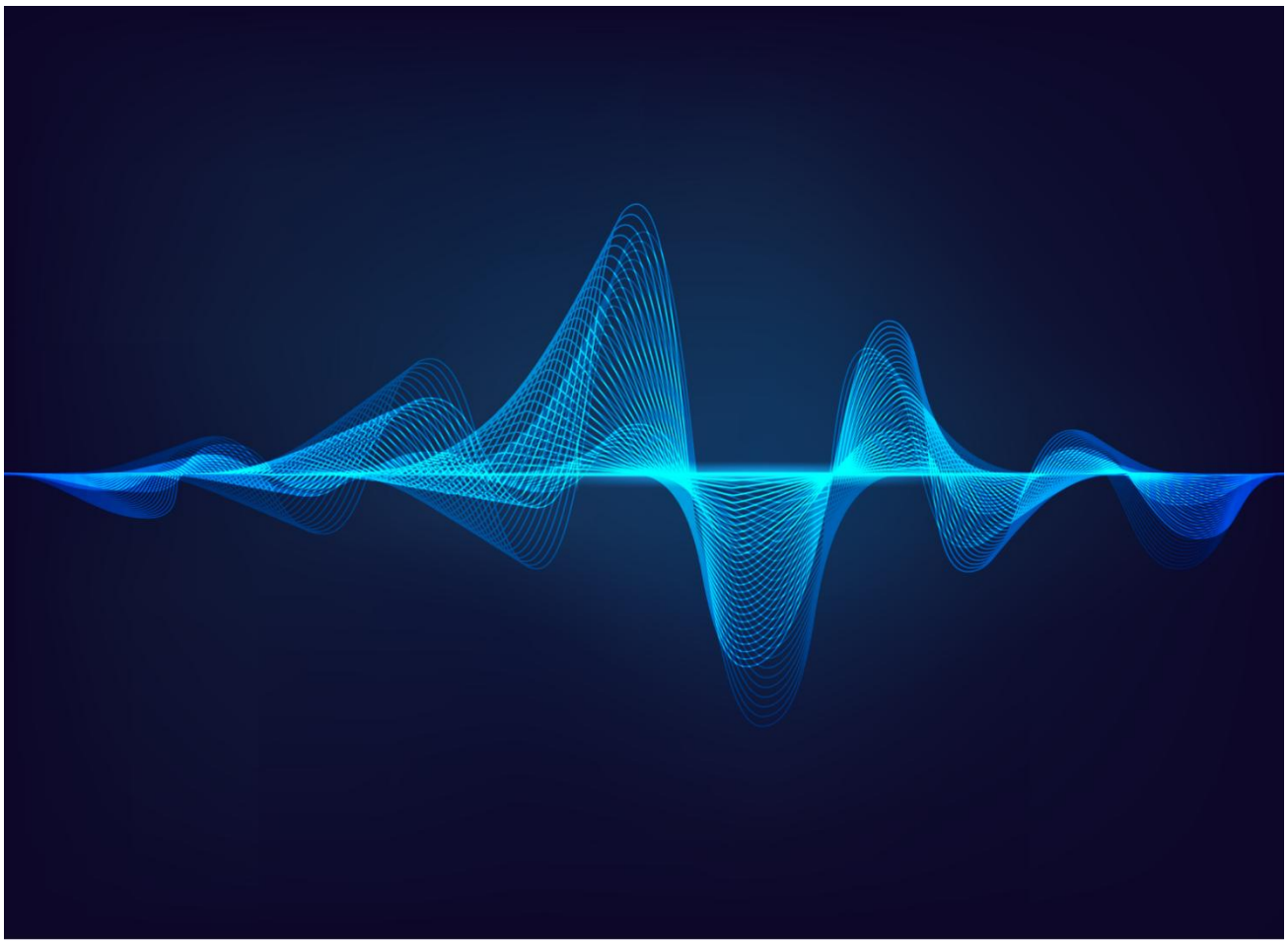

D4.3 – Software Validation Report

OpenMusE

An open, scalable data to-policy pipeline for European music ecosystems

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Project Factsheet

Acronym:	**OpenMusE**
Title:	**An open, scalable data to-policy pipeline for European music ecosystems**
Coordinator:	SINUS Markt- und Sozialforschung GmbH
Reference:	101095295
Type:	Research and Innovation Action
Program:	Horizon Europe
Start:	1st January 2023
Duration:	36 months
Website:	https://www.openmuse.eu/
Consortium:	**SINUS Markt- und Sozialforschung GmbH,** Germany (SINUS), Coordinator **TURUN YLIOPISTO,** Finland (UTU) **UNIVERSITEIT VAN AMSTERDAM,** Netherlands (UVA) **Scuola Superiore di Studi Universitari e di Perfezionament,** Italy (SSSA) **EKONOMICKA UNIVERZITA V BRATISLAVE,** Slovakia (EUBA) **Reprex B.V.,** Netherlands (REPREX) **SYNYO GmbH,** Austria (SYNYO) **MUSIC INNOVATION HUB SPA IMPRESA SOCIALE,** Italy (MIH) **Slovenský ochranný zväz autorský pre práva k hudobným,** Slovakia (SOZA) **Aloaded AB,** Sweden (ALOADED) **Music Export Ukraine,** Ukraine (MEU) **Muzikos Eksporto Fondas,** Lithuania (MXF) **ARTISJUS MAGYAR SZERZOI JOGVEDO IRODA EGYESULET,** Hungary (ARTISJUS) **MUSICAUTOR SDRUZHENIE,** Bulgaria (MUSICAUTOR) **HEARDIS! GMBH,** Germany (HEARDIS) **Fundacja Eksportu Polskiej Muzyki Music Export Poland**, Poland (MExP)

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Deliverable Factsheet

Number:	**D4.3**
Title:	**Software Validation Report**
Lead beneficiary:	SINUS
Work package:	WP4
Task:	T4.4
Dissemination level:	Sensitive
Submission date:	28.11.2025
Revision/resubmission:	15.05.2026
Main author(s):	Isabella Tautscher, SINUS
	James Rhys Edwards, SINUS
Contributor(s):	Leonie Regen, SINUS
	Daniela Vössing-Garcia, SINUS
Quality check:	Aleksi Lahtinen, UTU
	Olivia Kafka, SYNYO
	Leo Mörtenhuber, SYNYO
	Jakob Seper, SYNYO

Document history:

Revision	Date	Main modification	Author
0.1	25/11/2025	Draft sent to internal reviewers	Isabella Tautscher
0.2	26/11/2025	Comments from internal reviewers received	Aleksi Lahtinen, Olivia Kafka
0.3	27/11/2025	Comments from internal reviewers incorporated	Isabella Tautscher
1.0	27/11/2025	Submitted to EC	Isabella Tautscher
2.0	15/05/2026	Revisions in response to Review Consolidated Report	James Edwards

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Disclaimer of Warranties

This project has received funding from the European Union's Horizon Europe, research and innovation programme, under Grant Agreement No. 101095295.

This document has been prepared by OpenMusE project partners as an account of work carried out within the framework of the EC-GA contract no 101095295.

Any dissemination of results must indicate that it reflects only the author's view and that the Commission Agency is not responsible for any use that may be made of the information it contains.

Neither Project Coordinator, nor any signatory party of OpenMusE Project Consortium Agreement, nor any person acting on behalf of any of them:

(a) makes any warranty or representation whatsoever, express or implied,
- with respect to the use of any information, apparatus, method, process, or similar item disclosed in this document, including merchantability and fitness for a particular purpose, or
- that such use does not infringe on or interfere with privately owned rights, including any party's intellectual property, or
- that this document is suitable to any particular user's circumstance; or

(b) assumes responsibility for any damages or other liability whatsoever (including any consequential damages, even if Project Coordinator or any representative of a signatory party of the OpenMusE Project Consortium Agreement, has been advised of the possibility of such damages) resulting from your selection or use of this document or any information, apparatus, method, process, or similar item disclosed in this document.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Glossary

CCSIs: Cultural and Creative Sectors and Industries
CMO: Collective Management Organisation
CoLiDeS: Comprehension-based Linked Model of Deliberate Search
CSO: Civil Society Organisation
CTA: Concurrent think-aloud tests
CWW: Cognitive walkthrough for the web
CW3: The third variation of Wharton et al' cognitive walkthrough method
ECW: Enhanced cognitive walkthrough

- **U:** ECW problem category "User"
- **H:** ECW problem category "Hidden"
- **T:** ECW problem category "Text and Icon"
- **S:** ECW problem category "Sequence"
- **P:** ECW problem category "Physical Demands"
- **F:** ECW problem category "Feedback"

GDPR: General Data Protection Regulation
LMC: Live Music Census
LSA: Latent semantic analysis
NGO: Non-governmental organization
OMO: Open Music Observatory
RTA: retrospective think-aloud
SCW: streamlined cognitive walkthrough
SKCMDb: Slovak Comprehensive Music Database
SME: Small or medium enterprise
SUS: system usability scale

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Executive Summary

This deliverable summarises the findings of the OpenMusE usability testing sequence, which entailed the following activities:

- Workshop 1 (June 2025): online cognitive walkthrough of the early-release outputs with the majority of the consortium
- Workshop 2 (October 2025): online cognitive walkthrough of the early-release outputs with selected consortium partners (focus on semantic analysis)
- Individual end-user tests (September 2025): concurrent think-aloud tests of selected market-ready outputs conducted with N=13 external end-users

The two workshops, conducted with project partners, found a number of minor issues but generally suggested good usability at the internal review stage.

The individual usability tests, conducted with N=13 external end-users, identified several more basic challenges, consistent with the lower contextual familiarity of external users relative to consortium partners. Across the two tools, the key findings are as follows.

Live Music Census Dashboard and Maps. The dashboard and maps performed well overall. All 13 participants completed three of the four assigned tasks at a 100% completion rate; the fourth (locating a venue count via the map) reached 85% completion but was complicated by an unexpected display issue on some devices. The System Usability Scale (SUS) score of 76/100 places the tool in the "good" usability range. Positive user attitudes (engagement, liking, and perceived intuitiveness) predominated across all participants. The most frequently occurring problem types were sequencing issues (17% of tasks), hidden affordances (15%), and label confusability (15%), concentrated primarily in the audience and venue dashboard section rather than the maps. Participants consistently highlighted the value and novelty of the data, and expressed strong interest in re-use, particularly if an export function (Excel/CSV) were available.

Finno-Ugric Sampo. The Sampo presented greater usability challenges. The average task completion rate was 63% across five tasks, with one task (exporting a list of songs) failing entirely for all 13 participants due to a mislabelled function that launches a SPARQL query rather than a data export. The SUS score of 61/100 indicates above-average but improvable usability. Both negative and positive attitudes were expressed by all participants, with complexity (92% of users) and annoyance (85%) the most prevalent negative dimensions. The main structural issues are: a security warning on the domain blocking access for some users; inconsistent filter presentation; critical links hidden inside the back-end wiki rather than surfaced in the main interface; and an export function that does not perform as labelled.

