Project Deliverable 6.1

Evaluation Framework

Worker-Centric Workplaces in Smart Factories

www.facts4workers.eu

FACTS4WORKERS

Factories of the Future
Public Private Partnership

Series: Public Deliverable

Published by: FACTS4WORKERS: Worker-Centric Workplaces in Smart Factories.

FoF 2014/636778

Date: 31.7.2016

Volume 1.0: Evaluation Framework

Reference / Citation

Lacueva, F.J.; Mayo, S; Heinrich, P.; Moertl, P.; Hannola, L.. (2016): "Evaluation Framework". Project Report - FACTS4WORKERS: Worker-Centric Workplaces in Smart Factories.

Cover Design: Florian Ott, Cooperation Systems Center Munich

Worker-Centric Workplaces in Smart Factories

E-Mail: facts4workers@v2c2.at

Internet: www.facts4workers.eu

This document is published under a Creative Commons Attribution Non Commercial No Derives license. You are free to copy and redistribute the material in any medium or format. You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes. If you remix, transform, or build upon the material, you may not distribute the modified material.

http://creativecommons.org/licenses/by-nc-nd/4.0/

About this document

Executive Summary

This document represents Deliverable 6.1 ("Evaluation Framework") of the H2020 project "FACTS4WORKERS -Worker-Centric Workplaces in Smart Factories" (FoF 2014/636778).

The Evaluation Framework, as main tool used for reaching WP6 goals (to evaluate the impact of the project solutions on the workers), contributes to all other project's WPs, generating data for iterating initial requirements and for evolving the designed solutions. That's why we firstly point the relationship between the work to be performed in WP6 and the rest of WPs.

The evaluation of how the introduction of solutions (including ICT) in the workplace affects the daily work and impacts on the worker implies a very broad research scope. Very different and complementary research lines are involved in that purpose, and we establish the rationale of the framework in a wide range of methods and tools among which we will choose those most appropriate for the purpose of the framework.

The evaluation framework is defined then. Taking into account the available rationale and background, but with the project idiosyncrasy in mind, we establish our primary evaluation targets and process. FACTS4WORKERS tries to change the worker's practices, using the help of ICT tools (but not only leveraging on them). This is going beyond of just to evaluate the deployed solutions. That's why the evaluation framework is defined in terms of the validation and impact assessment of the introduced new practices (with the difficulty to separate this impact from other factors), which is going a step further of just using a subset of methods and tools detailed in the rationale.

Document authors and reviewers

The following individuals have made a direct contribution to the document. Please note that many others have also supported our work, and we thank them all sincerely.

Lead Authors

Name	Organisation	Role
Francisco J. Lacueva	Instituto Tecnológico de Aragón	WP 6
Sergio Mayo	Instituto Tecnológico de Aragón	WP 6 Lead
Dr. Peter Heinrich	University of Zurich	WP 6
Dr. Peter Moertl	Virtual Vehicle Research Centre	WP 6
Dr. Lea Hannola	Lappeenranta University of Technology	WP 6

Featuring Authors

Name	Organisation	Role
Miguel A. Gracia	Instituto Tecnológico de Aragón	WP 6
Marlene Schafler	Virtual Vehicle Research Centre	WP 6
Dr. Kirsi Kokkonen	Lappeenranta University of Technology	WP 6
Dr. Antero Kutvonen	Lappeenranta University of Technology	WP 6
Dr. Ville Ojanen	Lappeenranta University of Technology	WP 6

Reviewers

Name	Organisation	Role
Martin Wifling	Virtual Vehicle Research Centre	WP 8 Lead, Project Coordinator
Dr. Gianni Campatelli	Università degli Studi di Firenze	WP 3 Lead
Dr. Detlef Gerhard	TU Vienna	WP 5 Lead, Scientific Coordinator

Table of Contents

Table of Figures

Index of Abbreviations

AET .... Affective Event Theory

AS .... Activity Scenario

BB........................ Building Block

DES .... Differential Emotions Scale

DQ .... Data Quality

DS .... Design Science

ESM..................... Experience Sample Method

ESS .... Enterprise Social Software

F .... Fidelity

F4W..................... FACTS4WORKERS

HCI..................... Human Computer Interface

HED/UT .... Hedonic Utility Questionnaire

HMI..................... Human Machine Interface

ICT....................... Information & Communication Technologies

IP.......................... Industrial Partner

IS.......................... Information Systems

ISSA .... IS success Antecedents

ISS........................ IS Success

ISDT .... Information System Design Theory

IT.......................... Information Technologies

JDI........................ Job Descriptive Index

JDS....................... Job Diagnostic Survey

JS.......................... Job Satisfaction

JSS........................ Job Satisfaction Survey

K M .... Knowledge Management

KMS..................... Knowledge Management System

MPS..................... Motivating Potential Score

MESA.................. Manufacturing Enterprise Solution Association

PS......................... Problem Scenario

QNR .... Questionnaire

QoE...................... Quality of Experience

QoS...................... Quality of Service

R .... Reliability

RD........................ Routine Design

SotA .... State of the Art

SUS .... System Usability Scale

TAM .... Technology Acceptance Model

TRA .... Theory of Reasoned Action

UC .... Use Case

UI.......................... User Interface

UMUX .... Usability Metric for User Experience

UTAUT .... Unified Theory of Acceptance and Use of Technology

UX......................... User Experience

WE .... Workflow Engine

WIMP.................. Windows, Icons, Menus, Pointer

WO....................... Work Orders

WP .... Work Package

1 Introduction

FACTS4WORKERS project main goal is to develop, pilot and evaluate solutions that support the inclusion of elements of knowledge work on the shop floor. From the time that the 'evaluation' process is a main project objective (as the proper development of those solutions are), with a dedicated WP (WP6) planned jointly within the project scope, this process has to be very linked to every stage and task of the project, and this is the first point we need to remark.

Since the ICT solutions are evolving at every stage of the project, and different approaches, maturity levels and pilots are being reached in each IP, the evaluation framework must take into account this inherent characteristic of the project (also common in any agile and/or perpetual beta environment). Thus, the evaluation framework must have a solid anchor on the worker-centric solutions definition, for being able to evaluate the whole project solution-creation process. This is why we're leveraging many WP6 fundamentals and steps on the work performed on WP1 (Worker Needs, organisational requirements and Industrial Challenges), where an incipient evaluation framework is already defined.

In addition to the aforementioned, at FACTS4WORKERS project we have the chance to access a broad different data sources obtained from the developed ICT solutions. This means that we can enhance our evaluations tools and methods with these data (and their corresponding tools and methods of acquisition), in order to build a better evaluation framework.

The framework, as stated in the project proposal, has 2 main purposes, which are the ones we're focusing on: 1) To define metrics and key performance indicators as well as methods to measure the impact of the smart factory in the workers QoE (Quality of Experience) and 2) To plan and describe the proper methodologies for the iterative evaluation of the aspects defined with end users and experts.

This deliverable (D6.1) develops the evaluation framework definition, whose main goal is to be a usable and valuable tool for evaluating the FACTS4WORKERS solutions deployed at the factories, via the worker satisfaction and innovation skills assessment. Since, as mentioned, solutions will have different degrees of maturity and deployment throughout the project, and the evaluation process should be used at all stages, the framework provides an evaluation process and a set of tools and methods varied and considered as suitable for each phase: "One-size-fits-all" (in terms of set of evaluation tools and methods used for each IP and project phase) is not considered a proper approach, since assessment needs change as it does the type and maturity of the solutions (also taking into account that a sustainable evaluation framework should be performed, in order to be alive beyond the project end). Thus, the evaluation framework can also be seen as a process containing different methodologies, approaches and tools for each stage and solution of the project.

What not to expect in this document: As stated above, the perpetual beta approach and the heterogeneous set of IPs, Use Cases and solutions (ICT or not) to be deployed along the project, will lead us to establish a broad and different set of evaluation tools and methods to be used. That set is the framework. Thus, the framework is where the proper indicators and measurements are defined, and they are defined depending on the needs of each case to be evaluated. In this document there will not be detailed a set of measurements to be used in each Use Case or at each stage of the project, but the set of tools and methods that, using different possible measurements, will be used in each case.

Thus, deployment issues and how the framework will be used in each case (Use Case, IP, stage of the project...) will be treated in D6.2, since each case will take different items from the defined framework and will require different tools, methods and deployment needs for performing the evaluation properly.

This document is structured as follows: After this introduction and some definitions, chapter 2 explains the links and relationships between the evaluation framework and the rest of WPs. Chapter 3 explains the rationale of the framework, in terms of what kind of tools and methods we can use for evaluation purposes. Chapter 4 details the evaluation framework itself.

1.1 Guide for the reader

This deliverable (D6.1) is a large document, with much valuable information for, firstly, analysing different assessment options and then explaining the framework.

The framework rationale (Chapter 3) is a wide and detailed description of different tools and methods used for evaluation purposes, in the field of users of ICT solutions. The different methods, tools and measurements in this chapter are presented as both options/candidates and foundation for the framework description (chapter 4).

Thus, being a key baseline for the framework, the content of chapter 3 is not needed at all for a complete understanding of a) the framework goals and b) the framework elements, both described in chapter 4. This is why we propose two different paths for the proper reading of this document, depending on each reader's interests:

- **For a deep and comprehensive reading**: Read the document following the proposed structure, from the beginning to the end.
- **For a quick and fast approach to the framework**: Start by reading **Chapter** 2 to have a clear view of what is FACTS4WORKERS from the point of view of its Work Packages. This way, you will get to know what kind of tasks we are doing in the project and why/how the Evaluation Framework (developed in WP6) is connected to all of them. Follow with **Chapter** 4**,** the core of this document, in order to know the Evaluation Framework defined for the assessment.

  Once you are aware of the Evaluation Framework, and only as further reading, optionally finish with **Chapter** 3, in case you want to understand the different evaluation models used so far and as baseline and rationale for defining the framework.

2 The Framework within the FACTS4WORKERS Project

In order to clarify where WP6 (and hence, this deliverable) is placed within the project to readers without prior contact with FACTS4WORKERS, this is, in a very brief summary (not accurate enough, but right from the didactic point of view) the role of each WP in the project:

- WP1 – Workers needs and work practices information acquisition. Requirements definition
- WP2-WP4 – Technical solutions development
- WP5 – Deployment at IPs
- **WP6** – Evaluation of the solutions and their impact
- WP7 – Dissemination and exploitation
- WP8 – Coordination

Given that WP6 evaluates the work done (from the point of view of its impact on the workers) along the project, it has a very close connection with WP1-W5. The following picture describes the interplay among WPs as stated in the project proposal:

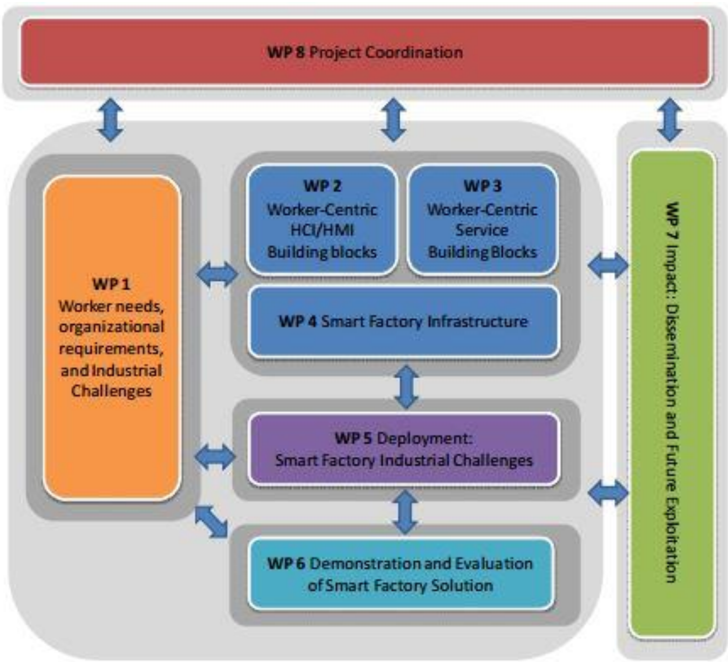

Figure 1. FACTS4WORKERS WPS schema.

Since the Evaluation Framework will feed the project with valuable data to iterate on the requirements and to evaluate the impact of the deployed solutions, a very brief outline of how the framework relates with every technical WP (and also WP 7) is detailed below.

2.1 WP 1

WP1 has a very close relationship with WP6, in general, and with the Evaluation Framework in particular. When ***WP1 explores the worker's practices and then defines requirements***, it also establishes a model to drive the evaluation process of those new practices along the project lifecycle. In fact, in D1.1 (Heinrich, 2015), an Evaluation Framework is already defined, with the description of the worker & organisational ***impact dimensions*** and the anticipated impact of the planned project interventions. Those impact dimensions are, indeed, the core indicators of the WP6 Evaluation Framework, the main aggregators of many of the data that the framework will produce. It makes total sense that the worker's practices description (both present and improved within the project) and the requirements definition are linked with the strategies for measuring the impact of the project interventions on the shop floor.

Similarly, the Evaluation Framework is seen as a process (where different stages of the project require different evaluation strategies), and use of a common research approach make WP1 and WP6 a kind of teamwork set of tasks. Indeed, there is a thin line between some planned evaluation tasks (i.e. mock-ups validations, focus groups) and the properness to place them in one or another WP.

This clear link and close connection makes WP1 the main source for WP6 evaluation framework definition and also the main feedback destination for the generated data.

2.2 WP 2

WP2 focuses on HCI/HMI building blocks and, thus, ***specifies the interfaces that workers will use to interact with the project ICT solutions***. These building blocks are both a source of data for evaluation purposes and also an item that influences in the worker satisfaction and innovation skills improvement (i.e. an item that really impacts on the worker daily life). That's why WP6 will use these building blocks and its associated devices for evaluation purposes.

2.3 WP 3

WP3 focuses on the service back end building blocks for the project ICT solutions. This means that WP3 defines and manages the main amount of data that the project solutions will gather. Thus, this is the main source of information that

WP6 will use when the evaluation framework requires using technological data (3.3) for evaluation purposes.

2.4 WP 4

WP4 develops the Semantic Workflow Engine that will be used to ***compose and coordinate the back end developed services***. Like stated for WP3, this can be a source of information about how the system is used and thus, when required and estimated proper, to help in the evaluation of the impact of the project interventions on the workers.

2.5 WP 5

WP5 is in charge of the deployment of the developed solutions on the different IPs shop floors. If the Evaluation Framework requires some deployment or set up (which would be detailed in D6.2) within the project's general architecture, this will have to be aligned with the guidelines established in WP5. Also, the designed architecture to develop and deploy the project ICT solutions will influence how any required 'data acquisition' functionality (for evaluation purposes) is developed in the scope of WP2-4.

2.6 WP 7

The outcomes of the evaluations will be disseminated (and ***dissemination is the main goal of WP7***). Thus it is crucial to be able to extract conclusions which can be backed up by data gained throughout the evaluation process, both evaluating the smart factory solutions as well as the methodology itself. The framework must ensure that personal data must be kept confidential. All the ethics regulations apply for internal (partners and projects) as well as external (public) dissemination.

3 Framework Rationale

3.1 Framework Elements: General Overview

The Evaluation Framework defined for FACTS4WORKERS (which is detailed in chapter 4) has as main goal to help the project to demonstrate and evaluate the benefits introduced in the factories via the solutions developed in WP2-WP5. This means that the framework is made of a very concrete set of tools and methods, taking existing ones from literature as a base, tailoring them and defining new approaches when considered. In the same way, the framework is developing a proper evaluation process for meeting our project needs and goals.

To better understand why we are building a framework like the one we are proposing later in this document, and why we're using and defining a certain subset of elements to build it, we consider a key issue to raise awareness on the 'bricks' that we have in order to create the framework.

In a very brief and general summary, we count on two different approaches (set of tools and methods but also different academic research backgrounds) to build the evaluation framework:

a) The one we call the "**classical approach**", which is the academia SotA of tools and methods for evaluating purposes in the field of introducing ICT tools (IS, in general) in the working environment. This approach is the fundamental pillar of the framework, and it will be used to define the evaluation process and main elements, tools and methods. The IS research in this field has traditionally used what we call from now on 'social sciences evaluation tools', just for better understanding (and to properly differentiate them from the tech. approach), which are composed basically of <u>qualitative</u> research data collection, analysis and field research design (i.e. observation, focus groups, expert evaluations, interviews, surveys, etc.). Use of other sources of data (logs, statistics of use, etc.) is less common in this environment, and usually not aiming at our evaluation purposes (i.e. worker satisfaction) but just IS success (i.e. in terms of extent of use).

b) The second approach takes advantage of the fact that FACTS4WORKERS is designing, developing and deploying ICT tools that will contribute to empower the workers on the shop floor. The use of these tools usually generates large amounts of data (logs, content), which can be used to analyse how the worker is interacting with them and, thus, to be able to extract valuable conclusions. This is what we call from now on the "**technological approach**", which does not necessarily rely on traditional 'way-of-doing' (in terms of using social sciences tools for evaluation but, like the classical approach, with a solid academic background), but on the data analysis chances allowed by ICT tools. We consider that this is a perfect complement to the classical approach, introducing different academic research knowledge (used mainly in different scenarios so far) than the classical one and allowing us to have different data sources to provide feedback to both the project and the IPs. Also, the type of data generated via this approach is quantitative, which complements the classical approach. In this sense, it's appropriate to highlight that this approach scope will depend mostly on the solutions developed and deployed (WP2-WP5).

Of course, the aforementioned segmentation is just a high level didactical approach (not exhaustive) with the only objective of let the reader understand what (low-level items) we are going to use to build the framework.

In chapter 4, where the framework is defined, the fix of both approaches within the framework definition will be clearer: We will use different **tools and methods**, adapted from the aforementioned approaches, in order **to feed the need of evaluation information that the framework requires** in each stage and case of the project. In the following picture (Figure 2) we outline the aforementioned fix:

Tools and Methods approaches: Feeding the Impact Dimensions

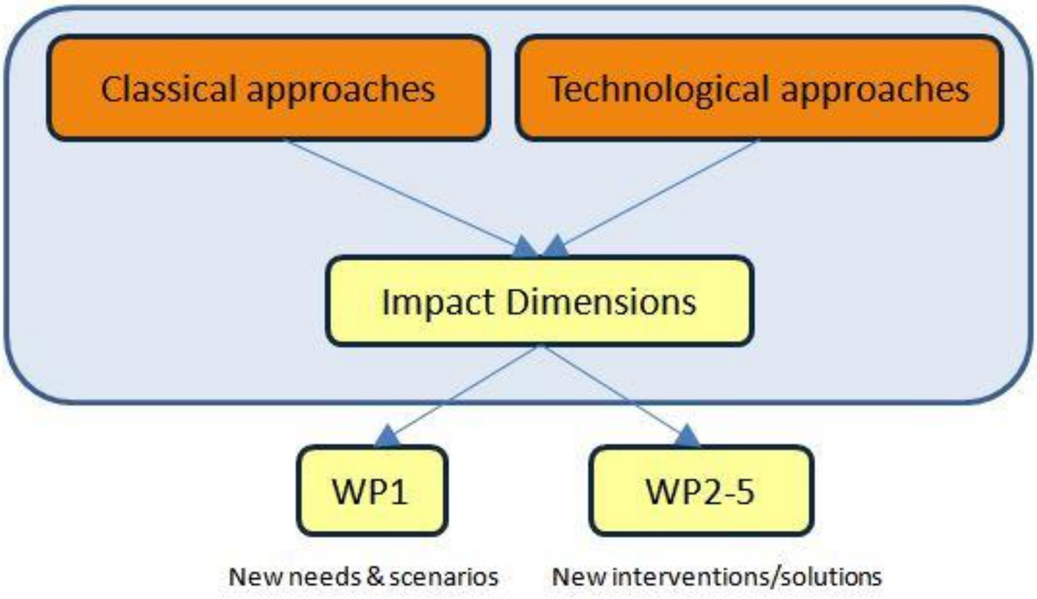

Figure 2. Framework approaches for tools and methods.

Classical approaches are expected to perform better providing tools and methods for early stages of the project (solutions development and deployment early stages), acquiring valuable qualitative information by interacting directly from workers, acquiring quantitative data for the solutions validation and providing still pictures (and feedback) of the solutions validation and its impact on the workers at any given stage of the project.

