WORKINGAGE

Smart Working Environments for All Ages

D2.5 Field Test Strategy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 826232

WP2 – User Centric Design

D2.5 – Field Test Strategy

Project number:	826232
Project acronym:	WorkingAge
Project title:	Smart Working Environments for All Ages
Author(s):	J. Czerniak, S. Comai, V. Rick, L. Sbattella
Partners contributed:	POLIMI, RWTH
Delivery date:	M8 (September 2019)
Document date:	17/10/2019
Version:	1
Revision:	1
Deliverable type:	Report
Remarks:	
Status:	⦿ PU (Public) O PP Restricted to other programme participants (including the Commission Services) O Restricted to a group specified by the consortium (including the Commission Services) (please specify the group) O Confidential, only for members of the consortium (including the Commission Services

Document Revision Log

VERSION	REVISION	DATE	DESCRIPTION	AUTHOR
0	0	12/9/19	Just a starting point to foster discussion.	S. Comai, L. Sbattella
0	1	23/9/19	Description of use cases	L. Sbattella
0	2	27/9/19	Info on lab tests	J. Czerniak, V. Rick
0	3	1/10/19	Info on tests	L. Sbattella
0	4	3/10/19	Info on field tests	L. Sbattella
0	5	3/10/19	Review	M. van Gasteren
0	6	3/10/19	Review	G. Pelosi
0	7	7/10/19	System integration in Lab tests	L. Sbattella
0	8	8/10/19	Sensor about "pose" for the Driving use case	L. Sbattella
0	9	15/10/19	Details about planning and size of the pilot. Pilot for "home environments". Fixes to tables on sensor estimations. Added the "smartband" sensor. Review by BS and TPZ. Added figures from DoA for supporting test schedule. Time schedule fixed to meet Tab 5 of DoA part B. Added the Skin Temperature and gesture sensors.	L. Sbattella, B. Valentin, V.Ronca, G. Borghini
1	0	17/10/19	TPZ sensor added to tab 5	L. Sbattella
1	1	18/10/19	Final review before submission	M. van Gasteren

Acronyms and Terminology

The following table reports the acronyms used in this deliverable.

Term	Definition
AR	Augmented Reality
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculogram
GA	Grupo Antolín
GSR	Galvanic Skin Response
HCI	Human Computer Interaction
ST	Skin Temperature

Executive Summary

This deliverable is one of the outputs of the project task T2.5 DEFINITION OF PILOT TESTS, whose main objective is to address the specifications of the pilot applications to be deployed.

D2.5 defines the how & when of the lab & company pilots: time schedule, workplaces, description of sample workers.

Notice that the protocol for system evaluation is given in D2.6. D2.5 provides guidelines for preparing and conducting the tests.

This document is the first release of D2.5, the second one being scheduled for M16. Thus, the document provides as we currently envision the pilot test strategy. Notice, however, that the second release of D2.5 could bring many changes due to refinements or more deep changes.

The output of the work depicted in D2.5 and D2.6 will be the basis for WP3, WP4, and WP9; in particular WP9 is in charge of providing document D9.1 Pilot Operational Manual specifying the field test protocol.

This document is structured as follows:

- Section 1 introduces the goals of the test strategy and its relationships with D2.6;
- Section 2 describes the use cases;
- Section 3 specifies what we are going to measure;
- Section 4 focuses on the description of Lab Tests;
- Section 5 introduces the Field Tests;
- Section 6 further specifies short-term Field Tests;
- Section 7 further specifies long-term Field Tests.

Table of Contents

1 Introduction

1.1 Goals of the test strategy

The goal of the test strategy is providing a guideline for conducting the testing activity about all the modules of the WA Tool.

In particular, the testing strategy is divided into three parts:

- Lab Tests
- Short-term Field Tests
- Long-term Field Test

For each of them, this deliverable provides an initial set of guidelines, and specifies the differences about the three case studies where the WA Tool will be deployed.

Such case studies are:

- Office
- Driving
- Manufacturing

1.2 Main interrelationships with other deliverables

The present deliverable is strictly bound to D2.6 "Intervention Protocol". It receives inputs and/or provides guidelines and requirements to:

- WP3, in particular for the definition of measures to be adopted at the different pilot sites;
- WP4 where Lab Tests will be performed;
- WP7, dealing with ethical, security and privacy issues;
- WP9, dealing with the final Pilot Operational Manual.

D2.5 version1 will be refined at M16 with the inputs from WP3 and WP4.

It will provide basic input to WP9 for the definition of the "Pilot Operational Manual" (D9.1) that will describe in detail the pilot application, the documents to be used during the pilot tests (e.g., informal consent, questionnaires, etc.), user recruitment rules at each pilot site, performance metrics and their assessment procedures.

Figure 1 depicts the interrelationships between D2.5 and other WPs.

Finally, D2.2 "Analysis of Available and Suitable Sensors" provides an in-deep analysis of the technologies adopted by sensors we are going to deploy, while D2.3 "Data Management Plan" provides details on the pseudo-anonymisation and encryption.