The **priority actions** arising from the testing are: (1) resolve the Sampo's domain security issue and correct the export function labelling; (2) add an export function (Excel/CSV) to both the Live Music Census maps and dashboards; (3) fix the display/zoom issue affecting the dashboard on some devices; (4) add an information button to the dashboards explaining the sample and filter logic; and (5) surface critical links directly within the Sampo interface rather than burying them in the wiki back-end. These interventions are directly applicable to the forthcoming Slovak Comprehensive Music Database (SKCMDb) and the Cultural Access and Participation (CAP) dashboards, which share the same underlying design. The findings therefore extend beyond the two tools tested and will inform the final configuration of the Open Music Observatory as a whole.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Table of Contents

Table of Contents 7
1 Introduction 9
2 Description of the object of evaluation 12
2.1 Formal name, release/version, and parts evaluated 12
2.2 User groups for which the object is intended 13
2.3 Brief description of the object and its purpose 13
2.3.1 The Live Music Census Dashboard and Maps 13
2.3.2 The Finno-Ugric Sampo 13
2.4 Intended context of use 14
2.5 Expected impact of the object 14
3 Purpose of the evaluation 15
3.1 Description of the purpose 15
3.2 Functions and components evaluated 15
4 Workshops 1 & 2 (June 2025 and October 2025): Cognitive walkthrough for the web (CWW) ... 16
4.1 Method 16
4.1.1 General: Type of evaluation used and sufficient information to replicate the evaluation 16
4.1.2 Evaluators/participants 20
4.1.3 Tasks 21
4.1.4 Evaluation environment 22
4.2 Procedure 22
4.2.1 Design of the evaluation 22
4.2.2 Data to be collected 23
4.3 Results 24
4.3.1 Data analysis 24
4.3.2 Presentation of the results 24
4.3.3 Interpretation of results and recommendations 27
5 Individual usability testing (September 2025): Concurrent think-aloud (CTA) 30
5.1 Method 30
5.1.1 General: Type of evaluation used and sufficient information to replicate the evaluation 30
5.1.2 Evaluators/participants 31
5.1.3 Tasks 32
5.1.4 Evaluation environment 33
5.2 Procedure 33

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

5.2.1 Design of the evaluation 33
5.2.2 Data to be collected 34
5.3 Results 37
5.3.1 Data analysis 37
5.3.2 Presentation of the results – Live Music Census Dashboard and Maps 38
5.3.3 Presentation of the results – Finno Ugric Sampo 40
5.3.4 Interpretation of results and recommendations 43
6 Ethical considerations 46
7 Conclusion 47
References 48
Annexes 50

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

1 Introduction

This task and deliverable adapt and develop a highly standardised, **ISO-compliant usability testing** method and workflow developed by task lead SINUS and successfully applied in two prior research and innovation projects[1]. **Usability** is defined in *ISO 9241-11:2018 Ergonomics of human-system interaction* 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" (2018). It is generally not treated as a single construct that can be measured directly, but rather as a composite of several constructs: in the ISO definition, these are:

- **Effectiveness**: "the accuracy and completeness with which users achieve specified goals"
- **Efficiency**: "the resources used in relation to the results achieved" (e.g., time, human effort, cost, and materials)
- **Satisfaction**: "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"

Another key dimension is **accessibility**, "the extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of user needs, characteristics and capabilities to achieve identified goals in identified contexts of use" (ISO 2018). Additional dimensions must be determined on a case-by-case basis depending upon the characteristics of the system being evaluated, the aims of the designers, the needs of the users, and the context of use, among other factors; examples are fun, aesthetics, sociability, and flow (Hornbaek 2006, pg. 80).

Based on the above definitions, *ISO/IEC 25066:2016 Systems and software engineering* suggests the following indicators for measuring the three core usability constructs:

Table 1. *Indicators for the dimensions of usability*

Indicators for the dimensions of usability		
Effectiveness	**Efficiency**	**Satisfaction**
Tasks completed	Task time	Overall satisfaction
Objectives achieved	Time efficiency	Satisfaction with features
Errors in a task	Cost efficiency	Discretionary usage
Tasks with errors	Productive time ratio	Feature utilisation
Task error intensity	Unnecessary actions	Proportion of users complaining
	Fatigue	Proportion of user complaints about a particular feature
		User trust
		User pleasure
		Physical comfort

[1] These were PERCEPTIONS (https://cordis.europa.eu/project/id/833870) and COVINFORM (https://cordis.europa.eu/project/id/101016247).

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

ISO/IEC 25066:2016 Systems and software engineering furthermore specifies minimum requirements for different types of usability evaluations:

Table 2. *Requirements for usability evaluations*

Requirements for usability evaluations				
Type of evaluation	**Inspection**	**User observation**		
		Observing user behaviour	**Measuring user performance and response**	**Information from users**
Physical environment and facilities	N/A	Required	Required	Optional
Technical environment	**Required**	**Required**	**Required**	**Required**
Evaluation administration tools	Recommended	Recommended	Recommended	Recommended

As the OpenMusE outputs are technical environments, this task utilises the following methods to meet the following requirements and measure the following constructs:

Table 3. *Requirements, methods, constructs, and data types*

Requirements, methods, constructs and data types			
Requirement	**Method**	**Construct**	**Data type**
Inspection	Cognitive walkthrough for the web (Blackmon et al. 2002)	Effectiveness, efficiency, satisfaction, accessibility	Qualitative
Observing user behaviour	Concurrent think-aloud (Ericsson and Simon 1993) with exit interview	Effectiveness, efficiency, satisfaction, accessibility	Qualitative
Measuring user performance	Quantitative performance metrics	Effectiveness, efficiency	Quantitative
Measuring user response	System usability scale	Satisfaction	Quantitative
Information from users	Socio-demographic questionnaire	User attributes	Quantitative

The tests were conducted in three iterations:

- Iterations 1 (June 2025) and 2 (October 2025) comprised two cognitive walkthroughs in the form of online workshops

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- Iteration 3 (September 2025) comprised online concurrent think-aloud tests with supplementary quantitative questionnaires and qualitative interviews

This deliverable describes the purpose of the evaluation and the evaluation methods, then presents the results, along with interpretation and recommendations. It adopts the report structure recommended in *ISO/IEC 25066:2016 Systems and software engineering*.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

2 Description of the object of evaluation

The objective of the software validation study was to test the Open Music Observatory (OMO), which is envisioned as a network of federated, interoperable music data sharing spaces and data discovery tools. The tests assemble developer and user feedback and help ensure that the outputs are usable by the target group.

REPREX described the intended Open Music Observatory infrastructure in D5.1 “Open Music Observatory Building an open data sharing space for the European music sector”, submitted on 31.05.2025. REPREX was requested by the Coordinator and WP1-4 leaders to upload into this infrastructure very small data subsets representative of the types of data collected during the project:

- Statistical data in tabular form: e.g., a music-relevant Eurostat table, e.g. re: NACE 32.2 Manufacture of musical instruments.
- Survey microdata: e.g., a subset of the SurveyHarmonies dataset collected by SINUS and REPREX in a separate project in 2022.
- Repertory data: e.g., an anonymised subset of the Slovak repertory data provided to REPREX by SOZA in 2023.
- Streaming data: e.g., a subset of the Spotify data collected by UVA and REPREX in 2024.

Such subsets could have allowed a walkthrough of the Open Music Observatory's functionality. In the Coordinator's and WP1-4 leaders' view, this would have supported iterative development. However, REPREX did not agree with this view, instead preferring to keep the Observatory empty and to refrain from demonstrating it for the Consortium until all full datasets collected by all partners had been provided and extensively documented. As an alternative, REPREX offered for usability testing a use case (the Finno-Ugric Sampo) which was built on the same infrastructure as the Observatory, and which used a very similar graphical user interface as the Observatory’s intended repertory data explorer. Due to this similarity, insight gained by testing this exploitation case is directly applicable to the Observatory itself.

In addition to the Finno-Ugric Sampo, it was decided to test the Live Music Census Dashboards and Maps. These dashboards were originally created as a mitigation strategy to balance and complement the REPREX Open Music Observatory infrastructure’s perceived focus on repertory data rather than other kinds of data, as well as its technical complexity. It therefore seemed to be usable only by a narrow part of the target group with highly specialized technical skills. Further dashboards will be developed by project end, especially on the results of the cultural access and participation (CAP) study in WP3. The tested Live Music Census dashboards are very similar in design to the CAP dashboards, and insights gathered from the software validation study will thus improve their usability and facilitate development.

2.1 Formal name, release/version, and parts evaluated

The Live Music Census Dashboard and Maps are integrated into the OpenMusE website. The dashboards have been described in “D4.4 -Survey Data Collection and Management Methods and Tools”.

The Finno-Ugric Sampo as the “front-end” of the Finno-Ugric Data Sharing Space is use case of the Slovak Comprehensive Music Database (SKCMDb). The SKCMDb is developed under T 2.3 as part of the OMO. Only the music-related parts of the tool were tested, as they will be almost identical in the

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

SKCMDb. At the time of testing, the Finno-Ugric Sampo and SKCMDb had not yet been integrated into the OpenMusE centralized website but should be in accordance with user suggestions (and the more general suggestion of online presence integration made in the interim review).

2.2 User groups for which the object is intended

The object is intended for the following user groups, as identified previously in OpenMusE D6.2 “Ethics Framework” (Section 2.1.1):

- TG1: Music industry – representative organisations (CMOs, music exporters; etc.)
- TG2: Music industry – commercial entities (labels; publishers; promoters; music tech SMEs)
- TG3: Music industry – individuals (artists, DJs, producers, engineers, etc.)
- TG4: Music CSOs and non-profits (educators; non-profits; musical amateurs; etc.)
- TG5: Policymakers & policy workers (EU; national; regional; municipal; third-country)
- TG6: Scientific community
- TG7: Adjacent industries (CCSIs; data venders; market researchers/consultants; etc.)
- TG8: General public

2.3 Brief description of the object and its purpose

2.3.1 The Live Music Census Dashboard and Maps

The **Live Music Census Dashboards** provide intuitive, web-based tools for exploring the results of the T3.3 Live Music Census pilot study. It provides the possibility to view response patterns, making demographic comparisons, and identifying trends without requiring advanced technical skills. There are two inter-linked dashboards:

- **The audience dashboard** enables comparisons across cities and gender for audience survey responses, such as spending data, information on transportation or barriers to attending live music events.
- **The venue dashboard** provides similar comparisons across cities for venue survey results, e.g. challenges faced by the venues, or attitudes concerning the live music scene in their city

The Dashboards are further described in D4.4 “Survey Data Collection and Management Methods and Tools”.

