil the user's physical and behavioral assessment.
DS11	DS11.ASER-01.HypotheticalData	*Temporary until further notice*

Table 2: Summary of Datasets of the Ageing@Work project

3.1 Dataset "DS01.CERTH-01.PersonalDataAndBasicDemographics"

1.	Dataset Name

DS01.CERTH-01.PersonalDataAndBasicDemographics

2.	Data Summary

The personal data refers to information which is related to people and describes their demographics profile. This data is essential to be collected for performing any meaningful statistical analysis to support the project objectives. The main purpose is to allow users to create and maintain a user profile on the online system to be used in order to be connected to the Ageing@Work services and tools. The username and password credentials will be used for user authentication, while other parts of the dataset will be used to personalize the productivity enhancement tools (WP6), the Emphatic Virtual Coach Mirroring Avatar (T5.1) and its smart behavior (T5.4), the UI and the contents of the behavior/workplace analysis dashboard (T5.2), the reward-system (T5.3) and the decision making support system for experts/clinics (T5.5).

This expected functionality is directly related to the first objective of the project to enable extensive personalization capabilities and the third objective that is to provide workers with personalized work ergonomics and process design services. Moreover, it is indirectly related to the research and development of the virtual assistant (objective 4) and the advanced personalized ICT-based workability and productivity enhancement tools (Objective 5). The personal and demographics data will be also needed in the pilot studies (Objective 7) in order to demonstrate and evaluate the framework under realistic conditions.

Type and format of the dataset

This dataset will contain a unique identifier representing each user, another field related to the user's credentials where the password will be stored using a hash function such as the Secure Hash Algortihm-2 (SHA2), a set of basic demographics parameters (such as gender, age, residency, spoken languages and education), contact information and living status. Moreover, it will include essential elements of job description as the current position, years of experience in the same job position and also computer or driving abilities.

Data Collection, expected size and usefulness

The aforementioned data will be collected during the profile development of the users and it will start just after the successful creation of a new user account. The size of the personal and demographics dataset will be a few kilobytes for each user, having the photo or avatar icon being the most demanding one. The avatar icon will be useful in social activities like AR telepresence and online discussions, while the sharing this info will be in user's hand (controlled user visibility). The data will be collected mainly during the T7.2 (Pilot trial definition and planning) and will be useful to researchers to conduct studies (such as comparisons between user age groups, or genders, or countries of origin), and to the online platform in order to offer personalized notifications and recommendations.

3.	FAIR Data

3.1 Making data findable, including provisions for metadata

The personal and demographics datasets will be findable by both users and the system modules (e.g. processing algorithms) in the online environment of Ageing@Work. Each user will have access to its own data and will be also be responsible for updating the information included there. A printed format to deliver this information to the offline users will be offered for those users willing to terminate their participation in the online platform. The same template could be used to present online to the users 'what the system knows about them'. The exported user profiles will be in PDF format (for participants) and an XML or JSON format for computer use. Both file formats will be saved externally to the system and will be named based on the UserId.

The identifiability of the data will be achieved by using unique identifiers for each user: the UserId and the SHA256 (a hash function of 32-bit output length and widely used in security applications and protocols). The salt key for the one-way hashing of password will be saved along with the hash, so upgrade or moving to another server will not make hashes useless. The salt will be generated randomly for every single password and will be stored as a part of the string the system will get back from the 'password hash' function. This process ensures that the output of the function will be different each time someone hashes a password. The reason to use salt in passwords is to get a new hash every time a new password is created. For instance, if two users use the same password "rockstar5" they will be finally stored as totally two different database values making rainbow tables useless (pre-computed tables for reversing cryptographic hash functions, usually used for cracking password hashes). This identification mechanism provides the system the ability to move to a new server without making the previously hashed passwords useless.

The personal and demographic datasets will be searched online based on a criteria-based search engine. Keywords and range values for various parameters (e.g. nicknames, dates of birth, etc.) will be used for searching.

3.2 Making Data openly accessible

Parts of the information included in the personal and demographics dataset will be available for research within Ageing@Work project and will be accessed by the system and by authorized researchers. The collected data will be available during the project lifecycle to consortium members only, while parts of the processed data might become available to the public after anonymization. Access can be granted onsite at the repository (visiting scientist) or – with sufficient clearance – through controlled remote data processing. The processed data (statistical analysis results) will be made openly available through scientific publications and project results distributed through the project site and the related social media.

However, in compliance with the requirements of the new regulation (GDPR), the raw personal and demographics data and especially sensitive data will not be openly available during the project lifetime, while after the termination of the project an informed consent will be required from users for making the anonymous data available to the public under a non-commercial license (as reuse is only permitted for non-commercial purposes). In such a case, the anonymized datasets could be made available through the project portal or the project profile in the coordinator's site (CERTH) at least for 5 (five years).

The datasets will be accessed by widely used general-purpose software tools like Microsoft Excel (CSV files), PDF readers (for printed user profiles) or a web platform for building and managing online surveys like REDCap for example.

The public deliverables of the project will describe the data format of the personal and demographics data and will provide information related to the associated metadata and documentation.

Files will be uniquely identifiable and versioned by using a name convention consisting of project name, dataset name, method, used ID, place and date.

3.3   Making Data interoperable

A new dataset will be created in the database each time a new user registers to the Ageing@Work platform and updates will be performed each time the user modifies his/her profile. Although the editing tool has not yet been defined, it is expected that data will be stored in a MySQL, NoSQL, MongoDB and JSON related database formats. Metadata will include information about the datetime of the user profile creation, the range of possible values for user category, etc. This metadata will be associated to each table and will follow the Common European Research Information Format (CERIF) metadata standard[1].

In terms of data interoperability, a shared metadata repository will be built which describes the content and intent of the datasets stored in the Ageing@Work information system. In addition, a good practice for making data interoperable will be to publish semantic agreements in that computing devices can interpret.