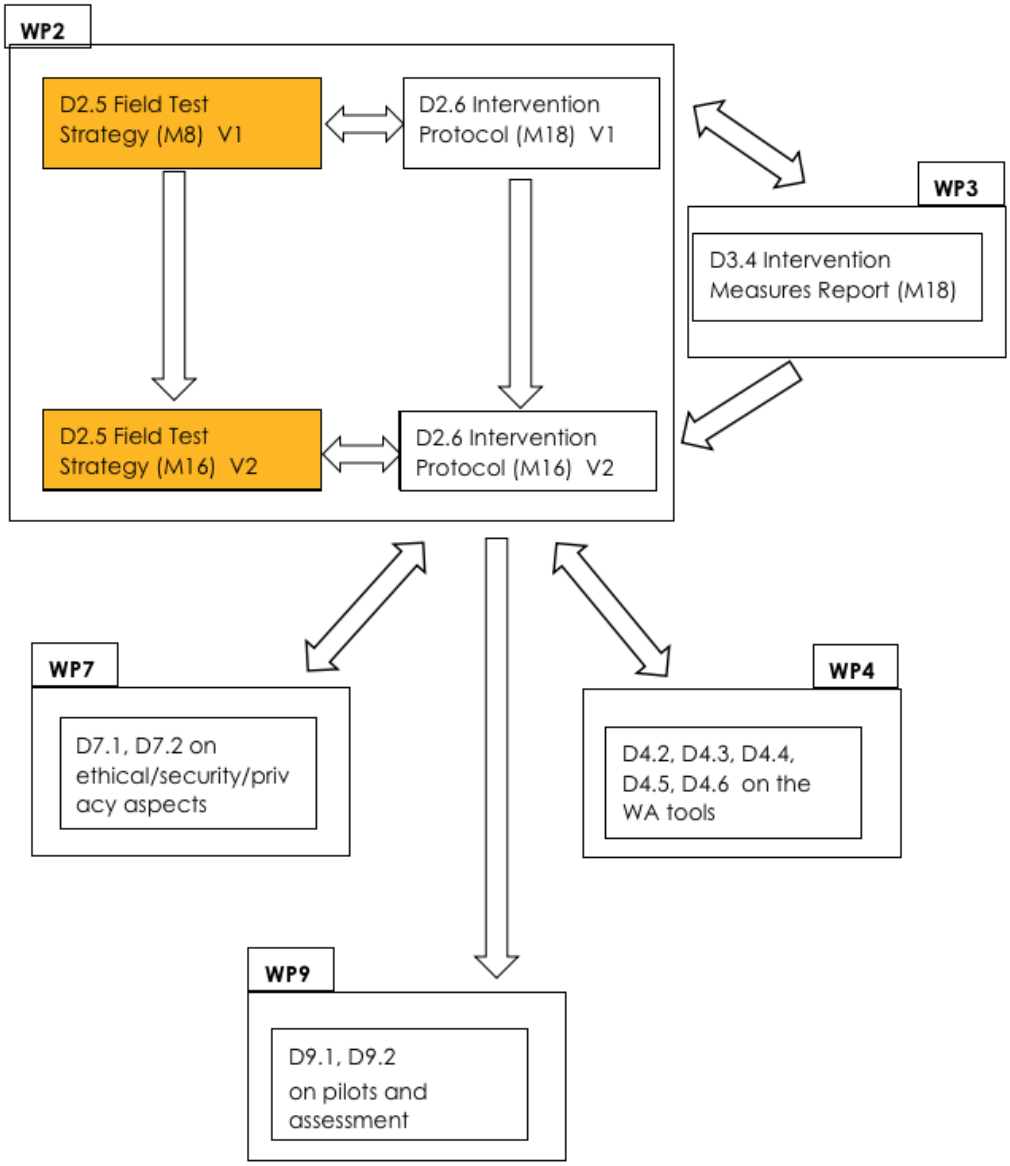

Figure 1 - Interrelationships between D2.5 and other deliverables/work packages

2 Use cases

The WA project will be tested in three reference use cases:

- Office
- Driving
- Manufacturing

Such use cases provide a good overview of the typical workplaces, and thus we believe they will permit to effectively testing the WA Tool.

2.1 Office

The Office use case will be at Mutua Universal in Barcelona. The Piraeus Bank offices mentioned in the DoA turned out to hardly have personnel over 50 years old.

The description of the office workplaces at Mutua Universal will be included in the M16 version of this deliverable.

2.2 Driving

The company IPLUSF, that had signed a Letter of Interest to provide the Driving use cases, unfortunately withdrew from the project.

The consortium entered in contact with an Athens-based ambulance company, First Aid Ambulance, who showed great interest in the project. Unfortunately, as occurred with the bank offices, they do not have sufficient drivers aged over 50 to be able to host the field tests.

The consortium actively searched for alternatives and we are currently very close to an agreement with a large company. Again, more information in the M16 version of this document.

Since general workflows at both office and driving workplaces are quite well defined, delays caused by these company switches are reduced.

2.3 Manufacturing

The factory, belonging to the GA RyA, which agreed to be the WA project pilot site for the "Manufacturing" use case, is in Valladolid (Spain). The factory manufactures car interior components such as dashboards, doors, etc.

2.3.1 Workplace

The factory is organised into several "islands" where specific jobs are performed by workers with the aid of specific tools[1]. Many different jobs are performed, such as assembling a dashboard, for example, which require several steps.

[1] As this is a public deliverable, no images of the specific workplace are included due to confidentiality requirements from GA RyA.

After observing the whole process we selected some feasible "island" that allowed, more easily than others, to place sensors without interfering too much with the worker's activity:

- Leather inspection G-Class
- Stitching G-Class
- Assembly G-Class
- Assembly XFB
- Welding XFB

In such settings, workers do not move a lot (which is good for sensors based on cameras) and the environment should not harm the electronics (it's not too humid or hot). Unfortunately, the whole factory is way too noisy for any kind of microphone (even noise-cancelling ones) to be employed to the porpoise of the WA Tool, so voice recording is not feasible.

Temperature and humidity could make the environment not conformable for workers, especially for some of the selected "islands", while light seems good (intensity and position).

2.3.2 Supporting infrastructure

Electric power supply is available.

A room could be used to deploy WA servers.

2.3.3 Possible issues affecting sensor deployment

In some areas of the factory, heat and humidity could harm sensors; thus, we selected the "islands" to avoid such risks.

2.3.4 Privacy concerns

Since audio recordings are not to be collected, privacy concerns are only related to the public disclosure of images containing sensitive information for the factory.

2.3.5 Health and stress problems

The major health and stress problems at PDS are due to the posture, and to the stress of keeping up to the pace required by the job procedure.

2.3.6 Human Machine Interaction

The job could be dangerous (some machines require high attention to avoid injuries), so an interactive HCI could feasible but requires special care. In other words, the WA Tool should be able to provide suggestions and recommendations anytime the worker needs it (for example, when she/he is getting overloaded, or overstressed); thus, the WA Tool should be able to interrupt the worker also during the working activities, but considering her/his status (for example, when the worker is taking a break).

The HCI could be based on a smartphone where a simple touch-based UI will be present. Interaction could be both WA tool-initiated (sort of "interrupt") and worker-initiated (a "request").

2.3.7 Notes and issues

None.

3 Measurements and sensors

The WA system is going to deploy several sensors and gather their measurements. The following sections provide an overview of what is going to be collected by the WA system.

Note that we have two measurement typologies:

- Utilising electronic sensors
- By means of questionnaires

These two different methodologies to gather data about workers permit to compare the behaviour of the WA tool, assessing its perceived effect on workers' life.

3.1 Sensors

The following body sensors will be analysed in Lab Tests and, if proving useful, deployed in Field Tests; this list was decided after the technical survey provided in D2.2.

- ELECTROCARDIOGRAM (ECG)
- ELECTROENCEPHALOGRAM (EEG)
- ELECTROMYOGRAPHY (EMG)
- GALVANIC SKIN RESPONSE (GSR) / SKIN TEMPERATURE (ST)[2]
- FACIAL EXPRESSION
- VOICE ANALYSIS: BLUETOOTH HEADSET WITH NOISE-CANCELLING MICROPHONE
- ELECTROOCULOGRAM (EOG)
- EYE MOVEMENT AND PUPIL DIAMETER
- BODY POSTURE
- GESTURE RECOGNITION

Additionally, the following environmental sensors will be tested:

- NOISE
- THERMO-HYGROMETRIC
- LIGHT
- USER LOCATION

Correlations between strain and environmental conditions will be analysed according to the best practices described in the current public academic literature.

As already derived in D2.1, the strain types mentioned in Table *1* can be investigated by means of the physiological parameters monitored by the sensors.

[2] GSR and ST are alternative; to be decided during Lab Tests.

Table 1. *Measurement of mental, emotional and physical strain with the Sensors used in the WorkingAge project*[3]

X:direct indicator O:indirect indicator	Electrocardiogram (ECG)	Electromyography (EMG)	Galvanic skin response (GSR)(*)	Electroencephalogram (EEG)	Electrooculogram (EOG)	Facial expression	Voice Analysis	Gesture recognition	Pupil diameter (PD)	Body Posture	Eye-Movement
psychological strain								o			
• mental	o		o	x					o		o
sensory				o	o				o		
discriminatory			o	o					o		
combinatory			o	x					o		
• emotional	o		x	o		o	o	o	o		
physical strain								o			
• muscular	o	x								o	
dynamic	x	x						o			
static	o	x	o							o	
cardiovascular	x		o	o							
• skeletal	o	o								o	

(*) ST will be tested as an alternative sensor, for collecting comparable indicators.