The Live Music Maps contain a geographical depiction of all live music venues as of October 2024 in the five target cities of the T3.3 Live Music Census pilot study, i.e. Heidelberg, Helsinki, Lviv, Mannheim and Vilnius. Users have the possibility to select a city, search by venue name or filter by venue size and type. When clicking on a venue, the user can access further information about that venue, such as contact information or links to their website or social media accounts.

2.3.2 The Finno-Ugric Sampo

The Finno-Ugric Sampo portal is an interactive gateway to the Finno-Ugric Data Sharing Space—a collaborative platform for exploring and organizing cultural data across domains such as music, film, literature, and traditional textiles.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

This semantic browser is intended to be used for the Open Music Observatory. It allows the search of datasets related to the Finno-Ugric culture together with related publications, question banks, and musical works in print or recorded circulation. It will be the convenience user interface for connected data that is placed on the EU Open Data Portal, Europeana, EOSC, and the OMO Wikibase instance.

The Finno-Ugric Sampo has the following functional sub-pages, which have been included in the software-validation study:

- Music: this subpage allows the user to browse through Finno-Ugric music, listen to the songs on Spotify or access further meta-data in the “Wiki”-backend, the latter two via external links. The users can filter the content by artist, language or location.
- Publications: This subpage gives users the possibility to access publications on the Finno-Ugric music and culture, as well as further meta-data in the “Wiki”-backend. As with the music-tab, multiple filter options exist to facilitate the search of publications.
- Feedback: The feedback-tab allows users to contact REPREX to provide feedback
- About: This page explains the use of and the technology behind the portal
- Project Blog: This links to a blog entry on the REPREX website

Further tabs exist, but have not been included in the study as they focus on cultural items outside the topic of music, and therefore go beyond the envisioned functions of the SKCMDb:

- Films
- Clothing
- Photos

Further information on the SKCMDb, and potentially the Finno-Ugric Sampo, will be available in D2.3 “Report on the pilot study on music diversity and circulation”.

2.4 Intended context of use

The LMC dashboard and maps, and the Sampo are best viewed using a personal computer, due to the complexity of their functions. Mobile usability is limited.

2.5 Expected impact of the object

The Live Music Census Dashboards and Maps, as well as the Finno-Ugric Sampo provide a novel service for music industry stakeholders, and the wider scientific and education communities. The tools offer interactive means of accessing information on live music consumption, live music venues and cultural heritage, as well as access to a wide range of data points that are otherwise inaccessible in a single place. The dashboards and maps could be used to explore data on numerous aspects of audience interaction with live music, challenges and opportunities facing the live music sector, etc. The Finno-Ugric Sampo could be used to radically improve the discoverability of underrepresented minority repertories and communicate their intrinsic cultural value and irreplaceable contribution to the rich diversity of musical life in Europe.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

3 Purpose of the evaluation

3.1 Description of the purpose

The purpose of the evaluation was twofold:

- To provide user input during an iterative development process, allowing for the continual improvement of the OpenMusE tools from the perspective of relevance and usability.
- To assess the functionality and usability of the final release.

3.2 Functions and components evaluated

In case of the Live Music Census dashboard and maps, all functions and components were evaluated. Concerning the Sampo, as the Finno-Ugric Sampo was an exploitation case going beyond the use of music, it was decided to only evaluate the music-related functions that would also be available in the Slovak Sampo.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

4 Workshops 1 & 2 (June 2025 and October 2025): Cognitive walkthrough for the web (CWW)

4.1 Method

4.1.1 General: Type of evaluation used and sufficient information to replicate the evaluation

The **cognitive walkthrough**, which is based on the *theory of facilitated learning by exploration* in the field of human-computer interactions, was developed as a means of evaluating "walk-up-and-use" interfaces, i.e., interfaces that do not require special training (Polson & Lewis 1990; cf. Lewis et al. 1990). It is a review process in which the developers and designers of an interface present the product to a group of peers representative of the stakeholder group for which the interface is intended, who then evaluate it using structured tasks and criteria (Wharton et al. 1994).

The developers of the cognitive walkthrough technique acknowledge that it is first and foremost a method of evaluating ease of learning, and that "use of the Cognitive Walkthrough as the only method for evaluating an interface would push design trade-offs in an interface in the direction of ease of learning"; they thus recommend pairing the cognitive walkthrough with other evaluation techniques throughout the development process (Wharton et al. 1993, pg. 4). For this reason, the cognitive walkthrough is supplemented with concurrent think-aloud (CTA) tests.

In part due to the recognition of such trade-offs, the cognitive walkthrough method has evolved over the years, generating several variations. The third variation on the original authors' cognitive walkthrough method – referred to as CW3 – responds to criticisms by adopting a highly structured flow, consisting of five sequences of steps (adapted from Wharton et al. 1993, pg. 2-3):

1. Define inputs to the walkthrough
    a. Who are the target users?
    b. What tasks will be analysed? Note that the benchmark tasks should be as realistic as possible and must include descriptions of context.
    c. What is the proper action sequence for each task, and how is it described? A rule of thumb is that the actions should be described at the same level of detail as a basic tutorial.
    d. How is the interface defined or implemented? The cognitive walkthrough process can be conducted using mock-ups or alpha implementations.
2. Convene the analysts. These peers can include other developers/designers/engineers, but also other stakeholders. One peer should act as a facilitator, and another should be responsible for taking memos; the others can contribute insight from their domains of expertise in a relatively open manner.
3. Walk through the action sequence for each task
    a. Tell a credible story, considering:
        i. Will the user try to achieve the right effect?
        ii. Will the user notice that the correct action is available?

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

iii. Will the user associate the correct action with the effect they are trying to achieve?

iv. If the correct action is performed, will the user see that progress is being made toward the solution of their task? Note that the worst-case scenario here is receiving no feedback on the status of the task.

4. Record critical information

   a. User knowledge requirements: should be defined at the onset of the walkthrough as a whole and at the onset of each task.

   b. Assumptions about the user population: should be defined at the onset of the walkthrough.

   c. Notes about side issues and design changes: should be recorded throughout the walkthrough.

   d. The credible success story, founded on common features of success, e.g.:

      i. Users may know what effect they want to achieve.

      ii. Users may know what actions are available.

      iii. Users may know what actions are appropriate for the effect they are trying to achieve.

      iv. Users may know things are going OK / as expected after an action.

***Note that failure stories are also a part of cognitive walkthroughs. Common failure points can include users not knowing prerequisite actions or states; users not associating the required action with the desired effect, in context; users not being aware of certain functions; users confusing two or more functions; users misrecognising visual elements in the UI; users not knowing that certain follow-up actions are needed; users not knowing how to determine whether a task was successful; users "timing out"; users being unable to physically execute an action (e.g., holding down three keys simultaneously); and users being unaware that terminator actions are required (e.g., pushing the pound key after entering a number).*

5. Revise the interface to fix the problems

Other teams of developers and HCI scholars have proposed variations on the original cognitive walkthrough. The heuristic walkthrough (HW; Nelson & Molich 1990), for instance, consists of two phases: phase 1 follows the question flow of CW3, while phase 2 is a free exploration of the system supported by "heuristics" derived from the heuristic evaluation method. The streamlined cognitive walkthrough (SCW; Spencer 2000) trims CW3's five questions down to two:

1. Will users know what to do at this step?
2. If users do the right thing, will they know that they did so and are making progress toward their goal?

The enhanced cognitive walkthrough (ECW; Bligard & Osvalder 2013) takes the opposite approach, supplementing CW3 by adding two sequences of analysis questions, one sequence on analysis of functions and one sequence on analysis of operations (pg. 7-8):

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

1. Analysis of functions
    a. Will the user know that the evaluated function is available? Does the user expect, on the basis of previously given indications that the function exists in the machine?
    b. Will the user be able to notice that the function is available? Does the machine give clues that show that the function exists?
    c. Will the user associate the clues with the function? Can the user's expectations and the machine's indications coincide?
    d. Will the user get sufficient feedback when using the function? Does the machine give information that the function has been chosen and to what position the user is in the interaction?
    e. Will the user get sufficient feedback to understand that the function has been fully performed? Does the user understand, after the performed sequence of actions, that the right function has been performed?
2. Analysis of operations
    a. Will the user try to achieve the right goals of the operation? Does the user expect, on the basis of previously given indications, what is to be performed?
    b. Will the user be able to notice that the action of the operation is available? Does the machine give clues that show that the action is available and how to perform it?
    c. Will the user associate the action of the operation with the right goal of the operation? Can the user's assumed operation and the machine's indications coincide?
    d. Will the user be able to perform the correct action? Do the abilities of the user match the demands by the machine?
    e. Will the user get sufficient feedback to understand that the action is performed, and the goal is achieved? Does the user understand, after the performed operation, that he/she has done right?

The ECW furthermore suggests utilising a structured typology of problems during the analysis: for instance, in the example case study of testing a home-care ventilator, the following problem types are identified (pg. 8):

- User (U): The problem is due to the user's experience and knowledge, possibly because the user is accustomed to different equipment.
- Hidden (H): The interface gives no indications that the function is available or how it should be used.
- Text and icon (T): Placement, appearance and content can easily be misinterpreted or misunderstood.
- Sequence (S): Functions and operations have to be performed in an unnatural sequence.
- Physical demands (P): The interface sets too high demands on users' physical speed, motoric skill and force.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- Feedback (F): The interface gives unclear indications of what the user is doing or has done.

Note that the analysis questions and problem types identified in the ECW are not all relevant to purely online platform use cases: e.g., the operations analysis question "do the abilities of the user match the demands by the machine" and the problem type "physical demands" are unlike to arise in an online platform use case, except with regard to vision-impaired users.

The need to more precisely tailor the CW method to online use cases gave rise to another variant, the cognitive walkthrough for the web (CWW). CWW is based on a learning model developed with reference to online search, CoLiDeS or Comprehension-based Linked Model of Deliberate Search (Kitajima, Blackmon, & Polson 2000). CoLiDeS holds that "problem solving processes, guided by users' goals and *information scent*, drive users' information seeking or search behaviours when exploring a new website or carrying out a novel task on a familiar website" (Blackmon et al. 2002, pg. 464). The CWW tailors the CW3 sequence of testing steps to tasks specifically related to online information seeking and search behaviours, including through the integration of analytical techniques suited to text-based interface analysis (pg. 465-466):

1. Compile a set of realistic user goals and intended selections, optionally from the perspective of defined user personas. This step corresponds to step 1a to 1c in the CW3.
2. Use latent semantic analysis (LSA), a technique in natural language processing, to estimate semantic similarity of user goals and page/subregion/widget headings/link labels. This "pre-analytical" step extends step 1d in the CW3. Blackmon et al. (2002) suggest using the LSA website developed by the University of Colorado Boulder (http://lsa.colorado.edu/), which can generate estimates of the semantic similarity of the user goals and the heading/link labels in cosine form. Note that this analysis requires some familiarity with the theory behind latent semantic analysis. In principle, other quantitative and/or qualitative semantic analytical methods could be substituted.
3. Identify problematic headings/link labels. This step also extends step 1d in the CW3.
4. Find goals-specific problems. This step picks up step 3a/4a of the CW3, the "credible success story." CWW modifies the CW3 sequence of analysis questions as follows:
    a. Will the correct action be made sufficiently evident for the user?
    b. Will the user connect the correct action's description with what he or she is trying to do?
        i. Will the user correct subregion of the page with the goal using heading information and her understanding of the site's page layout conventions?
        ii. Will the user connect the goal with the correct widget in the attended-to subregion of the page using link labels and other kinds of descriptive information?
    c. Will the user interpret the system's response to the chosen action correctly?