On the other side, technological approaches are expected to perform better providing tools and methods for longitudinal quantitative analysis (even beyond the end of the project), being less intrusive for workers and acquiring large quantitative data for further and deeper analysis.

Both are based on different research backgrounds, and we will see that both will converge on feeding D1.1 impact dimensions in order to assess the worker satisfaction and innovation skills.

3.2 Classical Approaches

The following points will explain the main sources considered by the participants in WP6 to define and develop the 'classical approach' of the framework.

3.2.1 Worker Satisfaction and Innovation Skills Evaluation

Since two of the objectives of the project (O.1 and O.2) focus on innovation skills and cognitive job satisfaction (JS), it is clear that we need to combine the efforts already made by researchers with the project particularities, in order to have a solid basis for the evaluation framework. In the course of the project we evaluated and reviewed the existing broad field of research in Academia regarding psychological theories and metrics of job satisfaction in order to investigate implications for socio-technical interventions.

The assessment of the theories shows a basic distinction between the following approaches (Weiss & Cropanzano, 1996):

- **Cognitive theories** base JS on cognitive judgments that workers make about their work experience and work conditions.
- **Dispositional theories** emphasize certain predispositions of worker toward expressing JS.
- **Motivational theories** focus on the factors in the work environment and the work itself that influence workers motivation and increase JS. We review two such theories: the "two factors theory" and the "job characteristics model".

Cognitive theories describe JS essentially as an outcome of cognitive assessments. An example of such theory is (Fishbein & Ajzen, 1975) who describe the conditions under which attitudes lead to intentions which in turn lead to behaviour. One example for such an extended theory is the *Affective Events Theory (AET)* which views JS as a result of evaluative judgments with affective as well as cognitive components (Weiss & Cropanzano 1996). Affective components consist of feelings that the work environment engenders whereas cognitive components consist of the workers' believes about the work environment. AET especially explains the intra-individual variability of JS measurements and implies that a single measurement of JS does not capture the whole range of possible JS states for a set of workers. Therefore, measurements have to be repeated over time to achieve accuracy.

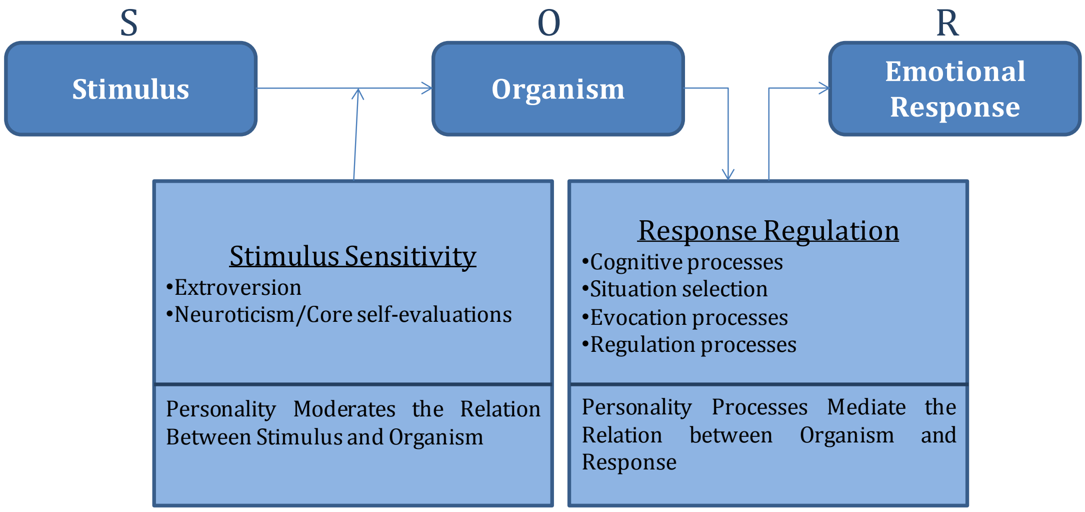

Figure 3. Judge & Larsen (2001): SOR Model of Personality Moderating Affective Responses.

A principle assumption of **dispositional models for JS** is that certain dispositional properties of workers influence JS (Judge & Larsen, 2001; Dugguh & Ayaga, 2014) such as the workers' self-evaluations ("core-evaluations": Judge et al., 1998; Srivastava, et al., 2010) and negative and positive affectivity. "Core self-evaluations" are fundamental and subconscious conclusions that people have formed about themselves, about other people, and about the world around them. These conclusions concern their self-esteem, their self-efficacy, neuroticism, and locus of control. Therefore, core-evaluations moderate the impact of work place experiences on the worker's emotional responses. Dispositional models do not negate the influences of cognitive components on JS but focus on the workers' dispositions and personality traits that trigger them. For example, the first two of the "Big Five" personality dimensions (*extraversion*, *neuroticism*, *agreeableness, conscientiousness, and openness*) were found to be especially relevant for JS. (Judge & Larsen, 2001). The model is illustrated in Figure 3.

Dispositional theories explain the inter-individual variability of JS measurements and imply that different workers will respond to socio-technical job interventions in different ways. Therefore, to meet the needs of all workers, socio-technical interventions should be adaptable to specific worker characteristics. Initial fielding of socio-technical interventions should also be sensitive to any dominant dispositional qualities of the involved work force. For example, it would be preferable to field prototypes of socio-technical interventions among workers with generally positive affective dispositions who would be emotionally more likely to engage with them.

In addition two motivational theories of JS are here reviewed, the **job characteristics model** by (Hackman, 1976) and the **two factors model** (Herzberg, Mausner & Snyderman, 1959). *The job characteristics model* was empirically confirmed by the Job Diagnostic Survey (JDS), which will be described in the next section. The model focuses on positive motivational incentives and is especially useful to determine desired work place changes that could strengthen JS. It is less suitable for potentially dysfunctional work aspects such as highly repetitive work. The model considers the relationship between work and individuals (i.e. not teams); it does not explicitly consider interpersonal, technical, or situational variables.

The *two factors model* of (Herzberg, Mausner & Snyderman, 1959) has important implications for the potential of socio-technical interventions to increase work satisfaction. For example, it disconfirms traditionally held notions that supervisor training or pure salary alone would increase workers JS. Also, merely decreasing technical or administrative inconveniences would not lead to increased JS but would only reduce perceived dissatisfaction. Instead, the theory implies that JS would grow as work is experienced in meaningful ways, results are recognized, and personal growth is achieved. The theory was criticized by pointing out that the distinction between motivator versus hygiene factors on JS may have only inter-individual applicability such that it is more valid for some workers than for others (Hackman, 1976). To determine socio-technical interventions that increase workers JS, ICT designers would need to investigate the intervention's relation to the worker. There, research has investigated empowerment (Spreitzer, 1995) that combines factors from Hackman's and Herzberg's theories: meaning, competence, self-determination (Deci et al. 1989) and impact.

Below, some of the main instruments used for JS purposes supporting the aforementioned argumentations are briefly explained.

3.2.2 Job Satisfaction Measurement Instruments

3.2.2.1 Job Satisfaction Survey

The JS survey (JSS) was developed by Paul Spector from the University of South Florida (Spector, 1985) to measure JS specifically to human services, public and non-profit sector organizations. The survey is based on an understanding of JS as evaluative feelings about the job which are measured by the survey. Questions are to be answered on five point Likert-style ratings scales, ranging from strongly agree to strongly disagree. The survey contains nine subscales that were extracted from a review of literature at the time: pay, promotion, supervision, benefits, contingent rewards, operating procedures, co-workers, nature of work, and communication. Reliability is reported as r = .91 for internal consistency, and between .34 to .74 for a long interval test-retest) based on a sample of 2,870 participants. Also various types of validity were assessed and are reported in (Spector, 1985). Access to the JSS can be gained from Spector`s website[1].

3.2.2.2 Job Descriptive Index

The job descriptive index (JDI) was developed by researchers at Cornell University in the late 1960's (Smith, Kendall & Hulin, 1969) and since then has been validated with large groups of participants. It assesses attitudinal aspects of JS without imposing specific structural or process models. Workers are assumed to relate their work environment to their internal frame of reference, representing an internal standard and adjusting their responses to their experiences, thereby reflecting their specific adaptation level (Smith, Kendall & Hulin, 1969).

There are 90 questions on the JDI that can be answered on a three point scale (yes, no, and undecided) and are grouped in following five factors (respective number of questions): Work (18); Pay (9); Promotions (9); Supervision (18); Co-workers (18); Job in General (18).

JDI validations (Brodke et al., 2009) indicate good internal consistency of items within each factor (all r ≥ .88, average r = .91) and good differentiation between factors (all r ≤ .67, average r = .42). External validity was assessed via correlations with intent to quit (average r = -0,42), feelings of job stress (r = .21) and a single measure of JS (r = .53). Various forms of the JDI and descriptive information can be accessed from the website at the Bowling Green State University[2] free of charge.

3.2.2.3 Job Diagnostic Survey

The Job Diagnostic Survey (JDS) is based on the Job Characteristics Model as described above and consists of 83 items in 7 subsections plus a short biographic questionnaire (Hackman & Oldham, 1974). Response scales are seven point scales except one scale that uses a five-point Likert scale (agree – disagree). The JDS is intended to diagnose the motivational properties of jobs prior to interventions or redesign as well as to assess the effects afterwards (Hackman & Oldham, 1975). It was validated with over 1.500 individuals in more than 100 jobs in about 15 different organizations. (Hackman & Oldham, 1976) report satisfactory reliability, that ranges for internal reliability (i.e. item consistency within a scale) between .56

[1] http://shell.cas.usf.edu/~pspector/

[2] http://www.bgsu.edu/arts-and-sciences/psychology/services/job-descriptive-index.html

and .88 and for discriminative reliability (i.e. differentiation between scales) between .12 and .28. (Hackman et al., 1975) report JDS validity; workers with a higher Motivating Potential Score (MPS) report lower absenteeism than those with low MPS (about 3 versus 7 days per year). Also workers with higher MPS show slightly higher job performance when rated by their supervisors than those with lower MPS.

Job Dimensions	Skill Variety Autonomy Dealing with others Task Identity	Feedback from Job Task Significance Feedback from Agents
Psychological States	Experienced Meaningfulness of work	Knowledge of Results
Affective Responses to the Job	General Satisfaction Internal Work Motivation	Specific Satisfaction: Job Security, Pay, Social, Supervisory, Growth
Growth Need Strength	Measured as “would like”	Measured as “job choice”

Table 1. JDS Scales

3.2.3 Technology Success and Acceptance

Most of the interventions of FACTS4WORKERS solutions are related to information systems (IS). IS serve as one of the knowledge bases in the question of how to measure and validate the technology acceptance of FACTS4WORKERS solutions. IS researchers have derived models to explain and measure success, taking various perspectives and system types into account. One of the most prominent approaches is the **Technology Acceptance Model (TAM)** (Fishbein and Ajzen, 1975), which explains why some information systems are more accepted by users than others. One of its most major adaption is the **Unified Theory of Acceptance and Use of Technology (UTAUT)** (Venkatesh et al., 2003). The UTAUT suggests four constructs (performance expectancy, effort expectancy, social influence, and facilitating conditions) as direct determinants of usage intention and behaviour. Another dominant model is **the IS Success Model** (DeLone and McLean, 2003), providing a taxonomy of IS success consisting of six variables: system quality, information quality, use, user satisfaction, individual impact, organisational impact, and service quality (FACTS4WORKERS, Description of action, 2014).

In addition, one of the challenges when developing new innovations, in this case new FACTS4WORKERS systems into production environments, is to get the new system adopted in the workplace. (Rogers, 1983) has introduced the well-known **theory of innovation diffusion** to explain this challenge. Coming up next, the above mentioned models are shortly described.

3.2.3.1 Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) is originated from the Theory of Reasoned Action, TRA (Fishbein and Ajzen, 1975), which has widely been used for prediction of behavioural intentions. According to the model, behavioural intentions are a function of beliefs about the likelihood that a particular behaviour leads to a specific outcome. These beliefs are divided to behavioural and normative, i.e. an individual's attitude towards performing the behaviour and subjective norm about performing the behaviour (Fishbein and Ajzen, 1975; Madden et al., 1992). TAM was introduced by Fred Davis in 1985, and it uses two technology acceptance measures in order to explain an individual's attitude. The model proposes *perceived usefulness (PU)* and *perceived ease of use (PEOU)* as the fundamental determinants of technology adoption, and examines their mediating role between systems characteristics and the probability of system use. Perceived usefulness means the degree to which a person believes that using a particular technology will enhance the job performance. Perceived ease of use refers to the degree of effort the utilization of a particular technology requires – the lesser effort is needed, the higher the perceived ease of use (Davis et al., 1989). TAM has been used and also extended in several studies. First and foremost, it is confessed as the traditional adoption theory in the field of IT (Legris et al., 2003; Awa et al., 2014) but it can be utilized in the investigation of technology acceptance in a broader scope as well. The traditional TAM model is introduced in Figure 4.

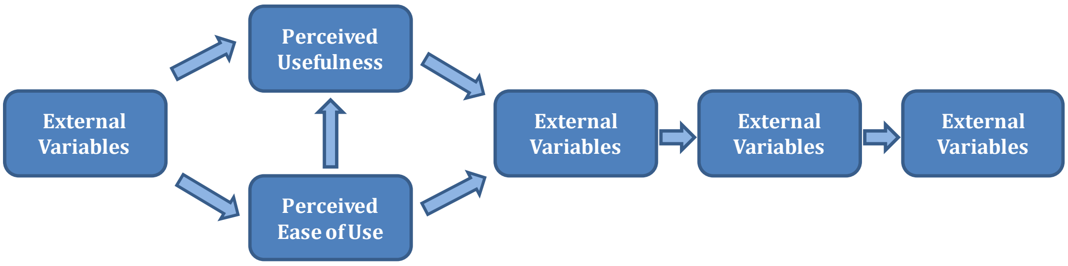

Figure 4. The original Technology Acceptance Model (Davis et al., 1989).

3.2.3.2 The Unified Theory of Acceptance and Use of Technology (UTAUT)

Primarily, the concept underlying the acceptance theory is based on the assumption that individual reactions to using IT have an influence on the intentions to use IT as well as the actual use of IT, which are different from each other. The actual use is influenced by the intentions to use IT. In the meantime, all experiences an individual has when using the system, later on also evoke positive or negative reactions. That is the common reason for doing semi-standardized surveys of user acceptance after an initial testing phase of the new system, and to predict the probability of acceptance and use. The basic concept underlying the theory of the acceptance of information technology is depicted in Figure 5.

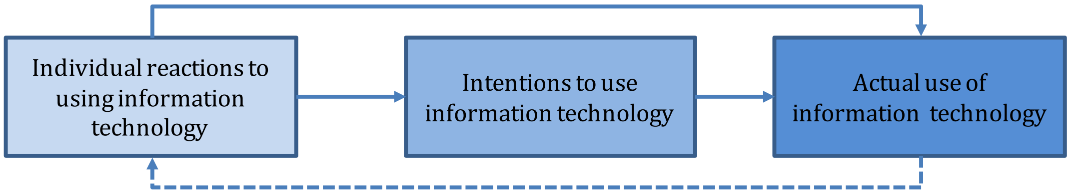

Figure 5. Concept underlying acceptance theory (Venkatesh et al., 2003).

The Unified Theory of Acceptance and Use of Technology (UTAUT) was formulated by (Venkatesh et al., 2003), which integrates the theory and research on individual acceptance of IT into a unified theoretical model that captures the essential elements of previously established theories and models like TAM. The four concepts play a significant role as direct determinants of user acceptance and usage behaviour, which are *performance expectancy, effort expectancy, social influence*, and *facilitating conditions*. The variables of the UTAUT are commonly used to explain user acceptance in the field of information systems. The UTAUT model is depicted in Figure 6.

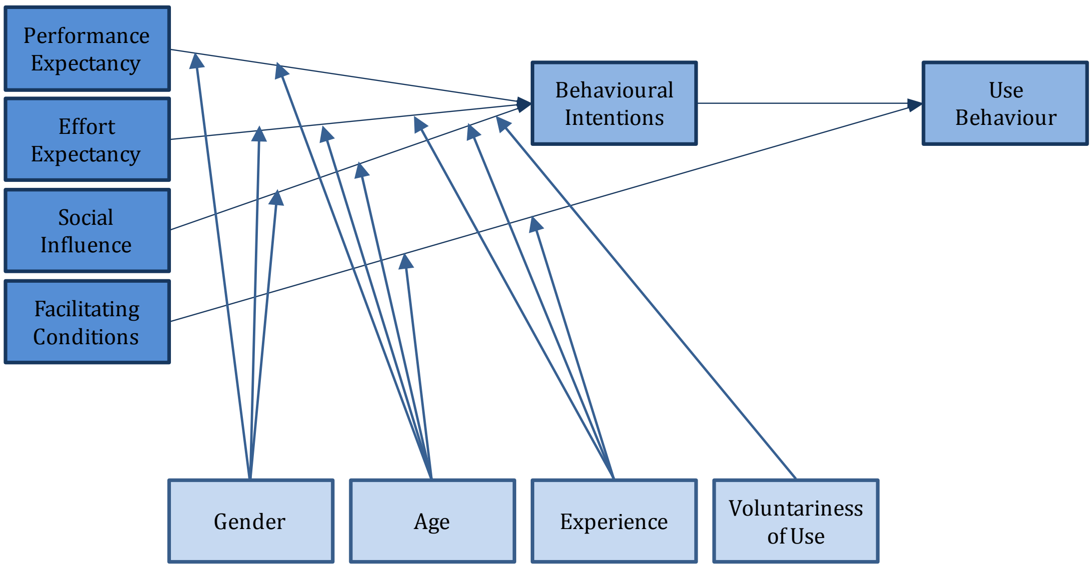

Figure 6. The Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003).

According to the Technology Acceptance Model (TAM) (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh, 2003) the direct determinants of technology acceptance are:

(1) **Performance expectancy** (perceived usefulness): the degree to which an individual believes that using the system in real life will help to attain gains in job performance. It is the strongest predictor of acceptance
(2) **Effort expectancy** (perceived ease of use): the expected degree of ease associated with the use of the system
(3) **Social influence**: the degree to which an individual perceives that important others believe he or she should use the system
(4) **Facilitating conditions:** the degree of support in terms of organizational or technical infrastructure perceived by an individual

It is posited that the impact of these four constructs is mediated by gender, age, experience and voluntariness of use.

3.2.3.3 The IS success model

DeLone and McLean (D&M) have introduced the original Information Systems (IS) Success Model in 1992. However, the role of IS has changed and progressed after that and they have published an updated version of the model in 2003. In Figure 7, the updated D&M IS Success Model is presented. In the model, there are three major dimensions of quality, i.e. information quality, systems quality and service quality. Each of the dimensions should be measured separately as they have an influence on the use and user satisfaction (DeLone and McLean, 2003). For example, information quality can be assessed based on how well it's organized, how effectively it's presented, how clearly it's written and is the information useful and up-to-date. System quality refers e.g. to the system's easiness to use, user friendliness, stability, security and speed. High quality service should be e.g. prompt, responsive, fair, knowledgeable and available (Holsapple and Lee-Post, 2006).

As a result, e.g. a high-quality system will be assumed to have more use, more user satisfaction, and positive net benefits. In other case, more use of a low quality system will be assumed to have more dissatisfaction and negative net benefits (DeLone and McLean, 2003).

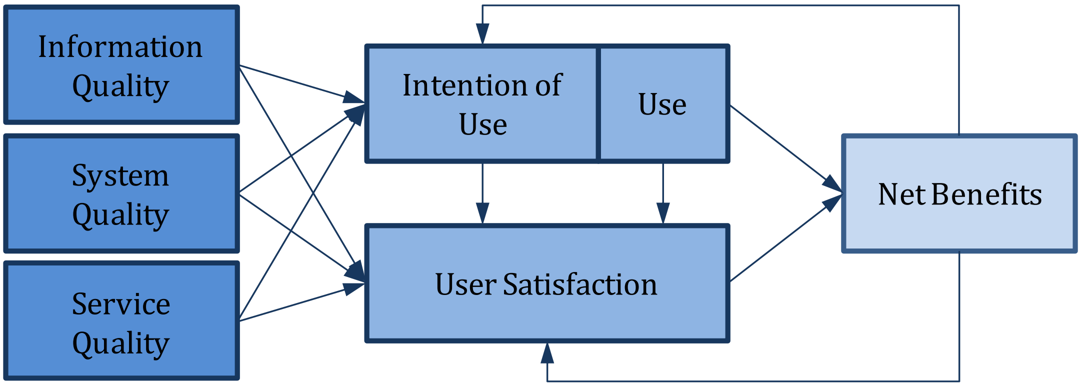

Figure 7. Updated D&M IS Success Model (DeLone and McLean, 2003).