The use of international open Web standards will be extended to linked data for data interchange (e.g., JSON-LD and RDF/XML). Linked Data modeling will take maximum advantage from the information management system, including an increased level of data abstraction. If needed to integrate data with external resources, this could be made by creating a mapping for a multitude of data sources and then use this mapping to create a canonical data model for data integration. Nonetheless, developing a vocabulary to serve as a common baseline to map different data sources is more preferable in order to avoid time-consuming processes and expertise in knowledge engineering.

Overall, the data, metadata and documentation will be compliant to disciplinary standards, open file formats and use controlled vocabularies and the standard metadata schema for easy interoperability and re-use.

3.4   Increase Data Re-use (through clarifying Licences)

Any part of the personal and demographics data that might be decided to become public (following proper anonymization) will be licensed under Creative Commons CC BY 4.0 International. This data will become available through the project portal during the last three months of project end (M33). Parts of these (anonymized) data may be further processed using statistical methods to create evidence to support papers published in scientific journals and presented in conferences. There will be no embargo period. The datasets will be available for re-use by other researchers in areas such as ageing, ergonomics and stress management, among others. Future research should also concentrate on the use of demographics to develop advanced VUM models for ergonomics assessment using available formats such as UsiXML. As open formats are used for data archiving, the data will remain re-usable until the repository withdraws the data or goes out of business.

4.   Allocation of Resources

---

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

Estimated costs for data preparation to be FAIR are around 5% of the total research budget, according to the Science Europe (2016) [2] . Expenses consist of preparation and application of the FAIR data management principles, additional publication and documentation costs of three months for one full time equivalent and publication costs of the repository. Associated costs for dataset preparation and data management during the project will be covered by the project itself. Each pilot site will be responsible for its own data management, while the pilot study leader will be responsible for data management plan updates and CERTH have the responsibility of data backup and storage, as well as of data archiving and publication within the repository.

Long term preservation will result in no additional costs other than repository charges for data submission. The dataset will increase in value over the years because of its fundamental impact in the Ergonomics assessment and VUMs field, also in the future. Long-term studies and interdisciplinary research are desirable and intended.

5. Data Security

During Ageing@Work, project data will be automatically saved frequently (daily) on a CERTH server with backup on a separate offsite institutional server. Backup will be checked manually at intervals of two weeks. CERTH – as project leader – will be responsible for backup and storage. Following consultations with the institutional IT team, no additional costs are expected for storage and backup.

Sensitive data will be separated as early as possible to create an anonymized dataset. Authorized access to sensitive data will be granted only for project members with clearance through non-disclosure agreements. In addition, rightful access will be ensured through secure passwords and encoding of the folders and files within cloud storage. Moreover, data transfer will be secured via HTTPS protocol.

Long-term preservation will be provided by the CERTH repository. Data will be stored locally for five years at the institution and no additional costs are associated with this local storage.

6. 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.

Informed consent for data sharing and long term preservation is included during data collection. Sensitive data will be separated as soon as possible and kept secure.

7. Other Issues

N/A.

---

[2] Science Europe (May 2016). Funding research data management and related infrastructures: Knowledge Exchange and Science Europe briefing paper, available at: https://www.scienceeurope.org/wp-content/uploads/2016/05/SE-KE_Briefing_Paper_Funding_RDM.pdf

3.2 Dataset "DS02.CERTH-02.InteractionHistory"

1.	Dataset Name

DS02.CERTH-02.InteractionHistory[plus Reward]

2.	Data Summary

This data set contains all the information related to the interactions generated by the user by its participation in the online platform, plus a gamification model of each user (gamification profile). Each personal interaction history will include timestamped interaction events, performance metrics, rules or earned rewards form the participation of the user to the gamification component. An initial gamification profile should be created – as an extension to the existing user profile – when a user enters in the Ageing@Work system and evolves with the participation of the user in the platform. There is an additional short gamification model to be used to report the gamification status of a user in brief.

Type and format of the dataset

The interaction history of each user will be series of timestamped data records containing the pages visited in the online platform, the self-administered scales and questionnaires fulfilled online, exposure-relevant online behaviors, the types of devices used by the user (including the IoT devices like smartwatches), online telepresence and training sessions, and the performance of the user on the online award system (gamification campaign).

Data Collection, expected size and usefulness

The traces of the online activity of the users will be collected during the use of the Ageing@Work platform. A set of targeted interaction events will be stored in log files and in the system database in real-time. Users will have access to their personal interaction history records through a 'personal timeline' view. The volume of the data collected will be analogous to the number of users who will participate in the pilot studies in ANEFA and SIEMENS. Those datasets will be useful to the recommender component in order to produce personalized notifications (accessed also thought he mirroring avatar), and to the researchers in order to perform comparative research and studies (behavioral analytics). Moreover, European Commission services and agencies might find those datasets useful.

3.	FAIR Data
3.1	Making data findable, including provisions for metadata

The data produced by the online behavior of the platform users will be treated by a metadata model used to describe user online monitoring collections. Those will include data collected as a result of the user's interaction with the platform and its peripheral components on a regular basis. An open source web analytics platform will be used on top of the Ageing@Work platform to track online visits, like Matomo[3] for example. Thus, the Report Metadata API Methods or similar will be used to return the report containing the processed metrics, the analytics data and all other additional information. The visual analytics page of the monitoring platform will support searches on the data collections using criteria provided by the platform administrators and consortium researchers. Specific filters, based on the metadata elements (like start and end dates, user categories, types of interaction events etc.), will allow to refine the search across datasets.

---

[3] https://en.wikipedia.org/wiki/Matomo_(software)

Additionally, it will be possible to download the raw data locally in XML file format for further statistical analysis, according to the needs of the project pilot studies. Moreover, a 'personal wall' will be available through the platform to help users explore their own interaction history.