Like the ECW, the CWW also suggests identifying specific types of problems relevant to the goals under evaluation, e.g.:

- Unfamiliar headings/link labels: the user does not know what the label means
- Confusable headings/link labels: the label could easily be confused with another label

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- Goal-specific competing headings/link labels: the similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal

The efficacy of the CWW in predicting actual user behaviour has been tested experimentally, e.g. by checking the first-click success rate of the unfamiliar, confusable, and competing labels identified in the CWW; the method's predictive power varies contextually but has generally been found to be good (Larson & Czerwinski 1998; Blackmon et al. 2002).

4.1.2 Evaluators/participants

Total number and segmentation of evaluators/participants

The workshops were conducted by interdisciplinary team of project partners. A representative from the majority of consortium partner (N=11) participated in Workshop 1 (June 2025), whereas a smaller group of N=3 partners were chosen to participate in Workshop 2 (October 2025), which represented several project roles and end-user fields.

Key characteristics of evaluators/participants and differences to the user population

The participants were aware of the project aims and generally aware of all project activities. Awareness of the specific aims, functions, and components of the object evaluated varied widely. In contrast, it is to be expected that users outside the consortium will not be aware of the project aims and activities. Hence, the individual usability testing will be conducted with external users completely unfamiliar with the project.

Table of participants by characteristics

Table 4. Workshop 1 participants (June 2025)

Workshop 1 participants			
Partner	**Country**	**Field**	**Workshop Role**
SINUS	Germany	Coordinator / Research	Facilitation
REPREX	Netherlands	Developer	Participant
UTU	Finland	Developer / Academia	Participant
SYNYO	Austria	Developer / Communication & Dissemination	Participant
SOZA	Slovakia	CMO	Participant
MEU	Ukraine	Music Export	Participant
MXF	Lithuania	Music Export	Participant
UVA	Netherlands	Academia	Participant
SSSA	Italy	Academia	Participant
MUSICAUTOR	Bulgaria	CMO	Participant
EUBA	Slovakia	Academia	Participant

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Table 5. Workshop 2 participants (October 2025)

Workshop 2 participants			
Organisation	**Country**	**Field**	**Role**
SINUS	Germany	Coordinator / Research	Facilitation
SYNYO	Austria	Developer / Communication & Dissemination	Participant
UTU	Finland	Developer / Academia	Participant

4.1.3 Tasks

Tasks used and task scenarios

During workshop 1 (June 2025) the use cases and possible tasks were discussed. After the first feedback generated during the workshop was adapted to the websites, SINUS defined specific tasks to reflect the discussions during the workshop.

Table 6. Workshop 1 tasks (June 2025)

Workshop 1 tasks		
Task	**Description**	**Scenario (use flow)**
Live Music Census Maps & Dashboard		
1	*Please find the Live Music Map of Helsinki.*	*Select "Live Music Census" → click "Live Music Maps" → Select "Helsinki → Select "bars" → Find the number of bars*
2	*Please, locate the bar "Filmriss" in Mannheim and visit its Instagram page.*	*Select "Live Music Census" → click "Dashboard" → Type "Filmriss" → click on the venue → click on the Instagram link*
3	*Please find out the gender distribution of live music audiences in Lviv.*	*Select "Live Music Census" → click "Dashboard" → Select "Lviv" → Select all genders → scroll down the audience survey results to find the gender distribution*
4	*Please compare live music venues in Vilnius and Helsinki: Which top three factors affect the respective venues most negatively?*	*Select "Live Music Census" → click "Dashboard" → Select "Vilnius" and "Helsinki" → scroll down the venue survey results to find the challenges and compare.*
Finno-Ugric Sampo		
5	*Please find out what this website is about.*	*Click on "Info" → Select "About this page"*
6	*Please find a song in Komi and open it on Spotify.*	*Select "Music" → Click on "Language" → select "Komi" → chose a song → click on "Spotify Track ID"*
7	*Please find out more information about a song, any song, by Hilda Griva.*	*Select "Music" → Click on "Performer" → select "Hilda Griva" → chose a song → Click on the title*
8	*Please export a list of songs from Talsi.*	*Select "Music" → Click on "Location" → select "Talsi" → Click on "Export"*
9	*Please find the Journal Article on the Revival of Finno-Ugric Studies and open or download it*	*Select "Publications" → Click on "Type of Publication" → select "Journal Article" → Click on "Revival …" → Scroll to URL → Open the link*

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Selection criteria and sources of the tasks

The workshop participants discussed the different functions of the “Live Music Census dashboards and maps”, as well as the “Finno-Ugric Sampo”. Possible use cases for each function were formulated, judged relevant to the likely target users. The workshop focused on the sections “Live Music Maps”, “Audience and Venue Dashboards” on the Live Music Census side, and on the sections “About”, “Music” and “Publications” in the Finno-Ugric Sampo.

Task data given to each evaluator/participant and/or inspector; criteria for task completion and abandonment for each task

Not applicable.

4.1.4 Evaluation environment

Physical and technical environments

Workshop 1 was conducted online on 30 June 2025. Workshop 2 was conducted online on 20 October 2025. Both workshops were recorded and transcribed.

Evaluation administration tools and administrators

The evaluation was administered, recorded, and analysed by SINUS.

4.2 Procedure

4.2.1 Design of the evaluation

Description of the evaluation design

The cognitive walkthrough lasted approximately 1.5 hours. The agenda is shown in the following tables.

Table 7. CWW agenda workshop 1 (June 2025)

CWW agenda workshop 1		
Time	**Phase**	**Phase name**
T	A	Confirm testing objects (i.e. the dashboards, maps and sampo)
T+20	A	Confirm target user groups
T+30	A	Confirm assumptions about target users
T+40	A	Confirm user knowledge requirements
T+60	B	Define user tasks
T+80	C	Identification of user-task-specific problems
T+90	End	Conclusion

Rather than integrating a latent semantic analysis (LSA) into the CWW agenda, a qualitative semantic analysis of the headings/link labels on the platform has been performed during the second workshop (October 2025) by a smaller sub-set of partners (SINUS, UTU, SYNYO). This fulfils the intended purpose of assessing the semantic similarity of page/subregion/widget headings/link labels – e.g., their potential confusability.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Independent variables; predefined criteria for inspection or observation; measures used in evaluation; and/or operational criteria of criteria/measures

Phase A and B of the walkthroughs did not entail inspection/observation/evaluation.

During Phase C, in accordance with the CWW format (Blackmon et al. 2002), for each task, the participants discussed the following "credible success story" criteria:

1. Will the user try to achieve the right effect?
2. Will the user notice that the correct action is available?
3. Will the user associate the correct action with the effect they are trying to achieve?
4. If the correct action is performed, will the user see that progress is being made toward the solution of their task? Note that the worst-case scenario here is receiving no feedback on the status of the task.

Interaction between individuals taking part in each evaluation session

There were no specific rules related to interaction between individuals taking part in the session, aside from standard norms of professional conduct.

General instructions given to participants and specific instructions on tasks

Prior to each evaluation phase, participants were informed of the aim of the phase. Participants were not given specific instructions on tasks.

Sequence of activities for conducting the evaluation

The outline of activities is provided above under "Description of the evaluation design." The full protocol template used in the evaluation is provided in Annex 2. The protocol generally follows the CWW format suggested by Blackmon et al. (2002), borrowing certain elements as needed from the CW3 (Wharton et al. 1993) and ECW (Bligard & Osvalder 2013).

4.2.2 Data to be collected

Usability defects in terms of deviations from predefined criteria

Phases A and B did not collect data on usability defects, nor deviations from predefined criteria; rather, these phases collected data on target user groups and knowledge requirements, then identified relevant tasks for the evaluation itself.

During Phases C, a structured problem typology was utilised that synthesises the CWW with problem types proposed in the ECW (Bligard & Osvalder 2013):

- User (U): The problem is due to the user's experience and knowledge.
- Hidden (H): The interface gives no indications that the function is available or how it should be used.
- Text and icon, including headings/link labels (T): Placement, appearance and content can easily be misinterpreted or misunderstood. For sub-categories, see above.
- Sequence (S): Functions and operations have to be performed in an unnatural sequence.
- Feedback (F): The interface gives unclear indications of what the user is doing or has done.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- Other (O)

During the qualitative semantic analysis of the platform (workshop 2), the headings/link labels were checked against the task descriptions, and unfamiliar, confusable, or goal-specific competing labels were flagged using a grid. Following Blackmon et al. (2002), the following categories of text and icon problem were identified (note that these categories sometimes overlap and classification is thus subjective):

- Unfamiliar text/icon: the user does not know what the label means
- Confusable text/icon: the label could easily be confused with another label
- Goal-specific competing text/icon: the similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal
- Other

Observed user behaviour; observed performance data; user-reported qualitative data; user-reported quantitative data

Not applicable.

4.3 Results

4.3.1 Data analysis

Approach used for the analysis of observed, measured, or collected data

The semantic legibility of the headings, link labels, and other critical text was assessed quantitatively (percentage of semantic pairs featuring problems of certain types) and qualitatively (description of the errors, generated through discussion).

The “credible success story” criteria per task (Phase C) were assessed quantitatively (percentage of criteria featuring a potential problem) and qualitatively (description of the use flow and any potential problems, generated through discussion).

Differences in planned and collected data

No differences arose between the planned and actual data collection.

Portion of data used in the analysis

All data produced was used in the analysis.

Data scoring; data reduction; statistical analyses

Not applicable.

4.3.2 Presentation of the results

Confirmation of target user groups, assumptions about target users, and knowledge requirements

During workshops 1, the participants narrowed down the target group described in section 2.2, to better match the functions of the two platforms. As the use cases largely differed between the Live Music Census dashboard and maps, and the Finno-Ugric Sampo, so did their relative target groups.

