H2020 - INDUSTRIAL LEADERSHIP - Information and Communication Technologies (ICT)
ICT-14-2016-2017: Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation

ICARUS:
"Aviation-driven Data Value Chain for Diversified Global and Local Operations"

D5.1 – ICARUS Demonstrators and Platform Evaluation Framework

Workpackage:	WP5 – ICARUS Data Value Chain Demonstration		
Authors:	CINECA, Suite5, UBITECH, SILO, AIA, PACE, ISI, CELLOCK		
Status:	Final	**Classification:**	Public
Date:	14/08/2019	**Version:**	1.00

ICARUS Project Profile

Grant Agreement No.:	780792
Acronym:	ICARUS
Title:	Aviation-driven Data Value Chain for Diversified Global and Local Operations
URL:	http://www.icarus2020.aero
Start Date:	01/01/2018
Duration:	36 months

Partners

UBITECH	UBITECH (UBITECH)	Greece
ENGINEERING	ENGINEERING - INGEGNERIA INFORMATICA SPA (ENG)	Italy
PACE a TXT company	PACE Aerospace Engineering and Information Technology GmbH (PACE)	Germany
Suite5	SUITE5 DATA INTELLIGENCE SOLUTIONS LIMITED (SUITE5)	Cyprus
University of Cyprus	UNIVERSITY OF CYPRUS (UCY)	Cyprus
CINECA	CINECA CONSORZIO INTERUNIVERSITARIO (CINECA)	Italy
OAG	OAG Aviation Worldwide LTD (OAG)	United Kingdom
SingularLogic	SingularLOGIC S.A. (SILO)	Greece
	ISTITUTO PER L'INTERSCAMBIO SCIENTIFICO (ISI)	Italy
CELLOCK	CELLOCK LTD (CELLOCK)	Cyprus
ATHENS INTERNATIONAL AIRPORT	ATHENS INTERNATIONAL AIRPORT S.A (AIA)	Greece
TXT e-solutions	TXT e-solutions SpA (TXT) – 3rd party of PACE	Italy

Document History

Version	Date	Author (Partner)	Remarks
0.1	27/09/2018	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Initial Table of Contents
0.11	18/12/2018	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Draft Version presented in the Limassol plenary meeting and gathering partners' feedback
0.12	31/12/2018	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Draft Version
0.20	07/02/2019	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Draft Version
0.30	22/02/2019	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Draft Version circulated to the management team
0.40	06/03/2019	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Draft Version to address comments by UBITECH for complete revision on 27/03/2019
0.50	15/04/2019	C. Arlandini, R. Turra, G. Pedrazzi (CINECA)	Draft Version to address comments by Suite5 for complete revision of section 4 on 08/03/2019
0.60	21/05/2019	D. Miltiadou, D. Alexandrou (UBITECH)	Contributions to sections 2 and 3.
0.70	22/07/2019	F. Lampathaki (Suite5)	Full restructuring of sections 3 and 4. Contributions to sections 1-5.
1.00	14/08/2019	F. Lampathaki (Suite5), D. Alexandrou (UBITECH)	Final version to be submitted to the EC

Executive Summary

The ICARUS Deliverable D5.1 "*ICARUS Demonstrators and Platform Evaluation Framework*" describes the outcome of the work performed in task T5.1 - "*Project Validation and Evaluation Framework Definition*", that aims at developing the ICARUS evaluation framework and validation methodology, defining the means for obtaining feedback from the end-users in a coordinated and unified manner, and finally, presenting a set of key performance indicators at technical and business level as part of the evaluation process.

The evaluation framework was developed based on the state-of-the art approaches in order to embed both technical and business aspects in a holistic manner and proposes indicative testing techniques spanning from questionnaires and interviews to automated tests and KPIs assessment, that can be used in each case in order to answer the following questions:

- *Does the operation of the ICARUS platform meet the defined objectives from the perspective of its users?* This question is tightly linked with the Business Validation and Product Validation and as such, it shall be answered by involving in the different test cases the demonstrator partners. Actual business benefits, usefulness of the platform, user acceptance, user satisfaction, and ease of use are key evaluation aspects.
- *Is the platform operating according to its specifications?* This question concerns the technical validation of the project (related to the Technical Validation and Product Validation steps) and has to be answered by conducting a quantitative technical evaluation, testing technical parameters of system availability, functionality, and performance.

The ICARUS evaluation framework intends to appropriately guide the testing and assessment activities for the ICARUS offerings both from a technical excellence and business impact perspective and proposes a combination of qualitative and quantitative techniques and templates to shed light on the evaluation aspects (within and beyond the ICARUS consortium) and feed the gained knowledge back to the software developers.

The ICARUS evaluation framework is built in accordance with the International Standards on System and Software Quality Requirements and Evaluation (SQuaRE), as defined in ISO 25010:2011 and ISO 25012:2008, that proposes three different evaluation models. The first model, namely the product quality model, is related to the evaluation of the ICARUS platform with respect to a set of properties of both the software (static) and the hosting IT system (dynamic). The second model, namely the quality in use model, assesses the usability of the platform in a particular context of use and the third model, namely the information quality

model, refers to the evaluation of the quality of the data that are provided through the ICARUS platform.

It is important to note that the ICARUS evaluation framework is instrumental to guide the evaluation activities in WP4 and WP5, yet it is placed in the context of a living process that shall continue to evolve until the end of the final evaluation phase providing the required coordination and guidance to the evaluation activities.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Purpose

The ICARUS Deliverable D5.1 "*ICARUS Demonstrators and Platform Evaluation Framework*" is charged to provide an inclusive demonstrators' evaluation framework, as well as a general guideline document to be used to monitor and align the different demonstrators' phases anticipated in the project. The framework needs to provision for the technical and business validation activities of the ICARUS platform and the ICARUS demonstrators and is expected to lead to valuable remarks and conclusions about the viability and the sustainability of the ICARUS platform.

In general, evaluation is a broad term that concerns the systematic determination of the benefits, added value and significance of a software product / platform / system, using well-defined criteria to effectively assess the user experience, as well as the pragmatic evaluation of the software usability by capturing the overall user experience in non-systematic interviews. It is often associated with the concept of validation and verification (V&V) which, according to ANSI/IEEE Std 1012-2012, aim at addressing (a) whether the software product / platform / system is built right (verification scope), and (b) if the right software product / platform / system is built (validation aspects). Although the international literature is generally replete with various Validation and Verification definitions, ICARUS adopts the ANSI/IEEE Std 1012-2012 definition of V&V:

- Verification is the process of providing objective evidence that the software and its associated products conform to requirements (e.g., for correctness, completeness, consistency, accuracy) for all life cycle activities during each life cycle process (acquisition, supply, development, operation, and maintenance); satisfy standards, practices, and conventions during life cycle processes; and successfully complete each life cycle activity and satisfy all the criteria for initiating succeeding life cycle activities (e.g., building the software correctly).
- Validation is the process of providing evidence that the software and its associated products satisfy system requirements allocated to software at the end of each life cycle activity, solve the right problem (e.g., correctly model physical laws, implement business rules, use the proper system assumptions), and satisfy intended use and user needs.

In this context, the ICARUS evaluation framework embraces the evaluation and validation concepts, as well as the verification aspects, and aims at addressing the question whether ICARUS offers a platform of sufficient value, with positive benefits to its intended users,

allowing them to do something they couldn't do before, or allowing them to do something better or faster than they could before.

Since the demonstration and evaluation activities are ongoing till the end of the ICARUS project, the ICARUS evaluation framework intends to serve the evaluation guidelines and cookbook to follow, yet it is considered as a "live document" that is expected to be revised, if necessary, to follow the project advancements.

1.2 Document Approach

The ICARUS evaluation framework is built based on an iterative process that included brainstorming and discussions on the following steps: setting the evaluation scope, defining the framework at high-level, elaborating on its technical aspects and diving into its business aspects. In practice, for the evaluation purposes of the ICARUS platform, a set of techniques effectively bringing together quantitative and qualitative methods shall be applied in the different evaluation phases to ensure that:

- The ICARUS platform is built according to the requirements and design specifications as expressed by the demonstrators and the aviation data value chain stakeholders (as externalized in the MVP interviews recorded in D1.3).
- The ICARUS platform actually meets the aviation data value chain stakeholders' needs, its specifications (described in D3.1-D3.3) were correct in the first place and it fulfils its intended use for data analytics, data sharing and data linking.

Both technical and non-technical aspects are handled according to recognized standards. The proposed technical validation strategy is based on ISO/IEC 25010:2011 and ISO/IEC 25012:2008. It allows to assess the quality of the ICARUS platform in three crucial axes: I. Its technical robustness; II. The end-users experience; III. The information quality. This final point is particularly important for ICARUS, since it targets the exploitation of an unprecedented volume, diversity and richness of aviation data.