3.2 Questionnaires

We also plan to collect information on workers by administering specific questionnaires to collect different type of data such as:

- Demographic data, e.g. age, gender
- Health status
- Cognitive and emotional situation
- Home time (nutrition, exercise, sleep, etc.)

Such questionnaires could be paper based or electronic. In particular, during Lab Tests questionnaires will be probably paper based, while during Field Tests, an electronic format will be adopted (the device used as system HCI will be used to administer and manage questionnaires).

[3] Adapted and extended following Kirchner (1986).

3.3 Measurement-use case matrix

Because each use cases has its specific constraints WA is not going to deploy exactly the same set of sensors for the three use cases. Table 2 shows what WA is going to deploy, for each use case, at the workplace and at home.

Note that the questionnaire (described in Section 3.2) is considered here as a measurement since they will also gather data related to each WA user.

Table 2 - Measurements and use cases (W=workplace, H=home)

		Use cases		
		Office	**Driving**	**Manufacturing**
Measurements	**ECG**	W	W	W
	EEG	W	W	w
	EMG (*)	-	-	-
	GSR / ST (**)	W	W	W
	Facial expression	W	W	W
	Microphone	W	W	-
	EOG	W	W	W
	Eye movement, Pupil diameter	W	W	W
	Body posture	W	W	W
	Gesture recognition	W	W	W
	Noise	W	W	W
	Thermo-hygrometric	W	W	W
	Light	W	W	W
	User location	W	W	W
	Smartband	H	H	H
	Questionnaires	W/H	W/H	W/H

(*) EMG sensor will be only deployed in Lab Tests, to derive theoretical considerations/concepts for the future integration of the EMG system into the WA Tool

(**) GSR and ST are alternative; Lab Tests will help in selecting the best one, in terms of comfort, signal stability, etc.

3.4 Measurement-data owner matrix

Each measurement is under the responsibility of one or more WA partners. Table 3 shows the partner(s) responsible for each measured data.

Table 3 - Measurements and data owners

		WA partners								
		POLIMI	AUD	ITCL	BS	INTRAS	RWTH	UCAM	TPZ	EXUS
Measurements	ECG				X					
	EEG				X					
	EMG						X			
	GSR / ST				X					
	Facial expression							X		
	Microphone	X	X							
	EOG				X					
	Eye movement, Pupil diameter						X			
	Body posture			X						
	Gesture recognition									X
	Noise		X							
	Thermo-hygrometric			X						
	Light			X						
	User location								X	
	Smartband			X						
	Question-naires					X				

4 Lab Tests

Lab Tests will assess system acceptability, usability, and validity. Each partner designs its own tests and manages its own documents, hardware, software, and data.

4.1 Goals

Lab Tests represent the proof-of-concept of the WA sensor components. At this stage, the sensor components for the final WA system will be selected, with respect to the study findings.

The main goal is to assess acceptability, usability, and validity. Each partner should indicate here its own specific:

- goals
- statistical methods
- evaluation methodologies
- Safety assessment

At the end of Lab Tests, acceptable, usable, and valid sensors will be considered for Field Tests.

An integration study, testing the whole WA system will be conducted at the end of Lab tests. During the integration study, we aim at finding:

- Possible interferences with other partners' sensors
- Acceptability, usability, and validity of the whole WA system

4.2 Recruitment

The WA project aims at a well-defined category of workers; the main requirements are:

- Possibly, age 50+
- Possibly, gender-balanced groups
- No disabilities

The Lab Tests, however, are a part of the development cycle, and the goal is to test the reliability of the subsystems that will compose the WA Tool. Therefore, the requirements mentioned above are not strictly enforced. In other words, each WA partner is free to select the more convenient set of users for testing its equipment.

4.3 Informed consent forms

It is under the responsibility of each WA partner to collect the informed consent forms from the users involved in Lab Tests. The information consent will involve the permission to:

- Measure
- Elaborate
- Store

the data collected by the WA partner.

The informed consent form will specify how long data will be retained by the WA partner, and how users can request to delete their own data.

Other WA partners won't be allowed to access the data, unless the informed consent form explicitly mentions that possibility, and the user explicitly allows it.

The consent form in paper format will be stored in the correspondent country in which they are generated.

4.4 Research questions

The research questions to be answered for Lab Tests refer to sensors and the appropriateness of physiological indicators as measure of mental, emotional, and physical strain itself.

About sensors, the objective is to prove that each of them is valid, feasible, and usable to measure its foreseen type of strain in the suggested way.

The general research questions for each individual sensor should be:

1. Validate the measurement parameter/sensor (Does the sensor measure what we want to measure?)
2. Evaluate differences between younger and older participants.

Then, Lab Tests will aim at confirming the appropriateness of the chosen physiological indicators to gauge the mental, emotional and physical strain of workers.

In order to determine the specific research questions, a literature survey for the above-mentioned sensors is necessary. This has to consider especially the following points:

1. Nature of the stimulus:
    a. Cognitive or informational task influence data collected by eye tracking, GSR, heartbeat measurement devices (sensors)
    b. emotional stimuli influence data collected by face and voice recognition devices;
    c. physical stimuli influence data collected by EMG as well as the body posture analyses;
2. Age-related differences of the measured parameters.

4.5 Study Design

The actual planning of the study depends on the kind of strain considered. As Figure 2 shows, uniform questionnaires are used in all settings to collect demographic and other study dependant data, e.g. control variables. To validate the sensors, the stimulus is changed in randomized cycles if it is required so by study design.

Closing questionnaires ask for parameters related to the measurement and the hypothetical system of the WA tool, such as acceptance and usability. The following list details the study workflow.

1. Questionnaires concerning participants are employed:

- To collect data about age, gender and health restrictions of the participants.