The participants agreed that the user groups are the following:

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Live Music Census Dashboard and Maps:

- TG1: Music industry – representative organisations
  - CMOs
- TG2: Music industry – commercial entities
  - Music managers
  - Concert promoters
- TG3: Music industry – individuals
  - Artists, etc.
- TG4: Music CSOs and non-profits
  - Teachers in music-related fields
  - NGOs or advocates in culture
- TG5: Policymakers & policy workers
  - Local politicians, especially on a municipal level
- TG6: Scientific community
- TG7: Adjacent industries
  - City planners
  - Market researchers

Finno-Ugric Sampo:

- TG1: Music industry – representative organisations
  - CMOs
  - Music export organisations
- TG6: Scientific community
- TG7: Adjacent industries
  - Radios, looking for local music
  - Librarians

In both cases, all participants furthermore agreed upon the following assumptions about the target users and their knowledge requirement:

- Users can be expected to be generally familiar with the concept of quantitative indicators
- Users must read English on a comparatively high level – this could be a barrier for some user groups, such as teachers.
- Users cannot be expected to have specialist knowledge in data science, statistics or similar fields.

Usability defects and potential usability problems likely to arise: Credible success stories per task

The participants talked through possible barriers of the tasks identified in Phase B and discussed areas of improvement. The qualitative inputs were subsequently analysed using the “credible success story” criteria developed by Bligard & Osvalder (2013).

The analysis found that of 45 total items, 89% featured no potential usability problems, whereas 11% did feature potential usability problems.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Table 8. Evaluation of credible success stories per task

Evaluation of credible success stories per task			
Code	**Problem type**	**Count**	**% of total**
NA	No usability problems	40	89%
U	The problem is due to the user's experience and knowledge.	0	0%
H	The interface gives no indications that the function is available or how it should be used.	1	2%
T	Placement, appearance and content can easily be misinterpreted or misunderstood. For sub-categories, see semantic analysis.	2	4%
S	Functions and operations have to be performed in an unnatural sequence.	1	2%
F	The interface gives unclear indications of what the user is doing or has done.	0	0%
O	Other	1	2%
Total items		45	100%
Total issues		5	11%

The full findings of the task evaluations are included in Annex 2 (T4.4_CWW_credible_success_stories).

Usability defects and potential usability problems likely to arise: Semantic analysis

The semantic analysis utilised a matrix by which all link labels present on each page were cross-checked for potential confusion or semantic clashes. For the Live Music Census Dashboards and Maps, the workshop 2 analysis found that of 711 total label pairs, 92% were understandable and did not clash, whereas 8% were either potentially confusing or potentially clashing. The specific breakdown of problem types, utilising categories developed by Blackmon et al. (2002), is as follows:

Table 9. Semantic analysis findings -Live Music Census

Semantic analysis findings – Live Music Census			
Code	**Problem type**	**Count**	**% of total**
OK	No semantic issue	711	92%
UT	Unfamiliar text/icon: the user does not know what the label means	0	0%
CT	Confusable text/icon: the label could easily be confused with another label	3	0%
GSCT	Goal-specific competing text/icon: the similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal	0	0%
O	Other	57	7%
Total pairs		771	100%
Total issues		60	8%

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

The full findings of the semantic analysis are included in Annex 2 (T4.4_CWW_semantic_analysis_LMC).

In the case of the Finno-Ugric Sampo, only the pages relevant to music and the OpenMusE project were analysed. The analysis found that of 433 total label pairs, 99% were understandable and did not clash, whereas 1% was either potentially confusing. The specific breakdown of problem types is as follows:

Table 10. Semantic analysis findings – Finno-Ugric Sampo

Semantic analysis findings – Finno-Ugric Sampo			
Code	**Problem type**	**Count**	**% of total**
OK	No semantic issue	430	99%
UT	Unfamiliar text/icon: the user does not know what the label means	0	0%
CT	Confusable text/icon: the label could easily be confused with another label	3	1%
GSCT	Goal-specific competing text/icon: the similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal	0	0%
O	Other	0	0%
Total pairs		433	100%
Total issues		3	1%

The full findings of the semantic analysis are included in Annex 2 (T4.4_CWW_semantic_analysis_Sampo).

4.3.3 Interpretation of results and recommendations

Interpretation of results– Live Music Census Dashboard and Maps

The workshop 1 participants generally agreed that the dashboard was useful and well-designed. The possibility to analyse the data or venue maps using different filters was appreciated specifically. The visual design of the live music maps was seen as positive, as they also show live music hubs, whereas the design of the dashboards could be modernised.

Two concrete usability errors were identified:

- There was too much information on one page, which might hide the dashboards at the bottom.
- The filters in the live music dashboards were too slow, as the dashboards were loading as soon as a filter was changed, which took time.

The semantic analysis showed very few issues in the main content page itself. However, the footer of the website was a repetition of all website functions in the headers. Instead of providing short cuts, this led to an overload of content and to confusion. What is more, the analysis also looked at other

parts of the project website, which was found to be very extensive. Simplification was needed, especially concerning the different media contents and the “team”-tab.

Although not part of the later tests, the mobile version of the website was not as functional as the desktop version. Certain features did not work or were distorted. Even if it is recommended to use the dashboard on a desktop, the mobile version should not be ignored, as it provides a first impression of the website and its contents.

Recommendations – Live Music Census Dashboard and Maps

The following recommendations were made in workshop 1:

- The page should be split into two: a) the Live Music Maps and b) the Audience and Venue Dashboards. This would make a clearer distinction between the two outputs and make them more findable.
- The text should be reduced to draw more attention to the outputs themselves.
- A “select all” or “unselect all” -function should be added to the filters in the live music maps.
- The venue definitions in the maps should be updated, as they still contained “working titles” in some cases, which led to confusion
- An automatic venue count could be a useful addition to the maps
- In the dashboards, it should be possible to change your filters before updating the dashboards to speed up the process.
- The loading time of the dashboards is quite slow, which makes them very cumbersome to use. This should be sped up. If that is not possible, an indicator should be added to show that the dashboards are loading. This could be done using an icon.

The following recommendations were made in workshop 2:

- The footer should be simplified to show the project facts instead of a repetition of the headers.
- Concerning other parts of the project website, several contents could be summarized under a “project outputs”-tab, such as the project newsletter, or the media centre. Other outputs could be put highlighted under the same tab, such as the project’s publications, the videos or the podcast.
- The different sub-pages under the tab “team” should be summarized or even deleted, to reduce redundancies and outdated content.

Interpretation of results– Finno-Ugric Sampo

The workshop 1 participants liked the simplicity of the landing page and the page structure, as it makes it easy for users to navigate. It was mentioned that as a first impression, the website seems more adequate for professional users and data scientists.

Two concrete usability errors were identified:

- There was no source of information on the page and its content, which makes it more difficult for users without prior knowledge of the project to understand what the website is about.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- The export function in the “music” or “publications” section did not work. It launched a query, which would be confusing for any non-technical user.

The semantic analysis in workshop 2 showed very few issues in general. Minimal confusing text found was in the filter options in both the “music” and the “publications”-sections, which offer both a filter for “location” and “location qualifier. Between workshop 1 and workshop 2, an “about”-page had been added, which also contained minimally confusing text, namely the header “Finno-Ugric-Sampo” versus the header “Finno-Ugric-Data-Sharing-Space”. Both seem to describe the same website, which caused some confusion about the actual name of the website.

Recommendations – Finno-Ugric Sampo

The following recommendations were made in workshop 1:

- Add an “about”-page to provide information on the website and the project.
- Either make sure the “export”-function lets the user export something, e.g. a csv- or excel-file, or rename the function to avoid confusion.
- Feature media that is collected to make the website more appealing for non-technical users, such as Spotify-links.

The following recommendations were made in workshop 2:

- There should be a clear name of the website, which remains constant throughout all pages, either “Finno-Ugric-Data-Sharing-Space” or “Finno-Ugric-Sampo”. Consistency is important to avoid confusion
- The filter “location qualifier” should be either renamed or deleted completely. Since during the time of the tests, the location qualifier only offered the option “unknown”, deleting this filter is recommended.
- Certain labels within the tables in the “music” or “publications”-sections could be renamed to make them more suitable for the target group. As an example, in the music-section, users can currently find the Spotify links under “Access point”. A more commonly used term, such as “Spotify Link” or “Music Link” could be more appropriate.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

5 Individual usability testing (September 2025): Concurrent think-aloud (CTA)

5.1 Method

5.1.1 General: Type of evaluation used and sufficient information to replicate the evaluation

Think-aloud methods are a well-established method of studying "task-based cognitive processes" in general, and human-computer interactions in particular (McDonald et al. 2012, pg. 3). In response to the criticism that thinking aloud is itself a cognitive process that violates the ecological validity of the task, Ericsson and Simon (1993) systematically investigated and categorised variations on the method, identifying two basic types: concurrent think-aloud (CTA), during which research participants narrate their cognitive processes while undertaking tasks, and retrospective think-aloud (RTA), during which participants describe their cognitive processes after undertaking tasks. Academic researchers and usability practitioners surveyed by McDonald et al. (2012) rated concurrent think-aloud tests as a more effective means of identifying usability problems and determining the root causes of such problems than post-task interviews, quantitative performance metrics (e.g., task completion rate, time on task, etc.), behavioural observation, or eye-tracking (pg. 15), though many also concurred that think-aloud tests are more suited to formative than summative evaluations (insofar as one aim is to provide feedback that can be used to improve the system, rather than simply measuring its quality).

In the "classic" concurrent and retrospective think-aloud methods, based on Ericsson and Simon's work, a single user undertakes the target tasks. Ericsson and Simon recommend taking standardised measures to maximise the utility and neutrality of the narrative data elicited: these include offering warm-up / practice opportunities; keeping written scripts as neutral as possible; keeping evaluator probes as general as possible (e.g., so-called Level 1 probes such as "keep talking", rather than so-called Level 2 or 3 probes such as specific instructions or questions about thought processes); and generally minimising interaction between the evaluator and the user. Among the alternatives to classic concurrent think-aloud method are "constructive interaction" or "co-discovery" approaches, in which two (or more) users undertake the target tasks together (Miyake 1986; Barnum 2010); "speech communication" approaches, in which an extended range of naturalistic communication cues are used by the evaluator to help the participant keep talking (e.g., giving feedback in the form of "um-hum" or "acknowledgement tokens" and/or briefly asking participants to clarify statements) (Boren & Ramsey 2000); and "coaching" or "active intervention" approaches, in which the evaluator provides verbal feedback and asks specific follow-up questions about the participant's testing activities and/or underlying thought processes (Dumas & Redish 1999).