The evaluation strategy is based on the demonstrators' scenarios and their associated test cases, that will be run by the ICARUS demonstrators, and will provide the necessary feedback for the development of the platform. Monitoring a set of key performance indicators (KPIs) is critical to understand the overall operation of the system from a technical and business perspective and the exploitation of data, while structured and unstructured communication with stakeholders within and beyond the ICARUS consortium will contribute in capturing the aviation industry stakeholder's feedback and experience in the ICARUS platform.

1.3 Relationship with other ICARUS Results

The ICARUS Deliverable D5.1 is prepared in the context of Task 5.1 "*Project Validation and Evaluation Framework Definition*" under WP5 "*ICARUS Data Value Chain Demonstration*".

The relevant input for the framework described in this deliverable have been produced in the context of WP1 "*ICARUS Data Value Chain Elaboration*", WP2 "*ICARUS Big Data Framework Consolidation*" and WP3 "*ICARUS Platform Design*". In particular, D3.1 "*ICARUS Architecture, APIs Specifications and Technical and User Requirements*" defines how the validation stage fits in the ICARUS Engineering Methodology and the other stages of the ICARUS implementation. D1.2 "*ICARUS Methodology and MVP*" and D1.3 "*Updated ICARUS Methodology and MVP*" provide the basis for the definition of the user-groups segmentation and of the usage scenarios that correspond to their needs and goals.

D5.1 is expected to lead the activities related to the demonstrators' evaluation in WP5, in T5.3 "Extra-Aviation Services in an Integrated Airport Environment", T5.4 "Routes Analysis for Fuel Consumption Optimisation and Pollution Awareness", T5.5 "Aviation Related Disease Spreading", T5.6 "Novel Airline Passenger Experience", and T5.7 "Demonstrators Evaluation and Impact Assessment".

In addition, D5.1 was prepared in very close collaboration with T5.2 "Demonstrators Baseline Activities, Operation Planning and Coordination" and T4.4 "Technical Verification and Integration Testing", and their corresponding deliverables D5.2 "Demonstrators Execution Scenarios and Readiness Documentation" and D4.2 "ICARUS Platform - Beta Version", respectively.

1.4 Structure of the Document

This deliverable is structured into the following sections:

- Section 1 provides an introduction of the deliverable purpose and its connection with other projects' outcomes and tasks.
- Section 2 describes at high-level the evaluation and validation framework scope and phases to be applied during the ICARUS project.
- Section 3 elaborates on the technical perspectives of the ICARUS evaluation framework.
- Section 4 dives into the business perspectives of the ICARUS evaluation framework.
- Section 5 provides a final overview and describes the future steps.

2 ICARUS Evaluation Framework at a Glance

The overall goal of the ICARUS evaluation framework is to ensure and maximise the success of the ICARUS platform through a holistic evaluation framework that clearly defines a series of coordinated evaluation actions. Towards this end, the ICARUS evaluation framework aims at: (a) ensuring that the ICARUS platform is built according to the requirements and design specifications, and actually meets the aviation stakeholders' needs, and (b) guiding the continuous evaluation of the ICARUS platform throughout the whole implementation phase of the project.

Building on the experience of the FITMAN verification and validation method (Lampathaki et al, 2014; FITMAN D2.1, 2013) that was successfully applied in the context of the FITMAN project (2018), the evaluation framework to be employed in ICARUS is depicted in the following figure.

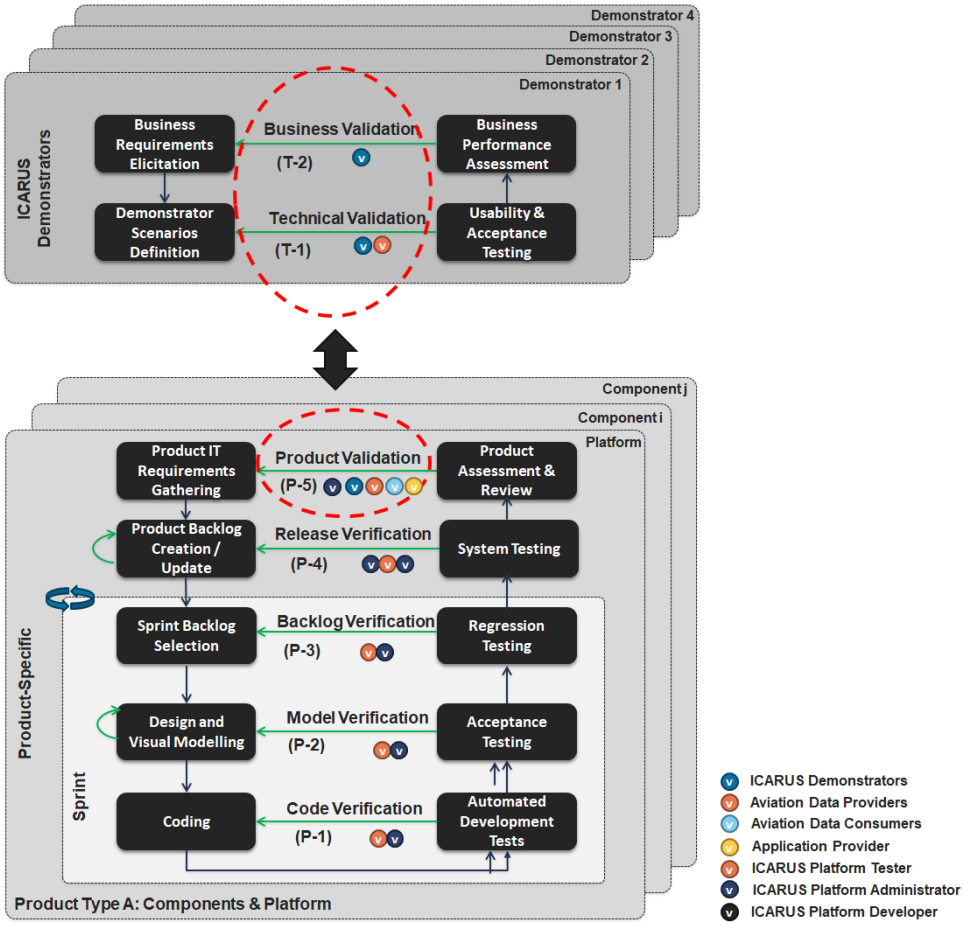

Figure 2-1: ICARUS Evaluation Framework[1]

[1] Adapted from (Lampathaki et al, 2014; FITMAN D2.1, 2013)

In brief (and keeping the terminology of the baseline FITMAN V&V method), the ICARUS evaluation framework bears two core phases, spanning over both the technical and the business perspectives:

- **Product-specific Perspective** that concerns solely the ICARUS platform and its individual components, including the following steps in an agile development approach: *Code Verification (P-1)* that ensures functionality, correctness, reliability, and robustness of code; *Model Verification (P-2)* that is responsible for the alignment between design and requirements, and design and code; *Backlog Verification (P-3)* to determine whether the requirements of the product after each sprint are met; *Release Verification (P-4)* to check whether the requirements of the each product release are met; and *Product Validation (P-5)* which investigates whether the ICARUS platform satisfies intended use and user needs both from a technical and a business view. It needs to be noted that the ICARUS Deliverable D4.2 has explained in detail the integration strategy and the automated tests that are performed to address steps P-1, P-2, P-3, and P-4 which are systematically assessed in WP4, T4.4 "Technical Verification and Integration Testing".
- **Demonstrator Perspective** that involves the ICARUS demonstrators to evaluate the ICARUS platform and the demonstrators' applications that are created on the platform depending on their scenarios, in the following steps: *Technical Validation (T-1)* to guarantee that the overall ICARUS platform and the application created for each demonstrator satisfies intended use and user needs from a technical and functional point of view only; *Business Validation (T-2)* to assess whether the overall ICARUS platform and the demonstrator's application eventually offers sufficient added value and has clear business benefits to the demonstrator, allowing it to operate more efficiently, and supporting it in the transition to the data sharing and analytics era that could not be done before.

Taking into consideration both the technical and the business aspects, the holistic evaluation of the ICARUS platform will be conducted incorporating two different "stakeholder" perspectives:

- **The ICARUS demonstrators' perspective**: The success of the ICARUS platform is tightly related to the successful implementation and execution of the ICARUS demonstrators that will act as success stories for the project. The evaluation framework will include the demonstrators' perspective in the evaluation in order to ensure that the demonstrators' requirements and objectives are fully met in the ICARUS platform and the demonstrators' applications.

- **The broader aviation industry stakeholders' perspective**: The success of the platform is not only related to the successful execution of the demonstrators of the ICARUS platform, but also lies in fulfilling the requirements of the different stakeholders from the three tiers of the ICARUS Data Value Chain that was presented in different ICARUS deliverables, such as D1.1 "Domain Landscape Review and Data Value Chain Definition", and D2.1 "Data Management and Value Enrichment Methods". Therefore, the evaluation framework should include an evaluation method mix to enable learning as much as possible from the broader industry.

