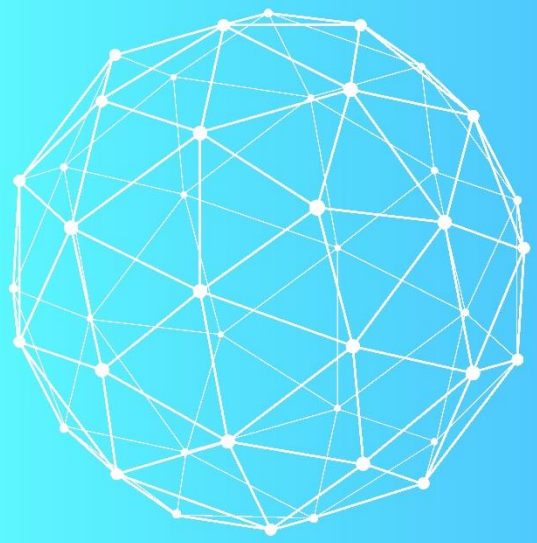

SPHERE

BIM DIGITAL TWIN PLATFORM

WP4 – Design & Construction Tools Development and Interaction

D4.5 WP4 Tools Usability Report

Authors

Marja Liinasuo VTT

Fausto Sainz COMET

Status	Released
Version	Final
Dissemination	Public

***SPHERE**: Service Platform to Host and sharE REsidential data, has received funding from the European Commission H2020 Programme under Grant Agreement No. 820805*

DISCLAIMER

ACKNOWLEDGEMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 820805.

Version	Date	Authors efforts	Description
V.0.1	07/10/2021	Marja Liinasuo (VTT)	Table of contents and some content addition
V.0.2	13/10/2021	Marja Liinasuo (VTT)	Chapter 4 and 5 content improvement
V.0.3	14/10/2021	Fausto Sainz Salces (COMET)	Corrections to expressions and copy-editing
V.1.0	15/10/2021	Rembrandt Koppelaar (EKO)	Internal review by work-package leader (copy-editing)
V.1.5	25/10/2021	Pekka Tuomaala (VTT)	Internal review by technical coordinator (expert assessment)
V.2.0	26/10/2021	Marja Liinasuo (VTT) + Fausto Sainz Salces (COMET)	Final corrections by deliverable authors
V.2.0	10/10/2021	Mikel Borras (IDP)	Final deliverable submission
V.2.1	24/08/2022	Marja Liinasuo (VTT)	The list of the tools tested by using heuristic evaluation added in Chapter 4.3
V.2.2	24/08/2022	Rembrandt Koppelaar (EKO)	Final check of the updated deliverable by the WP leader
V.3.0	26/08/2022	Technical coordinator	Internal review.
Re-submission	26/08/2022	Mikel Borras (IDP)	Updated final version.

Project			
Project Acronym	SPHERE		
Project Title	Service Platform to Host and SharE REsidential data		
Grant Agreement number	820805		
Call identifier	H2020-NMBP-EEB-2018		
Topic identifier	LC-EEB-06-2018-20 ICT enabled, sustainable and affordable residential building construction, design to end of life (IA 50%)		
Funding Scheme	Research and Innovation Action		
Project duration	48 months (From November 1st, 2018)		
Coordinator	IDP – Eduard Loscos / Mikel Borràs		
Website	http://sphere-project.eu/		
Deliverable			
Deliverable No.	D4.5		
Deliverable title	WP4 Tools Usability Report		
Description	This deliverable reports the studies regarding the usability WP4 tools (services and core services) implemented in SPHERE platform. Heuristic evaluation is performed to improve and optimise the implemented prototypes. The study methods, involving expert reviews, are presented as well as the results, i.e., the usability of the tools as revealed by heuristic evaluation. The detailed results of heuristic evaluation are delivered to the partner in question but are not reported in the deliverable due to the confidential nature of them. Instead, the conclusion of each heuristic evaluation is provided by hiding the company and tool name. This way the general level of the usability of the SPHERE tools can be verified.		
WP No.	WP4		
Related task	T4.5 Usability of Platform and Tools		
Lead Beneficiary	VTT		
Author(s)	Marja Liinasuo (VTT) Fausto Sainz (COMET)		
Contributor(s)	Jari Laarni, Esa Nykänen, Janne Porkka (VTT) Gisela Soley, David Martin (COMET) Rembrandt Koppelaar (EKO)		
Dissemination	**PU** Public		
Language	English – GB		
Due	M36	**Submission date**	26/08/2022 (First submission 29/10/2021)

Table of Contents

List of Figures

List of Tables

1 Executive Summary

The goal of SPHERE's project is the improvement and optimisation of buildings' energy design, construction, performance, and management, reducing construction costs and their environmental impact while increasing overall energy performance. The novelty of SPHERE project is the demonstration and validation the SPHERE platform working with real new and retrofitting construction cases.

SPHERE as a special Innovation Action assembles and integrates in a fully systematic, collaborative and integrated framework (IDDS) and corresponding Building Information Model (BIM) enabled set of tools (services and core services) with some diverse and complementary breakthrough Technology Readiness levels (TRL) 5-6 (Technology validated and demonstrated in relevant environment) technology assets. The demonstration activities in SPHERE will ensure that this platform, as well as the proposed tools, will finally reach TRL7 by the demonstration and validation on real residential construction and operational scenarios, both new and retrofitting.