3.2 Making Data openly accessible

In order to share the data collections with the consortium partners, a repository will be set up as part of the Ageing@Work and the online monitoring platform. This repository will offer access to the data trough web interfaces and will allow authorized users to view and share datasets under authors' interventions and research purposes. The data sharing services will be based on architectural components and tools that work with any kind of storage media.

The collected and processed data will be set up for Ageing@Work as it will aim to reach the highest level of GDPR compliancy by:

* Relying on security protocols for data sharing.
* Following a strict policy in granting access to the collected datasets.
* Logging of user identity during data access (for either view or download).

For making data openly accessible, data privacy and security protection methods and techniques as anonymization and pseudonymization will be initiated and applied on the interaction history datasets. For existing data (not generated by Ageing@Work) the data owners/providers (e.g. HR managers) will specify the level of granularity that data will be transferred. The data managers will indicate whether the existing data will be directly accessible for use within Ageing@Work, and if required the consortium will make requests to those data owners for accessing existing data related to their employees.

3.3 Making Data interoperable

Ageing@Work will build and widely reuse existing ontologies and data models for all datatypes, including the relations between them, allowing the solution to work with other external services or systems, at present or in the future, without any restrictions. The use of common data formats and protocols like XML and JSON for data transfers and SQL for data searches will achieve a high level of inter and cross-domain interoperability. Textual information will be stored in Unicode, static media information in JPG format and dynamic media (video form telepresence and training sessions) in MP4 format, which is a known sharing format for the internet.

3.4 Increase Data Re-use (through clarifying Licences)

The collected information will be highly reused to regularly update the worker Virtual User Models (wVUMs) to be used in the simulator for ergonomics assessment. Moreover, the interaction history of the users will be used for studies on behavioral analytics and for measuring the effect of the overall Ageing@Work solution. Several protocols will be followed to ensure a high data quality (data storage set up redundantly, quality preserving compression methods in media files, etc.).

After the termination of the project, a part of the available data will be published under license and thus it will be reusable for longer period of time and by a wider audience. The exact period during the data will remain accessible through the project's site, will be determined in the last version of the Data Management Plan (M36).

4. Allocation of Resources

Regarding datasets and the scientific publications which will be based on them, Ageing@Work will use repositories (like GitHub for example) and the infrastructure of CERTH. Regarding the part of the data that will be available in the project's web portal, the same policy will be applied.

5. Data Security

A wide part of the collected information related to online behavior of the users will contain personal information. All partners will comply with commonly agreed privacy regulations to ensure that the interaction history of the users will contain the minimum necessary personal and sensitive information (answers to online questionnaires, stress-diagnostic batteries, work-performance data and live observations when needed. Log files will be backed-up regularly in the knowledge repository (secure storage server) and secured with state of the art data protection mechanisms. Datasets will be stored in secured servers where only authorized personnel from the consortium will have access to and the data exported from the Ageing@Work system (in XML and/or JSON files) will be automatically anonymized by the online user monitoring platform. Moreover, the Ethics Responsible Person (ERP) of CERTH (Dr. Sofia Segkouli) will supervise the data maintenance and will take proper, constant and in time action for data quality assurance.

6. Ethical Aspects

The data collection process will be regulated by an ethical protocol as described in D2.4 (Ageing@Work Ethics manual). According to this, all partners who will collect data from workers and other users during the pilot studies and other data collection activities will have an ethics approval from their ethics committees. All human subjects will be asked to provide informed consent before their participation in online activities related to data collection.

7. Other Issues

N/A.

3.3 Dataset "DS03.CERTH-03.AgeingWorkerActivitiesMonitoring"

1.	Dataset Name

DS03.CERTH-03.AgeingWorkerActivitiesMonitoring

2.	Data Summary

Ageing@Work will monitor not only the online activities of the workers, but also the physical world activities using wearable sensors and IoT devices (accelerometers, biosensors, etc.), with the possibility to introduce also camera-based activity monitoring – if needed. This dataset will be produced as an outcome of the T4.2 (Unobtrusive worker activity and behavior monitoring).

This functionality is related to the second objective of the project which is to design an unobtrusive worker activity and behavior monitoring framework to be used for understanding the old worker's interests, behavioral factors in work and home environments.

Type and format of the dataset

This will be an anonymized dataset with annotated data during specific worker activities. More specifically, a sensing mechanism based on unobtrusive wearables such as a wristwatch or the user's smartphone will be set up. This way, health and working goals can be defined and progress can be monitored right from worker's wrist. The dataset will store data from sensors like accelerometer, gyro sensor, heart rate and/or blood pressure.

Data Collection, expected size and usefulness

The data collection rate will be variant and depending on the IoT device, but in any case it will be quite high to allow the extraction of meaningful results after applying machine learning approaches. The expected data volume will be high for biosensor parameters to allow close to real-time user monitoring. This information will be useful for behavior analysis and together with the online user activity monitoring platform ('Interaction History' dataset) will construct a complete online and offline user monitoring system capable of sensing activities with further contextual cues (e.g. time of day, place, workload, planned activity, performed activity and context).

3.	FAIR Data
3.1	**Making data findable, including provisions for metadata**

The data produced through the worker activities monitoring platform during the lifecycle of the project will be made discoverable through a centralized database, part of the knowledgebase. This component will be accessed via the web portal of the project and will allow accredited consortium members to have access to the data collections.

The data collection component will be decentralized and the Ageing@Work mobile app will be used for interfacing the data collections of the IoT devices. Each user will be responsible for storing the monitoring data on the local mobile device (tablet or smartphone) and also to share this data with the platform in order to receive the corresponding platform services (notifications, interventions, progress reports, awards etc.).

3.2	Making Data openly accessible

A part of the data collection (anonymized) will be available on the web after the conclusion of the pilot studies. Metadata for this public part of the dataset will be filled in online and will be made according to W3C Data Catalog Vocabulary (DCAT) standards for data catalogues published on Web.