As a number of researchers have observed, neither instructional texts nor usability testers in the field universally adhere to Ericsson and Simon's recommended practices (Boren & Ramsey 2000; Nørgaard and K. Hornbæk 2006; Shi 2008). For instance, numerous instructional texts recommend that evaluators user specific probes/interventions (i.e., other than "keep talking") (McDonald et al. 2012). This can be due to the need to adapt to cultural or personality characteristics of users (Clemmensen et al. 2009), the need to maintain the test flow, or because classic think-aloud narratives tend to be shallowly descriptive, lack relevance to the "why" of usability problems, and lack actionability with regard to iterative improvement (Makri et al. 2011). Of professionals surveyed by McDonald et al. (2012), academic researchers were equally likely to adhere to Ericsson and Simon's guidelines, while usability practitioners were notably more likely to take a flexible approach; the latter in particular are

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

likely to intervene when users are stuck on tasks, ask for help, mistakenly think tasks are complete, skip over features of interest to the evaluators, or complete tasks in "unexpected" ways (pg. 14).

Among professionals surveyed by McDonald et al. (2012), classic concurrent think-aloud tests are used much more frequently than constructive interaction approaches (96% vs. 22% among academic researchers and 98% vs. 36% among usability practitioners) (pg. 8). The argued benefits of classic concurrent approaches are efficiency, relative ease of recruiting participants, relative ease of explaining the tasks and method itself to participants, and relative ease of conveying findings to stakeholders; the argued benefits of constructive interaction approaches are that they are more ecologically valid and less "inquisitive," and that the conversational data that they elicit are less artificial than a narrative monologue elicited from a solitary user (pg. 9). Several studies have sought to experimentally compare the test characteristics of different think-aloud protocols. Alshammari et al. (2015), for instance, compared CTA with RTA on three levels: success in revealing usability problems while completing tasks on a ticket booking website; task performance; and participant satisfaction with the test protocol. Both CTA and RTA evaluations followed Ericsson and Simon's guidelines. Contrary to Alshammari et al.'s hypothesis, CTA proved more effective in uncovering usability problems.

Likewise, Olmstead-Hawala et al. (2010) compared classic CTA, in which Ericsson and Simon's guidelines were followed, with "speech communication" and "coaching" approaches, finding that the participants in the coaching tests were significantly more likely to accurately complete test tasks on the U.S. Census Bureau website, as well as to express satisfaction with the website. Olmstead-Hawala et al. state that "this is detrimental in a typical usability study because the coaching injects bias into the results: the results are skewed toward better performance than the participant would have achieved without help" (pg. 2388). However, no such biasing effect was found for the "speech communication" approach. Accordingly, a classic concurrent think-aloud protocol, supplemented by speech communication insights, has been utilised.

5.1.2 Evaluators/participants

Total number and segmentation of evaluators/participants

The CTA tests were conducted with a gender-balanced sample of representatives of the following target groups. The sample size is based on Alshammari et al.'s (2010) finding that a group of N=5 participants were able to find 80-85% of the website usability problems reported by a total group of N=30 participants, whereas N=12 participants were able to find 100% of the reported usability problems. In our case, N=13 participants took part in tests, due to overrecruiting.

Table 11. *CTA Sampling*

CTA Sampling			
Target groups	**Recruiting procedures**	**Inclusion & exclusion criteria**	**Sample**
Scientific community	Each partner organisation recruited participants within their network, with clear instructions on the target group	Each participant needs to: • have a minimum knowledge of statistics • speak English • work within the field of music	N=4
Music industry – representative organisations (CMOs; music exporters etc.)			N=4
Music industry – commercial entities (e.g. music managers, agents, promoters, etc.)			N=4
Policymakers & policy workers			N=1

Key characteristics of evaluators/participants

It was expected that users outside the consortium would not be aware of the project aims and activities. The task was intended to elicit input from this population.

Differences between sample and the user population

The sample was intended to represent the major user groups identified at the project start.

5.1.3 Tasks

Tasks used and task scenarios

The following tasks were identified during the concurrent think-aloud evaluations.

Table 12. CTA tasks

CTA tasks		
Task	**Description**	**Scenario (use flow)**
Live Music Census Maps & Dashboard		
1	*Please find the Live Music Map of Helsinki and count the number of bars in this city.*	*Select "Live Music Census" → click "Live Music Maps" → Select "Helsinki → Select "bars" → Find the number of bars*
2	*Please, locate the bar "Filmriss" in Mannheim and visit its Instagram page.*	*Select "Live Music Census" → click "Dashboard" → Type "Filmriss" → click on the venue → click on the Instagram link*
3	*Please find out the gender distribution of live music audiences in Lviv.*	*Select "Live Music Census" → click "Dashboard" → Select "Lviv" → Select all genders → scroll down the audience survey results to find the gender distribution*
4	*Please compare live music venues in Vilnius and Helsinki: Which top three factors affect the respective venues most negatively?*	*Select "Live Music Census" → click "Dashboard" → Select "Vilnius" and "Helsinki" → scroll down the venue survey results to find the challenges and compare.*
Finno-Ugric Sampo		
5	*Please find out what this website is about.*	*Click on "Info" → Select "About this page"*
6	*Please find a song in Komi and open it on Spotify.*	*Select "Music" → Click on "Language" → select "Komi" → chose a song → click on "Spotify Track ID"*
7	*Please find out more information about a song, any song, by Hilda Griva.*	*Select "Music" → Click on "Performer" → select "Hilda Griva" → chose a song → Click on the title*
8	*Please export a list of songs from Talsi.*	*Select "Music" → Click on "Location" → select "Talsi" → Click on "Export"*
9	*Please find the Journal Article on the Revival of Finno-Ugric Studies and open or download it*	*Select "Publications" → Click on "Type of Publication" → select "Journal Article" → Click on "Revival …" → Scroll to URL → Open the link*

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Selection criteria and sources of the tasks

The tasks were based on a set of tasks initially identified during the usability workshop as relevant to the target groups and performable within a comparatively short testing session. The initial selection was modified and expanded by SINUS to better reflect new functions made available between the date of the workshops and the date of the usability testing.

Task data given to each evaluator/participant and/or inspector

Not applicable.

Criteria for task completion and abandonment for each task

Criteria for task completion: as soon as the user believes the task has been completed.

Criteria for task abandonment: as soon as the user decides to abandon the task.

5.1.4 Evaluation environment

Physical and technical environments

As with the CWW, the CTA was conducted via teleconference, recorded, and transcribed.

Evaluation administration tools and administrators

The evaluation was administered, recorded, and analysed by SINUS.

5.2 Procedure

5.2.1 Design of the evaluation

Description of the evaluation design

The concurrent think-aloud test design is shown in the following tables. The full protocol template used in the evaluation itself is provided in Annex 3.

Table 13. *CTA Agenda*

CTA Agenda		
Time	**Phase**	**Phase name**
T	A	Description and warm-up task
T+5	B1	User tasks *LMC dashboard & maps*
T+15	C1	Exit interview *LMC dashboard & maps*
T+20	B2	User tasks *Sampo*
T+30	C2	Exit interview *Sampo*
T+35	D	System usability scale and socio-demographic questionnaire

The protocol template is provided in Annex 3. In keeping with Ericsson and Simon's guidelines, the protocol began with a warm-up during which the think-aloud concept was concisely explained, and the participant was given the chance to practice thinking aloud while completing a simple task (e.g., searching for a credible information source on Google). The participant was then given a selection of written scripts describing tasks. The task selection mirrors the tasks used.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

In this case, as two very distinct outputs were tested, the tasks and open questions were split up evenly between both sections. The test protocol was designed to take ca. 10-15 minutes per section. The participants were then asked to conduct a ca. 5-minute exit interview for each section, for a total length-of-interview of ca. 40 to maximum 45 minutes. The exit interview began with a few open questions about overall usability and ended with the administration of the system usability scale (SUS) questionnaire, which consists of nine general usability questions scored using a five-point Likert-type scale, as well as a socio-demographic questionnaire. After completing these questionnaires, the participants were given the option of sharing their SUS scores with the evaluator and explaining why they scored particular SUS items the way that they did; however, this was not mandatory

Independent variables, predefined criteria for inspection or observation, measures used in evaluation, and operational criteria of criteria/measures

Part A did not entail inspection/observation/evaluation.

Part B was recorded, transcribed, and coded for usability problem types, user attitudes, and quantitative performance metrics provided in Section 6.2.2.

Answers on parts C and D were recorded using a spreadsheet, entailing ordinal and categorical variables provided in Section 6.2.2.

Interaction between individuals taking part in each evaluation session

The scripts are kept as neutral as possible. During the test itself, the evaluators were instructed to minimise their interaction with the participants, whenever possible restricting their probes to Ericsson and Simon's Level 1 and 2 prompts and restrained "acknowledgement tokens". Level 3 prompts were only used if absolutely necessary to the successful completion of the test.

- Levels 1 and 2: reminders to "keep talking" and acknowledgement tokens (e.g., "um-hmm")
- Level 3: questions about specific actions or the thoughts behind them (e.g., "why did you click on that purple tab?")

General instructions given to participants

The participants were instructed to narrate their actions, cognitive processes, and reactions while undertaking tasks, freely and in detail.

Specific instructions on tasks

As the tasks are all relatively self-explanatory, users were not provided with task-specific instructions unless they specifically asked.

Sequence of activities for conducting the evaluation

The outline of activities is provided above under "Description of the evaluation design." The full protocol template used in the evaluation is provided in Annex 3.

5.2.2 Data to be collected

Usability defects in terms of deviations from predefined criteria

The following types of usability problems were coded:

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

- User (U): The problem is due to the user's experience and knowledge.
- Hidden (H): The interface gives no indications that the function is available or how it should be used.
- Text and icon, including headings/link labels (T): Placement, appearance and content can easily be misinterpreted or misunderstood.
    - Unfamiliar text/icon: the user does not know what the label means
    - Confusable text/icon: the label could easily be confused with another label
    - Goal-specific competing text/icon: the similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal
    - Other
- Sequence (S): Functions and operations have to be performed in an unnatural sequence.
- Feedback (F): The interface gives unclear indications of what the user is doing or has done.
- Other (O)

Observed user behaviour

The following observed attitudes toward the interface were also coded (Hornbaek 2006):

- Annoyance
- Anxiety
- Complexity
- Control
- Engagement
- Flexibility
- Fun
- Intuitive
- Learnability
- Liking
- Physical discomfort
- Want to use again

Observed performance data

The following quantitative performance metrics were measured per task:

- Effectiveness: tasks completed
- Effectiveness: errors in each task
- Effectiveness: tasks with errors
- Effectiveness: task error intensity (qualitative assessment by evaluator, 1-10 scale)
- Efficiency: task time
- Efficiency: unnecessary actions (i.e., deviations from shortest path to task completion)
- Satisfaction: proportion of users complaining during the CTA

User-reported qualitative data

Respondents were asked to provide freeform answers to the following qualitative questions:

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

1. What was your first impression of the platform?
2. After using the platform for a while, what was your impression of it in general?
3. Overall, how easy or difficult was it to complete the tasks? Why?
4. After completing the tasks, had your impression of the platform changed?
5. Compared to current platforms on the internet: does the platform offer anything new and valuable, or does it mostly duplicate what is already out there?
6. Overall, do you think you and your colleagues would use the platform and its content in the course of your work? Why or why not?