This deliverable describes how usability is evaluated in SPHERE, by means of heuristic evaluation with a systematic approach, and what the main results are among the SPHERE tools evaluated. Heuristic evaluation uses 'rules-of-thumb' or rapid value judgements in a structured manner to identify ease of usage flaws or required improvements, in this context from a tool usage experience perspective. It places the evaluators in the position of using the tools to evaluate not the technical share but the functionality of the tools.

The purpose of this deliverable is thereby to (i) describe how the tools developed in SPHERE as part of the SPHERE platform are evaluated from a usability perspective, and to (ii) report the main findings from the heuristic evaluation that benefit the tool developers to improve tools' usability. The end result of this work is improved SPHERE tools from the perspective of usability.

The template created in SPHERE for usability evaluation provided a clear and extensive enough format for usability evaluations to tools at this phase of tool development. The usability of the tools proved to be good or, alternatively, most flaws identified were found easy to correct.

The target audience of this report are SPHERE evaluation stakeholders, i.e., technical developers, in SPHERE, whose tools are evaluated and who have benefitted from the evaluation in practice. Also partners benefit who contribute to User Centered Design (UCD) practices, by being informed about the methodology of heuristic evaluation used at this phase in the SPHERE project.

On behalf of the Authors,

Marja Liinasuo, VTT

2 Introduction

2.1 Purpose and target group

The purpose of this deliverable is to describe how software tool (service or core service) usability is evaluated in SPHERE and report what the main results are. Correspondingly, the work benefits evaluation stakeholders, i.e., technical developers, whose tools are evaluated, as they receive improvement recommendations. Also partners contributing to User Centred Design (UCD) efforts are informed about the methodology of heuristic evaluation at this phase of the SPHERE project.

2.2 Contributions of partners

The work done for the report is performed by partners COMET and VTT. Both regarding the development of the methodology in carrying out the heuristic evaluation for usability as assessed in SPHERE, and the heuristic evaluations performed themselves. The main author of this deliverable report is VTT.

2.3 Baseline

The deliverable has been written as an outcome of a collaborative effort between VTT and COMET, who have consulted responsible partners for usability evaluations. The contributions of involved partners to various sections are as per table 1 below.

Table 1. Overview of tasks, chapters and involved partners.

Tasks carried out	Chapter	Involved Partner(s)
Preparation and writing of general parts	1, 2	VTT
Preparation and content proposal	3, 4, 5	VTT
Usability and heuristic evaluation	3	VTT, COMET
Heuristic evaluation in SPHERE	4	VTT, COMET
Conclusions	5	VTT
List of Acronyms and Glossary	6	VTT
References	7	VTT
Review with comments	ALL	COMET, EKO, VTT, Responsible tool developers

Usability is a known User Centric Design (UCD) related concept and it has been used also in EU projects abundantly. For example, EUDAT (2015-2018) is a H2020 European project that resulted in creating a data exchange infrastructure still functioning today; it offers common data tools and also deals with usability (Sitompul, 2016). Other examples of a more recent EU project in which usability

has been studied include Factory2Fit (2016-2019), where an essential usability perspective was integrated to evaluate new tools to support shop-floor workers in a factory, and Privacy.Us (2015-2020). Partners in the latter project faced usability challenges regarding privacy matters. Presently, usability is studied in for example BIMprove (2020-2023) to ensure that the digital twin related solutions are usable for all stakeholders needing digital twins for their work related to a construction site, from workers to managers.

The widespread use of usability of tools comes from usability evaluation results delivering practical value to the developers of the evaluated tools. In the form of understandings of the engagement practices of pontial future users and recommendations to improve functionality and tool design so as to make them easier and better to use. In delivering the works in task 4.5 other UCD efforts were studied in other H2020 as per the above, yet no active information exchange was carried out with other EU projects related to usability evaluation.

2.4 Relations to other activities

This deliverable has received input from various tasks in SPHERE and will affect, in turn, various future tasks in SPHERE project.

- Inputs:
  - T2.4 (User Centred Life Cycle Design Assessment): user-centric evaluation methodology perspectives are described, one being usability.
  - T2.6 (Usability of platform and tools): the first concept for usability testing is provided.
  - T4.5 (Usability of platform and tools): heuristic evaluation method is defined and evaluation is performed; this is the basis for the present deliverable.
  - T6.2 (Construction & Commissioning of Demo sites): the principle of stepwise evaluation in SPHERE is presented.
- Outputs:
  - T6.1 (Design of demo sites under a Digital Twin Environment): the tools are affected and improved by the results of heuristic evaluation, received by technology developers.
  - T6.4 (Social acceptance with final users): pilot trials with final users will be performed with more comprehensive usability testing than the one reported in the present deliverable, improving usability further.