In the ICARUS deliverable D3.1 "*ICARUS Architecture, APIs Specifications and Technical and User Requirements*", the ICARUS agile development methodology was presented. In this methodology, the platform is developed following multiple iterations (development lifecycles) in which a platform incremental version is produced with a set of new or enhanced features as defined in the Plan, Design, Build and Test phases. The last phase of each iteration dictates the constant evaluation and verification of each platform increment. Depending on the feedback collected from the evaluation and verification, additional requirements or features are added in the development backlog. The positioning of the ICARUS evaluation framework is clearly set on the last 2 phases, namely the Test and Review phases, as depicted in Figure 2-2.

In this context, the ICARUS evaluation framework will be in a position to identify any deviations from the requirements that were collected following the Requirements Engineering process also defined in deliverable D3.1. The main purpose of the evaluation framework in this case is to provide valuable feedback on each iteration to the development team that will result in new input in the Plan and Design phases of the upcoming iterations. Towards this end, the ICARUS evaluation framework will perform the assessment of each platform incremental version in order to ensure that the multiple iterations will gradually lead to a high-quality platform that provides added value to the identified stakeholders of the ICARUS aviation data value chain.

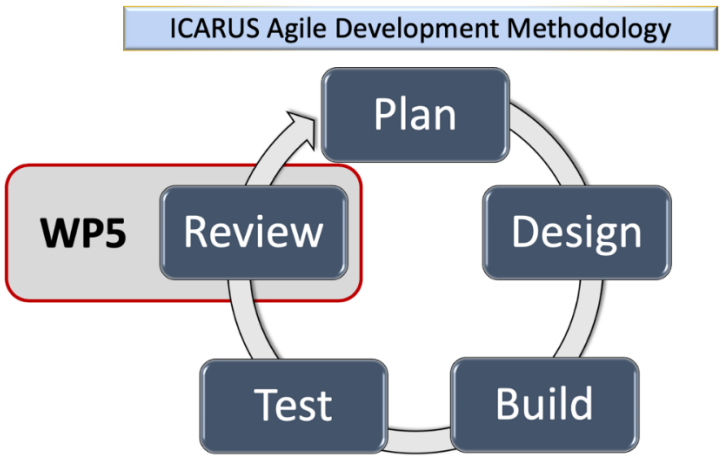

Figure 2-2: Positioning in the ICARUS Agile development methodology

However, due to the nature of the ICARUS project and the development activities that are performed collaboratively by multiple partners of the project, a set of structured activities must be set in order to ensure the correct and valuable execution of the ICARUS evaluation framework. Within the context of D5.2 "Demonstrators Execution Scenarios and Readiness Documentation", a set of scenarios will be formulated for each demonstrator of the ICARUS project and will be executed by the corresponding set of end-users defined in these scenarios. These scenarios anticipate a number of test cases that will be utilised as part of the evaluation framework. Through the ICARUS evaluation framework the demonstrators' phases will be monitored and aligned in order to provide structured and actionable feedback to development team of the ICARUS platform. The selected end users to be involved in the ICARUS demonstration and evaluation activities in WP5 need to cover all the related stakeholder groups acting as data providers, data consumers, application providers or application consumers, since each group has different needs that need to be considered and fulfilled.

The ICARUS evaluation framework intends to cover both verification, i.e. the discovery and elimination of malfunctions and possible security issues, and validation, meaning the capability to fulfil the stakeholders' needs. The achievement of these functions requires the involvement of the respective users in the multiple demonstration iterations that are performed. The project's demonstrators will be engaged in the evaluation and feedback loops since the beginning of the development activities, and will be exposed to preliminary versions of the platform services and APIs, allowing the developers to improve their components and the integrated platform, according to the continuous feedback that they will provide. Hence, in order to ensure the correct operation of the feedback cycle, the evaluation framework will become a structural aspect of the agile development lifecycle, where implementation and evaluation run in parallel. The benefits of this approach, is that the speed, quality and efficiency of the development process will be increased with the involvement and

collaboration of all interested parties in the process. At the end of each iteration, the results are evaluated by the demonstrators providing instant feedback that will drive the necessary adjustment and refinements in the next iteration.

Apart from the demonstrator scenarios that will be formulated and executed at technical level, the actual business impact will be monitored and quantified with concrete business key performance indicators. An additional set of specific technical performance indicators will be also defined and constantly measured in order to verify all the platform's aspects and features and provide insights on the objectives' achievement. Another relevant aspect that needs to be covered is the respect of the GDPR regulations whenever personal data handling is envisioned. Therefore, it is necessary that the evaluation activities also embrace high-level security performance indicators, to guarantee privacy respect and the minimization of risks of personal and private data leakages.

While the evaluation of the ICARUS platform will be constantly provided, through the direct inclusion of the evaluation framework in the development lifecycle as described above, the overall validation and evaluation phase will be documented into three phases as dictated by the ICARUS Description of Action:

- The first validation phase has started on M12 and will be completed on M24 of the project. The results of this phase will be reported in deliverable D5.3 "Demonstrators Operation Evaluation and Feedback–v1.00" in which both the first implementation phase of the demonstrators and the ICARUS Platform - Release 1.00 will be evaluated.
- The second validation phase will be start on M25 and will last till M30 of the project. The results of the second phase will be reported in deliverable D5.4 "Demonstrators Operation Evaluation and Feedback–v1.50" and will include the evaluation of both the second implementation phase of the demonstrators, as well as the evaluation of the ICARUS Platform - Release 1.50.
- The third and final evaluation phase will start on M31 and will be completed on M36. The results of the third phase will be reported in deliverable D5.5 "Demonstrators Operation Evaluation and Feedback–v2.00" and will include the evaluation of the third and final implementation phase of the demonstrators, as well as the evaluation of the final release of the platform, namely the ICARUS Platform - Release 2.00.

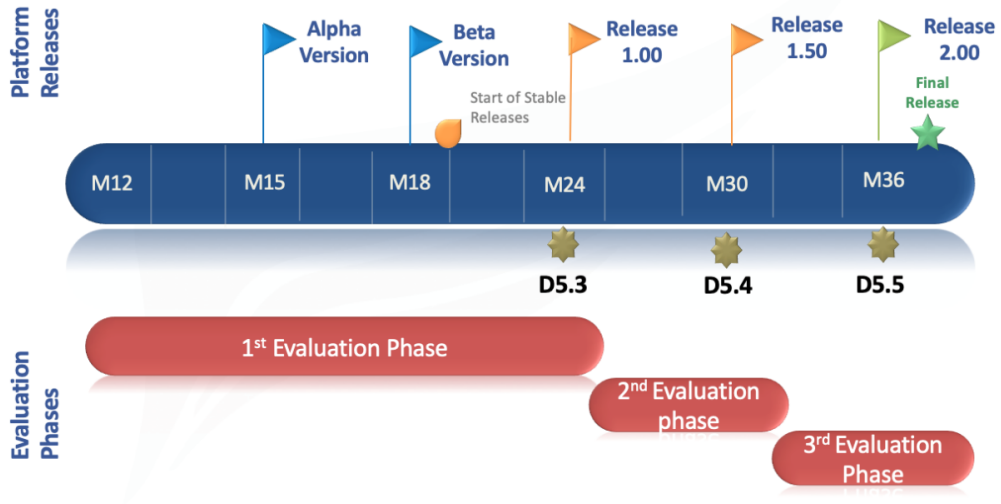

Figure 2-3: ICARUS Evaluation phases plan

3 Technical Perspectives of the ICARUS Evaluation Framework

As described in the previous section, the ICARUS evaluation framework incorporates the technical verification and validation aspects as perceived by the ICARUS demonstrators and the broader aviation stakeholders, in its holistic evaluation of the ICARUS platform. Hence, in the design of the technical evaluation framework, multiple different evaluation and validation factors need to be considered in order to ensure the proper performance evaluation and the identification of any deviation from the requirements and objectives of all identified ICARUS stakeholders.

The purpose of the evaluation framework from the technical perspective is to build the ICARUS platform in the right way, defining particular technical aspects and parameters that will be evaluated by the relevant stakeholders in order to assess the particular goal achievement. In this context, the related steps of the ICARUS technical evaluation framework are: Product Validation (P-5) and Technical Validation (T-1) as defined in section 2 (figure 2-1).

In order to guarantee the technical excellence, stability and reliability of the ICARUS platform, the following aspects are taken into consideration in the design of the ICARUS "technical" evaluation framework:

- The User profiles that are mapped the identified stakeholder groups.
- The ICARUS MVP, as defined in deliverables D1.2 and D1.3, as well as the detailed user and technical requirements as elaborated in D3.1.
- The holistic evaluation approach defined in section 2, that dictates the evaluation of the technical features, the users' experience and the quality of data assets.

The ICARUS integrated platform is a complex technological artefact consisting of a large number of individual components (as explained in the ICARUS Deliverables D3.1, D3.2 and D3.3). As in any complex software system, in order to assure its proper evaluation, a set of models can be retrieved, adapted and utilised from the literature. In these evaluation models, a set of key factors or characteristics are defined, each one having a set of corresponding sub-factors or sub-characteristics. However, as the models define the characteristics or sub-characteristics in a generic manner, the proper evaluation is performed per case, by assigning a set of metrics to each one of the sub-characteristics.