3.2.3.4 The diffusion of innovations

One of the challenges when developing innovations, in this case new FACTS4WORKERS solutions into production environments, is to get the new solution adopted in the workplace. (Rogers, 1983) has introduced the well-known theory of innovation diffusion to explain this challenge. Diffusion means the process when an innovation is communicated via particular channels among the members of a social community over time (Rogers, 1983). The social system in FACTS4WORKERS case would mainly mean the factory environment. Although (Rogers, 1983) mainly discusses on technological innovations and their adoptions, innovation diffusion as such is multi-disciplinary and studied and utilized in many contexts, e.g. in sociology and marketing, as well as in IS research.

The rate of adoption is the relative speed with which an innovation is adopted by members of a social system. The main variables that affect the rate of adoption are the five perceived attributes of innovations (statistically, from 49 to 87 % of the variance in the rate of adoption is explained by these attributes): (1) relative advantage, (2) compatibility, (3) complexibility, (4) trialability, and (5) observability (Rogers, 1983).

Relative advantage can be considered as degree to which an innovation or technology is perceived as better than the product is supersedes. The relative advantage of an innovation, as perceived by members of a social system, is positively related to its rate of adoption. It may be measured in economic terms, e.g. cost or financial payback, however, non-economic factors like convenience, satisfaction and social prestige may be equally important. The nature of the innovation and characteristics of potential adopters determine what specific type of relative advantage is important to adopters, i.e. which are the primary and secondary attributes of innovation.

There are also three other determinants of innovation diffusion, which, as perceived by members of a social system, are also positively related to its rate of adoption.

Compatibility is the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters. Trialability is the degree to which an innovation may be experimented with on a limited basis. Observability is the degree to which the results of an innovation are visible to others. It is closely related to the connections of the social system: Some ideas are easily observed and communicated to other people, whereas other innovations are more difficult to observe, to try or to describe to others.

The fifth determinant, complexity is the degree to which an innovation is perceived as relatively difficult to understand and use. As perceived by members of a social system, it is negatively related to the rate of adoption. It may not be as important as relative advantage or compatibility for many innovations, but for some new ideas complexity is a very important barrier to adoption (Tidd and Bessant, 2013; Rogers, 1983).

Although the theory of innovation diffusion is widely adapted in several disciplines, there is also some general criticism related to the theory (Tidd and Bessant, 2013; Peres et al., 2010; Macvaugh and Schiavone, 2010):

- Seeing diffusion as linear, unidirectional communication activity, while in most cases diffusion is and interactive process of adaptation and adoption,
- Viewing diffusion as a one-to-many communication system, although point-to-point transfer is also important,
- Preoccupying diffusion research as action-centred and issue-centred communication activity, although it is also a social process with interpersonal networks,
- Using adoption as the dependent variable (the decision to use the innovation), while other studies have used attitudinal change as the dependent variable,
- Suffering from a so-called pro-innovation bias, assuming that an innovation should be adopted by all members of society as rapidly as possible.

According to (Tidd and Bessant, 2013), one of the under-researched areas in diffusion research is also the pre-diffusion phase, i.e. what happens before the well-known S-curve of diffusion. Actually, the pre-diffusion phase can be relatively long period of time. Several conditions have to be met before actual diffusion: products have to be developed, produced, distributed and the necessary infrastructural arrangements have to be in place. Similarly, as in the diffusion of products in the market and also in factory environment in adopting new technologies, it can be assumed to be essential to early recognize the different types of adopters, especially technologically advanced early adopters who are crucial change agents in early and even pre-diffusion stages affecting the latter stages of diffusion.

The limits and inadequacies of diffusion theory may be overcome by considerations of complementing it with other theories and approached, or integrating them.

Similar types of terms and concepts exist in TAM, UTAUT, IS Success and innovation diffusion theories.

3.2.4 HMI Assessment

In user-centred design, the evaluation process and goals evolve together with the product development. Mock-ups and early functional prototypes can be used for validating the interaction and the user experience (the perceived ease of use and usefulness) at very early stage of development.

Although this is something that models presented on previous paragraphs already consider, HMI evaluation should be considered, somehow, independently (or, at least, shown independently) in order to highlight the impact that an application can have on users (workers) and their activities. Mock-ups and functional prototypes can be considered a communication tool. Because they are easy to improve, they can “implement“ user suggestions very quickly contribute to improve workers involvement and motivation within the project development and deployment. By separating these methods from the models, it would be possible to use them in different stages of the problem and also to use the results of the evaluation for different purposes (gathering requirements, refining design, assessment & worker acceptance, etc.).

HMI evaluation can evaluate three main issues: 1) physical interaction (which restrictions should be considered?, how the user feels?, etc.); 2) content (is the presented information useful?, is the content consistent?, etc.); 3) the user attitude and understanding (Angeletou & Graschall, 2013).

Several frameworks have been developed for the purposes of evaluate the HMI interfaces. They can be classified in two kinds: Usability frameworks and User Experience frameworks. The Usability ones deal with the user’s evaluation of the interfaces considering its Efficiency, Satisfaction, Learnability, Memorability and Errors. Expert evaluation (which does not include the user), Approaches based on tasks that can be performed on clickable mock-ups (Rettig, 1994), benchmarking with similar applications or questionnaires like SUS (Brooke at al., 1996) or UMUX (Finstad, 2010) are examples of these kind of framework some of which are introduced on next chapter.

User Experience frameworks deal with the sensory and emotional state of a user. They are basically based in questionnaires like HED/UT (Voss et al., 2003) and they are briefly introduced in chapter **Fehler! Verweisquelle konnte nicht gefunden werden.**.

As it happens with the evaluation of IS, all the proposed methods are based on external measurements, which are obtained on (more or less) regular “shots“ which are (to some extent) subjective (because they are not usually performed while executing the daily task, but on an "evaluation event"). As we already mentioned, more objective measurements would be provided by logging information while the applications is being used (Hashemi & Herbert, 2015). This real time measurements would include not only measurement to determine how the user is using the interface but also readings from sensors (accelerometer, gyroscope, physiological, etc.) defining the user (worker) digital Imprint (UDI). While these measurements can be used, for example, to determine the stress level of the cognitive load of a user (Setz et al., 2010), before considering their use it must be taken into account that the process is a time consuming process (and it can affect the system performance) and they are many legal issues to be considered.

3.2.4.1 Usability

(Nielsen, 1993) defined Usability as composed of five attributes: Efficiency, Satisfaction, Learnability, Memorability and Errors. He also identified utility as a global system attributes having great influence on its Usability.

More recently, the International Organization for Standardization (ISO 9241-11) defined usability as the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments. This definition identifies three factors that should be considered when evaluating usability: user, goal and context of use.

During product development, usability is measured to obtain a more complete understanding of users' needs and to improve the product in order to provide a better user experience (Bevan, 2008). However, it is also important to establish criteria for usability goals at an early stage of design, and to use summative measures to evaluate whether those goals are being achieved during development.

Summative measures are usually obtained from user performance and satisfaction; summative data can also be obtained from hedonic questionnaires or from expert evaluation. They can be used to establish a baseline, make comparisons among products or to assess whether usability requirements have been achieved or not. Measures need to be sufficiently valid and reliable to enable meaningful conclusions. One prerequisite is that the measures are taken from an adequate sample of typical users carrying out representative tasks in a realistic context of use. Any comparative figures should be accompanied by a statistical assessment of whether the results may have been obtained by chance.

Formative measures are used to identify usability problems, to obtain a better understanding of user needs and to refine requirements. The main data from formative evaluation is qualitative. However, some measures of the product obtained by formative evaluation, either with users or by an expert, such as the number of problems identified, may be useful, although they should be subject to statistical assessment if they are to be interpreted.

While comparing Summative and Formative measures, former are more expensive because of the need of a large quantity of samples to conclude its findings through discrete statistical distribution. Formative measures, on the other side, have a lower overhead because the results can be obtained among three to five different experienced evaluators (Cheng, 2015). Table 2 summarizes the summative/formative main methods:

Purpose	Description	When in Design Cycle	Typical Sample Size (per group)	Considerations
Early Formative Evaluations				
Exploratory	High level test of users performing tasks	Conceptual design	5-8	Simulate early concepts, for example with very low fidelity paper prototypes.
Diagnostic	Give representative users real tasks to perform	Iterative throughout the design cycle	5-8	Early designs or computer simulations. Used to identify usability problems.
Comparison	Identify strengths and weaknesses of an existing design	Early in design	5-8	Can be combined with benchmarking.
Summative Usability Testing				
Benchmarking/ Competitive	Real users and real tasks are tested with existing design	Prior to design	8-30	To provide a basis for setting usability criteria. Can be combined with comparison with other eSystems.
Final	Real users and real tasks are tested with final design	End of design cycle	8-30	To validate the design by having usability objectives as acceptance criteria and should include any training and documentation.

Table 2. Summative/formative methods.

Without considering the stage of the project when the usability measurement methods are applied (nor their purpose), there are three types of methods for usability evaluation: usability inspection, usability testing and usability inquiry (Rana, 2012).

In usability inspection methods, groups of experts (Nielsen, 1994) create the evaluation. Table 3 summarizes the types of evaluation methods designed to be performed by experts. These experts can use guidelines or they can work through task scenarios that represent what users would typically do with a system. Usability methods that do not use task scenarios are implemented via reviews or inspections, while task-based evaluations are implemented via walkthroughs. These methods have a reduce cost and they are able to discover a reduce number of errors (Hollingsed, 2007).

Guidelines	Task Scenarios	
	No	**Yes**
None	Expert Review	Usability Walkthrough
		Pluralistic Walkthrough
		Cognitive Walkthrough
General Guidelines	Heuristic Inspection	Heuristic Walkthrough
Detailed Guidelines	Guideline Inspection	Guidelines Walkthrough

Table 3. Types of Experts Evaluation Methods (Petrie, 2009).

Testing methods evaluate the product (or IS) by testing it on users while they are using the system or prototyping models. A minimum of assistance is given by those running the evaluation, except when participants get completely stuck or need information not readily available to them. Testing helps the evaluators to check how user interface helps users in their tasks. Testing methods include: Coaching Method, Performance Measurement, Question–asking Protocol, Retrospective Testing, Thinking Aloud Protocol, Co-discovery Learning, Teaching Method and Remote Testing.

Finally, usability inquiry methods involve experts to get information about the user requirements for the system, by communicating with them or observing them while users are operating the system. Inquiry methods include: Field Observation, Interviews/Focus groups, Proactive Field Study, Logging Actual Use and Surveys. Some of these methods were already used in WP1 (see FACTS4WORKERS deliverable D1.1). Table 4 shows a brief comparison between Inspection and Test:

	Inspection Methods			Test Methods		
	Heuristic Evaluation	**Cognitive Walkthrough**	**Action Analysis**	**Think Aloud**	**Field Observation**	**QNR**
Applicable Phase	All	All	All	Design	Final Testing	All
Required Time	Low	Medium	Low	High	Medium	Low
Needed Users	None	None	None	3+	20+	30+
Required Evaluators	3+	3+	1-2	1	1+	1
Required Equipment	Medium	High	High	Medium	High	Low
Intrusive	No	No	No	Yes	Yes	No

Table 4. Comparison of Inspection and Test Usability Methods (adapted from Holzinger, 2005)

A special group of methods are the SUS, UMUX, UMUX-LITE. They are usually used after the respondent has had an opportunity to use the system that is being evaluated, but before any debriefing or discussion takes place. Respondents should be asked to record their immediate response to each item, rather than thinking about items for a long time.

- SUS - the System Usability Scale is a simple, ten-item scale giving a global view of subjective assessments of usability (Brooke, 1995). SUS is a 5-point Likert scale. Selected statements cover a variety of aspects of system usability, such as the need for support, training, and complexity, and thus have a high level of face validity for measuring usability of a system. All items should be checked. If a respondent feels that they cannot respond to a particular item, they should mark the centre point of the scale.
- (Borsci, 2015) signalled that, when survey is administered to the users after a short period of product use, it is safer to consider the SUS a unidimensional scale survey, so he recommends again partitioning it into Usable and Learnable components in that context. Moreover, practitioners should anticipate that satisfaction scores of newer users will be significantly lower than the scores of more experienced people. When the SUS is administered to more experienced users, the scale appears to have bidimensional properties, making it suitable to compute both an overall SUS score and its Learnable and Usable components. The overall level of satisfaction will be higher than among less experienced users.
- UMUX - Usability Metric for User Experience (Findstad, 2010) is a ten-item questionnaire that was designed to produce scores similar to SUS. UMUX has a general question ("This system is easy to use") and three more questions from SUS associated with efficiency, effectiveness and satisfaction, which are evaluated in by a 7-point Likert scale.
- UMUX–LITE (Lewis, 2013) is a short version of UMUX which applies a two items questionnaire which proceed from UMUX and it also has a 7-point Likert scale.

The conclusions of the comparison performed by (Borcis, 2015) signalled that UMUX and UMUX-LITE show similar behaviours because of their correlation with SUS. When UMUX-LITE is applied with its adjustment formula, it provides results that are closer in magnitude to the SUS than the UMUX, making it the more desirable proxy. Even thought, the UMUX and UMUX-LITE are both reliable and valid proxies of the SUS but the authors recommended to use them in addition to the SUS rather than instead of the SUS. In particular they recommended avoid using only the UMUX for their analysis of user satisfaction because it seemed too optimistic. In the formative phase of design or in agile development, the UMUX-LITE could be adopted as a preliminary and quick tool to test users' reactions to a prototype. Then, in advanced design phases or in summative evaluation phases, we recommend using a combination of the SUS and UMUX-LITE (or UMUX) to assess user satisfaction with usability

3.2.4.2 User Experience

User Quality of Experience (QoE) is a subjective and difficult to measure concept. One important aspect of QoE, User Experience (UX), corresponds to the sensory and emotional state of a user. For a user interacting through a User Interface (UI), precise information on how the UI is used, can contribute to a better understanding the UX, and thereby understanding the QoE (Hashemi & Herbert, 2015).

UX is defined as a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service (ISO 9241-110:2010) (9241-210:2010, 2010). When defining the interaction with an IT system, UX states to take into account these three parameters: Process (what the user does), Outcomes (what the user achieves), and Affect (what the users feels). In addition, it's recognized that UX goes beyond usability in areas such as (Petrie, 2009):

- ***Holistic***: usability focuses on performance of and satisfaction with users' tasks and their achievement in defined contexts of use; UX takes a more holistic view, aiming for a balance between task-oriented aspects and other non-task oriented aspects (often called hedonic aspects) of system use and possession, such as beauty, challenge, stimulation and self-expression.
- ***Subjective***: usability has emphasised objective measures of its components; UX is more concerned with users' subjective reactions to systems, their perceptions of the systems themselves and their interaction with them.
- ***Positive***: usability has often focused on the removal of barriers or problems in systems as the methodology for improving them; UX is more concerned with the positive aspects of system use, how to maximize them, whether those positive aspects being joy, happiness, or engagement.

There are several methods for UX evaluation and measurement. Questionnaires, interviews, and surveys are used in HCI studies (Vermeeren et al., 2010). A complete list of methods classified using different criteria (availability, information source, location, product development phase, period of experience, type of data collected, applications/designs, time requirements, etc.) is presented by (Vermeeren et al., 2010). Some representative methods are: AttrakDiff, Differential Emotions Scale (DES), Experience Sample Method (ESM), Hedonic Utility Questionaire (HED/UT), Long Term Diary Study, PANAS, Premo or Timed ESM.

3.3 Technological Approaches

Chapter 3.2 introduces the rationale for being able to measure the worker satisfaction and the impact that an IS implementation, in our case a solution related to the Industry 4.0 trends, can have on the worker. They are based on the execution of surveys, interviews or the observation of the worker during the different phases of the project development. These methods require users to fill up questionnaires, attend to interview sessions, etc. Complicated, difficult, and confusing questions in an interview or a questionnaire can make it unpleasant for users. It is also not a good user's internal state indicator as determining emotions and moods are difficult. While they are well-tested methods, with a solid background, they are the only ones that can be used during the initial stages of the development of an IS project (Steinhueser et al., 2015). They are not easily applicable in practice (Richter et al., 2013) and several authors propose alternatives to measure the success based on the analysis of usage logs, data structure, etc., not replacing but complementing the aforementioned (3.2) tools and methods.

Industry 4.0 is a collective term for technologies and concepts of value chain organization (Hermann et al., 2015). Concepts such as CPSs, workflow engines, HCI/HMI, cloud, ERP, etc. are integrated under this umbrella. Despite of this multisystem integration under an Industry 4.0 solution, these solutions are usually seen (by the users) as just one system and, this way, it's not easy to determine the influence of each subsystem in a global evaluation. Thus, it becomes difficult to determine which actions should be considered in order to improve the system acceptance, usability or performance.

According to the FACTS4WORKERS proposal, and based on conclusions of issued (or about to be issued) deliverables D1.1, D1.2, D2.1 and D5.1, FACTS4WORKERS proposal for workers at shop floors could include these types of ICT solutions:

- Knowledge Management Systems (KMS): Knowledge can be divided into two types, tacit and explicit. Explicit knowledge is the knowledge that can be easily captured, codified, and shared through manuals, documents and standard operation procedures. As for tacit knowledge, it is the skill, experience and 'know-how' that is embedded in a person and cannot be easily expressed and shared (Wong, 2013). Within the shop floor environment tacit knowledge represents the knowledge which is provided to the worker by the ERP, MES and other systems providing "formal" and structured information. These systems (ERP, MES, etc.) are not KMS, but can feed them when a proper information management is designed. On the other hand, explicit knowledge can be gathered using social networks like, chat and audio/video conferencing or wiki system which are used by workers to easily sharing of information. The deployment of these Web 2.0 solutions in the shop-floor environment could be considered an innovative solution within the manufacturing sector but it has already probed an effective one in other fields such as ICT or for solving domestic problems.

- “Team Supporting Tools”: We consider under this category all the tools which are used to support communication, collaboration and, in general, connections among workers within a team scope. These tools include social networks but also more formal relations, for example the shift logbook or the documentation used during briefing meetings. It can seem that the consideration of Web 2.0 tools overlap with their consideration on the previous category. While they are included in the KMS because in some context they can are used to gather informal knowledge, here they are considering as worker relation enablers and we perform a review of the existing literature looking for measurements used for it. Even though, some of the proposed measurements will be shared both with the KMS but also with other systems categories.
- Data Management: We use this name for referring all the building blocks (BB) related with data management, from BBs gathering data from machines or existing management systems (i.e. ERP) to BB using this data to provide information and/or synthetic knowledge to the workers.
- Semantic Workflow Engine, which is the frontend between the user interfaces and the back-office of the project.
- HMI systems are the front end of FACTS4WORKERS solutions. Although the perception of the workers about them is highly dependent on the quality of the systems previously introduced, the quality of the interfaces is a big determinant of the perceived solution quality and, in consequence, of its success.

The type of data that FACTS4WORKERS solutions can provide, can be used for the framework evaluation goals and can be classified in the following groups: Information Quality; Service Quality (not expected to be used in the context of the project); System Quality.

Those groups are absolutely aligned with dimensions from classical approaches (DeLone and McLean, 2003), besides than having also a solid background for evaluating both organizational and individual impact (Gable et al., 2008).

Thus, this ‘technological approach’ is intended to be a valuable part for the framework, since it will let us to:

- Complement and benchmark the data obtained from classical approaches (the pillar of the framework) via new information inputs, but feeding the same dimensions for the evaluation.

- Launch a new and innovative approach for evaluation purposes, not just with a set of measurements (data), but within a process developed with a solid background, as the WP6 evaluation framework is.
- Establish the basis for an automated data collection methodology for evaluation purposes (empowered workers in Industry 4.0 environment).

In chapters 3.3.1 and 3.3.2, the different types of measures (in terms of information sources) that we get in this approach are presented. In addition, in 3.3.3 we show a Use Case based approach to describe chances to get more specific data from the possible ICT systems thought for each Use Case.