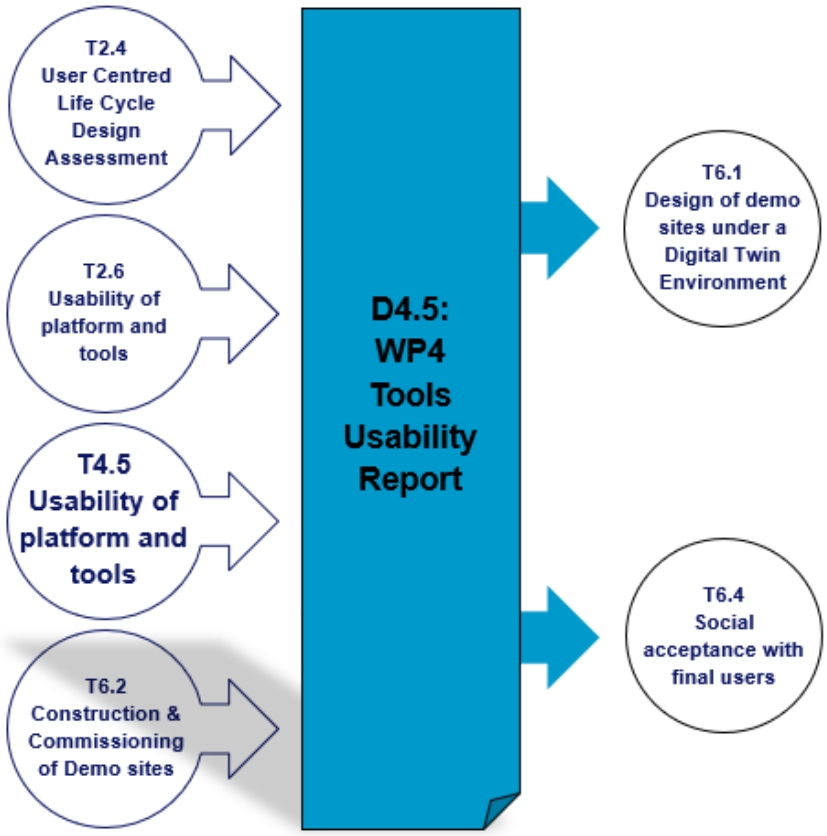

Figure 1. Illustration of tasks that affected this deliverable (input), and tasks that will be affected by it (output).

3 Usability and Heuristic Evaluation

3.1 The context of usability

Software tools and digital devices have become increasingly complex over time. It is important that the requirements and limitations of those using tools and devices, i.e., human beings and in relation different user groups, are taken into account during technological design. The worst possible outcome is that technical innovations are futile from the perspective of their value in the field and the contribution they provide because they are not usable even though they provide technically correct outputs. Tools which are hard to use are used reluctantly, and the work result is poorer than what it could be with better tools.

In the specific context of SPHERE, regarding tools designed to support construction, renovation, and building operation, the requirement of usability of tools by users is vital. In this sector work needs to be done according to a strict timetable with limited room for failures, which is both a short-term and long-term requirement. Therefore there are high quality requirements in tool usage with limited room for a lack of intuitive tool usage experiences. And there is limited time available requiring short-learning curves to learn how to use tools. If tools are not designed to result in low error use and rapid accession and usability this will not improve planning, building and operation of buildings. And every error in the process can cause costly delays, discomfort for building occupants and potentially added expenses to owners for several decades.

The 'taking the human into account' aspect is realised through User Centred Design (UCD) such as usability studies and the related amendments to technology. To ensure that technology is easy to use. In the following pages, usability and its assessment, in the form of heuristic evaluation, are presented to provide learning on these at the level of general concepts.

3.2 Usability

The concept of 'usability' is one cornerstone of this deliverable along with 'heuristic evaluation'. In the following, an introduction to usability is provided originating from the SPHERE deliverable D2.5 (Sainz Salces, Martin-Moncunill, Garcia, Cliftton, Koppelaar, Laarni, Liinasuo, Calvo, 2020).

Usability is defined in ISO 9241-11:2018:en on human-system interactions as '*the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use'*.

Thus, usability refers to a certain quality of some entity. It is a concept to be applied in situations where people use interactive systems as well as other types of systems (such as built environments), also including products (including industrial and consumer related ones) and services (including technical and personal ones).

The key elements of the above quoted definition of usability include:

- Effectiveness refers to accuracy and completeness with which users achieve specified goals.
- Efficiency refers to resources used in relation to the results achieved.
- Satisfaction refers to the extent to which the user's physical, cognitive, and emotional responses that result from the use of a system, product or service meet the user's needs and expectations.

Usability is not a stand-alone phenomenon. It is affected by users; for instance, a novice user may have different conception of system's usability than an experienced user. Different aspects play a role when using a system like a digital service for the first time in comparison to a user who has already used the system repeatedly.

Usability is also affected by the goals of the user. If a user needs the system only for simple tasks, the perceived usability may be different from the perception of a user who wants to use the system in many different more complex ways.

Finally, the context of use affects usability as well. For instance, if it is used under difficult circumstances, such as when simultaneously walking and using a tool, the perceived usability is different from using the system when not moving and sitting on a chair.

Appropriate usability is important. As ISO 9241-11:2018 on human-system interactions states, when usability is lower than expected, the intended users may not be able or willing to use the object of interest. The opposite is also true, when usability is higher than expected, the object can have a competitive advantage. Furthermore, when people are willing to use the object they get the full benefit of it.

Usability is relevant in the context of the built environment and the SPHERE project. The built environment encompasses external and internal environments and any element, component or fitting that is commissioned, designed, constructed, and managed for use by people (ISO 21542:2011, 3.10). In SPHERE, the built environment also includes Digital Twin based platforms with a Building Information Model (BIM) enabled set of services (software). In the projects context the SPHERE Digital Twin and the connected software service, which is the focus here in the context of measuring usability.

3.3 Heuristic evaluation

Usability can be measured in various ways. During a usability test, the observer(s) watch test participant performing tasks with the product in a test environment (Lewis, 2006). Field studies can involve researcher(s) observing user(s) performing tasks in real situations and in inspections such as heuristic evaluation or expert evaluations. A person uses a tool and studies his/her own experiences with the tool.