3.3	Making Data interoperable

The novel worker activity and behavior monitoring framework of the Ageing@Work platform will be extensible thanks to an open architecture. This approach will allow diverse sensing devices to provide the system with the necessary data and also third party service providers to be merged for future use (e.g. physical activity trackers, dietary apps and more). Moreover, Ageing@Work will provide an API for services integration with other third-party systems as outcome of the WP7 (Ageing@Work platform integration and validation).

3.4	Increase Data Re-use (through clarifying Licences)

The worker activity and behavior monitoring data collections will be highly reusable as the wVUMs will be regularly updated to follow recent changes in the user behavior and thus to better represent the user into the ergonomics 3D simulation and the working process optimization environment. Similar to the online user behavior data collections, several protocols will be followed to protect data quality (frequent data storage, lossless data compression methods etc.).

4. Allocation of Resources

(Same as 3.4)

5. Data Security

Similarly to the online user monitoring system, the Ageing worker activities monitoring datasets are of high value and may include personal information and possibly sensitive activities tracking, especially when applied in home environments.

All partners will comply with commonly agreed privacy regulations to ensure that the user's monitoring will contain the minimum necessary personal and sensitive information and that the data repository used by the project will include effective protection. The security approach will start with assessment of the security risks and their impact on the users and the system. The datasets will be posted from the end-device to the system after user's concern, while for the long term storage secured servers will be used. Similarly to the interaction history, the CERTH ERP will supervise the dataset collection and maintenance and will take action for data quality assurance.

6. Ethical Aspects

Partners within Ageing@Work need to comply with the Ethics protocols as described in the WP1 and more specifically in the D1.2 (POPD – Requirement No. 5), because this datasets will include sensitive information (biometric and/or health data). Moreover, the national legislations and law of the countries where the data collection and processing is taking place should be taken into consideration in respect to data processing.

7. Other Issues

N/A.

GA **#826299**

3.4 Dataset "DS04.CERTH-04.WorkerVirtualUserModel"

1.	Dataset Name

DS04.CERTH-04.WorkerVirtualUserModel.

2.	Data Summary

This dataset will be used to describe in detail the physiological, psychometric, behavioral, and social parameters of the users (potential disabilities or functional limitations will be included). The dataset will be a collection of wVUMs used to perform behavioral studies in the 3D simulation and in workplace optimization tasks. Each unique wVUM will represent a single user or a group of users who share common characteristics in order to produce evidence through experimental ergonomics studies.

The wVUMs which will represent single users will contain personal or sensitive information, but the wVUMs to be used for worker group representation (e.g. elderly machine operators) will contain average values of a group of people. Similarly to the demographics dataset (DS01.CERTH-01.PersonalDataAndBasicDemographics), this data is essential to be collected and regularly updated for performing meaningful simulation-based analysis (every 6 months through the results of an assessment using physical and psychometric scales). Parts of the personal and demographics data of the users will be used in wVUMs while performing an ergonomics or working processes simulation.

The purpose of the individualized wVUMs is to allow users themselves to maintain a computer file to represent them in ergonomics simulation studies and to generate evidence for their performance in future ergonomics designs. This expected functionality is directly related to the first objective of the project to enable extensive personalization capabilities and the third objective that is to provide workers with personalized work ergonomics and process design services. This compute file will not be shared with others without user's concern. On the other hand, the group wVUMs can be used by HR managers to perform company-level or shopfloor level simulations for decision making. This element, the 'Persona' will be a result of averaging a group of users who share similar characteristics and who previously gave permission to the Ageing@Work system to statistically process their personal wVUM.

This dataset will be used to personalize the productivity enhancement tools (WP6), the Emphatic Virtual Coach Mirroring Avatar (T5.1) including its smart behavior (T5.4), and the decision making support system for experts/clinics (T5.5). Unlike dataset DS01.CERTH-01.PersonalDataAndBasicDemographics, dataset DS04.CERTH-04.WorkerVirtualUserModel will not be used to personalize the UI and the reward-system (T5.3). Processed data from this dataset can be presented in the behavior/workplace analysis dashboard (T5.2).

Type and format of the dataset

This dataset will contain all information required for providing a structured machine readable representation of the worker (both ageing and young workers). Basic data containers (parts of the wVUM) will be: general attributes and preferences (like years of education, age, gender, etc.), a list of affected tasks (due to disabilities or a functional limitation), anthropometrics, working parameters, sensory information (further organized into visual and auditory information), cognitive parameters and some additional information to describe the ability to use of language.

D9.4   *Dissemination Level: Public*   **Page 23 of 49**

Data Collection, expected size and usefulness

The aforementioned data will be used for creating computer files able to represent a user of a group of users (Personas) in a simulated environment for ergonomics assessment. The size of each wVUM will be relatively small (few Kbytes) and the total number of wVUMs to be created in Ageing@Work is estimated around 50 to 60.

3. FAIR Data

3.1 Making data findable, including provisions for metadata

The personal wVUM will be findable only by its owner (the worker) in the local device, while the group wVUMs (Personas) will be findable through the Ageing@Work platform by HR managers, ergonomics designers and health and safety professionals. Both types of wVUMs will be XML files and will be generated by wVUM generator, a computer tool which will be available in the Ageing@Work toolet. The group wVUMs will be searched online based on a criteria-based search engine. Keywords and range values for various parameters will be used for searching. The private wVUMs will be stored and used locally by their users.

3.2 Making Data openly accessible

Only the list of group wVUMs (Personas) will be openly accessible to consortium members, HR managers and health & safety professionals. The wider public may have access to a small set of group wVUMs (2-3) to be used for demonstration purposes, statistical analysis studies and scientific publications. Those shared wVUMs will be distributed through the project site.

3.3 Making Data interoperable

The markup language XML (Extensible Markup Language) will be used to describe the wVUMs. The advantage of this open standard is that it provides a format which is readable both by humans and machines. Moreover, its textual data format will support different languages (Unicode support). This approach is standard in web services, thus lading to high interoperability and many application programming interfaces (APIs) have been developed to aid the processing of XML data[4].