3.3.1 Measurements from ICT systems

In this chapter we detail some of the possible measurements (in terms of information sources and/or data) that the expected ICT solutions for FACTS4WORKERS can help us to get the information required in the Evaluation Framework.

3.3.1.1 Knowledge Management System Metrics

Knowledge Management (KM) provides procedures and technology to help knowledge flow to the right people and at the right time, so they can act more efficiently and effectively. Knowledge Management is the art of transforming information and intellectual assets into enduring information and intellectual assets into enduring value for organization's clients and workers. The purpose of Knowledge Management Systems (KMS) is to foster the reuse of intellectual capital, to enable better decision making and to create conditions for innovation.

Several metrics have been developed to measure the performance of KMS and, in particular, when the measurements are done using the support of IS. For example, a very practical perspective is presented by (Knoco, 2014; Haghi, 2004) while realized an exhaustive review from the academia perspective (Wong, 2013). According to them, knowledge metrics can be classified in three categories:

- Knowledge Resources: They are intangible assets of an organization like human capital, knowledge and information capital and intellectual property being the two first within the scope of FACTS4WORKERS project and, thus, proper for this document.

- Human Capital refers, within our scope, to employees as the holders of most of the tacit knowledge, ideas, skills and abilities that add value to the company. Knowledge and information capital refers to the quantity and quality of knowledge that a company owns. Usually, this knowledge is stored in a company's data repository system (i.e. database) in various forms such as text, images, audios and videos.
- KM Processes: Several processes have been identified for performing Knowledge management: Knowledge acquisition; Knowledge internalization; Knowledge creation; Knowledge application and utilization; Knowledge codification and storing; Transferring and sharing of knowledge.
- Factors that affect KM: These factors support and drive KM activities such as culture, management, leadership, organizational infrastructure and technology.

From FACTS4WORKERS perspective, the most relevant metrics in this scope are the ones related to the KM Processes, since our solutions are supposed to support them. Table 5, adapted from (Wong, 2013), presents relevant metrics for the purpose of our project:

Category	Metric
Acquisition and Retrieval	Repeat usage of the repository items Employees search information for tasks from various knowledge sources administered by the organization Number of site accesses Number of downloads How often users are accessing the knowledge resources Internal training and the exchanges frequency Number of meetings for idea generation attended per employee per month Working hours per employee spent for inputting knowledge into KMS per month Number of new knowledge, ideas, and solutions created per employee per month Number of documents and articles accessed or downloaded per employee per month Number of documents and articles uploaded or updated per employee per month Development time for new products How many 'times' each employee brings up a proposal Number of meetings for idea generation attended per employee per month Number of new knowledge, ideas, and solutions created per employee per month Number of new products, inventions, and services generated per year
Creation and Generation	How often users are using the knowledge resources and practices The use of new knowledge and the ability to transform Number of new products, inventions, and services generated per year Number of problems solved and ideas implemented per employee per month
Application and Utilization	Amount of codification of available knowledge assets Amount of the organizational memory (OM) codified and included in the computerized portion of the OM. How often users are contributing to the knowledge resources Working hours per employee spent for inputting knowledge into KMS per month Number of documents and articles accessed or downloaded per employee per month Number of documents and articles uploaded or updated per employee per month
Codification and Storing	Amount of codification of available knowledge assets Amount of the organizational memory (OM) codified and included in the computerized portion of the OM. How often users are contributing to the knowledge resources Working hours per employee spent for inputting knowledge into KMS per month Number of documents and articles accessed or downloaded per employee per month Number of documents and articles uploaded or updated per employee per month
Transferring and Sharing	Number of team rooms and participants in each Level of interactions, discussions and collaborations among employees on important identified subjects Communication capability Employees share information and knowledge necessary for the tasks Employees improve task efficiency by sharing information and knowledge Employees promote sharing of information and knowledge with other teams Number of hours the employees participate in workshops/seminars/networks or other activities, per month Number of knowledge shared per measurement interval Number of users participating in knowledge sharing activities Level of information communication among the staff Level of inter-departmental information communication Level of information communication with customers Number of knowledge sharing sessions attended per employee per month Number of active communities of practice, research groups, and special interest groups Number of communications per employee per month
"Transversal" Metrics	Number of knowledge workers Number of frequent KMS users Number of knowledge assets generated per year

Table 5. Metrics for measuring the performance of KM processes. Adapted from (Wong, 2013).

3.3.1.2 Team Supporting Tools

Team Supporting Tools are all those tools which are used to support the communication, collaboration and in general relations between workers within a team scope but also other relationships. These tools include social networks but also more formal relations, for example the shift log book or the documentation used during briefing meetings.

It can seem that the consideration of Web 2.0 tools overlap with their consideration on the previous category. While they are included in the KMS because in some context they can are used to gather informal knowledge, here they are considering as worker relation enablers and we perform a review of the existing literature looking for measurements used for it. Even though, some of the proposed measurements will be shared both with the KMS but also with other systems categories.

(Behrend, 2014) performed a complete review on existing studies on social networks analysis. The review aimed to identify and categorize the scope of analytical measures and corresponding data sources to develop a framework with the data dimensions which can be applied for Enterprise Social Networks. They identified four dimensions, three of which can be obtained directly from the system:

a) Activities (usage data): the functions the user executed while interacting with the system, which create usage data that can be obtained from the logging information or from exporting the data from the underlying databases. Example of measurements can be: number of new enters (posts, group subscriptions, etc.), number of views, etc.
b) Content (user-generated data): applies sentiment analysis, text mining or genre analysis methods to user-generated data for trying to determine "Who says what, to whom, why, to what extent and with what effect?" One found of the authors was that microblogging in the enterprise context differed greatly from that in the private context.
c) Relations (structural data): this dimension studies the relations created when users interact with each other in an ESN. These relations can be create automatically from the information recorded in the underlying data bases and can be used to perform Social Network Analysis.

The fourth dimension -"Experiences"-, measures the UX and the attitude of the user when using a platform which is measured by interviews and questionnaires (3.2).

3.3.1.3 Data Management.

Data Management refers to all the building blocks (BB) related with data management, from BBs gathering data from machines or existing management systems (i.e. ERP) to BB using this data to provide information and/or synthetic knowledge to the workers. These BBs include both "traditional" management systems (such as relational databases) and new technologies supporting big data management.

While it is also important to consider some of the measurements we introduced before, for these building blocks the most important measurements are related with the quality of the raw data and the information or the knowledge that these BBs provide to the workers.

As it happens with other aspects related with the evaluation of a system, data quality assessment is a multidimensional concept dealing both with subjective perceptions of the individual involved with the data, and the objective measurements based on the data sets in question (Pipinno, 2002). Subjective perceptions can be evaluated using questionnaires as the one proposed in chapter 0. Here we introduce some measures related with the objective evaluation of data quality.

There are many papers trying to identify which are the main dimensions of Data Quality (DQ). For example, in (Strong, 1997) (Pipino, 2002) or (Sidi, 2012) more than thirty concepts related with DQ such as Accuracy, Objectivity/Objectively, Believability, Reputation, Accessibility, Access security/Security, Relevancy, Value-Added, Timeliness, Completeness, Amount of data, Interpretability, Ease of understanding/Understandability, Concise, Representation, Consistent representation, etc., are identified.

(Scannapieco, 2005) and (Batini, 2009) recognized that there are many discrepancies in the definition of these dimensions because of the contextual nature of DQ. Reviewing the more significant studies of the existing literature, the author identified the basic set of data that compose the 'quality' dimension:

- Accuracy: It is defined as a measure of the proximity of a value, v, to some other value, v', that is considered correct. Two type of accuracy can be distinguished, syntactic and semantic. The first –Syntactic- is measured by means of comparison functions that evaluate the distance between v and v' (i.e. because it is not correctly written Jhon and John), while the second captures the cases in which v is a syntactically correct value, but it is different from v', that is what is the closeness of a value, v, to the elements of the corresponding definition domain, D (i.e. Jane and John are names).
- Completeness: It is the extent to which data are sufficient breadth, depth and scope for the tasks at hand. In the research area of relational databases, completeness is often related to the meaning of null values. A null value has the general meaning of missing value, a value that exists in the real world but is not available in a data collection. In order to characterize completeness, it is important to understand why the value is missing.

- Consistency: This dimension captures the violation of semantic rules defined over data items. With reference to the relational theory, integrity constraints are an instantiation of such semantic rules. Integrity constraints are properties that must be satisfied by all instances of a database schema. There are two main categories of integrity constraints, namely: intra-relation constraints and inter-relation constraints. Intra-relation integrity constraints can regard single attributes (also called domain constraints) or multiple attribute of a relation. Inter-relation integrity constraints involve attributes from different relations.
- Time-related dimensions: these dimensions consider an important aspect of data, their update over time. The main time-related dimensions are currency, volatility and timeliness. They are defined as:

  - Currency is the degree to which a datum is up-to-date. A datum value is up to- date if it is correct in spite of possible discrepancies caused by timer-lated changes to the correct value. Currency is typically measured with respect to last update metadata, i.e., the last time in which the specific data have been updated. For data types that change with a fixed frequency, last update metadata allow to compute currency straightforwardly. For data types whose change frequency can vary, one possibility is to calculate an average change frequency and perform the currency computation with respect to it, admitting error rates.
  - Volatility describes the time period for which information is valid in the real world. Volatility measures the frequency according to which data vary in time. Volatility is a dimension that inherently characterizes types of data. Therefore, there is no need of introducing specific metrics for it.
  - Timeliness is the extent to which the age of data is appropriate for the task at hand. It tries to measure the delay between a change of a real world data and the resulting modification of the information system state. Timeliness measurement implies that not only data are current, but are also in time for a specific usage. Therefore, a possible measurement consists of (i) a currency measurement and (ii) a check if data are available before the planned usage time.

Table 6. Data Quality Metrics. Adapted from (Scannapiedo, 2005), (Batini 2009), (DAMA, 2013).

Table 6 describes some metrics that can be applied to the aforementioned DQ dimensions.

Dimension	Metrics Definitions
Accuracy	Syntatic Accuracy= Number of correct values/number of total values
	Number of delivered accurate tuples
	Number of duplicated values
Completeness	Number of not null values/Total Number of values
	Number of tuples delivered/Expected Number
Currency	Time data are stored in the system – time in which data are updated in the real world
	Time last update
	Request time – last update
	Age + (Delivery Time – Input date)
Timeliness	Max (0; 1-Currency/Volatility)
	Percentage of process executions able to be performed within the required time frame
Consistency	Number of consistent values/Total number of values.
	Number of tuples violating constraints/ number of coding differences
	Number of things in real world/Number of records describing different things

Table 6. Data Quality Metrics. Adapted from (Scannapiedo, 2005), (Batini 2009), (DAMA, 2013).

Previous introduced dimensions and metrics are defined for "traditional data management systems" which, within an industrial shop floor, we can consider a sub set of the Big Data management systems.

Big Data has emerged in the last decade as a new concept. Although there is not a clear definition of the term, it is considered as structured and unstructured datasets with massive data volumes that cannot be easily captured, stored, manipulated, analysed, managed and presented by traditional hardware, software and database technologies (Li, 2016).

(Liu, 2016) describes Big Data as the "4V" model: volume, variety, velocity and veracity, which can be extended with other "Vs" such as value, variability, visibility or visualization. Based on this description the author identifies the problems of big data: inauthentic data collection, information incompleteness, unrepresentativeness, inconsistency and unreliability, as well as ethical issues. Evaluating reported problems within the shop-floor inauthentic, unrepresentativeness can be discarded an, as (Li, 2016) stated for geospatial data the known methods and theories of quality assessment are still applicable.

3.3.1.4 Semantic Workflow Engine Metrics

In the past, Workflow Engines (WE) were identified as the computing models that enable a standard method of building Web-services applications and processes to connect and exchange information over the Web (Cardoso, 2004). They contribute to create new and innovative ISs, helping companies to be more competitive, efficient, flexible, and to integrate the value chain at different levels, including the IS level. Workflow Engine functionalities manage and streamline business processes. A person explicitly determines the flow and which are the services to be consumed (approaches including “intelligence features” are under development). Based on the semantic description of the services (the automatic accounting of values about some of their attributes) Semantic Workflow Engines, like the one developed in WP4, are able to determine in real time which services to consume, and even to determine which is the flow to be executed based on the description of the desired objectives.

Because of its central role in IS scenarios where a Workflow Engine is present, determining the quality of a workflow became an issue of research. Quality of Services (QoS) is a measure of the goodness of networking systems, real-time applications and middleware, and it was proposed by (Cardoso, 2004) as a way to determine the quality of a given Workflow (Management System). The author defined the QoS of a workflow representing the quantitative and qualitative characteristics of a workflow application necessary to achieve a set of initial requirements. Workflow QoS addresses the non-functional issues of workflows rather than workflow process operations. Quantitative characteristics can be evaluated in terms of concrete measures such as workflow execution time, cost, etc. Qualitative characteristics specify the expected services offered by the system, such as security and fault-tolerance mechanisms.

Based on the aforementioned research, the WE QoS dimensions usually considered are: time, cost, reliability and fidelity. From our point of view, as we are interested in assessment of the worker (user) satisfaction with the system, cost should not be considered as a dimension to be measured. Thus, the following kind of metrics could be useful for FACST4WORKERS evaluation purposes:

- Time to execute a workflow: The time needed by an instance to transform a set of inputs into outputs. Table 7. WE time dimension metrics describes some metrics that could be considered.

- Reliability (R): It corresponds to the likelihood that the components will perform for its users on demand; it is function of the failure-rate. (Cardoso, 2004) proposes two ways to determine this value. One follows a time-discrete modelling approach and it is defined as: R(t) = 1 – failure rate. Alternative approaches follow the continuous-time reliability models, and can be used when the failures of the malfunctioning equipment or software can be expressed in terms of times between failures, or in terms of the number of failures that occurred in a given time interval. Such reliability models are more suitable when workflows include tasks for equipment controlling, or for machines that have failure specifications determined by the manufacturer.
- Fidelity (F): It is a function of effective design; it refers to an intrinsic property(ies) or characteristic(s) of a goods produced or services rendered. Fidelity reflects how well a product is being rendered. Workflow tasks have a fidelity (F) vector dimension composed of a set of fidelity attributes (F(t).ar), that reflect and quantify task operations. Each fidelity attribute refers to a property or characteristic of the product being created, transformed, or analysed. Fidelity attributes are used by the workflow system to compute how well workflows, instances, and tasks are meeting user specifications. Depending on the task type, a task uses different strategies to set fidelity attributes. Three scenarios can be drawn: automatic tasks controlling hardware (automatic evaluation), automatic tasks controlling software (automatic evaluation), and human tasks (manual evaluation).

Metric	Comments
Workflow Response Time: The total amount of time that a workflow instance spends within a workflow process before it finishes.	It can be easily measure by logging the workflow start and finish time.
Workflow Delay Time (DT): Is the total amount of time that a workflow instance spends in a workflow, while not being processed by a task (aka waiting time).	It is a measurement that requires a very accurate logging of information by the WE, because it requires to store both the external invocation time and the real execution start time.
Minimum Workflow Response Time (minT) is the time required for a workflow instance to be processed, not accounting for any task delay time.	
Workflow Response Time Efficiency (E): The ratio of the minimum workflow response time and the workflow response time.	

Table 7. WE time dimension metrics

An important issue of modern workflows is their recursive nature: their tasks can also be considered workflows. That's why it is suggested (Cardoso, 2004) that, ideally, the proposed dimensions and metrics could be obtained in a more detailed level and then (by applying different aggregation criteria) they can be used for evaluating the workflow metrics.

3.3.2 HMI Measurements

In chapter 3.2.4 we introduced "classical" methods for evaluating the HMI techniques provided by an application. Most of these methods were created for evaluating traditional WIMP (Windows, Icons, Menus, Pointer) interfaces, and they can be applied more easily on lab environments than on real scenarios. These methods can be also applied for mobile applications, characterised by:

- Mobile context: Users are not tied to a single location; they interact with nearby people, objects and environmental elements.
- Connectivity: It can be slow and unreliable, impacting in the performance of applications using these features.
- Small Screen Size: It limits the information that can be displayed.
- Different Display Resolution, which may lead to different UX.
- Limited Processing Capability and Power.
- Data entry methods.

In other words, although "classical" methods can be used for analysing mobile application usability, and they provide good qualitative data, they are often expensive and time consuming. Moreover, they do not consider mobility and, in consequence, the result could not be completely correct (Lettner, 2012). This is even more important when we consider the evaluation of solutions for industrial environments, where -in many cases-, because of safety reasons, entering some areas is restricted. Another challenge is to flexibly manage variability for testing on different devices. It is also desirable that the implementation of usability testing is not intrusive (Enriquez, 2014).

(Lettner, 2012; Holzinger, 2005; Waterson, 2002) are some examples of authors proposing the use of data logging for usability evaluation. This approach involves statistics about the detailed use of a system. Data logging can provide extensive timing data, which is generally important in HCI and usability. Normally, logging is used to collect information about the use of a system after its release, but it can also be used as a supplementary method of collecting more detailed data during user testing. Typically, an interface log will contain statistics about the frequency with which each user has used each feature in the system, and the frequency with which various events of interest (such as error messages) have occurred.

(Lettner, 2012) introduced a set of low-level metrics which are based on the Android architecture of mobile applications, but which can be also be used for other mobile platforms. These low-level metrics can be used for identifying navigational errors or inefficient navigation concepts for existing known and unknown applications. It is proposed to gather information about the device, the operative system version and the application, and then to create a tree of the elements included in each activity (window). This data structure is used to aggregate the baseline data about session times, screen calls, button clicks, etc., which are the low-level data used to create usability metrics.

3.3.3 Industrial Measurements for Use Cases

In this chapter, we are reviewing common measurements that are applied in shop floor processes similar to the ones described in FACTS4WORKERS deliverable D1.2 as Problem Scenarios (PS) + Activity Scenarios (AS). Because of the nature of the shop-floor processes, most of the outlined measurements are based both on the performance perspective of the processes and on the quality improvement derived from them.

In order to get a clear view, just for evaluation purposes, we have classified the PS and the AS as:

- Batch Production Processes: Within our scope, Batch Production Processes are the manufacturing processes where a huge quantity of products units is created within a machine or line of machines. Within the project scope, examples of these processes can be found in SCA1-PS1, SCA1-AS1, THO-PS1, THO-AS. These processes are not exactly the same as the typically called MTS (Market-To-Stock), since no stock scenarios are observed.
- Project Based Production Processes (also known as ETO –Engineer-To-Order-Processes): These processes aim to create just one or a very little quantity of units of product (machine, machine line, etc.) for a given customer order satisfying a very concrete set of requirements. The most representative use case in FACTS4WORKERS project is the one introduced by the EMO1-PS1, EMO1-AS1.
- Maintenance Processes: Processes aiming to keep machines and other relevant equipment working in order to gather the compromised production levels. Examples of uses cases related with maintenance processes can be found in EMO2-PS1, EMO2-AS2, HID-PS2, HID-PS2, SCA2-PS1, SCA1-AS2, TKSE-PS1, TKSE-AS1, TKSE-PS2, TKSE-AS2, TKSE-PS3, TKSE-AS3, TKSE-PS4, TKSE-AS4.

- Set-Up Production Processes: These are processes which are related with the setting up of the machines for producing a new product (after retooling) of after the detection of errors to solve them. Example of use cased in the FACTS4WORKERs projects are HID-PS1, HID-AS1, HIR-PS1, HIR-AS1, SCA2-PS3, SCA2-AS3, THO-PS2, THO-AS2, THO-PS4, THO-AS4;
- Quality Control Processes: Those are all the processes which are related with the quality assurance of the products being manufactured. This class includes the processes described by EMO1-PS2, EMO1-AS2, HIR-PS1, HIR-AS1, SCA1-PS1, SCA1-AS1, SCA1-PS2, SCA1-AS2, SCA1-PS4, SCA1-AS4, THO-PS1, THO-AS1, THO-PS2, THO-AS2, , THO-PS3, THO-AS3;
- “Team Processes”: Under this umbrella we consider processes which are transversal to the previous ones, such as workers training, workers collaboration, etc. They are represented in the FACTS4WORKERs use cases SCA2-PS2, SCA2-AS2, SCA2-PS4, SCA2-AS4, THO-PS3, THO-AS3, TKSE-PS1, TKSE-AS1, TKSE-PS2, TKSE-AS2, TKSE-PS3, TKSE-AS3, TKSE-PS4, TKSE-AS4.