Heuristic evaluation is originally an informal method where evaluators are presented with a user interface (UI) design and are asked to comment it (Nielsen and Molich, 1990). However, individual

evaluators were found to be rather poor at performing such an evaluation. Only the aggregation of evaluations from several evaluators resulted in satisfactory results (Nielsen & Molich, 1990).

The responsibility for analysing UI is on the evaluator in the heuristic evaluation session. Usually, a heuristic evaluation session lasts from one to two hours (Nielsen, 1995). In practise, the evaluator goes through the UI in an iterative manner and inspects the UI elements from various perspectives according to a list of recognised usability principles, i.e., the heuristics (Nielsen, 1995).

To give an idea about the process, ten usability heuristics for UI design are briefly presented in the table below.

Table 2. *Ten usability heuristics by Nielsen.(Nielsen Norman Group, 2021).*

#1: **Visibility of system status**: The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time.
#2: **Match between system and the real world**: The design should speak the users' language. Use words, phrases, and concepts familiar to the user, rather than internal jargon. Follow real-world conventions, making information appear in a natural and logical order.
#3: **User control and freedom**: Users often perform actions by mistake. They need a clearly marked "emergency exit" to leave the unwanted action without having to go through an extended process.
#4: **Consistency and standards**: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform and industry conventions.
#5: **Error prevention**: Good error messages are important, but the best designs carefully prevent problems from occurring in the first place. Either eliminate error-prone conditions, or check for them and present users with a confirmation option before they commit to the action.
#6: **Recognition rather than recall**: Minimize the user's memory load by making elements, actions, and options visible. The user should not have to remember information from one part of the interface to another. Information required to use the design (e.g. field labels or menu items) should be visible or easily retrievable when needed.
#7: **Flexibility and efficiency of use**: Shortcuts — hidden from novice users — may speed up the interaction for the expert user such that the design can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
#8: **Aesthetic and minimalist design**: Interfaces should not contain information which is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes their relative visibility.
#9: **Help users recognize, diagnose, and recover from errors**: Error messages should be expressed in plain language (no error codes), precisely indicate the problem, and constructively suggest a solution.
#10: **Help and documentation**: It’s best if the system doesn’t need any additional explanation. However, it may be necessary to provide documentation to help users understand how to complete their tasks.

There are no systematic means to amend the usability related errors and weaknesses. However, as usability evaluation require inclusion of an explanation for each problem, it is often easy to improve the design according to the evaluation (Nielsen, 1995). Often each identified problem can be translated by a tool product owner in a task for improvement that is subsequently tackled by frontend software developers, with further feedback iterations as needed.

4 Heuristic evaluation in SPHERE

4.1 Evaluation goal and process

Heuristic evaluations were performed in SPHERE for the SPHERE software tools (services and core services) to support technical developers to design and improve the usability of their tools. The goal for heuristic evaluation was to provide information about flaws in UI design, so that after the heuristic evaluation there are no obvious usability related errors in the evaluated tool, or as few as possible.

The approach varied depending on the maturity status of each tool. This means that there were no common objectives to reach across tools, each tool was evaluated and the technical designers were provided feedback independently, in accordance with the quality of the tool. Therefore, if the tool User Interface (UI) was static and presented in a Power Point presentation, feedback was equally provided as for tools with a live user interface. It was not a pre-requisite to have the same functionalities for the heuristic evaluation. The principle was to give as much feedback as possible at any stage, based on the availability of the tool and the available information about the tool.

On the same principle no usability scoring was provided which would have set all tools in the same line. The purpose of heuristic evaluation was only to support the development of tools, not to compare them with each other.

The heuristic evaluations were performed by usability professionals in the partner organisations (VTT and COMET) in the SPHERE consortium. In practice, all tools were evaluated in a digital format. The evaluated tool was usually accessed through a web page, but in cases that this was not possible, due to, for example, the lack of an existing UI, some other documentation was used (such as a Power Point presentation). The evaluated tool could, correspondingly, be only evaluated based on a design combined with instructions on how to use the tool and its expected outputs, which will be eventually be provided after further development to tool users. Thus, in cases when no actual tool was ready enough for evaluation, it is possible that the performed heuristic evaluation only was performed preliminary features that may change when the actual tool is more finalised. Even in this case, the heuristic evaluation may have provided important information about how to design the UI in the context of the tool in question.

After the usability professional received some basic information about a tool (what its' purpose is, what user and their roles use it) its UI with some basic functionalities were evaluated (as per the evaluation process in Figure 2 below). The evaluation was written in a report providing feedback of mainly matters recommended for either minor corrections or larger improvements.

The heuristic evaluation output resulted in an evaluation feedback report, sent to technical developer(s), i.e., evaluation stakeholders. In some cases, the stakeholders wanted a second evaluation after correcting the weaknesses identified in the report. In these cases, new feedback was provided based on a second heuristic evaluation.