3.1 Background

In 2011, the ISO/IEC 25010 "Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models" was released and proposed a set of models that better address the evaluation of the software quality.

Moreover, this model is overcoming the barriers and limitations of other models, such as the model proposed by ISO/IEC 9126, that are either too generic or can be applied on specific cases or domains.

The ISO/IEC 25010 standard practically expands the key characteristics from previous models to eight (8) main characteristics with a widest range of thirty-one (31) sub-characteristics that capture all the fundamental aspects of a software evaluation.

The adoption of such a well-known standard (as ISO/IEC 25010:2011) is considered as the most suitable choice for the needs of the ICARUS evaluation framework, since it allows for the coverage of the key areas of the technical features and the end user's experience. In particular, the ISO/IEC 25010:2011 includes, as stated in its description:

- *A product quality model composed of eight characteristics (which are further subdivided into sub-characteristics) that relate to static properties of software and dynamic properties of the computer system. The model is applicable to both computer systems and software products.*
- *A quality in use model composed of five characteristics (some of which are further subdivided into sub-characteristics) that relate to the outcome of interaction when a product is used in a particular context of use. This system model is applicable to the complete human-computer system, including both computer systems in use and software products in use.*

As evident from the previous discussion, both quality models were considered as appropriate choices and are included in the ICARUS evaluation framework.

In order to evaluate also the aspects related to the quality of data, the ICARUS evaluation framework will complement the previous models with the adoption of a further ISO standard, namely the ISO/IEC 25012:2008 "Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model". In essence, both standards are adapted to ICARUS and combined into a holistic evaluation model based on a specific set of Key Performance Indicators (KPIs) to be measured and assessed either by the ICARUS development team and / or the end users. The two standards allowed the definition of the relevant characteristics and sub-characteristics while for each of them, additional ad-hoc metrics were defined. Their analysis allows a thorough understanding of the achievement level of the ICARUS platform at any time.

3.1.1 Technical evaluation - Product quality model

As stated above, the product quality model, as proposed by ISO/IEC 25010:2011, relates to measurable static properties of a software product and to dynamic properties of the computer

system where the software is running. The product quality model aims at evaluating the value that is provided by the system or software to the different stakeholders' needs, categorised in various characteristics and sub-characteristics.

This model is based on a structured set of characteristics[2] (Figure 3-1):

- **Functional suitability** - The degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions.
- **Performance efficiency** - The performance relative to the amount of resources used under stated conditions.
- **Compatibility** - The degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions, while sharing the same hardware or software environment.

Figure 3-1: ISO/IEC 25010:2011 - Product Quality Model

- **Usability**[3] - The degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.
- **Reliability** - The degree to which a system, product or component performs specified functions under specified conditions for a specified period of time.
- **Security** - The degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization.
- **Maintainability** - The degree of effectiveness and efficiency with which a product or system can be modified to be improved, corrected or adapted to changes in environment, and in requirements.

[2] Definitions of characteristics and subcharacteristics are taken from the ISO25000 standards family website: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010

[3] Note: the usability aspects are also considered in section 4.

- **Portability** - The degree of effectiveness and efficiency with which a system, product or component can be transferred from one hardware, software or other operational or usage environment to another.

Table 3-1 shows in detail the sub-characteristics of each category and indicates their suitability to the ICARUS platform.

Table 3-1: Technical characteristics, sub-characteristics and relevance to the ICARUS platform

Sub-characteristics	Definition of sub-characteristic	Suitability to ICARUS platform
Functional suitability		
Functional completeness	Degree to which the set of functions covers all the specified tasks and user objectives.	High
Functional correctness	Degree to which a product or system provides the correct results with the needed degree of precision.	High
Functional appropriateness	Degree to which the functions facilitate the accomplishment of specified tasks and objectives.	High
Performance efficiency		
Time behaviour	Degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements.	High
Resource utilisation	Degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements.	High
Capacity	Degree to which the maximum limits of a product or system parameter meet requirements.	High
Compatibility		
Co-existence	Degree to which a product can perform its required functions efficiently while sharing a common environment and resources with other products, without detrimental impact on any other product.	Medium
Interoperability	Degree to which two or more systems, products or components can exchange information and use the information that has been exchanged.	High
Usability		
Appropriateness recognisability	Degree to which users can recognize whether a product or system is appropriate for their needs.	High
Learnability	Degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use.	High
Operability	Degree to which a product or system has attributes that make it easy to operate and control.	High
User error protection	Degree to which a system protects users against making errors.	High
User interface aesthetics	Degree to which a user interface enables pleasing and satisfying interaction for the user.	High
Accessibility	Degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.	Medium
Reliability		
Maturity	Degree to which a system, product or component meets needs for reliability under normal operation.	High

Sub-characteristics	Definition of sub-characteristic	Suitability to ICARUS platform
Availability	Degree to which a system, product or component is operational and accessible when required for use.	High
Fault tolerance	Degree to which a system, product or component operates as intended despite the presence of hardware or software faults.	High
Recoverability	Degree to which, in the event of an interruption or a failure, a product or system can recover the data directly affected and re-establish the desired state of the system.	High
Security		
Confidentiality	Degree to which a product or system ensures that data are accessible only to those authorized to have access.	High
Integrity	Degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data.	High
Non-repudiation	Degree to which actions or events can be proven to have taken place, so that the events or actions cannot be repudiated later.	High
Accountability	Degree to which the actions of an entity can be traced uniquely to the entity.	High
Authenticity	Degree to which the identity of a subject or resource can be proved to be the one claimed.	High
Maintainability		
Modularity	Degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components.	High
Reusability	Degree to which an asset can be used in more than one system, or in building other assets.	Medium
Analysability	Degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, or to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified.	Medium
Modifiability	Degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality.	Medium
Testability	Degree of effectiveness and efficiency with which test criteria can be established for a system, product or component and tests can be performed to determine whether those criteria have been met.	Medium
Portability		
Adaptability	Degree to which a product or system can effectively and efficiently be adapted for different or evolving hardware, software or other operational or usage environments.	High
Installability	Degree of effectiveness and efficiency with which a product or system can be successfully installed and/or uninstalled in a specified environment.	Low
Replaceability	Degree to which a product can replace another specified software product for the same purpose in the same environment.	High

3.1.2 Information Quality Evaluation

In addition to the Product Quality and Quality in Use models that were adopted from ISO/IEC 25010:2011 as described in the previous sections, the Data Quality model as defined in the ISO/IEC 25012 standard is also adopted in the ICARUS evaluation framework for the data

quality evaluation dimension. In this model, a structured set of characteristics is defined organised in fifteen categories considering both certain intrinsic properties of data, and system-dependent attributes.

The term "inherent data quality" refers to the degree to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs when data is used under specified conditions[4]. In particular, the following aspects are considered:

- Data domain values and possible restrictions (e.g. business rules governing the quality required for the characteristic in a given application);
- Relationships of data values (e.g. for consistency purposes);
- Metadata (e.g. regarding alignment with the ICARUS metadata schema defined in D2.1 and D2.3).

System dependent data quality refers instead to the degree to which the data quality is reached and preserved within a computer system when data is used under specified conditions.

For the evaluation of the ICARUS data assets, the analysis of (Rafique et al, 2012), as well as the analysis of the different metadata standards in D2.1 and the final ICARUS metadata schema in D2.3, are taken into account.

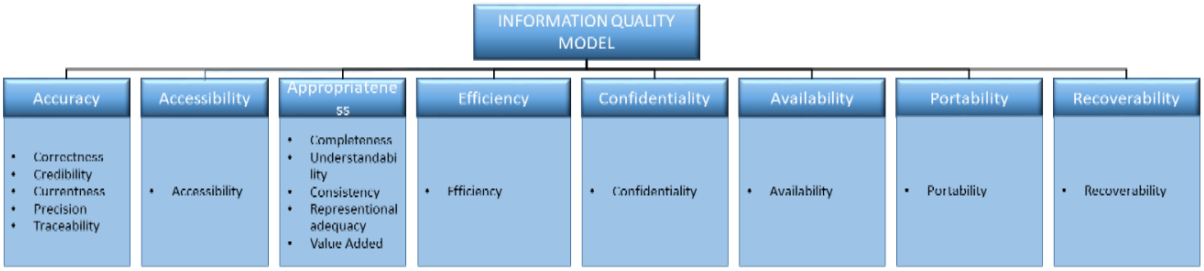

Figure 3-2: ISO/IEC 25012:2008 - Information Quality Model

The structured set of characteristics includes therefore[5] (Figure 3-3):

- **Accuracy** - The degree to which the delivered information is correct, precise, credible, traceable and current in a specific context of use.
- **Accessibility** - The degree to which information can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.