3.4 Increase Data Re-use (through clarifying Licences)

The wVUMs will be highly reused by the 3D simulator for the evaluation of new or existing ergonomics designs. The group wVUMs selected to be available to the wider public (following proper anonymization) will be licensed under Creative Commons CC BY 4.0 International.

4. Allocation of Resources

Associated costs for wVUMs preparation and the related data management will be covered by the project itself during the project.

---

[4] Extensible Markup Language (XML), https://en.wikipedia.org/wiki/XML

5. Data Security

The wVUMs will be automatically saved after creation or update on a secured CERTH server. Backup will be checked manually at 2 week intervals. CERTH will be responsible for backup and storage of the wVUMs which will become available to the consortium. The personal wVUMs will be stored only to the local device and typical data security principles will be followed to protect the file and user instructions with data protection advices will be given to individuals after installing Ageing@Work software on their devices. No additional costs are foreseen for storage and backup of the wVUMs.

6. Ethical Aspects

Informed consent for sharing or using a personal wVUM will be asked by data owners (workers themselves). For using group wVUMs which have been generated based on private wVUMs, the previously obtained consent from individual users should be enough.

7. Other Issues

N/A.

3.5 Dataset "DS05.CIOP-PIB-01.EmployeeDataset"

1.	Dataset Name

DS05.CIOP-PIB-01.EmployeeDataset

2.	Data Summary

State the purpose of the data collection/generation.

The data collection aims at defining the list of factors related to employees' work ability, quality of life and productivity to be included into the Ageing@Work system (WP3).

Explain the relation to the objectives of the project.

The collection of data on employees' work ability, productivity or quality of life is related to project's Objective 1: Enable extensive personalization capabilities to the Ageing@Work supportive approach Ageing@Work; Objective 3: Provide workers with personalized work ergonomics and process design services; and Objective 6: Co-design tools for managers and OSH specialists for improved age-friendly workforce management.

Specify the types and formats of data generated/collected.

Data will be collected through the questionnaire survey conducted at ANEFA and Siemens. The data will be stored in Excel and/or REDCap. In addition, data will be saved in SPSS, which is a standard format for the statistical analyses in social science.

Specify if existing data is being re-used (if any).

It is not. We will collect data in the group of employees working in pilot companies.

Specify the origin of the data.

Data will be obtained from employees during the questionnaire survey and individual interviews with employees working at ANEFA and Siemens.

State the expected size of the data (if known).

The expected data volume will be approximately 5 MB.

Outline the data utility: to whom will it be useful?

The data about employees obtained in task 3.1 would be useful to partners responsible for tasks 3.2, 3.3, and 3.4. We do not plan to share this data for secondary analysis.

3.	FAIR Data

3.1	Making data findable, including provisions for metadata

General remark:

We do not plan to share our dataset in a publicly accessible disciplinary repository. The results of statistical analysis will be available in the projects' Deliverable 3.1. This decision is based on the fact that the group of employees taking part in the questionnaire survey and individual interviews is quite small and easily identifiable in terms of the company they work for, and the data will include vulnerable

information on their health status, productivity or work ability. Finally, as the study group will be relatively small, the data collected will not fit into any metadata analyses.

Outline the discoverability of data (metadata provision).

We intend to share the results of statistical analysis in the project's Deliverable 3.1, which will be publicly accessible on the Ageing@Work project's website.

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?

N/A.

Outline naming conventions used.

The dataset will be named "DS04.CIOP-PIB-01.EmployeeDataset".

Outline the approach towards search keyword.

When applicable (e.g. in research or conference papers) we intend to use following search keywords: work ability, productivity, quality of life, working conditions, wellbeing, blue-collar, older employees.

Outline the approach for clear versioning.

N/A.

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.

N/A.

3.2   Making Data openly accessible

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so.

We do not plan to share the dataset but the results of the statistical analysis will be publicly available in the project's Deliverable 3.1. This decision is based on the fact that the group of employees taking part in the questionnaire survey and individual interviews is quite small and easily identifiable in terms of the company they work for, and the data will include vulnerable information on their health status, productivity or work ability. Finally, as the study group will be relatively small, the collected data will not fit into any metadata analyses.

Specify how the data will be made available.

N/A.

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)?

N/A.

Specify where the data and associated metadata, documentation and code are deposited.

N/A.

Specify how access will be provided in case there are any restrictions.

N/A.

3.3	Making Data interoperable

N/A.

3.4	Increase Data Re-use (through clarifying Licences)

N/A

4. Allocation of Resources

N/A.

5. Data Security

Address data recovery as well as secure storage and transfer of sensitive data.

Exchange of data with project members at different sites will be secure and redundant through use of available storage – SVN repository set private only for Ageing@Work authorized personnel.

6. Ethical Aspects

N/A.

7. Other Issues

N/A.

3.6 Dataset "DS06.Q-PLAN-01.NewsletterSubscription"

1.	Dataset Name

DS06.Q-PLAN-01.NewsletterSubscription

2.	Data Summary

State the purpose of the data collection/generation.

A subscription form hosted in the project's web portal will aid the collection of this data in which any interested stakeholder can freely provide his/her contact details in a dedicated sign-up form so as to receive the most up-to-date news and outcomes of the project. A newsletter will be sent to subscribers once per 6 months. With that in mind, this data will be collected so as interested stakeholders can be informed about the Ageing@Work project. Along these lines, the data will comprise a list of stakeholders along with their personal information. In this context, the data collected include the following information: (i) email address, (ii) first name and (iii) last name. A copy of this contact list will be stored to MailChimp's server (http://mailchimp.com), which is used for e-mail campaigns and newsletters distribution. All personal information included in this contact list will be used and protected according to MailChimp's Privacy Policy.

Explain the relation to the objectives of the project.

The collection of this dataset corresponds to Objective 8 of the Ageing@Work project, in the sense that it will be used to implement the newsletter distribution which is part of the communication and dissemination activities.