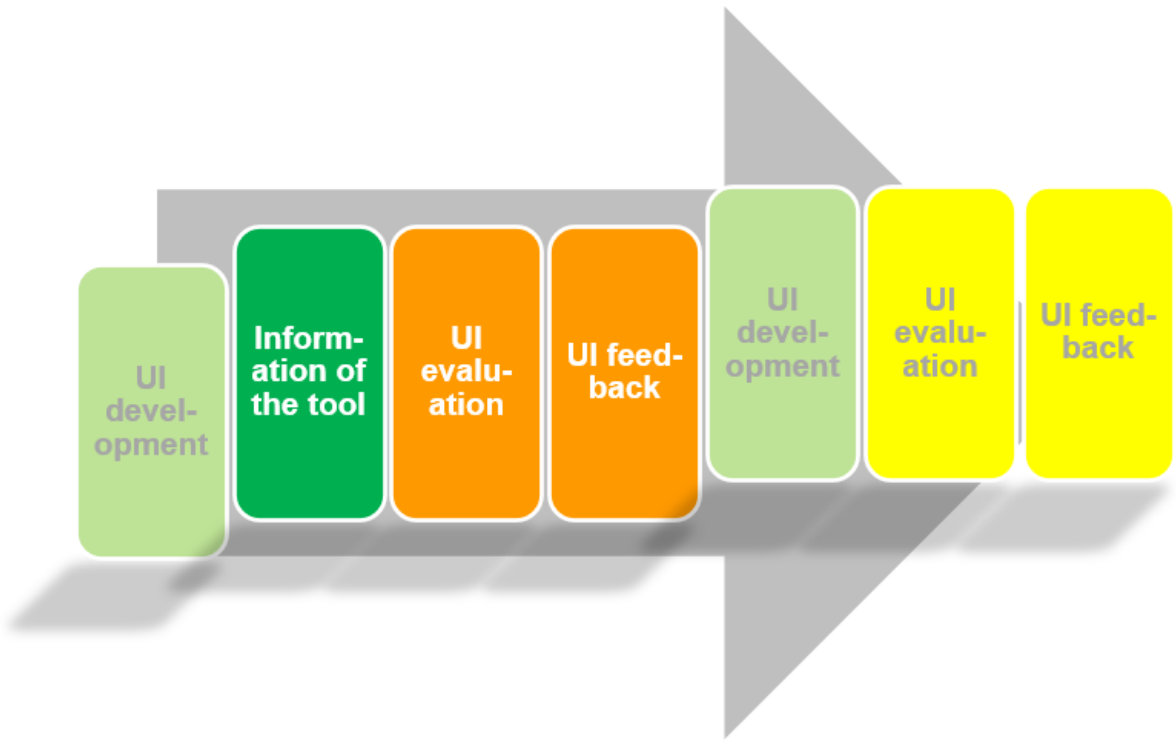

Figure 2. Process of heuristic evaluation during the tool technical development in SPHERE.

4.2 Method used for heuristic evaluation

The core idea of the evaluation is that it is performed from a user perspective, and that the evaluation is about usability; the technical content is not to be commented. However, it was evident that the UI of most tools was not finalised or only preliminary at this stage of the project. Hence, a classical or basic heuristic evaluation (see the 10 heuristics in section 3.3 table 2) would not have been feasible or even possible to conduct in this context.

The heuristics evaluation for each tool were therefore adjusted according to the needs of the SPHERE project and the assumed status of the SPHERE tools as described in general in section 4.1. Eventually, it was established that the initial expectation to carry out this adjustment was correct as no need emerged to alter the chosen evaluation format. Instead, if the tool was insufficiently developed for a full heuristic evaluation, the specific heuristic aspect of evaluation that was not applicable was not utilised and was left out.

A series of focus areas were developed for the evaluation according to traditional usability heuristics further complemented with additional heuristics from the SPHERE partners usability professionals. The focus areas can be classified in four categories (see Figure 3 below). These categories are presented next including a clarification for each category from the perspective of heuristic evaluation.

Focus Area A. Functionalities

a. **Core functionalities related to the purpose of the tool**
Clarification: This can be hard to evaluate in heuristic evaluation, without the specific expertise, which a real user is supposed to have, and without guidance on how to use the tool. However, if it is possible to do, evaluation can provide important information to evaluation stakeholders.

b. **Basic functionalities (logging in, adding and deleting things, cancelling etc.)**
Clarification: Here, the question is about general-level functionalities but they can be evaluated only if the evaluation is performed with a real tool, not relative to a Power Point presentation or the like.

c. **Operations in general; the intuitiveness of performing operations**
Clarification: See the clarifications above ; the operations can be related more to the purpose of the tool, i.e., core functionalities, or more the ones relating to the fluency of using the tool (basic functionalities).

Focus Area B. Visual details

d. **The understandability of symbols**
Clarification: The meaning of symbols is important to be understandable as in a professional tool, no symbol is used for the sake of illustration ; symbols should be used when they are important from the perspective of using the tool.

e. **The layout of the user interface**
Clarification: The layout of user interface affects the clarity and aesthetics of user interface; cluttered, too dense, and too empty user interface is more difficult to use

Focus Area C. Visual outlook as a whole.

f. **Aesthetics**
Clarification: Aesthetics is an aspect, which affects the pleasantness but also professionalism and the brand of the company selling the product.

g. **Professionalism**
Clarification: Professionalism is the style, which is needed, when designing tools for professional use.

h. **Consistency**
Clarification: Consistency supports the perceived level of logic in the user interface; an inconsistent user interface is more difficult and unpleasant to use.

Focus Area D. Language used

i. **Terms and concepts**
Clarification: As the user probably has some special expertise, the language should be in accordance with the special vocabulary of the profession or task the tool is to support.

j. **The structure and length of expressions**
Clarification: Language used in the user interface should be clear and understandable to the user. All expressions need to be clear and consistent by style (such as the usage of active or passive voice etc.).