There are several sources of measurements and metrics which can be applied to determine the performance of the listed processes. One complete set of measurements is the one provided by the MESA (Manufacturing Enterprise Solutions Association), that performed a study trying to identify the most utilized metrics by discrete, process, and hybrid/batch manufacturers (MESA, 2006). A second very complete set of measurements is provided by OpsDog (OpsDog, 2016) which creates an encyclopaedia of measurements definition and classifies them according to the different “areas of knowledge” where they can be applied.

From the referenced sources, the most relevant metrics for FACTS4WORKERS measurement purposes are selected and classified according to the kind of processes presented on the previous list, and they are presented in next paragraphs, just to evaluate whether (and how) to be included in the framework.

Finally, before introducing some candidate measures, we want to remark that these measures could be obtained either from the information of already deployed systems in the factories (i.e. ERP, MES, etc.) or from the logged information of FACTS4WORKERS BBs. The first approach will require opening the framework for integrating the required data from existing systems (and in consequence it would require some IT staff intervention). The second approach will make the framework independent of any existing system at the factory, but it would require more detailed logged information.

3.3.3.1 Batch Production Processes

From FACTS4WORKERS scope, Batch Production Processes are the manufacturing processes where a huge quantity of products units is created within a machine or line of machines. Some relevant measures that could be considered are (from MESA, 2006; OpsDog, 2016):

- Cycle Time: Manufacturing Process – The average number of days required to process a manufacturing work order from receipt of the customer's order at the appropriate manufacturing facility until the product is ready for packaging, including both standard and customized products.
- Manufacturing Cycle Time – Measures the speed or time it takes for manufacturing to produce a given product from the time the order is released to production, to finished goods.
- On-Time Delivery to Commit – This metric is the percentage of time that manufacturing delivers a completed product on the schedule that was committed to customers.
- On-Time On-Schedule Rate (OTOS) – The inverse of the target number of units to be produced minus the actual number of units produced divided by the actual number of units produced over a certain period of time, as a percentage.
- Production Attainment – Actual production (units or volume produced) divided by target production over a certain period of time, as a percentage.
- Throughput – Measures how much product is being produced on a machine, line, unit, or plant over a specified period of time.
- Yield – Indicates a percentage of products that are manufactured correctly and to specifications the first time through the manufacturing process without scrap or rework.

3.3.3.2 "Project Based" Production Processes

Project Based Production processes aim to create just one or a very little quantity of units of product (machine, machine line, etc.) for a given customer order, satisfying a very concrete set of requirements. These processes involve all the tasks from the engineering to the final assembly. Because they are production processes, metrics identified in the previous paragraphs apply also to these project based production processes. Even though, because of their particularities, additional measurements can be defined. Some of the most relevant found in the literature review are:

- Engineering Change Order Cycle Time – A measure of how rapidly design changes or modifications to existing products can be implemented all the way through documentation processes and volume production.

- Product Remanufacturing Rate – The number of products that are remanufactured over a certain period of time as a result of change(s) by the Design Team, Customer Engineering Department or Internal Engineering Department, in product specifications, supplies or other characteristics divided by the total number of products produced over the same period of time, as a percentage.
- Projected vs. Actual Project Hours – The variance between the hours scheduled for an employee over a certain period of time versus the amount of actual hours worked on the floor over the same period of time, as a percentage.
- Schedule Variance – The number of tasks performed over a certain period of time that were either unplanned (i.e., not scheduled) or did not conform to the production schedule divided by the total number of tasks scheduled over the same period of time, as a percentage.

3.3.3.3 Maintenance and Set up Production Processes

Industrial Maintenance Processes are those processes aiming to keep machines and other relevant equipment of the factories working correctly in order to gather the compromised production levels. Due to the advance of the technology, the maintenance strategies have evolved based on the support that ICT provides. As a consequence, nowadays very common maintenance strategies are condition-based maintenance, predictive maintenance, remote maintenance, preventive maintenance, e-maintenance, etc.

Set-Up Production Processes are processes which are related with the setting up of the machines for producing a new product (after retooling) of after the detection of errors to solve them. For some authors (Kumar, 2013) these processes (or tasks) are part of the maintenance processes as so we decided to treat them (maintenance and setting up) as different processes. We also based our decision on the clear differentiations of them in the use case definition performed in D1.2 and in the need to transfer some of these more routine tasks from expert workers (team leaders and maintenance workers) to less skilled workers.

We finally decided to treat them together after performing the literature review of possible measurements and indicators and realizing many of them can be applied for both initial types of processes. Using valid references for Maintenance Processes (Kumar, 2013; Parida, 2009) and for setting-up ones (Low, 2014), we create an initial list of potential common measurements and them two particular chapters for the ones applying to maintenance and setting-up processes, respectively. Next is the suggested list of common measurements:

- Breakdown frequency

- Downtime in Proportion to Operating Time – This ratio of downtime to operating time is a direct indicator of asset availability for production.
- Downtime as a Percentage of Uptime – The total amount of time a machine has spent not in operation over a certain period of time divided by the total amount of time a machine has been in operation over the same period of time, as a percentage.
- Equipment Failure Rate – The number of hours manufacturing equipment was not in operation due to failures over a certain period of time, divided by the total number of hours the manufacturing equipment was used for the same period of time, as a percentage.
- Machine Non-Operating Time – The amount of unplanned downtime, or NOT, for a particular machine (or group of machines) over a certain period of time.
- Machine Uptime – The average amount of time manufacturing equipment are in operation divided by the total amount of time in which the manufacturing equipment are scheduled for usage over the same period of time, as a percentage.
- Mean time to repair (MTTR)
- Mean time between failure (MTBF)
- Machine Utilization – The amount of time a machine is in operation over a certain period of time (i.e., 24 hours, etc.), as a percentage.
- Number of shutdowns.
- Overall Equipment Effectiveness (OEE) – This multi-dimensional metric is a multiplier of Availability x Performance x Quality, and it can be used to indicate the overall effectiveness of a piece of production equipment, or an entire production line.
- Total Lost Production Time – The total amount of time in which nothing is being produced due to one or multiple machines not being in operation because of either issues with the material or the equipment itself divided by the total amount of time the machines are scheduled to be running.
- Waste Rate per Machine – The total amount of waste produced (overproduction, waiting inventory, etc.) by a machine over a certain period of time divided by the total output of that machine over the same time period, as a percentage. Equipment Failure Rate – The number of hours manufacturing equipment was not in operation due to failures over a certain period of time, divided by the total number of hours the manufacturing equipment was used for the same period of time, as a percentage.

Considering just maintenance processes, some suggested measurements are:

- Cycle Time: Equipment Repair – The average number of hours required to repair equipment failure(s), either by internal employees or external repair/maintenance services, from the time the equipment fails until when the equipment is repaired.
- Equipment Repaired per Manufacturing Engineering Employees – The total number of equipment repaired over a certain period of time divided by the total number of Manufacturing Engineering employees.
- Number of work order requests in backlog.
- Percentage Available man hours used in proactive work.
- Percentage Planned vs. Emergency Maintenance Work Orders (WOs) – This ratio metric is an indicator of how often scheduled maintenance takes place, versus more disruptive/un-planned maintenance.
- Percentage WOs assigned for rework.
- Unplanned maintenance interventions.
- Unscheduled maintenance downtime.

Relevant measurements for setting-up processes are:

- Cycle Time to Make Changeovers – Measures the speed or time it takes to switch a manufacturing line or plant from making one product over to making a different product.
- Engineering Change Order Cycle Time – A measure of how rapidly design changes or modifications to existing products can be implemented all the way through documentation processes and volume production.

3.3.3.4 Quality Control Processes

Quality Control Processes are those processes which are related with the quality assurance of the products being manufactured. It is important to remark that the measurements which are included in next list are directly related with the product quality. It can be argued that some of measurements introduced of previous paragraphs (i.e. MTBF, Number of shutdowns, etc.) can be considered quality indicators of the processes under analysis (maintenance or setting-up) also impacting the product quality (for simplicity, we do not consider them here). Next are the candidate measurements:

- First Pass Yield (FPY) – The difference in units produced (output) by a manufacturing process over a certain period of time compared to the units that went into production (input) over the same period of time (i.e., input vs. output).
- Number of Non-Compliance Events – The total number of non-compliance incidents recorded over a certain period of time.

- Percentage of Products in Compliance – The number of units produced over a certain period of time that are in compliance with government regulations and internal guidelines after the first pass divided by the total number of units produced by the Manufacturing & Assembly Group over the same period of time, as a percentage.
- Percentage of Units Reworked – The number of units produced over a certain period of time that are reworked to make improvements or fix errors made during the production process divided by the total number of units produced by the Manufacturing & Assembly Group over the same period of time, as a percentage.
- Production Error Rate – The number of products produced with errors divided by the total number of products produced, as a percentage.
- Scrap Rate Due to Errors – The number of units produced over a certain period of time that must be scrapped because of product defects or errors divided by the total number of units produced by the Manufacturing & Assembly Group over the same period of time, as a percentage.
- Target Waste Amount Attainment Rate – The actual overall manufacturing process divided by the target amount of waste production over a certain period of time, as a percentage.
- Waste Rate per Machine – The total amount of waste produced (overproduction, waiting inventory, etc.) by a machine over a certain period of time divided by the total output of that machine over the same time period, as a percentage.

3.3.3.5 "Team Processes"

Team Processes are those processes which are transversal to the one we analysed previously such as workers training, workers collaboration, etc. Those processes where traditionally not performed on work-place but, because of the present ICT capabilities, and in accordance with the project objectives, it will be possible in the very near future. Next is a list of some relevant measures to be considered (Noble, 2003; Semler, 2014):

- Fraction of time correct team member is asked for information.
- Fraction of time information needed by others is provided in a way that could be understood without the need for clarification.
- Fraction of time "private information" needed by group is provided.
- #Number of training courses (and time to complete them).
- #Number of learners completing courses.

4 Evaluation Framework Description

4.1 Introduction

After positioning the Evaluation Framework within the project (chapter 2) and the elements that are used as baseline and rationale for its definition (chapter 3), this chapter proceeds with the Evaluation Framework description.

Taking into account both the Framework Rationale and FACTS4WORKERS project properties, needs and goals, we consider that the Evaluation Framework description should rely on these pillars:

a) It should consider the different stages of development and deployment of an ICT solution and, thus, the appropriate evaluation strategy and methods for each phase. With this approach, the framework may also be seen as an **evaluation process**, with a set of proper tools & methods being used depending on where we are in the process.

Evaluation Framework Mainstays

b) It should **rely (and leverage) on the work being performed in WP1**, particularly on the Worker Impact Dimensions (D1.1), since they compose an incipient evaluation framework based on a human-centred analysis of IPs workers practices. This is the main instrument for analysing the impact of the project solutions on the worker. Besides, a clear link with downstream tasks of the project (WP2-WP5) is established via the expected impact of each Context-Of-Use on the Dimensions.

c) It should provide **both a solid background and usable and extendable guidelines** of evaluating methods and tools. The framework should be perdurable in the sense that not only defined for meeting FACTS4WORKERS goals but also for being used and evolved beyond the project end.

Although no further versions of the framework definition are committed as project deliverable, the Evaluation Environment definition (D6.2, expected for M36) will complete what we are saying here, in order to have the remaining 'deploying' info for the framework that can be missed in this document.

The following chapters describe the framework in a detailed way (chapter 4.2) and also remark additional issues (chapter 4.3)

4.2 Framework Description

The central goal of the evaluation framework is to assess whether the FACTS4WORKERS project creates the intended impact to the work places. Therefore the dependent variables of the framework correspond with the main project objectives specified in the proposal. These are:

1. To **increase problem-solving and innovation skills** of workers.
2. To **increase cognitive job satisfaction** of workers participating in the pilots.
3. To **increase average worker productivity** by 10% for workers participating in pilots.
4. To **achieve TRL 5-7** on a number of worker-centric solutions through which workers become the smart element in smart factories.

Objective four might be the easiest one to evaluate. A TRL of five to seven means that the system's prototype can be used in the work environment. However, this direct use in the target environment is a strict precondition of reaching objective two and three. Therefore, if the solution can be embedded in the real world environment, and if objective two and three are met then objective four can automatically be considered met as well. For the other three dimensions a carefully designed evaluation framework is necessary. Figure 8 gives an overview of the framework's components and their (causal) relation to each other.

With respect to the overall project goals, the framework makes two core assumptions on the underlying causal relationships:

- Cognitive Job Satisfaction is positively influenced by motivation.
- Motivational factors are moderated by the individual characteristics of a worker as well as by the work environment itself.

Following established motivational theories, such as (Herzberg, Mausner and Snyderman, 1959; Hackman and Oldham; 1976) and more recently, the self-determination theory (Ryan and Deci 2000), motivation is positively influenced by the level of perceived "autonomy", "relatedness" and "competence". Task "variety" was also added as a further factor facilitating workplace motivation (Miner 2007; Turner and Lawrence 1965). Together with factors targeting the outcome of work, such as efficiency and quality, all goals can be causally related to intermediate factors. We argue that these factors are determined by the sustained change of work practices and are an emergent phenomenon resulting from the interventions (both technical as well as organizational) that the FACTS4WORKERS project introduces into these environments. Those factors are the worker impact dimensions identified in D1.1.

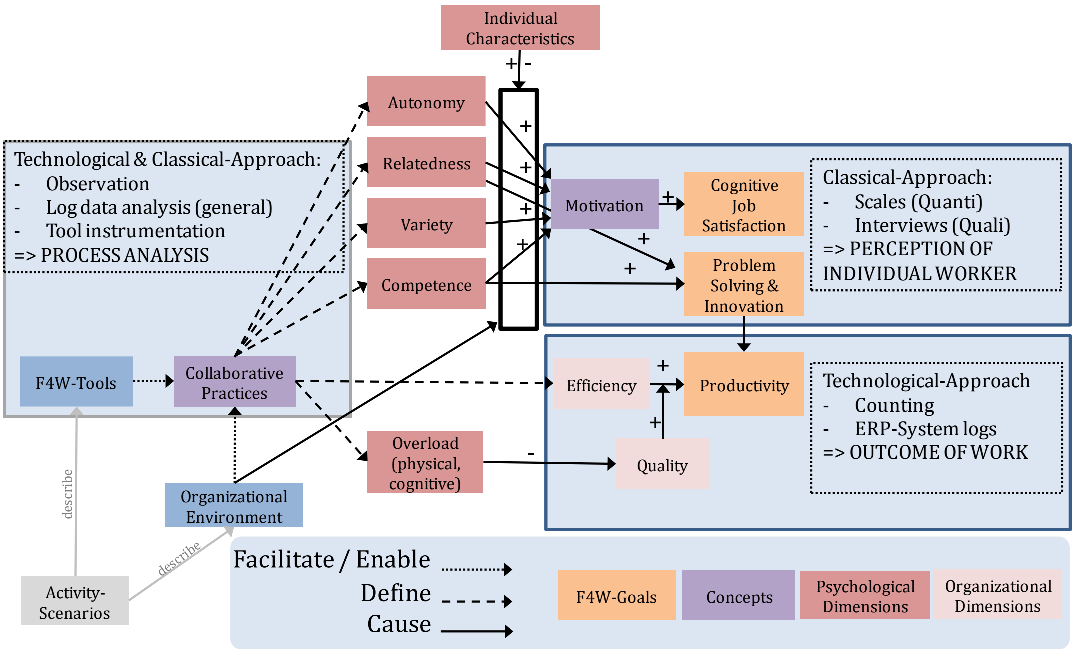

Figure 8. Evaluation Framework Overview and Causal Relationships.

As it was initially described in 3.1 and along the entire chapter 3, the main elements expected to be used to build the framework are **tools and methods** from both classical and technological approaches, which lead us to analyse the impact of the project solutions (socio-technical interventions) along the project life. Those tools and methods are based on a solid background and also provide measurements to feed the worker impact dimensions defined in D1.1 (as it will be explained in 4.2.2), which are the instruments that we're using to evaluate the impact of the solutions.

Thus, the framework is using a set of tools and methods (taking as source of knowledge and proven background the explanations in chapter 3) that will provide measurements that will be used to evaluate the impact of the project solutions on the workers (via D1.1 worker impact dimensions). These impact indicators will feed the project (WP1-WP5) in order to be able to redefine the interventions (following the perpetual beta paradigm explained in the project proposal). The following schema (Figure 9) summarizes this flow:

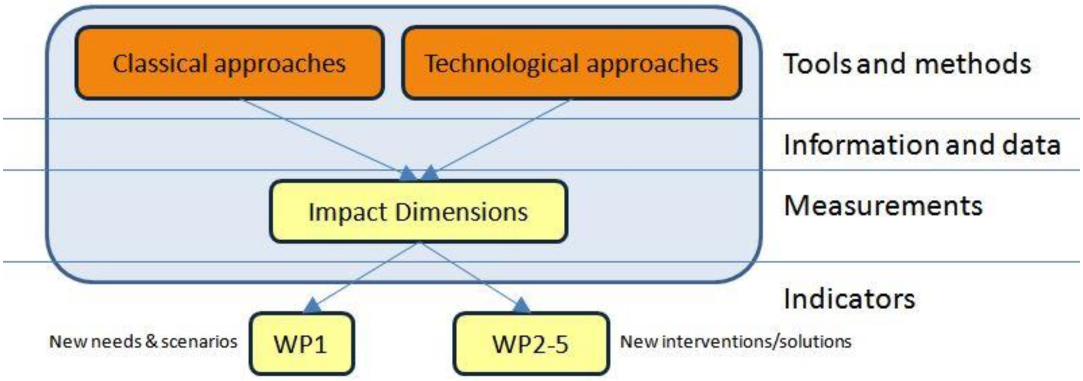

Figure 9. Evaluation Framework information flow (first approach).

Changing work practices (with the support of ICT) will impact the workers

The aforementioned scenario is based on just measuring the introduction of new tools and solutions as a means to impact the workers. This is necessary for having valid measurements for the solutions deployed (and also their connection with the D1.1 worker impact dimensions), but it is not sufficient for having the whole picture, since FACTS4WORKERS is not a software-centred project, but a worker-centred one. **The real impact on the workers will come from changing their work practices** and that is a process that:

a) Is being performed along the whole project lifecycle, with the contribution and feedback from workers, via different instruments and with different maturity levels for each project stage.

b) Involves the validation of the tools introduced in each intervention (again, in an iterative process and with different maturity levels). And this is something to be performed before measuring the real impact that the new work practice (facilitated by the tool) introduces.

c) Effectively, it uses ICT tools and solutions, the ones developed in WP2-WP5, and it leverages on them, but the process is ruled by a change on the work practices (where the introduction of ICT solutions is the second stage, in a needed but supporting role).

Thus, we need to complete the big picture of the framework with an approach that takes into account the project goal of putting the worker in the centre and change his/her practices.

Resuming the work performed in WP1, the worker practices identified in D1.1 (and to be evaluated via the worker impact dimensions) are focused on requirements of ICT solutions that support smarter work (D1.2). These requirements are detailed via different Use Cases (UC) –Context of use- that provide Activity Scenarios (AS) to solve the identified Problem Scenarios (PS).

The different AS propose the introduction of artefacts (basically composed of ICT tools and work processes) for each UC and in each IP, in order to support a consistent change of the worker practices. These artefacts are supposed to solve a given PS within a new AS. Thus, we firstly need to **validate** that the artefacts effectively carry out their task (as a proper means –and first step- to change the worker practices). This is the first mission of the framework.

As already said, in an iterative and perpetual-beta based process, like the one defined for FACTS4WORKERS, different stages of the project and different maturity level in the artefacts introduced during the process, will require a set of tools and methods to properly perform the validation.

Evaluation: Validation + Impact Analysis

The second mission (and final goal) of the framework is to effectively **analyse the impact** that the interventions (shaped like 'artefacts introduction') have on the workers. The core indicators of this analysis are, as it cannot be otherwise, the Worker Impact Dimensions (defined in D1.1). Since the impact on the workers depends on both the nature and extent of use (Steinhueser et al., 2015) of the artefacts, the tools and methods used for measuring both sides of use (nature and extent of use) must be effective and proven for those purposes but also connected to the already defined Worker Impact Dimensions, in order to complete a coherent and powerful measurement framework. In the same way than for the validation part of the framework, the impact analysis tools and methods should take into account, for a proper evaluation, the different maturity level of the artefacts and stages along the project lifecycle.