[4] https://iso25000.com/index.php/en/iso-25000-standards/iso-25012
[5] Definitions of characteristics and subcharacteristics are taken from the ISO25000 standards family website: https://iso25000.com/index.php/en/iso-25000-standards/iso-25012?limit=5&start=5

- **Appropriateness** - The degree to which the information is complete, consistent, understandable, represented adequately and have added value for the user, considering the specified user tasks and goals.
- **Efficiency** - The degree to which information has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.
- **Confidentiality** - The degree to which information has attributes that ensure that it is only accessible and interpretable by authorised users in a specific context of use.
- **Availability** - The degree to which information has attributes that enable it to be retrieved by authorised users and/or applications in a specific context of use.
- **Portability** - The degree to which information has attributes that enable it to be replaced or moved from one system to another preserving the existing quality in a specific context of use
- **Recoverability** - The degree to which information has attributes that enable it to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use.

Table 3-2 shows in detail the sub-characteristics of each category and indicates their suitability to the ICARUS platform**.**

Table 3-2: Information Quality model characteristics, sub-characteristics and relevance to the ICARUS platform

Sub-characteristics	Definition of sub-characteristic	Suitability to ICARUS platform
Information Accuracy		
Correctness	Degree to which data has reliable information	High – "Accuracy" in the ICARUS Metadata Schema
Credibility	Degree to which data has attributes that are regarded as true and believable by users in a specific context of use. Credibility includes the concept of authenticity (the truthfulness of origins, attributions, commitments).	High – "Veracity" in the ICARUS Metadata Schema
Currentness	Degree to which data has attributes that are of the right age in a specific context of use.	High – "Timeliness" in the ICARUS Metadata Schema
Precision	Degree to which data has attributes that are exact or that provide discrimination in a specific context of use.	Medium
Traceability	Degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use.	Medium – possible at dataset level only
Information Accessibility		
Accessibility	Degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.	YES
Information Appropriateness		
Completeness	Degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific	High – "Completeness" in

Sub-characteristics	Definition of sub-characteristic	Suitability to ICARUS platform
	context of use.	the ICARUS Metadata Schema
Understandability	Degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use.	High
Consistency	Degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use. It can be either or both among data regarding one entity and across similar data for comparable entities.	High
Representational Adequacy	Degree to which data or information is represented in a concise, flexible and organised way with due relevancy to the users' goals to help them to achieve their specified goals.	Medium
Value Added	Degree to which data or information are beneficial and provide advantages from their use.	High
Efficiency		
Efficiency	Degree to which data has attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.	Medium
Confidentiality		
Confidentiality	Degree to which data has attributes that ensure that it is only accessible and interpretable by authorized users in a specific context of use.	High
Availability		
Availability	Degree to which data has attributes that enable it to be retrieved by authorized users and/or applications in a specific context of use.	High
Portability		
Portability	Degree to which data has attributes that enable it to be installed, replaced or moved from one system to another preserving the existing quality in a specific context of use.	Medium – depending on policies and contracts terms
Recoverability		
Recoverability	Degree to which data has attributes that enable it to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use.	Medium

3.2 Technical evaluation criteria, targets and evaluation plan for the ICARUS platform

While the standards ISO/IEC 25010:2011 and ISO/IEC 25012:2008 specify the evaluation criteria, the specific list of indicators to measure them is left to the adopters. For this reason, a set of specific KPIs that are tailored to the needs of ICARUS project and the nature of the ICARUS platform were devised and discussed. These indicators will contribute in the technical evaluation of the ICARUS platform. Some indicators however are marked as optional since their measurement might not be feasible in the context of the ICARUS project due to the nature of the platform or might not provide added value in the evaluation process. Finally, it has to be noted that metrics relative to user experience and data quality need to be measured

in a qualitative manner, administering questionnaires to a set of users, collecting feedback with a typical five-level Likert scale.

Table 3-3 lists the defined technical evaluation KPIs related to the product quality model. Such KPIs are related to the Product Validation step and are measured by the ICARUS development team.

Table 3-3: Quantitative evaluation Metrics selected for the ICARUS platform

Sub-characteristics	Metric	Definition	Mandatory (Yes/N0)	Threshold Value
Functional suitability				
Functional completeness	Percentage of User Stories completed, covering the functional requirements	[Completed User Stories] / [Iteration Cycle of User Stories] * 100%	Yes	100%
Functional correctness	Percentage of User Stories completed successfully	[Completed User Stories without bugs] / [Iteration Cycle of User Stories] * 100%	Yes	>90%
Functional appropriateness	Straightforward task accomplishment	Are tasks completed without the use of unnecessary steps? [Yes/No]	No	Yes
Performance efficiency				
Time behaviour	Average latency required for the accomplishment of specific (sub-)tasks	[Total response time] / [Number of requests]	Yes	To be set in the ICARUS release 1.00
	Average Throughput during normal platform utilisation	[Total Number of Kilobytes] / [Total Time of Operation]	Yes	To be set in the ICARUS release 1.00
Resource utilisation	Mean CPU Utilisation	[Σ[%CPU utilisation probes]] / [Number of probes]	Yes	<60%
	Mean memory usage	[Σ[RAM Megabytes used in each probe]] / [Number of probes]	Yes	<60%
	Maximum memory usage	Maximum % RAM Memory utilisation recorded	Yes	<90%
	Maximum processing power used	Maximum % CPU utilisation recorded	Yes	<90%
Capacity	Maximum database/ storage size	Total number of Kilobytes of files	Yes	>10 TB
Compatibility				
Co-existence	Ability to Co-Exist	Can ICARUS platform operate in shared	No	Yes

Sub-characteristics	Metric	Definition	Mandatory (Yes/N0)	Threshold Value
		environment? [Yes/No]		
Interoperability	Exposure of APIs	Ability to expose information through well-defined APIs	Yes	Yes
	Ability to store different datasets	Can ICARUS platform store datasets of different formats (csv, xml, json, etc.)?	Yes	Yes
	Ability to handle different datasets	Can ICARUS platform process datasets of different formats (i.e. csv)?	Yes	Yes
	Ability to deliver different datasets	Can a user "download" from the ICARUS platform process datasets of different formats (csv, xml, json, etc.)?	Yes	Yes
Operability				
Technical Learnability	% Coverage of features with learning documents	[Unique number of help documents mentioning a feature] / [Total number of features available] * 100%	Yes	100%
Ease of Use	Dashboard availability	Is there a dashboard available with easy navigation? [Yes/No/Partially]	Yes	Yes
User error protection	% Coverage of input fields with error protection methods	[Number of error protected fields] / [Total number of critical input fields] * 100%	Yes	100%
Technical Accessibility	WCAG 2.0 Conformance Level	[None/ A/ AA/ AAA]	Yes	A
	Cross-Platform Accessibility	Is ICARUS Platform accessible and operational through different platforms (e.g. Windows, Linux, MacOS)? [Yes/No/Partially]	Yes	Yes
	Cross-Browser Accessibility	Is ICARUS Platform accessible and	Yes	Yes

Sub-characteristics	Metric	Definition	Mandatory (Yes/N0)	Threshold Value
		operational through different browsers (e.g. Chrome/ Firefox / Edge)? [Yes/No/Partially]		
	Cross-Device Accessibility	Is ICARUS Platform accessible and operational through different devices (i.e. PC/ Laptop)? [Yes/No/Partially]	Yes	Yes
Reliability				
Maturity	Simultaneous requests	Maximum number of simultaneous requests	Yes	>60
Availability	% Monthly availability	[1-[Downtime in minutes] / [Total month minutes]] * 100%	Yes	>90%
	Success rate	[Number of correctly completed requests] / [Total number of requests]	Yes	>95%
Fault tolerance	% of identified Software problems affecting the platform	[Critical Software Issues] / [Total number of Software faults detected] * 100%	Yes	<10%
	% of identified Hardware problems affecting the platform	[Critical Hardware Issues] / [Total number of Hardware faults detected] * 100%	Yes	<10%
Recoverability	Mean recovery time from Software problems	[Total recovering time from Software issues] / [Total number of Software issues in need of recovery]	Yes	<1h
	Mean recovery time from Hardware problems	[Total recovering time from Hardware issues] / [Total number of Hardware issues in need of recovery]	Yes	<24h
Security				
Confidentiality	Unauthorised access to information	Number of recorded incidents	Yes	0 (None)
Integrity	Unauthorised tampering with information	Number of recorded incidents	Yes	0 (None)
Non-repudiation	Successful identity	Number of	Yes	0 (None)

Sub-characteristics	Metric	Definition	Mandatory (Yes/N0)	Threshold Value
	modification	recorded incidents		
	System failure to properly identify / authenticate user	Number of recorded incidents	Yes	0 (None)
	Level of User authenticity	Can you identify that a subject (organization or user) is the one it claims to be? [Yes/ No/ Partially]	Yes	Yes
Accountability	User actions traceability	Are usernames included in each activity log entry uniquely? [Yes/No]	Yes	Yes
	Percentage of actions logged	[Actions logged] / [Total number of system operations]	Yes	100%
Maintainability				
Modularity	% of modularity	[Number of components that can operate individually] / [Total number of components] * 100%	Yes	>70%
Analysability	Level of analysability	Can the changes in the performance of the ICARUS platform be efficiently evaluated after each upgrade? [Yes/No]	No	Yes
Modifiability	% of update effectiveness	[Number of updates performed without operational issues] / [Total number of updates] * 100%	Yes	>90%
Testability	Level of testing	Are tests able to probe the behaviour of the ICARUS platform? [Yes/No]	Yes	Yes
Portability				
Adaptability	Mean number of errors per hardware change/ upgrade	[Total number of errors recorded] / [Total number of hardware changes]	Yes	<1
	Mean number of errors per software change/ update	[Total number of errors recorded] /	Yes	<1

Sub-characteristics	Metric	Definition	Mandatory (Yes/N0)	Threshold Value
		[Total number of software changes]		

Since the eventually ICARUS platform adoption is tightly related to the data assets it shares, the data quality aspects as part of the product validation activities cannot be disregarded. To this end, only the confirmed buyers of a data asset (that have an active data contract in the ICARUS platform) may provide a further data asset evaluation on a voluntary basis. Table 3-4 deals with an indicative set of Information quality model metrics and shall be finalized in the forthcoming development and evaluation activities considering the final ICARUS metadata schema (defined in D2.3).