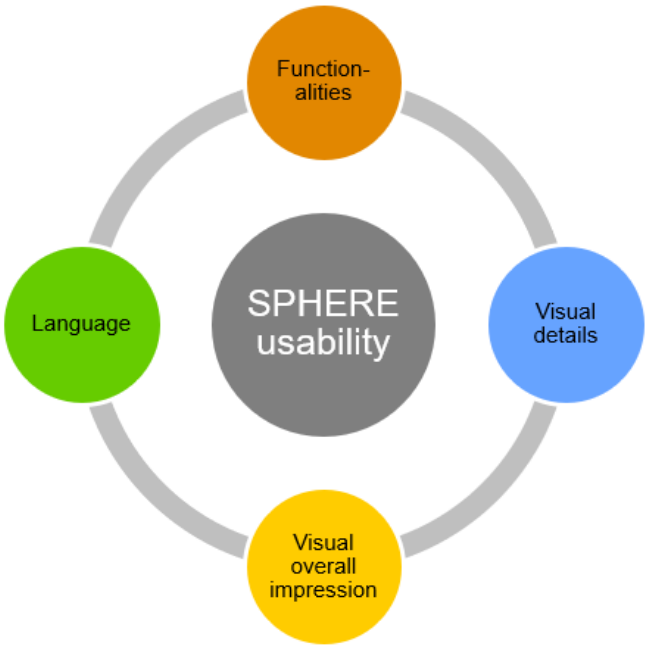

Figure 3. Four focus areas of the heuristic evaluation as performed in SPHERE.

The four focus areas above were included in the heuristic evaluation. To make the evaluation easy to the evaluators and understandable to the tool technical designers (evaluation receivers), an easy to understand evaluation template was designed. The template stated standardised questions which were answered by the usability professionals carrying out the heuristic evaluation. Comments clarifying evaluation related matters, such as limitations, were presented in the evaluation template as sentences in parentheses next to the evaluation question. In the following, key questions guiding the evaluation as expressed in the evaluation report are presented. Regarding vocabulary, the term 'system' is used instead of the one of 'tool'. They are used as synonyms in this context.

1. Does the system provide the needed functionalities (this depends on what the system is supposed to do, this may be also too hard to evaluate in the frame of brief heuristic evaluation)?
2. Are the main tasks easy to perform with the system (logging in, adding and deleting things, depending on what the system is for - so this is about complexity and learnability)?
3. Is the system intuitive so that it is easy to "guess" how to operate it?
4. Is the system visually clear?
    a. Is the meaning of symbols clear?
    b. Is the layout of the user interface clear?
    (not too crowded, logic in grouping user interface elements)
5. Is the system visually aesthetic and/or looks appropriate for professional use?

6. Is the system consistent (same style used throughout functionalities and user interface qualities)?
7. Does the system use appropriate language (language that users understand)?

   (this depends on who the intended users are, as some users may need special terms, not familiar to heuristic evaluators, so that this may be also too hard to evaluate thoroughly in the frame of brief heuristic evaluation)

In addition to answers to the above questions, some further explanation was sometimes provided to the evaluation stakeholders, for example, about the limitations to perform the evaluation due to the scarcity of information about the UI.

Answers to the questions above, i.e., the evaluation, sometimes resulted in long explanations on what could or should be amended. Describing what UI elements are missing or are not appropriate, and how they could be better, may take a lot of space even if the matter in itself is not serious. Pictures of the UI also take a lot of space; only in the cases of a Power Point presentation were detailed comments located in the slide set alongside the evaluation report. Evaluations pertaining to online tools tended to be longer since separate pictures about the UI needed to be included in the evaluation report itself.

Finally, a long report about errors and flaws could have misguided the reader – the level of usability can be good also when there is a lot to correct. That is why all heuristic evaluation reports include a conclusion at the end of the report, which sets the evaluation in proportion.

4.3 Heuristic evaluation results

The heuristic evaluation has been performed for all available SPHERE core services and SPHERE Services, that are together called as tools.

The heuristic evaluation was done slightly differently between the tools. Typically the evaluation was considering the graphical user interface, but since some of the tools do not provide that direcly or are operating in background and integrated to other tool, there is also a “lighter evalution” focusing merely on functionality testing. This is mentiond in table after the tool name in brackets. Due to the recent updates in last three tools, the testing procedure for those still ongoing and finishing soon.

As a result, 20 tools evaluated are listed in table 3 below with the corresponding company name. The list includes whole SPHERE ecosystem with core services and services.

Table 3. List of tested tools using heuristic evaluation

NUMBER	COMPANY	TOOL
1	ASC	Flink2Go
2	COMS	IMAN
3	EAI	EcosimPro
4	EKO	EPESUS (light evaluation, predecessor to CEAT)
5	EKO	CEAT
6	EKO	CEAT (Portal mockup)
7	EUT	Blockchain technology (light evalution, integrated to IDP DTwin)
8	EUT	iES_E
9	EUT	iES_W
10	EUT	iPredict (light evaluation, integrated to IMAN)
11	IDP	DTwin
12	MBSD	CMT (already in use, not evaluated)
13	MBSD	LCCCA (already in use, not evaluated)
14	MBSD	OPT
15	NEXT	Neanex portal
16	NUIG	ModSCO
17	R2M	En-MS
18	TNO	Robmos
19	VTT	HTM
20	VTT	Research Portal

In the following section, all the heuristic evaluation results are presented of the SPHERE tools, processed according to the heuristic evaluation template presented in section 4.2. Additionally, in the beginning of the evaluation process, some evaluation was performed more intuitively, just by looking for different perspectives during the evaluation. This was done because the idea of a general format to all evaluations emerged only later in a 'learning by doing' manner.