The following picture outlines main high-level activities of the Evaluation Framework (validation and impact analysis), and how they relate to Use Cases (UC) (solving them) and worker practices (improving them) identified in the Industrial Partners (IP) context (Figure 10):

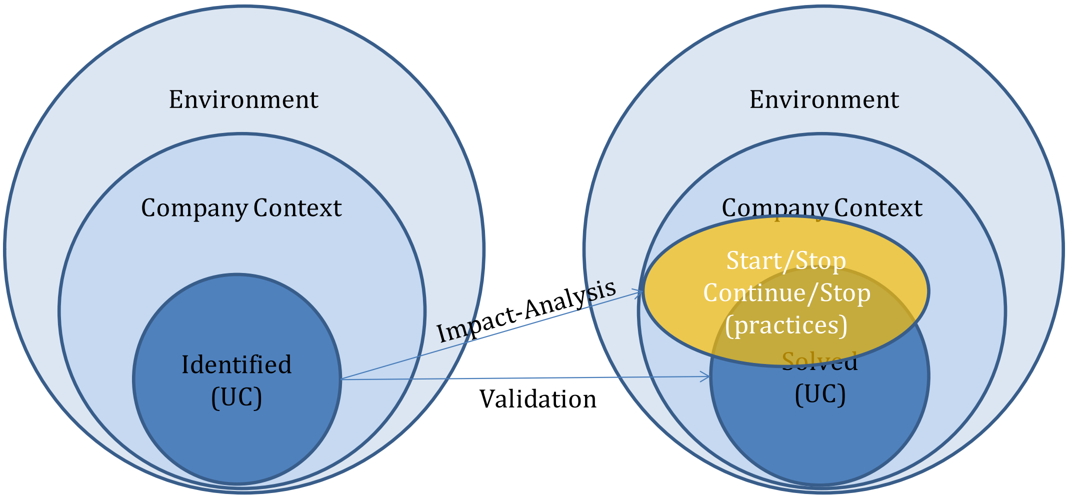

Figure 10. Evaluation Framework high level activities.

We now refine (Figure 11) to resume the Evaluation Framework information flow:

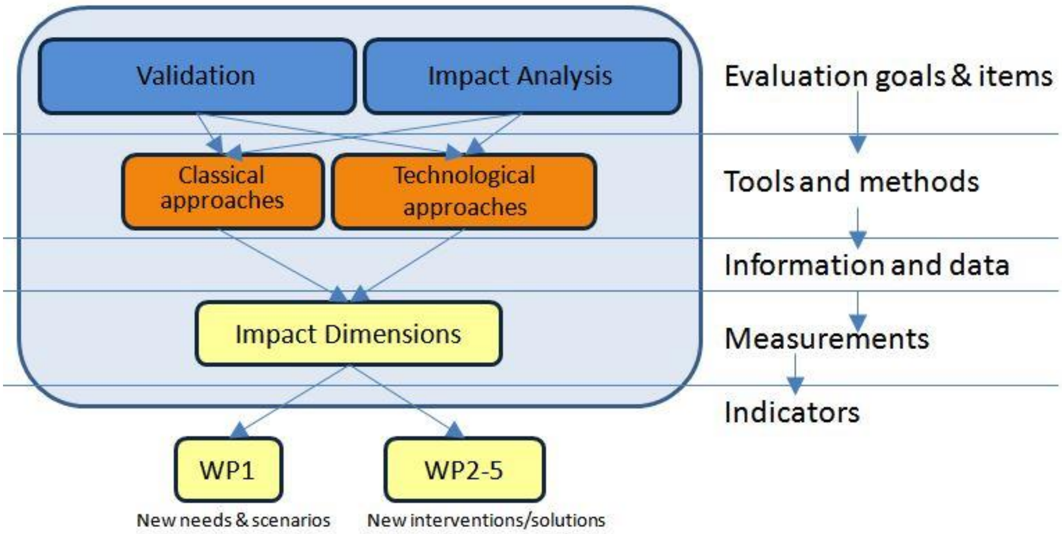

Figure 11. Evaluation Framework Information flow.

The previous figure outlines the Evaluation Framework information flow, and it can be explained as follows:

To successfully achieve the WP6 evaluation goals, we need to perform both validation and impact analysis activities, which will evaluate the project interventions (introduced artefacts) using the most proper tools and methods obtained from classical and technological approaches. The information generated by these tools and methods will feed and consolidate the worker impact dimensions, which will be used for both analysing the workers satisfaction and innovation skills improvement, but also to give feedback to the other FACTS4WORKERS WPs, in order to be able to continuously improve the worker practices and the solutions supporting them.

This Validation + Impact Analysis approach can be supported by 2 evaluation schemes (Figure 12): Formative evaluations and summative evaluations. While the formative evaluations aim to validate early system designs and provide new design insights throughout the development process, the summative evaluations aim at assessing the impact these systems can make in an improved work environment. The impacts however are not directly caused by ICT-systems but are affected by altered work practices and strategies instead (cf. Carr 2003). Nevertheless can formative evaluations focus on design-objectives of the technical systems and assess their fitness with respect to the specific context of use they are applied within. Further, general IS success factors, such as system acceptance, usability and case specific performance indicators are subject of these formative evaluations. This allows the project to get feedback/validation also on early designs from target users without the need to actually introduce them into the real work environment.

However, these formative evaluations are not sufficient for assessing the overarching project objectives. Evaluation criteria, such as productivity and or job satisfaction are a function of a sustained change in work practices rather than of momentary interventions. Therefore the framework supports longitudinal evaluation, typically referred to as Proof-of-Use (Nunamaker Jr et al. 2015). This "Proof-of-Use" aims to demonstrate the system's capabilities of supporting everyday work processes while demonstrating it fitness to address the problems and create specific value for the stakeholders (Nunamaker Jr et al. 2015).

Evaluation Object: Not the ICT solutions, but the new practices that they allow/enable

Summing this discussion up, the framework treats the ICT solutions as enabler and facilitators that foster certain collaborative practices to emerge given a specific work environment. Hence, the change in practices is the actual independent variable that causes the intended project goals (dependent variables) to emerge (Figure 8).

Assessing the different parts of this framework requires different data collection techniques and different scopes of data acquisition. The easiest and most reliable measure arguably is the assessment of change in productivity as the companies already measure these parameters. A pure technological approach might be sufficient to capture these changes. ERP systems for example can provide the required data on successfully completed parts as well as on parts that did not pass the quality control, on raw material wasted and on time the operations took. The scoping would in this case be the work area, directly affected by the FACTS4WORKERS solution. Both other targets (cognitive job satisfaction and problem solving & innovation skills) are highly dependent on the individual worker. Therefore these variables need to be assessed on a personal level for each individual worker to allow for meaningful interpretation of the data (e.g. Weiss and

Cropanzano 1996; Judge and Larsen 2001). Here, classical techniques, such as questionnaires and interviews are suggested as the primary instrument of data collection.

From the perspective of the epistemology of this evaluation framework it is necessary to proof that the FACTS4WORKERS interventions actually change the current work practices (i.e. change the independent variable). This change can either be directly observed at the workplace or indirectly through the IT system usage. In this case, the IT-systems would be instrumented accordingly to create log entries that allow the reconstruction of the applied processes and practices. So called process mining (discovering processes from log system logs) is a well research and widely applied data collection and analysis method (Van der Aalst, Weijters, and Maruster 2004) to analyse work practices and processes in technology supported work environments.

4.2.1 FACTS4WORKERS Evaluation Process Model

Evaluation as a Process

The evaluation framework as it was outlined above ensures that the project outcomes meet the project objectives in measurable ways and therefore represents the overall connecting structure of the FACTS4WORKERS project. It ties research, design, and development together and quantifies how the project objectives have been met. However, the framework does not in itself describe how it can be concretely realized in time and space. The process is therefore described in this section and depicted in Figure 12. The model matches recommended development and integration practices of iterative developments (see e.g. Walden et al. 2015).

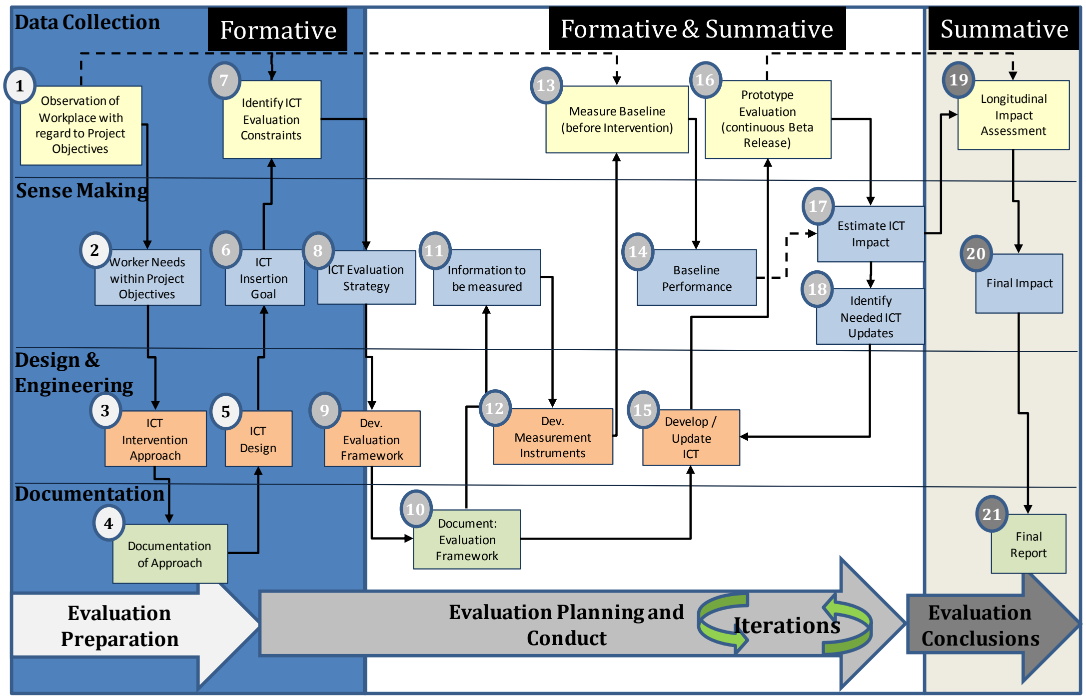

Figure 12: FACTS4WORKERS Evaluation Process Model

This model distinguishes three phases for the evaluation: First, the evaluation is prepared in steps 1 through 5, planned and conducted in steps 6 through 18, and conclusions derived in steps 19 through 21.

4.2.1.1 Evaluation Preparation

During the evaluation preparation, the project objectives are concretized for the work environments for which the ICTs will be developed: how could the project objectives be concretely realized? This information forms the foundation for the evaluation framework that was described above. Specifically, workplace observations are made from the perspectives of the project objectives (step 1), leading to the elicitation of concrete worker needs and technological opportunities (step 2). This information is used to determine the ICT interventions that meet these needs and opportunities (step 3) and that are documented in step (4). The ICT interventions are then designed in step (5). Other issues (see 4.3) must be also taken into account at this stage because of their impact on the evaluation strategy.

4.2.1.2 Evaluation Planning and Conduct

Once designed, insertion goals for the ICT interventions are derived: in what ways should the ICTs be "ideally" introduced into the work place to allow best-possible results, optimal design iterations, and maximal user and process acceptance? In what groups should the ICTs be ideally demonstrated, and for how long? How many workers and managers should be involved? What communication strategies should be used to inform the rest of the organization of these tests? Once formulated as ICT insertion goals in step (6), actual ICT insertion opportunities are identified in the concrete work environments (step 7). These reality checks would bound the actual insertion of ICTs into the work place: Not all insertion goals will be realizable in reality, for example, because only limited groups of workers may be available for an evaluation. Based on the gained understanding from steps 6 and 7, an evaluation strategy can be formulated in step (8), leading to an evaluation framework (step 9) that is documented in this present document (step 10), see Figure 8. The evaluation framework points to the information that needs to be measured for each of the use cases (step 11) and the selection and development of the needed measurement instruments (step 12). These measurement instruments are then used in the concrete work environments to quantify and describe the work situation prior to the introduction of any intervention. The baseline is necessary to assess the effectiveness of the ICTs based on the project objectives (step 14).

The evaluation framework furthermore informs the ICT development by providing specific evaluation information such as "this is where the intervention will be used, how it will be used, and what it is intended to achieve for the user" (step 15). This information is often not available per se to developers who can be isolated from the real world use environment and may make "ad-hoc" decisions that are not always aligned with the actual worker needs. Knowledge of the evaluation goals therefore allow to streamline the evaluation process and should increase the likelihood of positive evaluation results: concretizing and contextualizing the development goals should help the development.

Once a first prototype has been developed, it is presented in the work environment and evaluated (step 16). Such prototype evaluations occur within relatively constrained settings that limit the influence of other work activities. This allows a cleaner comparison of the performance with ICT to the baseline performance (step 17) and identify needed ICT updates (step 18) for the next ICT iteration in step (15).

4.2.1.3 Evaluation Conclusions

The cycles of evaluation conduct and iterative improvements continue for the foreseen period of time or until the estimated impact has reached the expected results. After that, longitudinal assessments of the ICT intervention (step 19) under less constrained evaluation conditions, lead to an estimation of its final impact (step 20) that is documented in step (21).

4.2.2 Evaluation framework background

The reviewed theories (chapter 3) and evidence lead us to expect that holistic frameworks of socio-technical interventions that address workers job satisfaction (JS) should not only consider the work and work environment but also consider worker inherent dispositions and how workers experience work related events. Specifically, there exist considerable intra-individual variations of JS as a function of affective events that workers naturally experience at their work place, as well as of their dispositional background of positive or negative affectivity. Both, affective and dispositional factors influence JS without being directly attributable to socio-technical interventions. Therefore, such factors need to be specifically measured so that they can be accounted in the evaluation of socio-technical work interventions. Across the reviewed literature, following factors were determined to influence JS:

	JSS	JDI	JDS	Herzberg Motivator Factor	Herzberg Hygiene Factor
Achievement				☑	
Recognition				☑	
Task variety			☑	☑ "Work itself"	
Skill variety			☑	☑ "Work itself"	
Autonomy			☑	☑ Responsibility	
Experienced meaningfulness of work			☑	☑ "Work itself"	
Compensation schemes	☑	☑	☑		☑
Supervision	☑	☑	☑		☑
Relation with co-workers			☑		☑
Feedback			☑		
Opportunities for growth	☑	☑			☑
Growth Need Strength			☑		
Operating Procedures	☑				☑
Security			☑		☑

Table 8: Influencing Scales

To impact workers job satisfaction, socio-technical interventions need to address **worker**-related dispositional, motivational, and affective needs, not just **work-**

related needs. Validation frameworks therefore need to investigate whether and how worker related needs are addressed by the socio-technical interventions.

These worker needs and empowerment concept (Spreitzer, 1995; Deci et al. 1989) have also been used as rationale to define the D1.1 Impact Dimensions. Thus, in this way, we have both the link with the Worker Impact Dimensions (as final measurements for impact analysis) and the support for our definition of the evaluation framework elements.

4.2.3 Measurements typology

In this chapter we briefly outline some of the ***candidate measurements*** that, using both the classical and technological approaches, will serve us as information sources to make evaluation tasks.

Note that this is not an exhaustive list, since each case (UC in IP) will require different information needs and, thus, different tools and methods to be used for acquiring the needed information.

That's why we just try to outline some of the commonplaces expected to be shared among the most cases. Of course, other measurements will be obtained from the process defined in 4.2.

4.2.3.1 Integrated model of Technology Acceptance for FACTS4WORKERS

In this evaluation framework, any specific model is not used for examination of acceptance of a certain technology, but we take a step towards taking a broader scope and using an integrated model for examination of the technology acceptance of FACTS4WORKERS solutions. So the aim is to further develop and enhance the technology acceptance models in order to measure success of the whole process and application within the production environment.

The models presented in (3.2.3) differ from each other, but they also include overarching and related elements. These elements have been assessed and reorganized in order **to form a model which combines the central viewpoints of each model** but aims to avoid overlapping. The integrated model is discussed below and presented in Figure 13.

In order to test technology acceptance of FACTS4WORKERS solutions, following five indicators from the previous theories and models were selected as the key categories:

- **Perceived usefulness** (TAM, UTAUT, Innovation diffusion theory) *(similar to performance expectancy and relative advantage)*

- **Perceived ease of use** (TAM, UTAUT, Innovation diffusion theory) *(similar to effort expectancy and complexity)*
- **Social influence** (UTAUT)
- **Facilitating conditions** (UTAUT)
- **Compatibility** (Innovation diffusion theory)

In addition, five indicators were selected as related elements for the key categories:

- **Trialability** (Innovation diffusion theory) → relates to facilitating conditions
- **Observability** (Innovation diffusion theory) → relates to social influence and facilitating conditions
- **Information quality** (IS model) → relates to perceived usefulness and perceived ease of use
- **System quality** (IS model) → relates to perceived usefulness and perceived ease of use
- **Service quality** (IS model) → relates to perceived usefulness and perceived ease of use

As stated above, perceived usefulness of TAM, performance expectancy used in UTAUT and relative advantage discussed in Innovation diffusion theory can all be seen to indicate the same idea – the advantage a person achieves by using the technology. The first main category "Perceived usefulness" is chosen for the integrated model to cover this idea. Correspondingly, perceived ease of use (TAM), effort expectancy (UTAUT) and complexity (Innovation diffusion model) all refer to the user's expected effort of using the system. These are combined into the second main category called "Perceived ease of use" in the integrated model. Moreover, the elements of IS Success model, Information, system and service quality are seen to have a remarkable influence on the perceived usefulness and ease of use of the system, and are thus selected as related elements in the integrated model. The third and fourth main categories: "Social influence" and "Facilitating conditions" are chosen from UTAUT model in order to indicate the social and technological preconditions for the system. The elements of trialability and observability from Innovation diffusion theory are seen to be included in these preconditions and are thus selected as related elements in the integrated model. The fifth main category, "Compatibility" indicates the system's fit to existing processes and routines in the organization as well as the existing tools and systems. Finally, the actual use of the system modifies the perceptions of usefulness and ease of use as well as the social influence factors over time thus creating a feedback loop and a dynamic effect to acceptance levels.

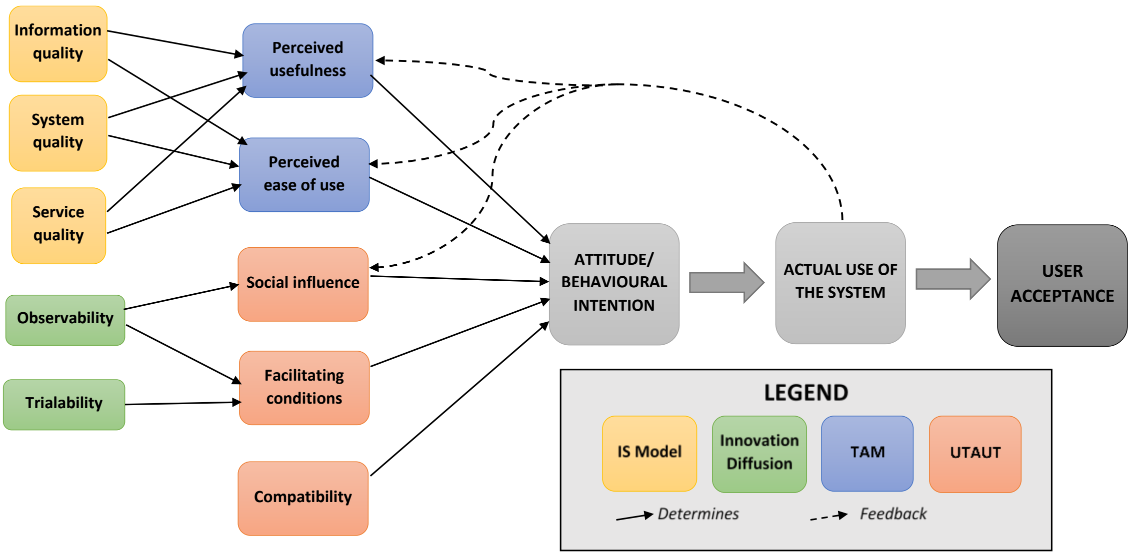

Figure 13. The integrated Technology Acceptance model for FACTS4WORKERS.