2. Measurement:
   - alternating strain
   - approximately about 3 to 5 different strain levels (two levels, for mental stress)
   - several cycles
3. Test Persons:
   - Female and male test persons equally distributed
   - Two age groups: younger than and older than 50 years, dependant on the requirements
   - The number of participants is related to the study design. For the pre-studies only a small number of participants is enough to recognize tendencies.
4. Additional subjective strain measurements, e.g.:
   - NASA-TLX: To evaluate the relationship of measured and subjective strain
   - Rating Scale of Mental Effort (RSME)
5. Statistical Analyses, for instance:
   - Within-subject-design
   - Mean value difference and two-way repeated measure of variance analysis tests for:
     i. Significance of mean value difference between no strain and strain condition, and between different strain levels
     ii. Significance of mean value difference between 50+ and younger test subjects (as stated above, the requirement about subject recruitment are not strictly enforced, as the goal is to find the more convenient set of users for testing the equipment),
   - Correlations between the evaluated physiological and subjective parameters
6. Closing questionnaires

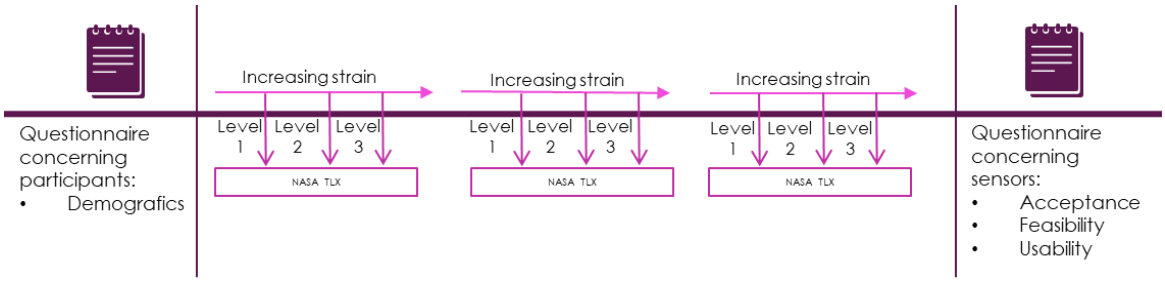

Figure 2 - Study design example

4.6 Time schedule

The time schedule for the Lab Tests can be derived from the Gantt chart in the DoA. A recommendation for testing and data analysing periods is addressed in the following sections. The final time schedule will be discussed in consultation with the work package leader of WP2 (RWTH).

According to the DoA the Lab Tests will be prepared between Month 11 and Month 16 (Dec 19 – May 20). During this period the study design will be developed and pre-tests for the preparation of the pilot tests will be completed including the analysis of the results. During this phase, we will involve a total of **30** subjects.

The conduction of further Lab Tests will take place between Month 20 and Month 25. Results of these tests will be used as basis for the pilots. During this phase, we will involve a total of **60** subjects (30 being the same involved in phase I).

The following time schedule can be estimated:

M11 (Dec 2019): Planning of the Lab Tests & developing study design
M16 (May 2020): End of Lab Tests phase I
M20 (Sept 2020): Planning of further Lab Tests phase II
M25 (Feb 2021): End of Lab Tests phase II

5 Field Tests: general info

Field Tests aim at assessing the system reliability and effectiveness. There are two Field Test typologies: short-time Field Tests and long-time Field Tests. In the following, the characteristics that these typologies share will be introduced.

Note that the protocol for system evaluation is given in D2.6. D2.5 provides guidelines for preparing and conducting the tests.

5.1 Informed consent forms

Each WA partner will adopt a standardised set of informed consent forms:

- Involving users
  - Document addressing a specific kind of data collected by a specific WA partner. Such documents will involve the permission to:
    - Measure
    - Elaborate
    - Store

    data collected by the WA partner. Each user involved into Field Tests must agree with these documents.
  - Document addressing elaboration of data by means of a centralised Agent, in change of providing users with personalised advice. Each user involved in Field Tests must agree with these documents.
- Involving the company
  - Document addressing a specific kind of data collected by a specific WA partner. Such documents will involve the permission to:
    - Measure
    - Elaborate
    - Store

    data collected by the WA partner. Each company involved into Field Tests must agree with these documents.
  - Document addressing elaboration of data by means of a centralised Agent, in change of providing users with personalised advices. Each company involved into Field Tests must agree with these documents.

The consent forms in paper format will be stored in the correspondent country in which they are generated (D2.3).

5.2 Test setting at work

Each sensor needs to meet a set of requirements, due to the very nature of the sensor and its technological limitations. Therefore, the test settings should be selected in a way that permits to all sensors to operate correctly.

Notice that sensors listed in Table 4 will be actually deployed after the study in Lab Tests shows they are acceptable, usable, and valid.

5.2.1 Common to all use cases

In general, a static setting is preferable, as cameras and eye trackers cannot "follow" users while they move. Moreover, a "reasonably quiet" environment is needed for the microphones to collect usable vocal samples[4]. Finally, the environment should not harm the sensors (no excessive humidity, dust, etc.)

The company hosting the use-case should provide a room for deploying some WA servers (see Section 5.3). The company should also provide places where the WA "smart Hot Spots", provided by GC, will be deployed.

5.2.2 Use case: Office

Sensors will be deployed on the worker's body and on the desk. The sensors we plan to deploy are:

- ECG, EEG, GSR, EOG, microphone, user location: on the worker's body
- Facial expression: a camera standing in front of the worker, on the monitor or on the desk
- Eye movement, pupil diameter: a device put on the desk, in front of the worker (a device mounted on glasses is under evaluation, too)
- Body posture: a camera put sideways, e.g. on a wall, observing the whole worker's body
- Gesture recognition: a device put in front of the worker
- Noise, thermo-hygrometric, illumination: sensors put in the office, not far from the worker
- Questionnaires: provided by means of the system HCI

The environments should not pose risks about sensor integrity.

5.2.3 Use case: Driving

Sensors will be deployed on the driver's body and on the dashboard. The sensors we plan to deploy are:

- ECG, EEG, GSR, EOG, microphone: on the worker's body[5]
- Facial expression: a camera put in front of the worker, on the dashboard
- Eye movement, pupil diameter: a device put in front of the worker, on the dashboard (a device mounted on glasses is under evaluation, too)
- User location: a sensor put inside the vehicle
- Body posture: a camera put in the cockpit, observing the worker's body
- Gesture recognition: a device put in front of the worker
- Noise, thermo-hygrometric, illumination: sensors put in the vehicle
- Questionnaires: provided by means of the system HCI

The environments could pose some risks about sensor integrity, as some of them will be installed under direct sunlight (the dashboard). During the field short-term study, these aspects will be investigated.

[4] We are going to use noise-cancelling microphones, but this technology has limitations.
[5] We will check regulations about traffic, in the Country where the Field Test will be conducted, to ensure sensors can be deployed on a driver's body. Moreover, during the Lab Test, we study how to ensure appropriate safety conditions.

5.2.4 Use case: Manufacturing

Sensors will be deployed on the worker's body and on the machine operated by her/him. The sensors we plan to deploy are:

- ECG, EEG, GSR, EOG, user location: on the worker's body
- Facial expression: a camera put in front of the worker, on the desk or machine she/he is operating
- Eye movement, pupil diameter: a device put in front of the worker, on the desk or machine she/he is operating (a device mounted on glasses is under evaluation, too)
- Body posture: a camera put on a wall, observing the whole worker body
- Gesture recognition: a device put in front of the worker
- Noise, thermo-hygrometric, illumination: sensors put in the workplace, not far from worker
- Questionnaires: provided by means of the system HCI

Sensors we currently do not plan to deploy:

- Microphone: the environment is too noisy even for a noise-cancelling microphone

The environments could pose some risks about sensor integrity, in particular dust, temperature, and humidity could harm some sensors.

5.2.5 Augmented Reality at the Manufacturing use case

The AR specific application will be a module of the WA Tool. This module will only be available for the workers in specific workplace(s) selected for it. Due to the fact that this type of application has to be specifically developed for each case, the objective of this part of the study is to understand whether this type of content can contribute to the reduction of the mental overload or the worker.