User-reported quantitative data

The following quantitative data, derived from the system usability scale, were collected:

- I think that I would like to use this [project] frequently.
- I found the [project] unnecessarily complex
- I thought the [project] was easy to use.
- I think that I would need the support of a technical person to be able to use this [project].
- I found the various functions in this [project] were well integrated.
- I thought there was too much inconsistency in this [project].
- I imagine that most people would learn to use this [project] very quickly.
- I found the [project] very cumbersome to use.
- I felt very confident using the [project].
- I needed to learn a lot of things before I could get going with this [project].

The following socio-demographic data on the respondents were furthermore collected:

- Type of organisation What type of organisation do you work for?
- Level of organisation’s operation (international, national, etc.)
- Sector of organisation
- Years working in current field
- Gender identity
- Age
- Highest level of education completed

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

5.3 Results

5.3.1 Data analysis

Approach used for the analysis of observed, measured, or collected data

During analysis, the CTA transcripts were coded using the same problem categories utilised in the CWW.

Differences in planned and collected data

No differences arose between the planned and actual data collection.

Portion of data used in the analysis

All data produced was used in the analysis. Due to the two platforms being rather different, it was decided to keep the data analysis separate between the Live Music Census outputs and the Finno-Ugric Sampo.

Data scoring; data reduction; statistical analyses

The following data scoring procedures were used:

- Observed usability defects: a count was taken of all usability problems; the percentage of users encountering usability problems and percentage of tasks in which usability problems were encountered (by at least one user) were calculated
- Observed user attitudes and behaviour: the transcripts were coded and a count was taken of user attitude codes; the percentage of users who appeared to express each attitude was calculated
- Performance data based on measurements: the following were calculated:
    - Effectiveness: users completing each task
    - Effectiveness: tasks completed per user
    - Effectiveness: users encountering errors per task and in average
    - Effectiveness: errors per task and in average
    - Efficiency: time per task and in average
    - Satisfaction: percentage of users complaining during the CTA
- Problems, opinions and impressions reported by users: the qualitative exit interview transcripts were coded inductively
- Measured level of user satisfaction or perception:
    - Quantitative data derived from the system usability scale: the average SUS score, SUS score per user, and average score per SUS item were calculated
    - Sociodemographic data on the respondents: a count was taken of all sociodemographic variables

Note that the SUS scoring calculation is based on a specific equation as follows:

$$\Sigma odd\ numbered\ questions\ -\ 5\ =\ X$$

$$25\ -\ \Sigma even\ numbered\ questions\ =\ Y$$

$$(X + Y)\ *\ 2.5\ =\ Z\ (where\ max\ Z = 100)$$

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

5.3.2 Presentation of the results – Live Music Census Dashboard and Maps

Observed usability defects and user attitudes and behaviours

In the case of the Live Music Census Dashboard and Map, the analysis found that of 52 total tasks assigned during the testing (four tasks assigned to each of N=13 users), 44% featured potential usability problems. The breakdown of problem types is as follows:

Table 14. Observed usability defects per problem type – Live Music Census

Observed usability defects per problem type – Live Music Census			
Code	**Problem type**	**Count**	**% of tasks assigned**
NA	No usability problems	29	56%
U	The problem is due to the user's experience and knowledge.	8	15%
H	The interface gives no indications that the function is available or how it should be used.	8	15%
UT	The user does not know what the label means	1	2%
CT	The label could easily be confused with another label	8	15%
GSCT	The similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal	8	15%
OT	Other text-related problems	5	10%
S	Functions and operations have to be performed in an unnatural sequence.	9	17%
F	The interface gives unclear indications of what the user is doing or has done.	2	4%
O	Other	5	10%
Total tasks assigned	*Note: four tasks assigned x N=13 users*	52	100%
Total tasks with any problem	*Note: as multiple problem types can be encountered within a single task, the column percentages for all problem types under "% of tasks assigned" add up to more than the percentage of total tasks with any problem.*	23	44%

For the full analysis of usability errors, see Annex 2 (T4.4_CTA_usability_errors).

The analysis furthermore found that users explicitly voiced or implicitly expressed the following negative and positive attitudes toward the platform. All users expressed some positive attitudes, and the vast majority expressed some negative attitudes. Overall, the positive expressions predominated by a significant margin.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Table 15. Observed user attitudes and behaviour – Live Music Census

Observed user attitudes and behaviour – Live Music Census		
Code	**No. of codes**	**% of users**
Annoyance (negative)	27	62%
Anxiety/uncertainty (negative)	5	15%
Complexity (negative)	42	77%
Control (positive)	14	69%
Engagement (positive)	17	77%
Flexibility (positive)	5	31%
Fun (positive)	7	31%
Intuitive (positive)	123	100%
Learnability (positive)	32	85%
Liking (positive)	73	100%
Physical discomfort (negative)	1	8%
Want to use again (positive)	12	62%
Any negative expression	75	85%
Any positive expression	283	100%

For the full analysis of user attitudes and behaviours, see Annex 2 (T4.4_CTA_user_attitudes).

Measured performance data

On average, 13 out of 13 users completed each task, with each user completing an average of 4 out of 4 tasks. Across all tasks, an average of 6 out of 13 users encountered errors (44%). The average number of errors encountered per user, per task was 0,81, and the average time per task was 2 minutes and 11 seconds.

Table 16. Performance data – Live Music Census

Performance data – Live Music Census					
Task	**Description**	**Users completing task**	**Users with errors**	**Average errors per user, per task**	**Average time on task**
1	*Please find the Live Music Map of Helsinki and count the number of bars in this city.*	11 (85%)	8 (62%)	1,23	00:02:25
2	*Please, locate the bar "Filmriss" in Mannheim and visit its Instagram page*	13 (100%)	3 (23%)	0,31	00:00:56
3	*Please find out the gender distribution of live music audiences in Lviv.*	13 (100%)	6 (46%)	0,92	00:02:20
4	*Please compare live music venues in Vilnius and Helsinki: Which top three factors affect the respective venues most negatively?*	13 (100%)	6 (46%)	0,77	00:03:02
Average		13 (96%)	6 (44%)	0,81	00:02:11

For all quantitative performance data, see Annex 2 (T4.4_CTA_performance_data).

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Measured level of user satisfaction

The Live Music Census dashboard and maps received a total system usability scale score of 76, indicating good overall usability. The SUS scoring calculation is based on a specific framework as follows:

Table 17. System usability scale score – Live Music Census

System usability scale score – Live Music Census		
	Item	**Average score (out of 5)**
1	I think that I would like to use this platform frequently.	3,77
2	I found the platform unnecessarily complex	2,23
3	I thought the platform was easy to use.	4,00
4	I think that I would need the support of a technical person to be able to use this platform.	1,54
5	I found the various functions in this platform were well integrated.	3,62
6	I thought there was too much inconsistency in this platform.	2,15
7	I imagine that most people would learn to use this platform very quickly.	4,31
8	I found the platform very cumbersome to use.	1,69
9	I felt very confident using the platform.	4,08
10	I needed to learn a lot of things before I could get going with this platform.	1,85
System usability scale score		**76 / 100**

5.3.3 Presentation of the results – Finno-Ugric Sampo

Observed usability defects and user attitudes and behaviours

When looking at the Finno-Ugric Sampo, the analysis found that of 65 total tasks assigned during the testing (five tasks assigned to each of N=13 users), 58% featured potential usability problems. The breakdown of problem types is as follows:

Table 18. *Observed usability defects per problem type – Finni-Ugric Sampo*

Observed usability defects per problem type – Finni-Ugric Sampo			
Code	**Problem type**	**Count**	**% of tasks assigned**
NA	No usability problems	27	42%
U	The problem is due to the user's experience and knowledge.	3	5%
H	The interface gives no indications that the function is available or how it should be used.	10	15%
UT	The user does not know what the label means	0	0%
CT	The label could easily be confused with another label	12	18%

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

GSCT	The similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal	10	15%
OT	Other text-related problems	2	3%
S	Functions and operations have to be performed in an unnatural sequence.	10	15%
F	The interface gives unclear indications of what the user is doing or has done.	2	3%
O	Other	8	12%
Total tasks assigned	*Note: five tasks assigned x N=13 users*	65	100%
Total tasks with any problem	*Note: as multiple problem types can be encountered within a single task, the column percentages for all problem types under "% of tasks assigned" add up to more than the percentage of total tasks with any problem.*	38	58%

For the full analysis of usability errors, see Annex 2 (T4.4_CTA_usability_errors).

Regarding user attitudes and behaviour, the analysis found the following positive and negative attitudes toward the platform. All users expressed some negative and some positive attitudes, with a higher number of positive (207) than negative attitudes (97).

Table 19. *Observed user attitudes and behaviour – Finno-Ugric Sampo*

Observed user attitudes and behaviour – Finno-Ugric Sampo		
Code	**No. of codes**	**% of users**
Annoyance (negative)	31	85%
Anxiety/uncertainty (negative)	10	54%
Complexity (negative)	51	92%
Control (positive)	15	69%
Engagement (positive)	5	38%
Flexibility (positive)	4	31%
Fun (positive)	0	0%
Intuitive (positive)	111	100%
Learnability (positive)	31	100%
Liking (positive)	32	69%
Physical discomfort (negative)	5	23%
Want to use again (positive)	9	54%
Any negative expression	97	100%
Any positive expression	207	100%

For the full analysis of user attitudes and behaviour, see Annex 2 (T4.4_CTA_user_attitudes).

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Measured performance data

On average, 8 out of 13 users (63%) completed each task, with each user completing an average of 3 out of 5 tasks (63%). Across all tasks, an average of 7 out of 13 users encountered errors (57%). The average number of errors encountered per user, per task was 0,82, and the average time per task was 2 minutes and 13 seconds.

Table 20. Performance data – Finno-Ugric Sampo

Performance data – Finno-Ugric Sampo					
Task	**Description**	**Users completing task**	**Users with errors**	**Average errors per user, per task**	**Average time on task**
5	*Please find out what this website is about.*	10 (77%)	5 (38%)	0,38	00:01:45
6	*Please find a song in Komi and open it on Spotify.*	12 (92%)	5 (38%)	0,62	00:02:41
7	*Please find out more information about a song, any song, by Hilda Griva.*	11 (85%)	2 (15%)	0,15	00:01:35
8	*Please export a list of songs from Talsi.*	0 (0%)	13 (100%)	1,69	00:02:41
9	*Please find the Journal Article on the Revival of Finno-Ugric Studies and open or download it.*	8 (62%)	12 (92%)	1,23	00:02:21
Average		8 (63%)	7 (57%)	0,82	00:02:13

For all quantitative performance data, see Annex 2 (T4.4_CTA_performance_data).