The content of the evaluation is confidential. That is why results are presented here anonymously, without the name of the tool nor the company or organisation developing it. The order of evaluations is not corresponding to the order of the tool names in the list above. Furthermore, the heuristic evaluations are highly detailed. Revealing the evaluation report as a whole would also result in the identification of the tool and the company in question. For the sake of confidentiality, neither the detailed results are provided but only a general-level conclusion. This conclusion was part of the feedback provided to technical developers in the heuristic evaluation report. In such cases where the conclusion would have revealed too much about the tool in question, the conclusion is modified to hide such facts. In the cases where text is added to increase the understandability of the conclusions or hidden, not to reveal the identity of the tool and company in question, square brackets are used.

Slide show illustrating the future user interface of the tool, report 3 pages, conclusion:

The display appears rather usable but as there is still quite a lot to design, it is not possible to present an extensive conclusion.

There is more detailed [evaluation] information in the slide set, both in the slide (text with brown colour) and, occasionally, also in the notes page.

Online tool, report 6 pages, conclusion:

The evaluated pages are of good usability and the found deficiencies are easy to correct. Also, the functionalities provided by the tool are useful and they can help professionals in carrying on with their professional work.

_____

Conclusion after the second evaluation, report 9 pages:

The application evaluated is quite good in general terms regarding usability, and the minor mistakes found are easy to correct. Also, the functionalities provided by the tool are useful and they can help professionals in carrying on with their professional work. The reports produced by the tool are also very clear and easy to understand.

Slide show illustrating the future user interface of the tool, report 8 pages, conclusion:

The web system (tested through a Safari browser) looks clear and professional and is easy to use. The windows for preparing the uploading the different files are consistent and look good, with only some exceptions that are easy to correct, those of possible actions components.

Certain minor changes regarding colours and objects position can be easily made.

Slide show illustrating the future user interface of the tool, report 15 slides, conclusion:

(No conclusions. Detailed evaluation comments are provided in the slide show representing the outlook of the future user interface.)

Online tool, report 5 pages, conclusion:

The evaluated service appears professional, is easy to use and the content of the display is easy to understand. The main issue is some inconsistency in the menus, not preventing fluent usage of the service and probably easy to fix.

Online tool, report 8 pages, conclusion:

The tool is mostly quite usable except for minor glitches that probably can easily be fixed. Carefully prosented error messages can also help users understand what is going on and, very importantly, how to proceed when actions have not been achieved or properly completed.

Set of files used in the tool, report 10 pages, conclusion:

The tool works well in Excel. This has its advantages for those familiar with the application, as well as its limitations regarding the limited possibilities to modify its outlook and the way it works. Certain design issues can be improved by enlarging or adding colour or other cell components (borders, background colour, etc.). Carefully presented error messages can also help users understand what is going on and, very importantly, how to proceed. The instructions, small sections on top of the tabs, help in the tool usage. Instructions, which describe the general features of the tool as well as how to use it in detail are needed as a written document.

Screenshots and notes illustrating the future user interface, report 4 pages, conclusion:

The slide set looks clear and professional and easy to use. A definitive evaluation about the usability of the instructions provided in the slide set (or the simulation tool itself) cannot be given as it is not known how the real simulation would function compared with the instructions - are there points that are not shown nor clarified here etc. The slide set is also very consistent and looks professional, with only some exceptions that are easy to correct. Additionally, it would be good to add a slide or two, which explain a bit about the simulation and especially the predefined matters that can and cannot be modified by the user.

The slides from the presentation could be used, with add-ons and modifications, as a tool manual for new users.

Comment about the corrections made based on the evaluation:

Changes are fine, the slide set is ready from the heuristic evaluation point of view.

Online tool, report 5 pages, conclusion:

The evaluated [x] appears professional, is easy to use and the content of the display is easy to understand by a professional user. The informative content of the display is consistent. There is only some inconsistency in the menus, not preventing fluent usage of the [x].

Online tool, report 8 pages, conclusion:

The evaluated [x] appears professional, to be used by professionals acquainted with the content and objects of the [x]. [x] appears consistent. Its visual appearance could be made a bit more solid by, e.g., using a bit stronger font. The present one uses unnecessarily thin lines. [x] already uses features supporting usability (appearance of symbol names when having the cursor close a symbol; functionality, which enables easy access to previous displays [x] etc.). [x] is, however, a complex entity and its usability could be further strengthened. For instance, usability could be supported by using more titles on the display. Based on the complexity of activities user can perform in [x], training and/or manual is needed.

Online tool, report 4 pages, conclusion:

The web system looks clear and professional and is easy to use. The windows for preparing the uploading the [xxx] are consistent and look rather professional, with only some exceptions that are easy to correct, those of possible actions components.

However, windows and documents should have some descriptive name, telling what they are about. Additionally, it would be good to add some information on the landing page about the system. Apart from the preparing windows, visual appearance is not consistent especially regarding the usage of fonts and colours. Also, the result option of browser has too small fonts, apart from the titles, and units are missing. To conclude, the system is simple to use and there are matters to correct, but corrections are easy to make.