Table 3-4: Data qualitative evaluation Metrics selected for the ICARUS platform

Sub- characteristics	Metric	Definition	Mandatory (Yes/No)
Information Accuracy			
Correctness	Error-free data	To what degree do you find the data provided reliable and errorless? [1-5]	Yes
Credibility	Acceptable ranges	To what degree do you find the data provided containing values within the range of known or acceptable values? [1-5]	Yes
	Credibility of the data source	To what degree do you find the originating source of the provided data (experts or organisation of a country, field, or industry) credible? [1-5]	Yes
Currentness	Update of the data source	To what degree do you believe that the data are up-to-date or regularly updated? [1- 5]	Yes
Traceability	Datasource verification	To what degree do you find that the originating source of the provided data is verifiable? [1-5]	No
Information Accessibility			
Accessibility	Data access	To what degree do you find to which the provided data can be accessed from people with specific disabilities? [1-5]	Yes
	Data policy	To what degree do you find that access control should be enforced to the provided data depending on the assigned access level (private, public)? [1-5]	No
Information Appropriateness			
Completeness	Context of use fitness	To what degree do you find that the provided datasets contain information relevant to their	Yes

Sub- characteristics	Metric	Definition	Mandatory (Yes/No)
		described context of use? [1-5]	
	Missing information expected	To what degree do you find that the provided datasets are complete and with no missing entries or missing values? [1-5]	No
Understandability	Definition/ Documentation	To what degree do you find that the provided datasets are accompanied by appropriate metadata? [1-5]	Yes
Consistency	Duplicate instances	To what degree do you find that the provided datasets are free of repeated instances of the same property that is not allowed [1-5]	No
Representational Adequacy	Adequate visualization features	To what degree do you find that the ICARUS platform offers the suitable visualisation options for the provided datasets depending on the intended context of use? [1-5]	Yes
	Normative data definition	To what degree do you find that the provided data (content, format, etc.) are clear and understandable? [1-5]	Yes
Value Added	Benefit for the user	To what degree do you find that the provided data are beneficial and provide added value to the user? [1-5]	Yes

3.3 Definition and Assessment of the ICARUS Test Cases

In addition to the definition of the KPIs, it is necessary to identify a set of suitable test cases, which are structured scripts of actions to be performed initially by the testers if the ICARUS development team, and subsequently by real end-users involved in the demonstrators. Such test cases will map realistic users' journeys on the platform, and therefore will provide insights both on platform usability and robustness.

For homogeneity purposes, all the test cases to be executed on the ICARUS platform are defined following the template presented in Table 3-5.

Table 3-5: ICARUS Test Cases Template

Test Case	*The title of the test case*
Actors	*The actors involved in the specific test case along with their role (i.e. data provider, data consumer).*
Importance	*High / Medium / Low*
Pre-conditions	*A brief reference to any pre-conditions that are necessary to be met prior to initiating this test case.*
Post-conditions	*A brief reference to any post-conditions that come into effect after this test case is completed.*
Workflow	*Either as sequence diagram or as steps (1-x)*
Alternative Flows	*Reference to the steps to which an alternative flow is provided.*

Related ICARUS Phase	*From the methodology in D1.2*
Success Indication	*The condition that suggests that the test cases was successfully completed.*
Failure Indication	*The indication that suggests that the test cases was completed with failure.*
Notes	*Any additional notes to highlight on the specific test case.*

In alignment with the specific test cases template, the initial list of test cases for the ICARUS demonstrators is presented in the ICARUS Deliverable D5.2 "Demonstrators Execution Scenarios and Readiness Documentation". In total, 37 test cases have been defined spanning all demonstrator scenarios and the different roles, i.e. Data provider, Data consumer, Application provider and Application consumer, with more emphasis on the data provider / consumer role for the first demonstration activities.

In addition to the qualitative feedback collected during the execution of the user stories, it is important to collect specific data on the outcome, namely:

- If the use story has been completed by the test participant with or without specific help;
- If the use story could NOT be completed by the test participant even after receiving specific help;
- If the user story could be completed, but with bugs. The report should include which and how many bugs were identified, describing also the context that triggered the bugs;
- If the user story could NOT be completed, due to bugs. A detailed report on the blocking bug should be included;
- The usefulness of the user story perceived by the test participant, using a 5-steps Likert scale (1: "*not at all*" to 5: "*extremely useful*")
- The ease of use in completing the user story perceived by the test participant, using a 5-steps Likert scale (1: "*extremely hard*" to 5: "*extremely easy*")

Table 3-6 describes a template for collecting the above-mentioned test cases evaluation information.

Table 3-6: Template for test cases evaluation

Release Number	Test Case ID	Test Case Title	Completed without Help	Completed with Help	Completed with bugs	Not completed due to bug	Not completed after help	Useful (1-5)	Easy (1-5)
Stakeholder Group Name									
…	Code#	…	…	…	…	…	…	…	…
…	Code#	…	…	…	…	…	…	…	…

It needs to be highlighted that such test cases intend to fully address the evaluation activities of the ICARUS platform on its beta and first official releases. As the project development activities advance, the test cases are also expected to evolve in the forthcoming evaluation phases to test any additional functionality that was not originally foreseen. The feedback collected will be also used to update the requirements, possibly including also new features, but also the test cases themselves. This allows a higher flexibility level of the development workflow, and a better alignment with the demonstrators' needs.

Finally, it needs to be noted that technical testing, bug reporting and analysis will be typically performed during all the development phases, not involving directly the demonstrators' stakeholders. Table 3-7 suggests a list of tests to be performed in this phase, with the evaluation framework characteristics they are related to. These tests will be developed and performed in the context of task T4.4 "*Technical Verification and Integration Testing*".

Table 3-7: Test cases types for the ICARUS development phase

Sub-characteristic	Test cases type	Description
Functional suitability	User Stories Check	The user stories will be analysed to check if they remain functional for all the considered stakeholders
	Integration tests	Automated integration tests will be executed in the platform prior to each release
Performance efficiency	System analytics	Analytics of the system's operation will be studied
	Stress tests	Extreme cases will be executed to ascertain boundaries in operability
Compatibility	Compatibility tests	KPIs will be measured during and after the system's implementation
Operability	Actual usage tests	Feedback from actual users will be collected as defined in section 4
	Concept tests	Feedback about the concept will be collected, after dissemination
	Unit tests	Automated unit tests will be executed in the platform
	System audit	The operation will be checked through a thorough walk through the system
	Aesthetics tests	Different devices will be tested to check consistencies and accessibility
Reliability	System analytics	Analytics of the system's operation will be studied
Security	Security Tests	Security will be tested both at system and at the software level.
Maintainability	System audit	The behavior will be evaluated performing a thorough walk through the system
Portability	Adaptability tests	The identified KPIs will be measured during and after system updates/upgrades

4 Business Perspectives of the ICARUS Evaluation Framework

In order to address the business validation aspects of the ICARUS evaluation framework and properly record all feedback of the end users whose organizations obtain legitimate access to the ICARUS platform (since it will operate on a know-your-customer basis), a number of methods dedicated to business performance evaluation proposed by researchers and practitioners in the literature have been meticulously studied. Usage of the ICARUS platform from the aviation data value chain stakeholders and the ICARUS demonstrators will be assessed on the basis of structured and unstructured feedback in a predefined set of Performance Indicators (PIs) as also explained in section 3. The ICARUS demonstrators will also quantify their expected benefits in a limited set of business performance indicators that are defined depending on their envisaged demonstration scenarios. In this context, the related steps of the ICARUS "business" evaluation framework are: Product Validation (P-5) and Business Validation (T-2) as defined in section 2 (figure 2-1).