Agreement to indicators will be measured on a five point Likert scale. The technology acceptance survey addresses these indicators. In addition, it may contain several open ended questions looking for critical incidents during the demonstration phase, for implementation barriers, and for suggestions for the improvement of the FACTS4WORKERS solution. The survey questionnaire is in Appendix A, which also contains a proposed survey for assessing the innovation skills and the perceived satisfaction. The survey questions have to be partially customized to the specific context of use and industry partner.

4.2.3.2 Measurements provided by ICT solutions

Chapter 3.3 of this document introduces a set of candidate measurements that can be obtained from the data logged when using the system and from the data (either shown or internally generated) during these interactions.

Next table (Table 9) matches the types of measurements defined in chapter 3.3. with the framework proposed by (DeLone and McLean, 2003).

	System Quality	Information Quality	Service Quality	Intention of Use	Use	User Satisfaction	Net Benefits
Measurements From ICT systems							
Knowledge Management System Metrics		☑			☑		
Team Supporting Tools	☑	☑			☑		
Data Management	☑	☑			☑		
Semantic Workflow Engine Metrics	☑				☑		
HMI Measurements	☑	☑			☑	☑	
Industrial Measurements for Use Cases							
Batch Production Processes							☑
"Project Based" Production Processes							☑
Maintenance and Set up Production Processes							☑
Quality Control Processes							☑
Team Processes	☑	☑			☑	☑	☑

Table 9. Matching of Technological Measurement Types to DeLone& MacLean Dimensions.

It must be remarked than the Service Quality (SQ) is not evaluable with the provided measurements. Although it could be important in other scenarios, we consider that SQ is not a key dimension within the FACTS4WORKERS project development and evaluation framework.

When reading the former table, next conclusions are remarkable:

- Measurement From ICT systems, which are based on logs data, can be used to determine the System Quality, the Information Quality, the use (and patterns of use), and, to some extent, the user satisfaction with the system.
- Industrial Measurements for Use Cases are dependent variables, as they can be used to measure Net Benefits (either individuals or organizational). This requires being able to use the data used by the different applications and, in order to determine the veracity of the impact, to compare these data with data from workers not involved in the FACTS4WORKERS. Moreover, it seems the only way to obtain this (business related) measurements.
- The "Team Processes" measures are the only exception of Team Processes measures that can be used to measure Information Quality, and to some extent, the System Quality.

- Although User Satisfaction can be obtained from the logged data by the HMI measurements, another source of data should have to be considered instead (i.e. classical methods).
- Subjective measures, such as Intention of Use or User Satisfaction, cannot be obtained from logged data.

Similar conclusions follow from the examination of Table 10, by the fact that Industrial Measurements for Use Cases are close related with Organizational Dimensions. On the other hand, measurements from ICT Systems are also related with Individual Impact Dimensions.

	1. Autonomy	2. Competence	3. Relatedness/communication	4. Variety	5. Protection	6. Quality	7. Time efficiency
Measurements From ICT systems							
Knowledge Management System Metrics	☑	☑	☑	☑	☑	☑	☑
Team Supporting Tools	☑		☑		☑		☑
Data Management				☑	☑	☑	☑
Semantic Workflow Engine Metrics					☑		☑
HMI Measurements	☑	☑	☑		☑		☑
Industrial Measurementes for Use Cases							
Batch Production Processes						☑	☑
"Project Based" Production Processes						☑	☑
Maintenance and Set up Production Processes						☑	☑
Quality Control Processes						☑	☑
Team Processes	☑	☑	☑	☑	☑	☑	☑

Table 10. Matching Technological Approaches Measurements and D.1.1 Impact Dimensions.

4.2.4 Quantification Strategies

The Evaluation Framework described above details the background, process, methods and tools to be used for the evaluation data acquisition. Once the data is acquired, both analysis and quantification are needed to properly evaluate the real impact on the workers (via the impact dimensions) of the interventions. The impact evaluation will help us to check the project goals accomplishment, as described in Figure 8.

With the proposed framework description and rationale we have the project goals determined by the impact dimensions (Figure 8) and these composed by the measurements from classical/technological approaches (see again the information flow in Figure 11). This bottom-up connectivity enables to design a proper quantification of the impact.

Flexible Impact Quantification

The 'amount' of the impact (i.e. the variation on the dimension measured item) will be determined by the measurements that feed and compose a given dimension. Chapter 4.2.2, Table 8, Table 9 and Table 10 outline part of the composition of some of the impact dimensions. Any given dimension composition and, therefore, its data feeding, will depend on the methods and tools used for the data acquisition and, thus, on the project stage, Activity Scenario (AS) being measured, maturity of the intervention (ICT solution and/or process) and Industrial Partners (IP) constraints (see 4.3). This is a key asset of the framework as it has been defined: Its flexibility to count on different data sources and acquisition methods and tools to validate and measure the impact, depending on the aforementioned project constraints.

Thus, at this stage it is only proper to summarize different quantification strategies that are likely to be used:

- Dimensions composition: Since different information sources can feed a given dimension in each AS, as described above, it may not be proper to establish a closed or fixed measurements mix (i.e. formula) to compose the impact dimension. Instead, different approaches will be applied depending on each case needs and/or constraints:
    - Simple and powerful approaches can be used when many occurrences of each measurement can be obtained (technological approaches): Using all the available measurements influencing each dimension in each AS to see the evolution of the impact dimension (like scatter-plot alike visualization tools). Longitudinal analysis can be supported via this approach. Of course, normalization should be performed to benchmark and compare impact among different AS.
    - Using the aforementioned background (4.2.2) and example tables, more detailed composition of the dimension will be defined, going beyond of just links (as stated in the tables) but proposing weights and 'roles' (as the causal relationships in Figure 8) only if it makes sense and when possible for each measurement (Nardo et al., 2005).

As stated in (Saisana and Tarantola, 2002), when we are dealing with composite indicators (indicators based in sub-indicators that have no common meaningful unit of measurement and there is no obvious way of weighting them), as impact dimensions are, trying to compose them in a single expression or formula may be not useful nor correct. Thus, quantification will be needed but just to evaluate the "amount of impact" of each intervention on each dimension, not for having a single number to describe the impact. Again, this quantification task will be highly dependent on the AS characteristics, stage of the project and maturity of the deployed solution.

- For each AS in each IP, a subset of the candidate measurements stated in chapter 3 will be selected for evaluation purposes. Some of them are more likely to be used (4.2.3), which also will determine the tools and methods for measurement (e.g. Appendix A) and, thus, the aforementioned Dimensions composition and the following statistical methods.
- Statistical analysis methods and tools (explorative/confirmative analysis; regression models, etc.) will be applied to most of the measurements obtained from social sciences tools and methods (Faul et al., 2007). Validation strategies used for mixed methods (qualitative and quantitative) as well as the scheme for building composite indicators (Saisana and Tarantola, 2002) can be applied for both composing and quantifying the dimensions (Venkatesh et al., 2013). These strategies and methods allow having a solid baseline for the date obtained from the social sciences tools (the so-called classical approach).

Finally, it is remarkable that quantifying the impact is not only a key step for evaluating the accomplishment of the project goals: It will be also a valuable internal tool for providing feedback to other WPs in FACST4WORKERS project (Figure 11).

4.3 Issues to Be Considered

4.3.1 Legal Issues

The fact of releasing an ICT solution on a shop floor has many implications to be taken into account, also legal ones, most of them described in ICT literature and practitioners' best practices. But, when in addition to that, a deep assessment to the impact of the new solutions on the workers must be done, new legal issues arise, and they are dependent not only on the ICT solution properties and, but also on a broader scope scenario. Since it's not possible to describe all legal issues to be taken into account from a general perspective (each case should be analysed individually), we briefly summarize some of the most important ones identified in the context of FACTS4WORKERS project. Also, our strict ethic guidelines can be found in the Project Handbook. Please, note that detailed information regarding the framework environment setup will be described in D6.2 ("Evaluation Environment Definition and Setup"):

- *Regional/National/European legal framework*: Different legal frameworks for workers may be implied in an assessment like the one performed in FACTS4WORKERS project. Several times there's a complex network of legal implications, even possible discussions may arise when not understandable (or even apparently contradictory) questions can be concluded when analysing legal issues. This deliverable is not the proper place to solve these questions but, from a project level perspective, we always need to answer why, what, how, where and to whom to apply each properly identified legal item.

- *Company level rules*: After having a clear view of the different legal frameworks that apply to any given project that implies the assessment of workers practices, we need to take into account the company normative framework. This set of rules should be observed from a dual perspective:

    - The normative one, which takes into account the company established rules and best practices, and that must be respected with the same level of conscientiousness than the aforementioned legal frameworks.
    - The 'sponsorship' one or, what it is the same, the propensity to collaborate with the project goals: It's key to have this kind of sponsorship from the company top management, since aligning goals will notably increase the chances of project success.

- *Unions notice and agreement*: Once we have the external (regional/national/European level) and internal (company level) legal framework, the workers concerns regarding the project implications are still present. Thus, Unions must be involved in the legal discussions, to take into account their concerns and requirements from the same dual perspective than the one identified for the company level rules.

- *Occupational risks/hazards prevention issues*: Finally, and even though these issues can be normally found in the previous points, an additional review of the risks/hazards prevention issues may be needed (legal but also a deeper analysis), since any innovative new ICT solution or practice deployed can also imply new risks to be observed.

4.3.2 Human Issues

ICT and Workers are two of the pillars of the Industry 4.0 vision. ICT is used to improve internal efficiency and to enable higher value-creation through the use of information, in particular, by the shop-floor workers. Workers are the base on which factories of the future will increase their flexibility, agility and competitiveness (EFFRA, 2013). Within Future Factories, monotonous and repetitive tasks will be automated or executed by robots while workers will execute tasks requiring more intelligent approaches (modification of parameters, use of previous experience to solve new problems, etc.). In other words, workers will have to be more dynamic, they will have to be able to improve their skills and competencies while they are working or they will have to be able to share and communicate the knowledge they acquire. It seems clear ICT is going to be the solution to these new worker needs.

Because of this close relation between workers and ICT, we think it is necessary to take a more detailed view on how ICT performance and workers interaction with them (via HMI/HCI) is going to influence in worker satisfaction.

Next paragraphs briefly introduce new issues to be considered when measuring worker satisfaction in an Industry 4.0 context. Based on existing experience in related fields (such as Education) and on our experience within FACTS4WORKERS project, we have identified some factors to be considered: Workers ICT-Literacy; Worker Involvement in the Design of the Solutions; Industry 4.0 Solution Acceptance and Success; assessments of HMI mock-ups and prototypes; Assessments of the worker sentiment when using the provided solutions.

Workers ICT-Literacy

Since more than 15 years ago (Prensky, 2001), we all are classified as Digital Native or Digital Immigrant. A rough criterion for classifying us is based on our year of born (before or after eighties). A more objective (and convenient) criterion is based on the evaluation of our ICT-Skills. Based on this second criterion workers can also be classified.

ICT-Skills are defined as the capacity to solve problems of information, communication and knowledge (Ananiadou & Claro 2009). In other words, the capacities a human being should have to have for being a 'productive citizen' in the information society. Since it was realized that ICT is going to play a transversal role supporting our life, it became a matter of interest for education researches in order to determine which skills will be needed by future citizens, how they must be taught, how student's levels could be evaluated.

Based on the 21st century skills (Trilling & Fadel, 2009) framework, the term ICT-Literacy is used to define our skills based on the use of contemporary technologies for information processing (computer literacy skills) and information and communications skills (Wilson et al. 2015) within the context of everyday complex cognitive problem solving. Several evaluation frameworks have been developed by education researches in order to determine which level the students have (Wilson et al. 2015). These frameworks are based on the design of experiments to solve problems using available technology (from basic search to develop a collaborative work). While they are not probably be used to determine which is the ICT-Skills of workers, because of the costs, these strategies can be used for training purposes and more pragmatic approaches can be used for determining workers ICT-Literacy.

One example is performed within the project T&TNet (Angeletou & Graschall, 2013) where the evaluation of the HMI (at different stages of their development) was preceded of a basic questionnaire in order to determine the ICT-Skills of the users. A questionnaire at early stages of an Industry 4.0 project could help to scan the kind of problems workers may have with the technologies to be deployed, and to anticipate the training solutions. Paradoxically, it seems the problem to be the same as within factories of the future and the education field: younger workers/students would have the ICT-Skills while older workers/teachers would have the knowledge and the experience.

Worker Involvement in the Design of the Solutions

When a consultancy firm is contracted for solving a problem on a company the solution frequently consists on the (minor) adaptation of existing software to the problem view and the previous experiences the executives and/or the consultants have. These top-down approaches are many times perceived as imposed solutions by the users. Moreover, because they do not consider the experience of the front line workers, they do not really support workers on their daily tasks. As a consequence, which can be seen as a cheap and effective solution isn't it and although the system is used not all the potential improvement can be obtained.

Deliverable D1.1 introduces a methodology for defining Industry 4.0 solutions requirements by co-designing them with workers. Following an agile iterative process requirements are gathered on a bottom-up approach. While initially the management of the company is asked to describe the company and the problems to be solved, once the Context of Use is established, the problem definition is created in collaboration with the worker using different methods (interviews, observations, questionnaires, etc.) resulting in the definition of representative Personas, the determination of the Problem Scenario and the Collaboration Diagrams (workflows) representing the information exchanges of the Personas (actors) involved in the Problem Scenario.

References

Angeletou, A., Garschall, M. 2013. T&T Net. D3.1 Pilot Assessment and Evaluation Plan. 2013. Available at: http://ttnet-aal.eu/sites/default/files/%5BTTnet%5DD3.1_Evaluation_Plan.pdf.

Ananiadou, K. & Claro. M. 2009. 21st Century Skills and Competences for New Millenium Learners in OECD Countries. OECD Education Working Papers, 41, OECD Publishing

Awa, H.A., Ojiabo, O.U. and Emecheta, B.C. (2014) Integrating TAM, TPB and TOE frameworks and expanding their characteristic constructs for e-commerce adoption by SMEs. *Journal of Science & Technology Policy Management*, 6, 76-94.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM computing surveys (CSUR), 41(3), 16.

Behrendt, S., Richter, A., & Trier, M. (2014). Mixed methods analysis of enterprise social networks. *Computer Networks*, 75, 560-577.

Bevan, N. (2008). Classifying and selecting UX and usability measures. *International Workshop on Meaningful Measures: Valid Useful User Experience Measurement*. 2008. p. 13-18.

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. International Journal of Human-Computer Interaction, 31(8), 484-495.

Brodke, M. R., Sliter, M. T., Balzer, W. K., Gillespie, J. Z., Gillespie, M. A., Gopalkrischnan, P., .. Yanekelevich, A. (2009). The Job Descriptive Index and Job in General Quick Reference Guide. Job Descriptive Index Office: Bowling Green, OH.

Brooke, J. et al. 1996. SUS-A quick and dirty usability scale. Usability evaluation in industry, 1996, vol. 189, no 194, p. 4-7.

Cardoso, J., et al. (2004) Quality of service for workflows and web service processes. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 1, no 3, p. 281-308.

Carr, N. G. (2003). “IT Doesn’t Matter.” Educause Review 38: 24–38.

Cheng, L. C.; Mustafa, M. (2015). A Reference to Usability Inspection Methods. International Colloquium of Art and Design Education Research (i-CADER 2014). Springer Singapore, 2015. p. 407-419.

Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. (1989) User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science, 35 (8), 982-1002.

DAMA UK Working Group on "Data Quality Dimensions". (2013). The six primary dimensions for data quality assessment: defining data quality dimensions. From: https://www.dqglobal.com/wp-content/uploads/2013/11/DAMA-UK-DQ-Dimensions-White-PaperR37.pdf

Deci, E. L., Connell, J. P., Ryan, R. M. (1989). Self Determination in a Work Organization. Journal of Applied Psychology, Vol 74 (4), 580-590.

DeLone, W.H., McLean, R. (2003) The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, *Journal of Management Information Systems*, 19:4, 9-30

Denner et al. (2015): " First version of requirements of workers and organisations ". Project Report - FACTS4WORKERS: Worker - Centric Workplaces in Smart Factories .

Dugguh, S., I., Ayaga, D. (2014). JS theories: Traceability to employee performance in organizations. IOSR Journal of Business and Management (IOSR-JBM), Volume 16, Issue 5. Ver. I (May. 2014), 11-18.

EFFRA 2013. Factories of the Future 2020 Roadmap. (http://www.effra.eu/attachments/article/129/Factories%20of%20the%20Future%202020%20Roadmap.pdf)

Enriquez, J. G.; Casas, (2014) S. *Framework para implementar pruebas de usabilidad flexibles para aplicaciones móviles*. En XX Congreso Argentino de Ciencias de la Computación. Buenos Aires.

Faul, F., Erdfelder, E., Lang, a.G., Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior research methods*, 39(2), 175-191.

Finstad, K. 2010. The usability metric for user experience. Interacting with Computers, 2010, vol. 22, no 5, p. 323-327.

Fishbein, M. and Ajzen, I. (1975) *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading Addison-Wesley, MA.

FACTS4WORKERS (2014) Description of Action, FACTS4WORKERS project.

Gable, Guy G.; Sedera, Darshana; Chan, Taizan (2008). Re-conceptualizing information system success: The IS-impact measurement model. *Journal of the association for information systems,* 2008, vol. 9, no 7, p. 377.

Gerhard, D.; Dumss, S.; Rosenberger, Patrick (2016): Deliverable 5.1: Blueprint Architecture and Integration Plan. Project FACTS4WORKERS: Work-Centric Workplaces in Smart Factories.

Hackman, J. R., & Oldham, G.R. (1974). The Job Diagnostic Survey: An Instrument for the Diagnosis of Jobs and the Evaluation of Job Redesign Projects. Technical Report No. 170-744.

Hackman, J. R., & Oldham, G.R. (1975). Development of the Job Diagnostic Survey. Journal of Applied Psychology, Vol 60, 159 – 170.

Hackman, J., R., Oldham, G. R., Janson, R., Purdy. K. (1975). A New Strategy for Job Enrichment. California Management Review, Vol XVII, No. 4, p. 57 – 71.

Hackman, J., R. & Oldham, G., R. (1976) Motivation through the Design of Work: Test of a Theory. Organizational Behavior and Human Performance, 16, 250 – 279.

Haghi, G. (2004). Measuring knowledge management at HP services consulting & integration. *Proceedings of I-KNOW'04*.

Hashemi M. & Herbert J. (2015). Baran: An Interaction-centred User Monitoring Framework.In Proceedings of the 2nd International Conference on Physiological Computing Systems. 52-60, 2015, ESEO, Angers, Loire Valley, France

Hermann, M., Pentek, T., Otto, B. (2015). Design principles for Industrie 4.0 scenarios: a literature review. Technische Universität Dortmund, Dortmund (2015).

Herzberg, F., Mausner, B., Snyderman, B. B. (1959). The Motivation to Work. New York, NY, USA: Wiley.

Heinrich, P.; Richter, A. (2015): "Captured and structured practices of workers and contexts of organizations". Project Report - FACTS4WORKERS: Worker-Centric Workplaces in Smart Factories.

Holsapple, C.W. and Lee-Post, A. (2006) Defining, Assessing and Promoting E-Learning Success: An Information Systems Perspective. *Decision Science Journal of Innovative Education*, 4:1, 67-85.

Hollingsed, T.; Novick, D. G.(2007) Usability inspection methods after 15 years of research and practice. *Proceedings of the 25th annual ACM international conference on Design of communication*. ACM, p. 249-255.

Holzinger, A. (2005). Usability engineering methods for software developers. Communications of the ACM, 2005, vol. 48, no 1, p. 71-74.

ISO 9241-11 Ergonomic Requirements for Office Work with Visual Display Terminals (VDT)s-Part II Guidance on Usability" ISO/IEC 9241-11,1998 (E)

ISO 9241-210: 2010–Ergonomics of human-system interaction–Part 210: Human-centred design for interactive systems. ISO. org,.

ISO FDIS 9241-210 (2009) Human-centred design process for interactive systems. ISO.