After the visit to the facilities of GA RyA, the preselected workplace for the AR experiences is the Kitting XFB: the worker needs to pick up several car door parts, in a 10-12 meter corridor containing several stacks. She/he has to pick up the material needed for other workplaces and, in some cases, mount some parts together, add tags, etc.

In this workplace we envision to provide guidance to the worker exploiting several devices: Hololens, Magic Leap, etc.

A new visit to the plant and a more detailed work description will be made to gather all the information needed and start with the development.

5.3 Test setting at home

About data collection at home, we plan to collect information by means of:

- Administering specific "home time" questionnaires, as specified in Section 3.2, using a mobile device
- Providing a smartband (e.g. the inexpensive Xiaomi MiBand) to obtain indication of sleeping habits and exercise measuring sleep and steps. Data stored into the band will be periodically uploaded into the WA Tool system

5.4 Privacy, anonymisation, and obfuscation

Privacy, anonymisation, and obfuscation are very important for the WA project. In the following sections, we explain how we plan to tackle these matters (technical detains can be found in D2.3).

5.4.1 Privacy

The WA project must be GDPR compliant. For that reason, privacy was one of the most important goals, from the very beginning of the project.

To be GDPR compliant, we plan to:

- Manage collected data so that only the WA partner that collected them can access them (encryption)
- Pseudo-anonymise users by means of an ID, and all recorded data refer to that ID. For each use case, a WA member (the WA use case manager) will maintain an encrypted file containing the mapping worker's name/ID; only that person will be able to access the content of the encrypted file.
- Prepare, administer, collect, and store informed consent forms

5.4.2 Anonymisation, and obfuscation

Companies, in general, do not allow us to elaborate sensible information about their workers, outside their facilities. This issue particularly affects voice recording and cameras. From those sensors, in fact, it is possible to recognise the person under measurement. Moreover, voice recordings could contain very private information (person names, bank accounts, etc.)

To cope with that serious issue, we envision a public-key based encryption workflow where data collected by sensors are sent to a centralized cloud storage, managed by BS, where only the WA partners in charge of further elaborating them can access them.

This way, each WA partner is allowed to access only the data legitimately intended for him by the data collector. The data pseudo-anonymisation will be performed on an in-company deployed server. Each data type will be anonymised and obfuscated in a specific way. A detailed description of such methodologies will be provided in the second release of the present Deliverable. Such scheme works by means of a locally deployed server, which is in charge of applying anonymisation and obfuscation to sensible data (for example, voice and video recordings). Each data typology will be anonymised and obfuscated in different ways. A detailed description of such methodologies will be provided in the second release of the present Deliverable.

5.5 Logistics and management

This aspect is about how to install and remove hardware and software:

- Bringing hardware on the field
    - Servers
    - Sensors

    - Telecom infrastructure
    - Device for HCI (a tablet or a smartphone), if workers can't use their own devices
- Installing hardware and software, and testing everything
- Removing hardware once the test is over

Moreover, during the test:

- Helping workers to equip with the sensors
- Moving sensors to a workplace to another (for example, from a desk to another one)
- Managing the WA servers
    - Deployed at the company facility
    - Deployed at WA partners
- Managing the WA Tool modules
    - Installed at the company facility
    - Installed at WA partners

We envision the following roles: a **company manager**, one or more **WA field managers**, several **WA home managers**, and a **WA use case manager**.

The WA field managers:

- Bring all the WA servers, WA telecom infrastructures, and devices for HCI to the company facility
- Administer the informed consent forms, for workers and company
- Install the WA Tool modules on the servers
- Bring the sensors, install the environmental sensors,
- Instructs the company manager about how to deploy wearable sensors, and about the WA HCI
- Test everything
- At the end of the test, remove everything.

The company manager:

- Helps workers to equip with the sensors
- Instruct workers about how to use the WA HCI
- Moves the environmental sensors, when needed
- Fixes small hardware/software issues

The WA home managers (one for each WA partner deploying sensors or telecom infrastructure, or running the Agent):

- Monitor software installed in their home server (if any), and fix errors
- Monitor the WA telecom infrastructure

The WA use case manager (incorporating the role of local Data Manger):

- Monitors that the whole WA Tool works well
- Maintains an encrypted file containing the mapping worker's name/ID, for that use case

The WA field managers are not required to stay at the company facility during the long-term Field Tests, whereas during the short-term Field Tests they (the whole group of WA field managers or part of it) will supervise the test.

Finally, note that the same person could play several roles.

5.6 Time schedule

The time schedule for Field Tests can be derived from the Gantt chart in the DoA (see Figure 3) and the methodology for involving companies described in Figure 4.