Specify the types and formats of data generated/collected.

The data will be stored in Microsoft Excel file (.xls).

Specify if existing data is being re-used (if any).

N/A.

Specify the origin of the data.

The data will be collected through a subscription form in the project's web portal.

State the expected size of the data (if known).

The expected size of data is 250 KB based on the adjusted size of data generated via similar activities in the past.

Outline the data utility: to whom will it be useful?

The data will be useful to the dissemination manager in order to distribute the newsletters.

3.	FAIR Data

3.1	Making data findable, including provisions for metadata

N/A.

3.2	Making Data openly accessible

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so.

This dataset will remain closed (accessible only to consortium members) as any anonymization will leave no data on the dataset.

Specify how the data will be made available.

N/A.

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)?

N/A.

Specify where the data and associated metadata, documentation and code are deposited.

A copy of the newsletter contact list will be stored to MailChimp's server (http://mailchimp.com), which is used for e-mail campaigns and newsletters distribution. All personal information included in this contact list will be used and protected according to MailChimp's Privacy Policy.

Specify how access will be provided in case there are any restrictions.

N/A.

3.3	Making Data interoperable

N/A.

3.4	Increase Data Re-use (through clarifying Licences)

N/A.

4.	Allocation of Resources

N/A.

5.	Data Security

Verification of subscribers (email confirmation) is mandatory in order to not allow third parties to sign up emails that do not belong to their legitimate proprietary.

6.	Ethical Aspects

It is required that the subscribed users can unsubscribe anytime via direct link on the email communications or either this link opens a website where the users can state their desire of not receiving communications anymore.

7. Other Issues

N/A.

3.7 Dataset "DS07.Q-PLAN-02.DataForDisseminationAndCommunicationReporting"

1.	Dataset Name

DS07.Q-PLAN-02.DataForDisseminationAndCommunicationReporting

2.	Data Summary

State the purpose of the data collection/generation.

This data will be collected through a periodic monitoring of the project's miscellaneous dissemination activities such as publications in relevant journals, social media activity, participation in events, etc. The purpose of collecting this data is to assess the outreach and efficiency of the dissemination activities during the implementation of the project. For this purpose, a template has been shared with all partners to log the activities they performed. The template is provided through the Tortoise SVN repository so as the partners can directly update their input.

Explain the relation to the objectives of the project.

The collection of this dataset corresponds to Objective 8 of the Ageing@Work project, in the sense that it will be used to evaluate the outreach of the project's dissemination activities.

Specify the types and formats of data generated/collected.

The data will be stored in Microsoft Excel file (.xls).

Specify if existing data is being re-used (if any).

N/A.

Specify the origin of the data.

This data will be collected through the periodic monitoring of the project's miscellaneous dissemination activities.

State the expected size of the data (if known).

The expected size of data is 200 KB based on the adjusted size of data generated via similar activities in the past.

Outline the data utility: to whom will it be useful?

The data will be useful to project partners in order to be aware of the outreach of their activities.

3.	FAIR Data

3.1	Making data findable, including provisions for metadata

N/A.

3.2	Making Data openly accessible

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so.

This dataset will remain closed (accessible only to consortium members) as it is useful only for reporting within the scope of the project.

Specify how the data will be made available.

N/A.

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)?

N/A.

Specify where the data and associated metadata, documentation and code are deposited.

The data are deposited in the Tortoise SVN repository.

Specify how access will be provided in case there are any restrictions.

N/A.

3.3	Making Data interoperable

N/A.

3.4	Increase Data Re-use (through clarifying Licences)

N/A.

4.	Allocation of Resources

N/A.

5.	Data Security

N/A.

6.	Ethical Aspects

N/A.

7.	Other Issues

N/A.

3.8 Dataset "DS08.UPAT-01.StressDetection"

1.	Dataset Name

DS08.UPAT-01.StressDetection

2.	Data Summary

State the purpose of the data collection/generation.

This dataset serves the aim of monitoring worker affective traits in realistic workplace environments, be it at the factory or at home.

Explain the relation to the objectives of the project.

The dataset will be required for task T4.3. Its use will enable personalization capabilities to the Ageing@Work supportive approach (Objective 1) and more specifically the design of a novel unobtrusive worker activity and behavior monitoring framework, coupling work, on the move and home -based tracking elements (Objective 2).

Specify the types and formats of data generated/collected.

The data will be acquired by monitoring the worker affective traits in realistic workplace environments, with the main emotion of interest in this respect being stress. Literature review (see [5] and [6]) has shown that the most common physiological signals used for stress detection, that can be measured unobtrusively in ambulatory settings, are the heart rate variability, and the Galvanic Skin Response (GSR), also named Electrodermal Activity (EDA) and Skin Conductance (SC). The analysis of this input data (acquired biosignals) will lead to the production of output (annotation) data that will reflect the stress state of the individuals during the monitored time period. Input and output data will have the same dimensionality and format. Their storage will be linked in order to facilitate future meta-analysis.

Specify if existing data is being re-used (if any).

We could not find any datasets collected in a working environment, but there exist a few datasets, which are freely shared, such as the WESAD and SWELL datasets, collected during a lab study. Since these datasets are annotated they will be used to construct our baseline prediction models.

The multimodal SWELL dataset (see [3]) was collected in an experiment, in which 25 people performed typical knowledge work, with their working conditions being manipulated with stressors (for data annotation purposes). A varied set of data was recorded: computer logging, facial expression from camera recordings, body postures from a Kinect 3D sensor and heart rate (variability) and skin conductance from body sensors. The WESAD (see [4]) is a new publicly available multimodal dataset for wearable stress and affect detection which features physiological and motion data, recorded from both a wrist- and a chest-worn device, of 15 subjects during a lab study.

Specify the origin of the data.

The biosignals will be acquired using unobtrusive wearables such as a smartwatch.

State the expected size of the data (if known).

The exact nature and specification of the dataset and their expected size is not yet available.