Judge, T. A., Larsen, R. J. (2001). Dispositional Affect and JS: A Review and Theoretical Extension. Organizational Behavior and Human Decision Processes Vol. 86, No. 1, September, pp. 67–98.

Judge, A. J., Locke, E. A., Durham, C. C. (1998). Dispositional Effects on Job and Life Satisfaction: The Role of Core Evaluations. Journal of Applied Psychology, Vol. 83, No 1, 17-34.

Knoco (2014). Knowledge Management Metrics. http://www.knoco.com/knowledge-management-metrication.htm

Kumar, U, et al. (2013) Maintenance performance metrics: a state-of-the-art review. *Journal of Quality in Maintenance Engineering*, vol. 19, no 3, p. 233-277.

Legris, P., Ingham, J. and Collerrette, P. (2003) Why do people use information technology? A critical review of the technology acceptance model. *Information & Management*, 40, 191-204.

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013, April). UMUX-LITE: when there's no time for the SUS. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2099-2102). ACM.

Lettner, F; Holzmann, C. (2012) Automated and unsupervised user interaction logging as basis for usability evaluation of mobile applications. En Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia. ACM, 2012. p. 118-127.

Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., ... & Cheng, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 119-133.

Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 134-142.

Low, Shye-Nee, et al.(2014) Measurement of Overall Performance Effectiveness in Setup Improvement. *Journal of Industrial Engineering*, vol. 2014.

Macvaugh, J. and Schiavone, F. (2010) Limits to the diffusion of innovation: A literature review and integrative model. *European Journal of Innovation Management,* **13**(2), pp. 197-221

Madden, T.J., Ellen, P.S. and Ajzen, I. (1992) A Comparison of the Theory of Planned Behavior and the Theory of Reasoned Action. *Personality and Social Psychology Bulletin*, 18:1, 3-9.

Mesa (2006). Metrics that Matter Guidebook & Framework - International Guidebook.

Miner, John B. (2007). Organizational Behavior 4: From Theory to Practice. M.E. Sharpe.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on constructing composite indicators. OECD publishing 20015

Nielsen, J. (1993). *Usability Engineering*. Academic Press, Boston, ISBN 0-12-518405-0 (hardcover).

Nielsen, J.(1994) Usability inspection methods. En Conference companion on Human factors in computing systems. ACM. p. 413-414.

Noble, D.; Letsky, M.(2003) Cognitive-based metrics to evaluate collaboration effectiveness. *Evidence Based Research Inc. Vienna* VA, 2003.

Nunamaker Jr, Jay F., Robert O. Briggs, Douglas C. Derrick, and Gerhard Schwabe. (2015). "The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research." Journal of Management Information Systems 32 (3): 10–47

OpsDog (2016) The Manufacturing & Production Kpi Enciclopedia. https://opsdog.com/industries/production/production-kpis-and-benchmarks

Parida, A.; Kumar, U. (2009) Maintenance productivity and performance measurement. En Handbook of maintenance management and engineering. Springer London, 2009. p. 17-41.

Peres, R., Muller, E. and Mahajan, V. (2010) Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing,* 27(2), pp. 91-106.

Petrie, H.; Bevan, N. (2009) The evaluation of accessibility, usability and user experience. The universal access handbook, 2009, p. 10-20.

Prensky, M. 2001 Digital natives, digital immigrants part 1. On the horizon, 2001, vol. 9, no 5, p. 1-6.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211-218.

Rettig, M. (1994) "Prototyping for tiny fingers." Communications of the ACM 37.4 (1994): 21-27.

Rana, Ajay; Mridu, Mridu. (2012). Analytical Comparison of Usability Measurement Methods. International Journal of Computer Applications, vol. 39, no 15, p. 11-18.

Richter, A., Heidemann, J., Klier, M., & Behrendt, S. (2013). Success measurement of enterprise social networks. Paper presented at the Wirtschaftsinformatik 2013, Leipzig, Germany.

Rogers, E.M. (1983) *Diffusion of Innovation*, Third Edition, A Division of Macmillan Publishing, New York.

Ryan, Richard M., and Edward L. Deci. (2000). "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being." American Psychologist 55 (1): 68.

Saisana, M. and Tarantola, S., (2002). "State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development", European Commission, Joint Research Centre, Ispra, Italy, EUR 20408 EN.

Scannapieco, M., Missier, P., & Batini, C. (2005). Data Quality at a Glance. Datenbank-Spektrum, 14, 6-14.

Setz, C., Arnrich, B., Schumm, J., Ehlert.U. (2010). Discriminating stress from cognitive load using a wearable EDA device. Information Technology in Biomedicine, IEEE Transactions on, 2010, vol. 14, no 2, p. 410-417.

Semler, S. (2014) Learning Metrics Building Your Training Scorecard. Cornerstone on Demand Foundation.

Sidi, F., Panahy, P. H. S., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2012, March). Data quality: A survey of data quality dimensions. In Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on (pp. 300-304). IEEE.

Smith, P. C. (1969). The Measurement of Satisfaction in Work and Retirement: A Strategy for the Study of Attitudes. Chicago, IL: Rand McNally.

Spector, P. E. (1985). Measurement of Human Service Staff Satisfaction: Development of the JS Survey. American Journal of Community Psychology, Vol. 13, No. 6, 693 – 713.

Spreitzer, G. M. (1995) Psychological Empowerment in the Workplace: Dimensions, Measurement, and Validation. Academy of Management Journal, 48(5), 1442-1465.

Srivastava, A., Locke, E. A., Judge, T. A., Adams, J. W. (2009). Core self-evaluations as causes of satisfaction: The mediating role of seeking task complexity. Journal of Vocational Behavior, 77, 255 – 265.

Steinhueser, M., Herzog, C., Richter, A., Hoppe, U. (2015). A Process Perspective on the Evaluation of Enterprise Social Software. Proceedings of the 2nd European Conference on Social Media ECSM 2015.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. Communications of the ACM, 40(5), 103-110.

Tidd, J. and Bessant, J. (2013) *Managing Innovation – Integrating Technological, Market and Organizational Change*, 5th ed., Wiley.

Turner, Arthur Nicholson, and Paul R. Lawrence. (1965). Industrial Jobs and the Worker: An Investigation of Response to Task Attributes. Harvard University, Division of Research, Graduate School of Business Administration.

Trilling, B., Fadel, C. 2009. 21st century skills: Learning for life in our times. John Wiley & Sons, 2009.

Van der Aalst, Wil, Ton Weijters, and Laura Maruster. (2004). "Workflow Mining: Discovering Process Models from Event Logs." IEEE Transactions on Knowledge and Data Engineering 16 (9): 1128–1142.

Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D. (2003) User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly,* Vol. 27 No. 3, pp. 425-478/September 2003.

Venkatesh, V., Brown, S. A., Bala, H. (2013). Bridging the Qualitative-Quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly Vol. 37 No. 1, pp. 21-54 (2013)*.

Vermeeren, A. P., et al. (2010) User experience evaluation methods: current state and development needs. En Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. ACM, 2010. p. 521-530.

Voss, K. E., Spangenberg, E. R., Grohmann, B. (2003). Measuring the hedonic and utilitarian dimensions of consumer attitude. *Journal of marketing research*, 2003, vol. 40, no 3, p. 310-320.

Walden, David D., Garry J. Roedler, Kevin Forsberg, R. Douglas Hamelin, and Thomas M. Shortell. (2015). Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities.

Waterson, S; Landay, J. A.; Mathews, T. (2002). In the lab and out in the wild: remote web usability testing for mobile devices. En CHI'02 extended abstracts on Human factors in computing systems. ACM, 2002. p. 796-797.

Weiss, H. M., Cropanzano, R. (1996). Affective Events Theory: A Theoretical Discussion of the Structure, Causes, and Consequences of Affective Experiences at Work. Research in Organizational Behavior, Vol 18, 1-74.

Wilson, M., Scalise, K., Gochyyev, P. (2015). Rethinking ICT literacy: From computer skills to social network settings. Thinking Skills and Creativity, 2015, vol. 18, p. 65-80.

Wong, K. Y., et al. (2013). Knowledge Management performance measurement: measures, approaches, trends and future directions. *Information Development. 2015, Vol. 31(3) 239–257*

Wu, J-H. and Wang, S-C. (2015) What drives mobile commerce? An empirical evaluation of the revised technology acceptance model, *Information & Management*, Vol. 42 (2005) pp. 719–729.

A. Data Acquisition tools – Survey models

TECHNOLOGY ACCEPTANCE

(PLEASE READ CAREFULLY)

The goal of this survey is to capture your current perception about the new FACTS4WORKERS technology.

We will neither assess your performance nor will the data be used later on to do so!

Some tips to fill out the questionnaire:

The individual aspects are specified by a descriptive text. You can give your answer by crossing one of the five boxes beneath the description.

Example 1

I know a lot about soccer and its rules:
I strongly disagree ☐ ☐ ☐ ☐ ☒ I strongly agree

In this example the person strongly agrees i.e. she knows a lot about soccer.

Please fill out the questionnaire completely and carefully without omitting any answers!

The analysis of the results will be carried out in anonymized form only!

Perceived usefulness

	I strongly disagree	I disagree	Neither agree nor disagree	I agree	I strongly agree
Overall, the system is useful for daily operations	❑	❑	❑	❑	❑
The system decreases my workload (if negative, implies added effort due to the system)	❑	❑	❑	❑	❑
The system improves the chance to do something that make use of my abilities	❑	❑	❑	❑	❑
The system improves the chance to develop new and better ways to do the job	❑	❑	❑	❑	❑
The system gives a good overview of the workflow	❑	❑	❑	❑	❑
The system improves my level of situational awareness	❑	❑	❑	❑	❑
[BUILDING BLOCK XXX] is useful for my daily work (replace [] by use case relevant activity - e.g. Checking part availability through the system is useful for my daily work)	❑	❑	❑	❑	❑

*** Note to survey implementation: Items shaded with grey colour are from the dimensions survey. If the acceptance survey is given at the same time as the dimensions survey, these items may be excluded and the results from the dimensions survey to these items utilized instead (to avoid repetition and reach a lower count of answerable items. Items shaded with yellow need to be customized for individual use cases to cover the most important building blocks / features piloted.*

Perceived ease of use

	I strongly disagree	I disagree	Neither agree nor disagree	I agree	I strongly agree
Overall, the system is easy to use	❑	❑	❑	❑	❑
The system displays an appropriate amount of information	❑	❑	❑	❑	❑
Customizing the displayed information is easy	❑	❑	❑	❑	❑
The information displayed is easy to read in all conditions	❑	❑	❑	❑	❑
Messages for interaction with the user are clear and easily comprehensible	❑	❑	❑	❑	❑
The system triggers an acceptable number of notifications	❑	❑	❑	❑	❑
The system swiftly recovers after loss of signal or breakdown	❑	❑	❑	❑	❑
It's easy to find the information that I need	❑	❑	❑	❑	❑
Getting used to the system was easy (training effort was low)	❑	❑	❑	❑	❑
[DOING XXX] is easier by using the system (replace [] by use case relevant activity - e.g. Checking part availability is easier by using the system)	❑	❑	❑	❑	❑

Social influence

	I strongly disagree	I disagree	Neither agree nor disagree	I agree	I strongly agree
My colleagues feel that the system is useful	❑	❑	❑	❑	❑
Most of my colleagues are happy to use the system	❑	❑	❑	❑	❑
My superiors encourage using the system	❑	❑	❑	❑	❑
I am able to solve problems that arise in my daily tasks on my own or with the help of coworkers	❑	❑	❑	❑	❑
I am able to propose new ways of doing or new solutions to existing needs or problems	❑	❑	❑	❑	❑

Social interaction*	Number ……. of comments by the user to others and their content
Social interaction*	Number ……. of comments to the user and his content
Social interaction*	Number ……. of logged peer interactions in system
Social interaction*	Number ……. of new solution ideas to problems

*Data from the FACTS4WORKERS system log or with survey

Facilitating conditions

	I strongly disagree	I disagree	Neither agree nor disagree	I agree	I strongly agree

I know that some improvements are planned to be deployed in my workplace	❑	❑	❑	❑	❑
I know the type of technical solutions planned to be deployed	❑	❑	❑	❑	❑
I know the type of organizational improvements planned to be deployed	❑	❑	❑	❑	❑
I am aware of what's going on, in general, in my company	❑	❑	❑	❑	❑
Communications within this organization is good	❑	❑	❑	❑	❑
I am involved in my daily tasks closely with my colleagues	❑	❑	❑	❑	❑
The spirit of cooperation among my coworkers is good	❑	❑	❑	❑	❑
The open exchange of ideas between you and your peers has increased since you started to use the system?	❑	❑	❑	❑	❑
Training was helpful	❑	❑	❑	❑	❑
Support was readily available during the testing	❑	❑	❑	❑	❑
The testing did not interfere too much with other duties	❑	❑	❑	❑	❑

Compatibility (with processes and routines or with other tools)

	I strongly disagree	I disagree	Neither agree nor disagree	I agree	I strongly agree
I deal with a manageable amount of information and inputs in my daily tasks	❑	❑	❑	❑	❑
I see added value replacing current XX system (e.g. manual machine book) with this new system	❑	❑	❑	❑	❑
The system fits our working practices and processes	❑	❑	❑	❑	❑

How willing you are to incorporate new ways-of-doing in your daily work?						
Absolutely reluctant	☐	☐	☐	☐	☐	Absolutely willing

Background information

I am currently working as		
I am working there since		years
I am	....	years old.
I am	☐ Female ☐ Male	
Frequency of system use*	Average .... of times per hour that I interact with the FACTS4WORKERS system	

*Data from the FACTS4WORKERS system log or with survey

SATISFACTION AND INNOVATION SKILLS

(PLEASE READ CAREFULLY)

The goal of this survey is to capture your current perception about your job practices.

We will neither assess your performance nor will the data be used later on to do so!

Some tips to fill out the questionnaire:

The individual aspects are specified by a descriptive text. You can give your answer by crossing one of the five boxes beneath the description.

Example 1

I know a lot about soccer and its rules:
I strongly disagree ☐ ☐ ☐ ☐ ☒ I strongly agree

In this example the person strongly agrees i.e. she knows a lot about soccer.

Please fill out the questionnaire completely and carefully without omitting any answers!

The analysis of the results will be carried out in anonymized form only!

Willingness to include new ways of doing

How willing you are to incorporate new ways-of-doing in your daily work?						
Absolutely reluctant	❑	❑	❑	❑	❑	Absolutely willing

FACTS4WORKERS project awareness

To what extent you know what FACTS4WORKERS project proposes for your daily work?					
	I strongly disagree	**I disagree**	**Neither agree nor disagree**	**I agree**	**I strongly agree**
I know that some improvements are planned to be deployed in my workplace	❑	❑	❑	❑	❑
I know the type of technical solutions planned to be deployed	❑	❑	❑	❑	❑
I know the type of organizational improvements planned to be deployed	❑	❑	❑	❑	❑

Innovation skills

- Approximately what percentage of your or your team's weekly time is made available to pursue creative ideas?

	%

- Has the open exchange of ideas between you and your peers increased since you joined the company?

YES ❑

NO ❑

	Frequently	Regularly	Sometimes	Occasionally	Never
How often do you have a vibrant exchange of ideas between individuals within your organization?	❑	❑	❑	❑	❑
How often do you take any risk by implementing a new idea/solution/decision in your daily work?	❑	❑	❑	❑	❑
How often do you share your workplace ideas with others	❑	❑	❑	❑	❑
How often do you turn you (or your team) new ideas into new or modified products, processes or services?	❑	❑	❑	❑	❑

Job practices and satisfaction with them

Ask yourself: How satisfied am I with this aspect of my job?

Very Dissatisfied: I am very dissatisfied with this aspect of my job
Dissatisfied: I am dissatisfied with this aspect of my job
Neutral: I can't decide whether I am satisfied or not with this aspect of my job
Satisfied: I am satisfied with this aspect of my job
Very Satisfied: I am very satisfied with this aspect of my job

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
The chance to develop new and better ways to do the job	❑	❑	❑	❑	❑
The chance to do something that make use of my abilities	❑	❑	❑	❑	❑
The chance to be responsible for planning my work	❑	❑	❑	❑	❑
The chance to make decisions on my own	❑	❑	❑	❑	❑

The spirit of cooperation among my coworkers	❑	❑	❑	❑	❑
The chance to work independently of others	❑	❑	❑	❑	❑
The chance to do something different everyday	❑	❑	❑	❑	❑
I can take decisions in my daily job based on information acquired or on my own experience	❑	❑	❑	❑	❑
I am able to solve problems that arise in my daily tasks on my own or with the help of coworkers	❑	❑	❑	❑	❑
I am able to propose new ways of doing or new solutions to existing needs or problems	❑	❑	❑	❑	❑
I am involved in my daily tasks closely with my colleagues	❑	❑	❑	❑	❑
I am aware of what's going on, in general, in my company	❑	❑	❑	❑	❑
I deal with a manageable amount of information and inputs in my daily tasks	❑	❑	❑	❑	❑
Lack of stress and manageability with my job and daily tasks	❑	❑	❑	❑	❑
The way I enjoy my coworkers	❑	❑	❑	❑	❑
Communications within this organization	❑	❑	❑	❑	❑
Explanations about my job assignments	❑	❑	❑	❑	❑
Number of tasks that I have to perform daily	❑	❑	❑	❑	❑
The way I enjoy my job	❑	❑	❑	❑	❑
The diversity of tasks I can perform during my daily work	❑	❑	❑	❑	❑

Background information

I am currently working as		
I am working there since		years
I am	....	years old.
I am	☐ Female ☐ Male	

ABOUT THE PROJECT

- Personalized augmented operator,
- Worked-centric rich-media knowledge sharing management,
- Self-learning manufacturing workplaces,
- In-situ mobile learning in the production.

PROJECT PARTNERS

The FACTS4WORKERS project is composed of 15 partners from 8 different European countries:

Virtual Vehicle Research Center	Austria
Hidria TC Tehnološki center d.o.o.	Slovenia
Universita degli Studi di Firenze, Department of industrial Engineering	Italy
Technische Universität Wien	Austria
ThyssenKrupp Steel Europe AG	Germany
Hidria Rotomatika d.o.o., Industrija Rotacijskih Sistemov	Slovenia
iMinds VZW	Belgium
Sieva d.o.o.	Slovenia
University of Zurich, Department of Informatics	Switzerland
Thermolympic S.L.	Spain
EMO-Orodjarna d.o.o.	Slovenia
Evolaris Next Level GmbH	Austria
Itainnova - Instituto Technologico de Aragon	Spain
Schaeffler Technologies AG & Co. KG	Germany
Lappeenranta University of Technology	Finland

PROJECT COORDINATOR / CONTACT:

VIRTUAL VEHICLE Research Center
Inffeldgasse 21A
8010 Graz, AUSTRIA
Tel.: +43-316-873-9077
Fax: +43-316-873-9002
E-Mail: facts4workers@v2c2.at

FOLLOW US AT:

FACTS4WORKERS
@FACTS4WORKERS
facts4workers-project

Evaluation Framework

This document represents Deliverable 6.1 ("Evaluation Framework") of the H2020 project "FACTS4WORKERS -Worker-Centric Workplaces in Smart Factories" (FoF 2014/636778).

The Evaluation Framework, as main tool used for reaching WP6 goals (to evaluate the impact of the project solutions on the workers), contributes to all other project's WPs, generating data for iterating initial requirements and for evolving the designed solutions. That's why we firstly point the relationship between the work to be performed in WP6 and the rest of WPs.

The evaluation of how the introduction of solutions (including ICT) in the workplace affects the daily work and impacts on the worker implies a very broad research scope. Very different and complementary research lines are involved in that purpose, and we establish the rationale of the framework in a wide range of methods and tools among which we will choose those most appropriate for the purpose of the framework.

The evaluation framework is defined then. Taking into account the available rationale and background, but with the project idiosyncrasy in mind, we establish our primary evaluation targets and process. FACTS4WORKERS tries to change the worker's practices, using the help of ICT tools (but not only leveraging on them). This is going beyond of just to evaluate the deployed solutions. That's why the evaluation framework is defined in terms of the validation and impact assessment of the introduced new practices (with the difficulty to separate this impact from other factors), which is going a step further of just using a subset of methods and tools detailed in the rationale.