4.1 Background

Over the years, a number of frameworks that evaluate usage and impact to software users have emerged and can be classified to two categories:

- *Usage evaluation standards* that include the "Quality in Use" model of the ISO/IEC 25010:2011 standard that was analyzed in section 3.1.2, as well as an extension to the Technology Acceptance Model (TAM2) (Venkatesh & Davis, 2000), the UTAUT model (Unified Theory of Acceptance and Use of Technology) (Venkatesh et al, 2003), and the User Experience Questionnaire (UEQ).
- *Industrial business performance measurement frameworks*, such as BSC (Balanced Score Card) (Kaplan & Norton, 2000) and ECOGRAI (Doumeingts et al, 1995). BSC is oriented to the determination of Key Performance Indicators (KPIs) at the strategic level of the enterprise, while ECOGRAI can be used at all levels of the decisional structure of a system: Strategic, Tactical and Operational and suggests a limited number of indicators.

4.1.1 Usage Evaluation - Quality in use model

The second model proposed in the ISO/IEC 25010:2011 standard and utilised in the ICARUS evaluation framework is the "Quality in Use" model that considers the user's point of view to measure the perception of the quality of the system. Hence, in this model, the outcomes of the interaction of a user with the corresponding system and the deployed software are effectively measured.

This model is based on a structured set of characteristics[6]:

- **Effectiveness** - The accuracy and completeness with which users achieve specified goals.
- **Efficiency** - The resources expended in relation to the accuracy and completeness with which users achieve goals.
- **Satisfaction** - The degree to which users are satisfied with the experience of using a product in a specified context of use.
- **Freedom from risk** - The degree to which a product or system mitigates the potential risk to economic status, human life, health, or the environment.
- **Context coverage** - The degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified.

Table 4-1 shows in detail the sub-characteristics of each category and indicates their suitability to the ICARUS platform.

Table 4-1: Quality in use model characteristics, sub-characteristics and relevance to the ICARUS platform

Sub-characteristics	Definition of sub-characteristic	Suitability to ICARUS platform
Effectiveness		
Effectiveness	Degree of accuracy and completeness with which users achieve specified goals when using the system.	High
Efficiency		
Efficiency	Degree to which the users find that the software is efficiently covering its intended purpose.	High
Satisfaction		
Usefulness	Degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use.	High
Trust	Degree to which a user or other stakeholder feel that they can trust the system and have confidence that a product or system will behave as intended.	High
Pleasure	Degree to which a user finds the software's functions a pleasure to use (emotionally).	Medium
Comfort	The degree to which users think that the system provides the comforts needed (physically)	Medium
Freedom from risk		
Economic risk mitigation	Degree to which a product or system mitigates the potential risk to financial status, efficient operation, commercial property, reputation or other resources in the intended contexts of use.	High
Health and Safety risk mitigation	Degree to which a product or system mitigates the potential risk to people in the intended contexts of use.	Low
Environmental risk mitigation	Degree to which a product or system mitigates the potential risk to property or the environment in the intended contexts of use.	Low
Context coverage		

[6] Definitions of characteristics and subcharacteristics are taken from the ISO25000 standards family website: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010

Sub-characteristics	Definition of sub-characteristic	Suitability to ICARUS platform
Context completeness	Degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in all the specified contexts of use	High
Flexibility	Degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in contexts beyond those initially specified in the requirements.	Low

4.1.2 ECOGRAI adaptation

As described in (Doumeingts et al, 1995; FITMAN, 2013a), ECOGRAI is a participative method to design and to implement Performance Measurement Systems for industrial organizations, independent from the application domains. ECOGRAI builds on the concepts of Decision Variables (DV) and Action Variables (AV) that represent the drivers of the objectives to be achieved, while it links the Performance Indicators (PIs) with the reaching of the objective by evaluating the effect of such variables. Taking into consideration the broader "enterprise" view and the objectives defined by the decision makers, ECOGRAI aims at defining a limited set of customized, "smart" indicators, that are specific, measurable, attainable, relevant and time-bound.

In brief, as defined by (Doumeingts et al, 1995), ECOGRAI is designed on the basis of six phases including: (a) modelling of the system to be evaluated and determination of the decision centres where specific Performance Indicators need to be defined; (b) identification of the objectives at the enterprise level and per decision centre; (c) identification of the drivers (named Decision Variable (DV) or Action Variable (AV)) and of possible conflicts; (d) identification of the Performance Indicators and performing an internal coherence check; (e) clear definition of each performance indicator, containing: its identification (name, decision centre responsible for it, horizon, periodicity), the related objectives and drivers, any possible negative effects, the data required for its implementation; (f) integration of the performance indicators into the entreprise information system.

Taking into consideration a simplified version of ECOGRAI with only three phases (as proposed in (FITMAN, 2013b)) that has been successfully applied in different projects, ICARUS intends to follow a similar mentality regarding how the demonstrators' key performance indicators are defined, without following all steps in detail, as explained in section 4.3.

4.2 Business evaluation criteria, targets and evaluation plan for the ICARUS platform

Business Validation depends mostly on meeting criteria set by users, which will help them to identify and measure the impact and the usefulness of the ICARUS platform in their everyday operations. For each demonstrator and "aviation data value chain stakeholder, a different set

of documents will be produced, to gather the questions asked in a standardized form and capture the experience gained from the ICARUS platform.

Table 4-2 presents the quantitative and qualitative evaluation metrics which correspond to the evaluation of the ICARUS platform operation phase for Product Validation (P-5) purposes. In general, many of the key performance indicators (KPIs) that are adopted refer to the Quality in Use model (presented in section 4.1.1) and the usability aspects of the ISO/IEC 25010:2011 (presented in section 3.1.1) and are measured in a qualitative manner, either by measuring AS-IS and TO-BE values, or in case of more qualitative answers, by using a 1-5 scale.

Table 4-2: Business Validation - Evaluation Metrics selected for the ICARUS Platform

Sub-characteristics	Metric	Definition	Mandatory (YES/NO)
Business Value			
Clarity	Clarity level	How clear was it for you what the ICARUS platform is about? [Scale 1 (Little) -5 (Very)]	Yes
Added Value	Added value level	How much added value do you feel that the ICARUS platform provides to your operations while using it? [Scale 1 (Low) -5 (High)]	Yes
Need Importance Level	Need importance level	How important is for you the need that the ICARUS platform covers for you? [Scale 1 (Little) -5 (Very)]	Yes
Need Coverage	Need coverage level	To which degree does the ICARUS platform covers your needs? [Scale 1 (Low) -5 (High)]	Yes
Innovation	Innovation level	How innovative do you find the idea of the ICARUS platform? [Scale 1 (Little) -5 (Very)]	Yes
Intention to use	Intention level	To what extent do you intend to use the ICARUS platform? [Scale 1 (Low) - 5 (High)]	No
Virality	Virality level	How probable is it for you to recommend the ICARUS platform to other aviation stakeholders? [Scale 1 (Low) -5 (High)]	Yes
Effectiveness			
Effectiveness	Effectiveness level	Is the ICARUS platform enabling you to accurately achieve your goals for data sharing and data analytics? [Scale 1 (Low) -5 (High)]	Yes
Efficiency			
Efficiency	Efficiency level	Is the ICARUS platform efficiently fulfilling its intended purpose? [Scale 1 (Low) -5 (High)]	Yes
Satisfaction			
Usefulness	Usefulness level	Do you find the ICARUS platform	Yes

Sub-characteristics	Metric	Definition	Mandatory (YES/NO)
		useful? [Scale 1 (Low) -5 (High)]	
Trust	Trust level	Do you trust the ICARUS platform and its provided functionalities? [Scale 1 (Low) -5 (High)]	Yes
Pleasure	Pleasure level	Does the ICARUS platform please you when you use it? [Scale 1 (Low) -5 (High)]	Yes
Comfort	Comfort level	Do you feel that the ICARUS platform provides a comfortable user interface and workflows? [Scale 1 (Low) -5 (High)]	Yes
Freedom from risk			
Economic damage risk	Level of economic damage risk	How sure are you that ICARUS protects you from exposing you on economic damage? [Scale 1 (Low) -5 (High)]	Yes
Privacy harm risk	Level of data privacy damage risk	How sure are you that ICARUS is on protecting your data privacy? [Scale 1 (Low) -5 (High)]	Yes
Usability			
Learnability	Learnability level	How easy it was for you to learn how to use the ICARUS platform? [Scale 1 (Not) -5 (Very)]	Yes
Flexibility	Flexibility level	How much do you believe the ICARUS platform can be used for other applications than the demonstrator ones? [Scale 1 (Low) -5 (High)]	Yes
Content Conformity	Content quality	How useful do you find the data and the applications found in the ICARUS platform in terms of quality? [Scale 1 (Little) -5 (Very)]	Yes
	Content quantity	How satisfied are you from the quantity of the data and the algorithms found in the ICARUS platform? [Scale 1 (Little)-5 (Very)]	Yes

In addition, it is important to identify the exact customer segments in the aviation industry that are really interested in the ICARUS platform. To this direction, a correlation of the users accessing the platform with their characteristics (gathered through dedicated questionnaires) can help the project identify what are the target segments of the platform. In alignment with TAM2 that puts emphasis on the user context during the evaluation process, the results should be examined for correlation based on the following user characteristics:

- **Experience**: how many years the user has been doing this job, or using similar "data" systems
- **Image**: how influential the user is considered in general and within his/her organization
- **Job Relevance**: the relevance of the user's job with the ICARUS platform
- **Output Quality**: how the user perceived the quality of the output in total, even if he is interested in the platform or not
- **Result Demonstrability**: if the user is willing to show to someone the results obtained through the ICARUS platform

Although running behavioural tests, like those proposed in the quality in use and data quality models, is complex, ICARUS shall adopt the following workflow for controlled tests:

- Users should enter the ICARUS platform with only a minimum of information about they can expect or do.
- Users are left free to explore the platform, trying to devise how to accomplish the requested test cases or fulfill a need they have (e.g. to find data or acquire data and run an analysis).
- Users should be given also goals and tasks, outside the specific user stories, to see how they behave (e.g.: "*If you need to create a new application, how would you do it?*").
- Every failure to complete a task should be documented, and analysed with specific follow-up questions.
- In case of malfunctions, bug reports should be collected with collaboration with the users (like screenshot capturing).
- Users demographics should be carefully collected (it is more efficient doing it at this phase, instead of the beginning);
- The users should fill in the generic questionnaire built based on the metrics of Table 4-2 (that may be further detailed depending on the type of the user, e.g. data provider or data consumer), collecting data from the user;
- The ICARUS consortium should schedule a short interview with each user to provide them with a more thorough description of the platform and its objectives;
- Collect from the user a more general feedback about the platform. The preceding discussion should enable deeper comments.

Such evaluation information will be duly gathered in the framework of Task T5.7 "Demonstrators Evaluation and Impact Assessment" and discussed in D5.3-D5.6.

4.3 Definition and Assessment of the ICARUS Demonstrators' Key Performance Indicators

In general, Business Validation (within step T-2 as defined in Figure 2-1) is the final and perhaps the most crucial step of the ICARUS evaluation framework since it answers the question whether the ICARUS platform offers sufficient added value to the demonstrators against their stated requirements. Business validation entails demonstrating that the platform developed provides tangible benefits and measurable impact to the demonstrators' operation through concrete and quantified performance indicators. In order to appropriately extrapolate such key performance indicators for each of the ICARUS demonstrators, a very simplified version of the ECOGRAI method is adopted in the context of the ICARUS evaluation framework, anticipating the following:

- *Phase I*: Description of the detailed demonstrator scenario in which the performance indicators will be defined. Such a description is provided in the ICARUS Deliverable D5.2 and includes: (a) the current challenges that the demonstrator partners currently face, (b) the business objectives within ICARUS, (c) the expected interactions with other aviation stakeholders, (d) the as-is processes compared to the expected to-be situation with the ICARUS platform, (e) test cases on how each demonstrator plans to technically validate the ICARUS platform, and (f) data availability and needs status.
- *Phase II:* Determination of the potential actions to reach the objectives defined per demonstrator.
- *Phase III:* Elaboration on the key performance indicators for each demonstrator scenario in order to measure the accomplishment of the business objectives. Such KPIs provide a way to compare the efficiency of the operation of the demonstrator before and after the use of the ICARUS platform.
- *Phase IV*: Calculation of the target and actual values of the KPIs by collecting required data before and after the application of the ICARUS platform during the demonstration activities (in WP5).

It needs to be noted that all required data for the KPIs are to be collected by the demonstrators in three steps: (a) Before starting using the ICARUS platform in order to set the baseline and target values; (b) A short period after the use of the ICARUS platform, in order to confirm that the values of the performance indicators are moving towards the expected directions (e.g. time decreasing, quality increasing etc.); and (c) At specifically defined time periods for each evaluation phase in ICARUS in order to iteratively measure whether specific targets for the business indicators values are met. Although the business validation of a demonstrator is considered as successful when the specified targets are met, it can be also considered as successful in the case that specific changes in the internal or

external environment of the demonstrator lead to the need of re-examining and amending the targeted values.

For homogeneity purposes, all the key performance indicators to be defined by the ICARUS demonstrators are defined following the template presented in Table 4-3.

Table 4-3: ICARUS Key Performance Indicators Template

Key Performance Indicator	*The title of the key performance indicator (KPI)*
Scenario	*The related demonstrator scenario for the specific KPI.*
Calculation Method	*The applicable formula for measuring the KPI (or simple metrics to be measured)*
AS-IS Value	*The value of the AS-IS Scenario. If not measurable, say why, or if confidential, provide estimation in terms of percentage in the TO-BE value.*
TO-BE Value	*Estimated target in the demonstrator scenario that will be executed in the ICARUS platform.*

In alignment with the specific KPIs template, the initial list of Key Performance Indicators per ICARUS demonstrator is presented in the ICARUS Deliverable D5.2 “Demonstrators Execution Scenarios and Readiness Documentation”. In total, 20 key performance indicators have been defined spanning all demonstrator scenarios and put into context the business impact and expected benefits by the demonstrators.

5 Conclusions and Next Steps

The scope of the ICARUS Deliverable D5.1 "ICARUS Demonstrators and Platform Evaluation Framework" is to describe the outcomes of Task 5.1 - "*Project Validation and Evaluation Framework Definition*", aiming at developing the evaluation framework from a technical and business viewpoint, defining ways to obtaining feedback from demonstrators' partners and end-users, and finally, providing a complete set of test cases.

The successful implementation of the proposed evaluation framework to be applied in the ICARUS platform and demonstrators relies on a 6-step procedure (that shall be applied in each iteration):

I. **Preparation** of the Product Validation (P-5), Technical Validation (T-1) and Business Validation (T-2) activities by adapting the methods to be employed according to the latest project's developments, selecting the stakeholders to be involved and creating the necessary material (e.g. test cases, questionnaires, training material, etc.).

II. **Initiation** of the Product Validation (P-5) and Technical Validation (T-1) activities by involving the appropriate "test" users (within and beyond the ICARUS consortium) and running the designed test cases.

III. **Communication** of the preliminary evaluation results to the platform developers to address any critical issues that were encountered by the users.

IV. **Completion** of the Product Validation (P-5) and Technical Validation (T-1) activities involving the necessary stakeholders through interviews and online questionnaires (as designed in step I) and calculating the related technical evaluation KPIs.

V. **Involvement** of key stakeholders in the Business Validation (T-2) activities, by collecting all necessary data to assess the business KPIs per demonstrator.

VI. **Interactive discussions** with the platform developers to discuss and assess the complete evaluation results and plan – to the extent it is feasible – the necessary updates of the ICARUS platform in an efficient manner.

It is important to note that such a validation and evaluation framework is instrumental to guide the evaluation activities, yet it is placed in the context of a living process that shall continue to evolve until the end of the final evaluation phase (M30) providing the required compass needle for the evaluation activities. The evaluation framework will start to be applied once the demonstrators reach an adequate development phase, allowing the demonstrators' end-users and the project's experts to provide feedback on their interaction with the ICARUS platform. Data will be collected according the guidelines of the evaluation framework to ensure the feedback is aligned to the standards and the templates defined. If the feedback shows that certain requirements are not fulfilled in accordance with the users'

needs, then the ICARUS development plan, considering all its aspects (individual platform features, integration plan, evaluation plan), might need a reassessment. Only when a cycle of evaluation is properly complete and its outcome has been properly assessed and acted upon, a new cycle can start as anticipated in the ICARUS evaluation framework.

Annex I: References

Doumeingts G., Clave F., Ducq Y. (1995) ECOGRAI — A method to design and to implement Performance Measurement Systems for industrial organizations — Concepts and application to the Maintenance function. In: Rolstadås A. (eds) Benchmarking — Theory and Practice. IFIP Advances in Information and Communication Technology. Springer, Boston, MA

FITMAN (2013a) D2.1 "FITMAN Verification & Validation Method and Criteria". August 2013.

FITMAN (2013b) D2.2 "FITMAN Verification & Validation Business and Technical Indicators Definition". October 2013

FITMAN Project (2018). Accessed on May 3rd, 2018 from: http://www.fiware4industry.com/

IEEE. "1012-2012 - IEEE Standard for System and Software Verification and Validation." 2012.

ISO/IEC 25010:2011, Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models

ISO/IEC 25012:2008, Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model

Kaplan, R.S. and D.P. Norton (2000) The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment, Harvard Business School Press

Lampathaki, F., Panopoulos, D., Kokkinakos, P., Bompa, C., Koussouris, S., & Askounis, D. (2014). Infusing Verification and Validation in ICT Solutions in Manufacturing: The FITMAN V&V Method. In Enterprise Interoperability VI (pp. 307-317). Springer, Cham.

Rafique I., Lew P., Abbasi Q.M., Li Z. (2012). Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model. International Journal of Computer, Electrical, Automation, Control and Information Engineering, 6, 568-573

Venkatesh, V. & Davis, F. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Vol 46(2), pp. Management Science 200046:2, 186-204

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Towards a unified view. MIS Quarterly, 27 (3), 425-478.

User Experience Questionnaire. https://www.ueq-online.org