Outline the data utility: to whom will it be useful?

The data will be useful to the researchers to develop the affective traits recognition tool in the system development phase and also to the final users (workers) in the testing phase.

3. FAIR Data

3.1 Making data findable, including provisions for metadata

According to the planned schedule of tasks, this information is not yet defined.

3.2 Making Data openly accessible

Personal data will never be publicly available and will be used only for research purposes after anonymization. The anonymized data will not be made publicly available during the duration of the project while their exploitation is ongoing. The access rights of anonymized data, or secondary outcomes, after the completion of the projects are still to be defined and will follow the data protection regulations. Since more partners in the consortium are involved in the acquisition, analysis and exploitation of the data, a common decision must be reached.

3.3 Making Data interoperable

Not relevant yet.

3.4 Increase Data Re-use (through clarifying Licences)

Not relevant yet.

4. Allocation of Resources

UPAT is not involved in data allocation.

5. Data Security

UPAT will receive anonymized data.

6. Ethical Aspects

All the data will be used only by and for the projects purposes and accessed and exploited by the data scientists for anonymous research and statistical analysis. Personal data won't be used and no data sharing will take place outside the consortium. The end users are the only authorized entities to access their very own data and they do hold the rights for its rectification when desired according to the GDPR.

7. Other Issues

Not relevant yet.

3.9 Dataset "DS09.UPAT-02.VirtualWorkplace"

1.	Dataset Name

DS09.UPAT-02.VirtualWorkplace

2.	Data Summary

State the purpose of the data collection/generation.

This dataset concerns the establishment of a virtual workplace model.

Explain the relation to the objectives of the project.

The dataset will be required for task T3.3 focusing on the development of workplace models (VWMs) coupling the workplace 3D representation with interrelated semantics (Objective 1). Such models will be used to provide workers with personalized work ergonomics (Objective 3).

Specify the types and formats of data generated/collected.

We will apply 3D scanning technologies to reconstruct objects of interest in the workplace. The data will be in the form of 3D point clouds and 3D meshes and some of them will have a semantic/task annotation.

Specify if existing data is being re-used (if any).

Besides the user-specific workplace environment parts, we will use existing 3D representation models to visualize common (across workplaces) objects.

Specify the origin of the data.

Depth sensors will be used to capture the domestic working environments and create the 3D workplace model. The necessary links from the space to the work processes performed in it, coupled with semantics on the machines and tools used, as well as on the collaborative resources in more general, will be performed manually based on information provided by the end-users.

State the expected size of the data (if known).

The exact nature and specification of the dataset and their expected size is not yet available.

Outline the data utility: to whom will it be useful?

The data will be useful to the researchers to develop the Ageing@Work ergonomics optimization tools, from which the end-users (workers) will later benefit.

3.	FAIR Data

3.1	Making data findable, including provisions for metadata

According to the planned schedule of tasks, this information is not yet defined.

3.2	Making Data openly accessible

Although the amount of personal information is very limited in this dataset, since the 3D models to be developed are not necessarily linked to a particular worker, we clarify that any possible personal data will never be publicly available and will be used only for research purposes after anonymization. The anonymized data will not be made publicly available during the duration of the project while their exploitation is ongoing. The access rights of anonymized data, or secondary outcomes, after the completion of the projects are still to be defined and will follow the data protection regulations. Since more partners in the consortium are involved in the acquisition, analysis and exploitation of the data, a common decision must be reached.

3.3	Making Data interoperable

Not relevant yet.

3.4	Increase Data Re-use (through clarifying Licences)

Not relevant yet.

4. Allocation of Resources

UPAT is not involved in data allocation.

5. Data Security

The workplace models will not include any personal information. If such a case arises, the data will be anonymized before their use.

6. Ethical Aspects

All the data will be used only by and for the projects purposes and accessed and exploited by the data scientists for anonymous research. Personal data won't be used and no data sharing will take place outside the consortium. The end users are the only authorized entities to access their very own data and they do hold the rights for its rectification when desired according to the GDPR.

7. Other Issues

Not relevant yet.

3.10 Dataset "DS10.UPM-01.BackendData"

1.	Dataset Name

DS10.UPM-01.BackendData

2.	Data Summary

This dataset allows the support of the end-user website that will be developed in T5.2. The data models being implemented are being used for the visualization of certain statistics and the algorithmic processing in order to fulfil the user's physical and behavioral assessment. This data and final product will be useful for the end-users.

The data will be originated by the different sensors/apps/devices being integrated in the pilots and defined in the T3.2 (Data model) while the resulting products of its processing will be defend in the T4.2. Both the raw data and the processed data is what the Backend dataset consist on.

Due to the intrinsic dependency among the different tasks and the planned schedule, the exact nature and specification of the dataset and their expected size is not yet available.

3.	FAIR Data

3.1	Making data findable, including provisions for metadata

According to the planned schedule of tasks, this dataset is not yet defined.

3.2	Making Data openly accessible

No data of the dataset will be publicly available as for just simple research publications using specific and anonymous statistics.

The access of the data will be restricted to the official client (either web or app) so it has to be performed with a mandatory authentication of the interested user; being capable of accessing their own and only data sandbox and their own processed results.

The access will consist on a single factor authentication (extensible to double in the future) based on username and password.

3.3	Making Data interoperable

The dataset won't be re-used by any other partner since this is the end-user tool. Also, according to the data protection regulation, this data should not be shared. In case specific requirements arises that require the data sharing, this implementation will be aligned with the GDPR and will be shared using JSON based web services operating over a well-documented web-API.

3.4	Increase Data Re-use (through clarifying Licences)

N/A. In case specific requirements arises that require the data sharing, this implementation will be aligned with the GDPR and will be shared using JSON based web services operating over a well-documented web-API.

4. Allocation of Resources

Data allocation will take place in a webserver property and located in the UPM. The UPM assumes all costs of this development and deployment of the in-production data center and related costs and maintenance. A cold redundant database for backup purposes will be stored physically as well in the UPM office.