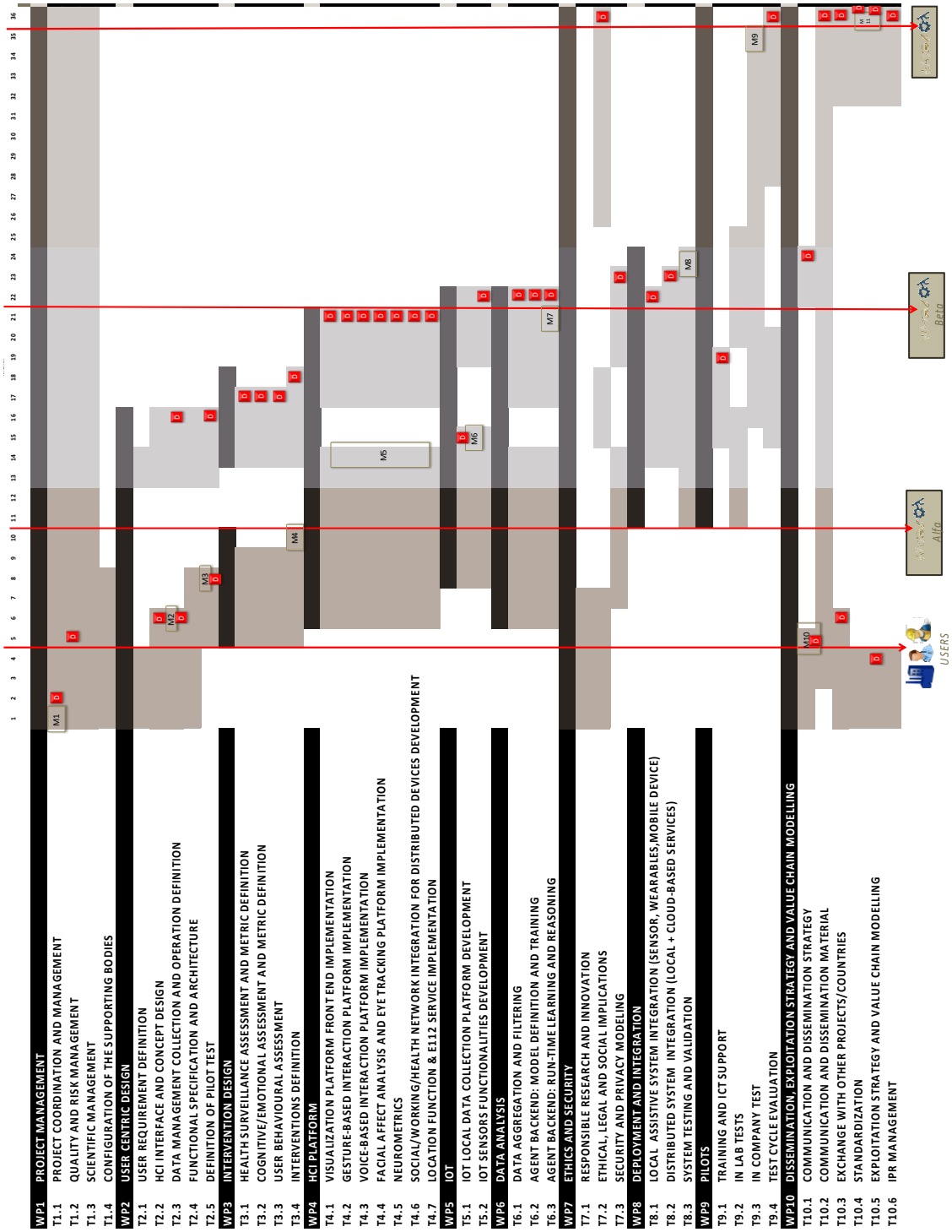

Figure 3 - WA general Gantt

Details about the schedule of the testing activities (including Lab Tests) can be found in Figure 5.

Notice that field studies on the three use cases will be conducted in parallel.

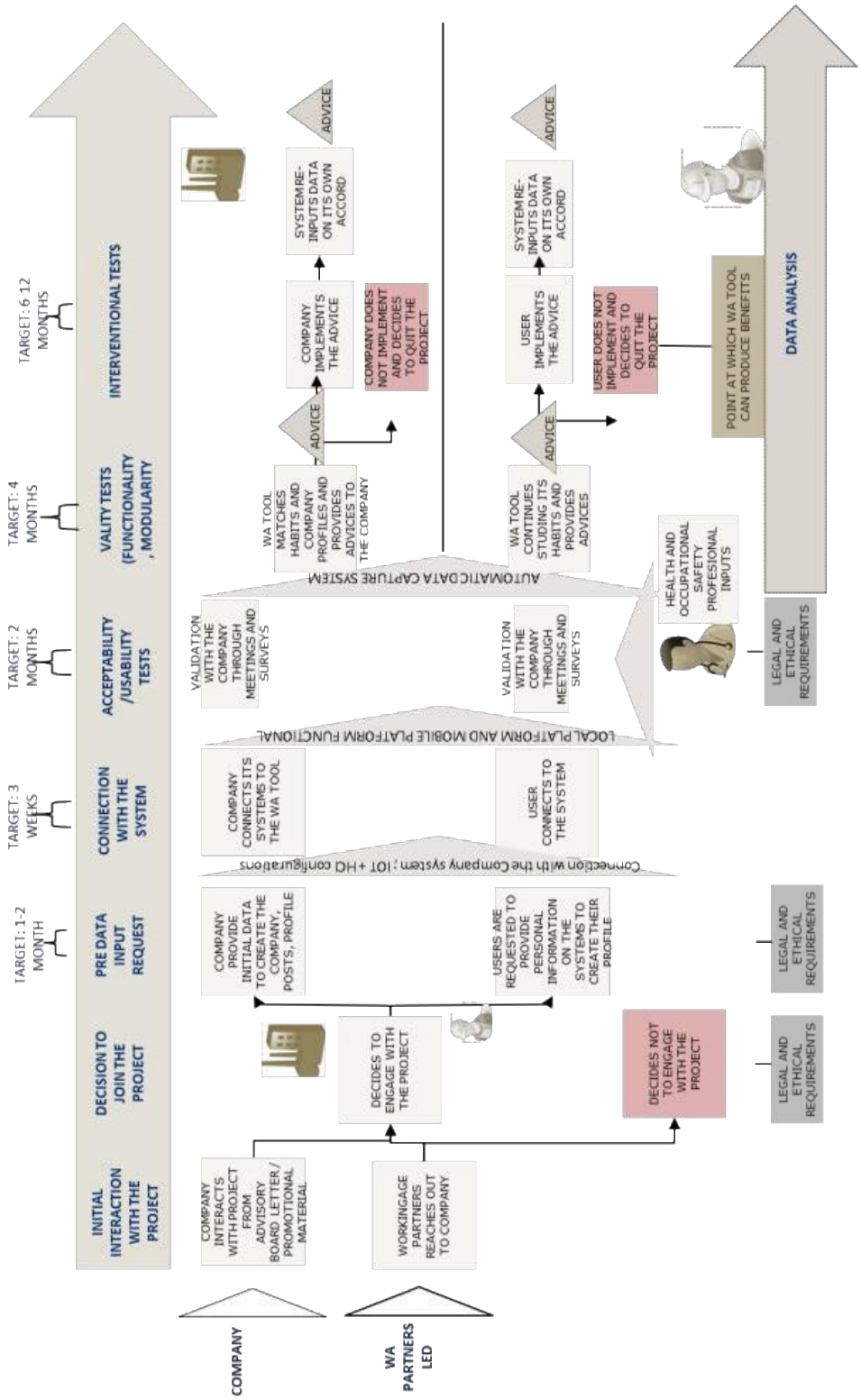

Figure 4 - Company implication in WA

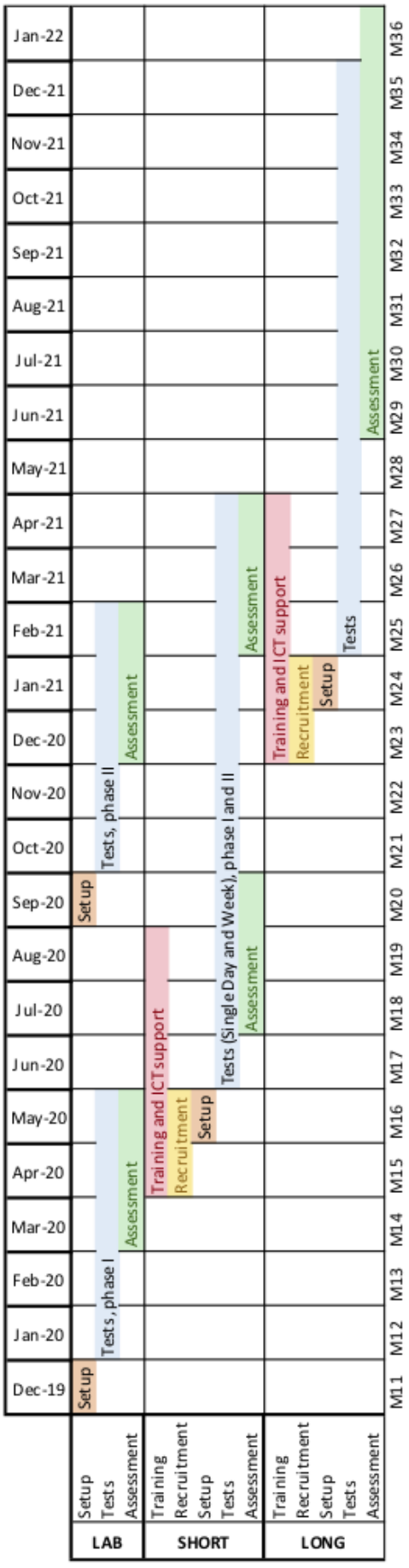

Figure 5 – Schedule

6 Field Tests: short-term study

Field tests aim at assessing the WA system reliability, when deployed on the field.