Online tool, report 5 pages, conclusion:

[The system looks adequate; the evaluation document provides detailed suggestions for improvements. No conclusons provided in the report.]

5 Conclusions

5.1 Summary of achievements

This deliverable presents an evaluation format usable for SPHERE services, which are not in the final development phase. The format using four focus areas, functionalities, visual outlook as a whole, visual details, and language, is of the kind that services in various lower maturity levels can be evaluated. The format can also be extended to instructions for using a service. In the cases of very low maturity tools and instructions, only one perspective of the four focus areas, the one of functionality testing, must be left out.

The design of such a tailored template for heuristic evaluation, fit for tools with low-level UI maturity, is not usually performed. To the author's knowledge no such template has been designed to present. This template can be used in other projects so it will likely have practical value beyond SPHERE, either directly for future works by COMET and VTT, and indirectly from orther organisations utilising the methodology in this report as part of the open science innovation transfer process.

Performing the heuristic evaluation for core services and services (tools) in SPHERE ecosystem provided great feedback for their developers on the user interface and the functionality of the tools. We developed a procedure to go through them all. This feedback included practical comments on improvements and verified, by using a template with SPHERE logo, that the evaluation was performed as part of the SPHERE project. The challenge from evaluation perspective is also related to the nature of research projects. Tools to develop in a project are usually based on existing work and ICT solutions adopted by the developers already before the project. This makes the evaluated tools more complex, complicating also the evaluation.

5.2 Relation to continued developments

The performed heuristic evaluation supports technical development of the tools and paves the way to final demonstrations in WP6. As part of the piloting works in this work-package, further UCD based heuristic evaluations can be carried out to iteratively improve the usability of SPHERE tools.

5.3 Other conclusions and lessons learned

This study strengthened the principle of preparing a systematic approach to evaluation whenever possible. The trade-off is the flexibility in evaluation, which may suffer in a systematic approach, but if the format is appropriate and adaptable enough (e.g. leaving out one of the four focus areas if needed), it can still include many important aspects and provides a solid and extensive enough evaluation to be of use to all evaluation stakeholders.

6 Acronyms and glossary

6.1 Acronyms

API: Application Programming Interface.

BIM: Building Information Modelling.

D: Deliverable.

DT: Digital Twin.

GUI: Graphical User Interface.

PaaS: Platform as a Service.

SaaS: Software as a Service.

T: Task.

UCD: User Centered Design.

UI: User interface.

WP: Work Package.

6.2 Glossary

Digital Twin: Digital Twin is a dynamic, virtual representation of a building, its assets, processes and systems (SPHERE project 2019).

Evaluation: The systematic collection of information about the activities, characteristics, and outcomes of programs for use by specific people to reduce uncertainties, improve effectiveness and make decisions with regard to what these programs are doing (Patton, 1990)

Evaluation stakeholders: The term "stakeholder" within an evaluation context, refers to those who have a vested interest in that which is being evaluated, and thus, would be in a position to use the evaluation results in some way. Depending on their role relative to the program or initiative being evaluated, stakeholders are positioned to use evaluation findings in different ways (Preskill & Jones, 2009)

Usability: The ability of a product to be used with effectiveness, efficiency, and satisfaction by specified users to achieve specified goals in particular environments (Hvannberg, 2007).

SPHERE Digital Twin ecosystem: a selection during the development of the SPHERE platform to call approach as ecosystem, to highlight its integrated nature and tools integrated as services (SPHERE project 2019).

SPHERE Digital Twin platform: The SPHERE platform offers tools to create, update and manage building information from early design to operation phase, enabling efficient utilization of Digital Twin. Project decided to call this as ecosystem (SPHERE project 2019).

7 References

Nielsen, J. (1995) *How to conduct a heuristic evaluation*. Nielsen Norman Group, 1, 1-8. http://www.ingenieriasimple.com/usabilidad/HeuristicEvaluation.pdf

Nielsen, J., & Molich, R. (1990, March) *Heuristic evaluation of user interfaces*. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249-256).

Nielsen Norman Group (2021). Web site https://www.nngroup.com/articles/ten-usability-heuristics/ 10 Usability Heuristics for User Interface Design. Accessed in October 7, 2021.

Lewis, J. R. (2006) *Usability testing. Handbook of human factors and ergonomics*, *12*, e30. http://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo:modelagem-e-simulacao-de-sistemas-humano-computacio/usabilitytesting-ral.pdf

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, inc.

Preskill, H., & Jones, N. (2009). *A practical guide for engaging stakeholders in developing evaluation questions*. Robert Wood Johnson Foundation, Evaluation Series. https://folio.iupui.edu/bitstream/handle/10244/683/091022.stakeholder.involvement.fullreport.draft.pdf?sequence=2&sa=U&ei=IwNWU5WAFIqL4gTK-4DgAg&ved=0CEQQFjAH&usg=AFQjCNGl7ktAkP6jWjiIcuAznrqvj5b3Lg

Sainz Salces, F., Martin-Moncunill, D. , Garcia, A., Cliftton, Koppelaar, R., Laarni, J., Liinasuo, M., Calvo, A. (2020) *User Centred Dynamic Design methodology*. D2.5, project Service Platform to Host and SharE Residential data (SPHERE). pp. 76.

Sitompul, T. A. (2016) *Usability and User Experience Evaluation of EUDAT Services*. Master's thesis, Aalto university, Degree Programme in Computer Science and Engineering. pp. 159.