5. Data Security

The data will be anonymized so no personal data is related to the meta-information and values stored. The incoming transactions will be protected with SSL, while the endpoint and APIs will be audited and safely designed in order to comply with known security risks and weaknesses such the ones documented on the OWASP Top 10.

6. Ethical Aspects

All the data will be used only by and for the projects purposes and accessed and exploited by the data scientist for anonymous research and statistical analysis. Personal data won't be used and no data sharing will take place outside the consortium. The end users are the only authorized entities to access their very own data and they do hold the rights for its rectification when desired according to the GDPR.

7. Other Issues

N/A.

3.11 Dataset "DS11.ASER-01.HypotheticalData"

1.	Dataset Name

DS11.ASER-01.HypotheticalData

Remark:

No data collection by ASER is planned at present. It is possible that data collected by other project partners will be transferred to ASER in anonymized/pseudonymized form during the later course of the project and evaluated by us. At present, however, no statement can be made as to the type and scope of the data or the possible objectives of the evaluation. As soon as more details are known, the information will be sent to the responsible persons for data management within the Ageing@Work project.

2.	Data Summary

State the purpose of the data collection/generation.

Investigation of questions, e.g. within the framework of T2.4.

Explain the relation to the objectives of the project.

Investigation of questions, e.g. within the framework of T2.4.

Specify the types and formats of data generated/collected.

Probably data in MS-Excel format or for IBM SPSS.

Specify if existing data is being re-used (if any).

Possibly, but it is not yet possible to conclude at this stage.

Specify the origin of the data.

Probably data collected from other project partners.

State the expected size of the data (if known).

Unknown.

Outline the data utility: to whom will it be useful?

Currently no statement possible.

3.	FAIR Data

3.1	Making data findable, including provisions for metadata

Currently no statement possible.

3.2	Making Data openly accessible

Currently no statement possible.

3.3	Making Data interoperable

Currently no statement possible.

3.4	Increase Data Re-use (through clarifying Licences)

Currently no statement possible.

4. Allocation of Resources

Currently no statement possible.

5. Data Security

Address data recovery as well as secure storage and transfer of sensitive data.

Regular backups are created, full backups and incremental backups. It is ensured that only authorized persons have access to the IT systems (user ID, password protection) and the server is protected by firewall, access from outside only possible via VPN and defined data directories.

6. Ethical Aspects

Currently no statement possible.

7. Other Issues

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

All employees involved are obliged to maintain the confidentiality of the data and are instructed in the safe handling of the data.

References

[1] TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)
Annotated version for the use of participants under Societal Challenge 1
http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf

[2] General Data Protection Regulation (GDPR)
REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 April 2016
on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
All Languages: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679

[3] Kraaij, Prof.dr.ir. W. (Radboud University & TNO); Koldijk, MSc. S. (TNO & Radboud University); Sappelli, MSc M. (TNO & Radboud University) (2014):
The SWELL Knowledge Work Dataset for Stress and User Modeling Research. DANS.
https://doi.org/10.17026/dans-x55-69zp

[4] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018):
Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection;
ICMI 2018 – Proceedings of the 20th ACM International Conference on Multimodal Interaction
(Pages 400-408)
https://doi.org/10.1145/3242969.3242985

[5] Smets E, De Raedt W, Van Hoof C. Into the Wild: The Challenges of Physiological Stress Detection in Laboratory and Ambulatory Settings. IEEE journal of biomedical and health informatics. 2019 Mar;23(2):463-73.

[6] Panicker SS, Gayathri P. A survey of machine learning techniques in physiology based mental stress detection systems. Biocybernetics and Biomedical Engineering. 2019 Feb 23.

Annex A:
Dataset Identification Template

In order to develop the deliverable above, a template for dataset identification presented below in this Annex was first drafted on the basis of the EC guideline document "Template Horizon 2020 Data Management Plan" (see [1]) in accordance with the table structure in Annex I to that EC document. This was distributed to all project partners so as to collect all relevant information concerning the datasets that they are planning to develop during the course of the project. On the basis of the partner feedback received, the preliminary Data Management Plan of the project was then developed, as described in the deliverable above.

Ageing@Work

Smart, Personalized and Adaptive ICT Solutions for Active, Healthy and Productive Ageing with enhanced Workability

Project Acronym:	Ageing@Work
Project Full Name:	Smart, Personalized and Adaptive ICT Solutions for Active, Healthy and Productive Ageing with enhanced Workability
Grant Agreement:	No 826299
Project Duration:	3 years (starting 1 January 2019)

Data Identification Template
Internal Report

(To be filled in by all consortium partners)

Work Package:	WP9: Project Coordination and Management
Task:	T9.3 Data Management, Ethics, and Standardization
Authors:	ALL PARTNERS
Status:	Final
Style:	Internal Report
Distribution:	All Partners
Document ID:	AAW_DataIdentiticationTemplate_PARTNER-NAME.doc

Executive Summary

This *Ageing@Work* form template for identifying dataset(s) has been built on the basis of the EC document *"Template Horizon 2020 Data Management Plan"*[5] and is structured according to the template tables provided in *Annex I* of that EC document.

The purpose of this form template is to facilitate each consortium partner of the *Ageing@Work* project in identifying and properly describing the dataset(s) that are planned to be generated during the project, toward the development of the project's overall *Data Management Plan*.

---

[5] http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf

The table(s) below should be filled in by each consortium partner according to the given questions, if applicable. One table instance should be provided for each expected dataset.

Dataset "PARTNER-DX.DatasetName"

1.	Dataset Name

2.	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?

3.	FAIR Data
3.1	**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.

3.2	**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.

3.3	**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?

3.4	**Increase Data Re-use (through clarifying Licences)**

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.

4.   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.

5. Data Security

Address data recovery as well as secure storage and transfer of sensitive data.

6. 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.

7. Other Issues

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