6.1 Goals

The main goal is to assess the WA reliability; in particular:

- Validating the acceptance of the tool by users (user-friendly, non-intrusive, not problematic for proper work conduction, etc.)
- Collect data to validate the data processing algorithms (for instance, validate that no false alarms are raised by the tool)
- Validate reliability of sensors, in particular in potential harmful environments (for example, under direct sun light)
- Validate the architecture of the system (local clouds, global clouds, databases managements, etc.) when deployed on the field; in particular: reliability and robustness

6.2 Recruitment

The WA consortium will need to organise a meeting at each company office to explain the system and the goals of WA and ask for their cooperation.

The main requirements for the recruitment are:

- Age 50+
- No disabilities

Starting from such requirements, each company will provide a list of possible candidates; then, the WA consortium will select the persons who will be involved into the project.

Note that gender aspects will be considered in the data analyses. However, because it is not straightforward to find a large enough sample of workers, it is decided not to put gender restrictions in the recruitment requirements. As a 'nice-to-have' requirement, in case the luxury of choosing participants exists, a gender-balanced group will be sought.

6.3 Time schedule

The goal of the short-term Field Tests is to assess that all the WA Tool components are working as expected. Fields Tests are divided into:

- Single Day, aiming at testing the WA components;
- Week, aiming at testing the whole WA system.

The general time schedule is:

M15 (Apr 2020):	Beginning of recruitment and training
M16 (Mar 2020):	Beginning of activities for short-term, Single-Day Field Tests, phase I

M20 (Sep 2020): End of activities

M23 (Nov 2020): Beginning of activities for short-term, Single-Day Field Tests, phase II
M27 (April 2021): End of activities

M20 (Sep 2020): Beginning of activities for short-term, Week Field Tests
M24 (Jan 2021): End of activities

6.4 Protocols

Each component requires a different testing protocol; in particular:

1. Sensor hardware (e.g., the microphone)
2. Server hardware and software, for machines deployed at the company facility (for example, machine and software needed for voice obfuscation anonymisation, and encryption)
3. Server hardware and software, for machines deployed at some WA partner (for example, machine and software needed to infer emotional state, from an utterance, installed at POLIMI)
4. Network infrastructure installed at the company facility
5. Device for HCI
6. Questionnaires
7. The whole Agent loop: sensors → information → agent → advice

Such protocols will be provided by:

- For components 1, 2, and 3, each WA partner will provide a protocol to be adopted for effectively testing everything.
- For component 4, GC will provide a protocol.
- For component 5, UCAM will provide a protocol.
- For component 6, INTRAS will provide a protocol.
- For component 7 (i.e., for testing that the whole WA Tool is working properly), the WA consortium will agree on a protocol.

Notice that here as "protocol" we mean a document specifying the test procedures to apply, the expected results, and how to address issues.

Small issues will be addressed by the WA field managers and the company manager, whereas more complex problems (in particular, issues affecting machine/software installed at WA partners) will be addressed by WA home managers.

We aim at involving **30** workers, per use case, during short-term Field Tests. Each WA partner will provide, as a part of its protocol, how many workers will be measured, and how many hours of data recordings will be collected.

7 Field Tests: Long-term study

Field tests aim at assessing the system effectiveness, when deployed on the field.

7.1 Goals

The main goal is to assess system effectiveness, when deployed on the field; in other words, understanding whether the WA Tool is able to reduce user's strain, by advices about stress management, body postures, and environmental control.

All the use cases will permit to highlight issues related to:

- Mental stress
- Physical strain
- Environment conditions

However, each use case, due to its very nature, will be more suitable for one or more of these issues.

7.1.1 Use case: Office

This use case is focused on body posture and mental stress.

Advices generated by the WA Tool could be in "real time" (i.e., proactively provided as soon as the system discovers that they are useful) or "offline" (for example, as a report provided at the end of the shift). The worker could also ask the WA Tool for any "pending" advice.

7.1.2 Use case: Driving

This use case is focused on mental stress and environment conditions.

Advices generated by the WA Tool will be "offline" (for example, as a report provided at the end of the shift), to avoid distraction. However, the worker could ask the WA Tool for any "pending" advice.

7.1.3 Use case: Manufacturing

This use case is focused on body posture, physical strain, mental stress, and environment conditions.

Advices generated by the WA Tool could be in "real time" (i.e., proactively provided as soon as the system discovers that they are useful) or "offline" (for example, as a report provided at the end of the shift). The worker could also ask the WA Tool for any "pending" advice.

7.2 Testing approach

The long-term Field Test requires measuring 30 workers per use case, along a period of one year. This means that we should provide enough sensors and computational power, and ensure that the WA tool stays up and running for one year.

The following factors must be taken into account:

- Each group deploying sensors has its own constraints in terms of devices available for the test (mainly due to user invasiveness/acceptance, e.g. wearing an EEG head cap continuously, or the cost of the device) and measuring periods (from few hours a day to a full day).
- Any failure will be resolved as soon as possible. Additionally, instability of sensors can be a reason to exclude them from the in-company tests

Thus, to comply with the claims stated in the DoA, we designed the following strategy, highlighted by Figure 6.

The rationale is monitoring each worker for a few hours per day (depending on the sensor), for several days (we envision a week). Then, after a while, the same worker will undergo another testing period under another data collection period, and so on until we reach the required measuring time.

This way, we ensure that we have enough data about every single worker, even if we cannot measure all of them at the same time. Moreover, we can manage to monitor all the workers we planned (30 for each use case) without deploying too many sensors.

Summing up, we plan to:

- Split one year into several so-called ***data collection periods***.
- Data collection periods are separated by a ***technical update iteration*** in which the gathered data and feedback will be used to improve WA Tool performance and stability. Note that whenever possible, of course we will do any maintenance on the go during the tests.
- Each data collection period is split up into several ***weeks***, where a given number of workers will be monitored.
- Weeks are separated by a ***technical inter-week iteration***
- Rotation schemes will define per sensor which workplaces will use the sensor, when and for how long; and how many sensors will be deployed in parallel. Some examples:
    - It is not necessary that an illumination sensor at an office measures 8 hours a day every day as office lighting is expected to be fairly constant.
    - EEG head caps are too uncomfortable to wear 8 hours a day.
    - Microphones are not invasive (and inexpensive) and can be used for the full working shift.

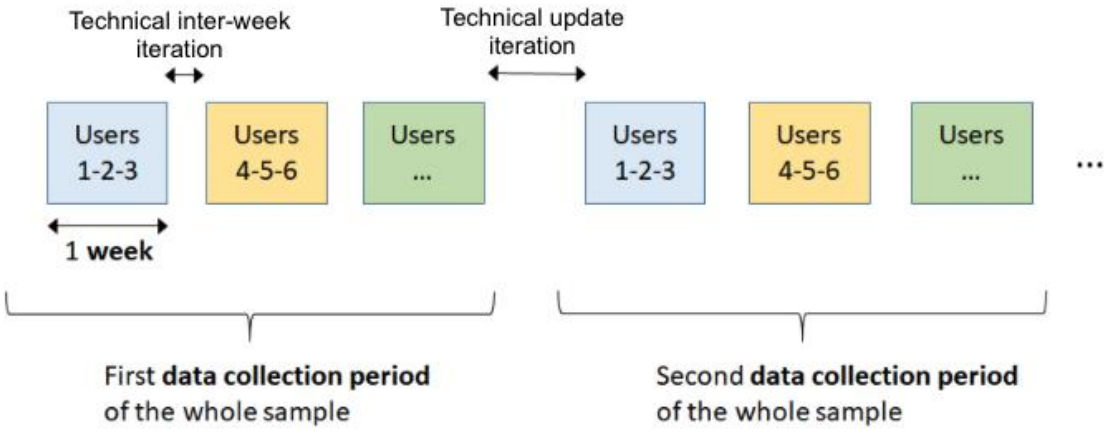

Figure 6 - Testing strategy

7.3 Recruitment

The WA consortium will need to organise a meeting at each company office to explain the system and the goals of WA, and ask them for their cooperation.

The main requirements for the recruitment are:

- Age 50+
- Gender-balanced group
- No disabilities

Starting from such requirements, each company will provide a list of possible candidates; then, the WA consortium will select the persons who will be involved into the project.

7.4 Time schedule

The general time schedule for the long-term Field Test is:

M23 (Dec 2020): Beginning of recruitment and training
M24 (Jan 2021): Beginning of activities related to long-term Field Tests
M35 (Dec 2021): End of activities
M36 (Jan 2021): Final assessment

Each WA partner provided initial estimations about (see Table 4):

- How many workers, at the same time, can be measured. Limitations can be due to:
  - Sensor availability (for example due to cost)
  - Software constraints about the number of concurrent data flows
- How long the measurement time is. Limitations are mainly due to battery life, recharging options/schemes necessary to guarantee operation
- How long it takes to move (environmental) sensors from working place to another one
- How long it takes to wear (personal) sensors
- The suggested working period before maintenance is needed (to the sensor, to software or to some server)
- How long it takes to complete maintenance

From such information, the following schedule details have been estimated:

- How many users we can monitor at the same time
- How long the "week" is (see Figure 6)
- How long the "technical inter-week iteration" is
- How long the "technical update iteration" is
- How many "data collection periods" we have

Table 4 shows initial estimations about each sensor typology; note that figures are *per use case*.

Table 4 – Information about sensors; initial estimation. Figures are per use case

			Sensor information				
		Home / Work	**# workers**	**Measurement time [h]**	**Time to move or wear**	**Working period before maintenance [h]**	**Maintenance**
Measurements	**ECG (BS)**	W	3 (2)	5/6 (1)	5 min	5/6 (1)	recharge battery
	EEG (BS)	W	3 (2)	5/6 (1)	20 min	5/6 (1)	recharge battery
	GSR / ST (BS)	W	3 (2)	5/6 (1)	20 min	5/6 (1)	recharge battery
	Facial expression (UCAM)	W	7/8	8	negligible	undefined (3)	none
	Microphone (AUD, POLIMI)	W	7/8	8	negligible	8 (1)	recharge battery
	EOG (BS)	W	5 (2)	8	20 min	5/6 (1)	recharge battery
	Eye movement Pupil diameter (RWTH)	W	1 (4)	8	20 min	undefined	none
	Body posture (ITCL)	W	8	8	hours	undefined	none
	Gesture recog. (EXUS)	W	3	8	10 min	undefined	none
	Noise (AUD)	W	(*)	8	minutes	undefined	none
	Thermo-hygrometric (ITCL)	W	(*)	8	minutes	1 day	recharge battery
	Light (ITCL)	W	(*)	8	minutes	1 day	recharge battery
	User location (TPZ)	W	7/8	8	negligible	8 (1)	recharge battery
	Smartband (ITCL)	H	7/8	16	negligible	3 weeks	recharge battery
	Question-naires (INTRAS)	H/W	N/A	N/A	N/A	N/A	N/A

(*) Environmental sensors; we'll deploy as many as needed to cover the environment.

(1) Due to battery limitation.

(2) Sensor availability (due to cost).

(3) We assume that the facial expressions are captured by webcams, and the webcams are connected to the power all the time.

(4) The Eyetracker sensor, which provides these measurements, needs a laptop for the measurement. Moreover, we have to observe the measurement because it is very time-consuming and error prone. We assume the laptop is connected to the power all the time.

(5) The platform consists of: a) the facial recognition module for verification and to allow the simultaneous recording of multiple verified users' gestures and b) the gesture-based interaction module.

Note that the measuring time can vary a lot among the sensors; for example, the EEG sensors can work for 3/4 hours, while a microphone can collect data for the whole shift. Even the number of workers we can measure at the same time depends on the particular sensor: we can't deploy a lot of EEG sensors due to cost, while microphones are not expensive.

As a result, the Agent will have to cope with incomplete information (even if we actually assume that most of the time almost all sensors will be available). As a side effect of this requirement, the Agent will have to be as robust as possible with respect to data loss (due, for example, to sensor malfunctions).

An initial estimation of the schedule is:

- Following **30** workers, for a given use case
- Monitoring **7 or 8** workers at the same time
- For one week (actually, **5** working days)
- For the entire shift (we assume **8** hours/day)
- **4** weeks per month (20 days, not considering Saturday and Sunday)
- No "technical inter-week iteration"
- **12** "data collection periods"
- The "technical update iteration" is **2/3 day long** (from the last measurement week to the end of the month)

Thus, we have 4 groups, composed of 7 or 8 workers (so that the sum is 30); every month all 30 workers will be monitored for one week.

Each worker is monitored for a total of 60 working days (480 hours) along the long-term Field Test. The amount of collected data – in hours – is 30×480=14400.

Again, some sensors will be able to measure workers for 480 hours, while others will be available for a subset of that time period (see Table 4). In particular, the Eyetracker sensor, which provides eye movement and pupil diameter, needs WA personnel to attend the measurements; thus, we'll give that sensor a "special status", treating it as an external component without automatic integration with the rest of the system (something similar to the home measurements; see Section 5.3). RWTH, the partner managing that sensor, will provide a special schedule for conducting experiments with the Eyetracker.

Considering all the use cases, we'll be following 90 users, monitoring maximum 21/24 users at the same time.

For the Manufacturing use case at Grupo Antolín RyA, the measuring time could be 16h a day because of the **two shifts**. Thus, for that use case, the 7/8 workers to be measured each day could be divided into two groups.

7.5 Protocol

The WA consortium will agree on a common protocol for testing the WA system; as "protocol" we mean a document specifying the test procedures to apply, the expected results, and how to address issues.

The final protocol will be described in D9.1 Pilot Operational Manual.

Small issues will be addressed by the WA field managers and the company manager, whereas more complex problems (in particular, issues affecting machine/software installed at WA partners) will be addressed by WA home managers.