Measured level of user satisfaction

The Sampo received a score of 61 indicating an above average usability. The SUS scoring calculation is based on a specific framework as follows:

Table 21. System usability scale score – Finno-Ugric Sampo

System usability scale score – Finno-Ugric Sampo		
	Item	**Average score (out of 5)**
1	I think that I would like to use this platform frequently.	2,77
2	I found the platform unnecessarily complex	2,77
3	I thought the platform was easy to use.	3,62
4	I think that I would need the support of a technical person to be able to use this platform.	2,31
5	I found the various functions in this platform were well integrated.	3,31
6	I thought there was too much inconsistency in this platform.	2,46
7	I imagine that most people would learn to use this platform very quickly.	3,69
8	I found the platform very cumbersome to use.	2,23
9	I felt very confident using the platform.	3,15
10	I needed to learn a lot of things before I could get going with this platform.	2,23
System usability scale score		**61 / 100**

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

5.3.4 Interpretation of results and recommendations

Interpretation of results – Live Music Census Dashboard and Maps

- **General Feedback**: The overall feedback of the live music census dashboard and maps was positive. Participants liked the content and found it useful and novel. They would use it in their work, especially if they could download the data sets as an excel or csv file (see below). They generally liked the design, although some found that it could be more colourful. On a less positive note, the website appears distorted on certain devices, as if zoomed in. This resulted in a reduced usability of the website and its features. The menu items at the top right were not visible in these cases, which made certain tasks more difficult for participants as important features were hidden. It was also suggested to add a search function at the top of the page, to make it easier to find certain website contents.

- **Live Music Maps**: Most users navigated the live music maps quite intuitively. They understood how to select the cities and use the filters and search function. Some smaller issues did occur, but they did not significantly deteriorate the overall usability of the maps. One of these was that users not always found the automatic venue count, but counted manually, as a first reaction. However, they usually found the feature quickly after this initial count, which suggests that the feature is not hidden, but rather unexpected and that counting manually is the default option. Some participants suggested to include an export feature, to be able to export a list of live music venues. This would allow them to make better use of the data in their day-to-day work, e.g. for research purposes.

- **Audience and venue dashboards**: In this section of the page, some more usability errors occurred. Users did not always find the correct tab but continued looking for the audience or venue data within the live music maps. This suggests that the tabs are not linked well enough. What is more, they were at times overwhelmed with the dashboard filters, such as the possibility to add and compare multiple cities, or the necessity to click on “update data plots” for the page to load. The expand button, which enlarges the dashboards, was rarely used, as most participants did not perceive it. Finally, the participants also mentioned, that further information about the samples, as well as an export function would be positive.

Recommendations – Live Music Census Dashboard and Maps

- **General recommendations**: It is important to fix the display issues, most importantly the “zoomed-in” effect, which significantly harms usability. Integrating a search function would also largely improve the overall usability of the website.

- **Live Music Maps:** Most importantly, a download function should be added to be able to use the data in an excel or csv format. It is also recommended to link between the “live music maps” subpage and the “Audience and venue data dashboard” to connect the outputs and guide users to additional contents.

- **Audience and venue dashboards**: The most efficient and effective solution to fixing the usability issues in this section of the website, would be to add an information button. The information button should include an explanation on how to use the dashboard, as well as on the sample and survey method. As with the live music maps, an export function would be useful, and linking to the live music maps would make other interesting outputs more visible.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Interpretation of results – Finno-Ugric Sampo

- **General Feedback:** The overall feedback of the Finno-Ugric sampo was positive. Participants understood the use of the website, and it made them curious to learn more about the music and culture of the Finno-Ugric people. Certain participants, such as collective management organisations or music export offices, would state that they could use such a website in their work, if it was available for their country. The layout, however, was deemed rather old-fashioned, and could be improved. A more significant problem was the domain itself, that issued a security warning. This made it difficult for some users to open the website at all, as their security settings blocked the website. In general, participants found this website more complex and slightly less easy to use than the Live Music Census outputs.
- **About page:** Most participants found this page quite intuitively and understood what the website was about. However, they critiqued that there was too much text and that it was too technical, which made it difficult for them to understand parts of the about page.
- **Music-tab**: The use of the filters was rather intuitive for most participants, but it was confusing in some cases, that the “location”-filter was already open, but the others were not. Consequently, most participants were looking in locations first and used the search option there, thinking that it would search in all filters. Treating all filters equally, all closed or open, and adding an overall search option would improve the usability. Participants also did not always understand the back-end wiki. For most it appeared to be empty, as they only looked at the references, which were usually zero. They did not understand the metadata statements. Finally, the export function did not work and generated some frustration among users. In fact, the function seems to be mislabelled as it launches as query instead of exporting a data file. This was despite the fact that the developer had been informed about the issue prior to the start of the individual usability tests.
- **Publications**: As the users were already familiar with the filters and overall page functions, they navigated this part of the website more intuitively than the “music”-tab. However, they often did not find the link to the publication, as it was hidden in the “wiki”-backend. This was different for different publications. For some, a direct link is available under “access point”, for others, one needs to open the wiki first and then scroll to find the link. This was a rather unnatural sequence and made it more difficult for users to find the publications that they were looking for.

Recommendations – Finno-Ugric Sampo

- **General Feedback:** Security issues should be resolved as soon as possible, which includes the use of a secure domain. The layout can be modernized, although this is secondary, as the Slovak sampo will use the official OpenMusE- layout and will vary from the Finno-Ugric sampo.
- **About page:** The text should be adapted to the target group. It should be less technical, and the overall amount of text should be reduced. Images or graphics will improve readability even further.
- **Music Tab:** First, an overall search function, that searches across all filter categories, would largely improve usability. Second, the filter options should be depicted equally, i.e. they should either be all closed or open, when first opening the page. This would reduce confusion. Third, concerning the wiki, we suggest that non-technical users should not be seen

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

as the target group for this “back-end”, as they would not normally need the functions. This, however, leads to the recommendation that necessary links should not be “hidden” in the wiki, but already available and visible in the sampo itself. Finally, the current export function should be labelled “sparql query" instead of “export”, and an export function should be added.

- **Publications:** As with the music tab, important links should be easily available within the sampo page. In the case of publications, this is already true for some under “access point”. This link should be added for all publications. It is also suggested to change the label into “publication link” which is less technical and easier to understand by a larger number of users.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

6 Ethical considerations

Only anonymous data has been collected. While no personal data has been collected, a project information sheet and GDPR-compliant (digital or paper) informed consent document was presented asking participants to consent to participate in the usability test.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

7 Conclusion

This deliverable has reported the findings of the OpenMusE usability testing sequence, comprising two cognitive walkthrough workshops with consortium partners and a round of individual concurrent think-aloud tests with N=13 external end-users, covering the Live Music Census Dashboard and Maps and the Finno-Ugric Sampo. The findings are largely consistent with expectations: internal workshop participants, who are familiar with the project context, encountered few problems, while external end-users surfaced more fundamental navigational and labelling issues. This divergence is methodologically informative and validates the decision to complement internal walkthroughs with genuine end-user testing.

Findings: The Live Music Census Dashboard and Maps demonstrated good usability (SUS 76/100, 96% average task completion), with problems concentrated in the dashboard section rather than the maps. The Finno-Ugric Sampo showed greater scope for improvement (SUS 61/100, 63% average task completion), with one task failing entirely due to a known but unresolved developer-side bug, and broader challenges around navigation, label clarity, and domain security. Both tools generated authentic user interest, with participants across professional profiles recognising the value and novelty of the data they provide.

Actionability: The findings translate into a set of prioritised development actions.

- For the **Live Music Census Dashboard and Maps**, the highest-priority actions are: (1) diagnosing and fixing the device-specific display/zoom rendering issue, which is the single greatest usability barrier for this tool; (2) adding an information or help button to the audience and venue dashboards explaining the sample design, filter logic, and the “update data plots” workflow; (3) adding an export function for venue lists and dashboard data in Excel or CSV format, which was the most consistently requested feature and would significantly extend the tool’s utility for professional users; and (4) strengthening the navigational link between the Maps and the Dashboards sub-pages to reduce user disorientation.
- For the **Finno-Ugric Sampo**, the highest-priority actions are: (1) resolving the domain security issue, which currently prevents some users from opening the website at all; (2) correctly labelling the SPARQL query function and implementing a genuine data export; (3) normalising the filter display so all filter categories are consistently open or closed on initial page load, and adding a cross-category search function; and (4) surfacing critical links — Spotify identifiers, publication access points, and other external URLs — directly within the sampo interface rather than routing non-technical users through the Wikibase back-end. The About page should be simplified and supplemented with visual elements to lower the entry threshold for non-specialist audiences.

Beyond the two tools directly tested, the findings carry implications for the Open Music Observatory as a whole. The insights from the Sampo testing are transferable to the forthcoming Slovak Comprehensive Music Database (SKCMDb), which is built on the same infrastructure and will use a closely related interface. Similarly, the dashboard design lessons are applicable to the Cultural Access and Participation (CAP) dashboards currently in development under WP3./WP4 Implementing the recommended changes across both tool families before final release will reduce the risk of usability barriers affecting the project’s exploitation objectives and the long-term uptake of the Observatory by its intended user communities.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

References

Literature

Alshammari, T., Alhadreti, O., & Mayhew, P.J. (2015). When to ask participants to think aloud: A comparative study of concurrent and retrospective think-aloud methods. International Journal of Human Computer Interaction (IJHCI) 6(3): 48-64.

Barnum, C.M. (2010). Usability Testing Essentials: Ready, Set, Test!. San Mateo, CA: Morgan Kaufmann.

Blackmon, M.H., Polson, P.G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the web. Proceedings of CHI 2002: 463-470.

Bligård, L-O. & Osvalder, A.L. (2013). Enhanced cognitive walkthrough: Development of the cognitive walkthrough method to better predict, identify, and present usability problems. Advances in Human-Computer Interaction 2013: Article ID 931698.

Boren, M.T. & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication 43(3): 261-277.

Clemmensen, T., Hertzum, M., Hornbæk, K., Shi, Q., & Yammiyava, P. (2009). Cultural cognition in usability evaluation. Interacting with Computers 21(3): 212–220.

Dumas, J.S. & Redish, J. (1999). A Practical Guide to Usability Testing: Revised Edition. Exeter, UK: Intellect.

Ericsson, K.A. & Simon, H.A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. Mind, Culture, and Activity 5(3): 178-186.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human-Computer Studies 64: 79-102.

ISO/IEC (2016). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for Usability — Evaluation report. ISO/IEC 25066:2016.

ISO (2018). Ergonomics of human-system interaction. ISO 9241-11:2018.

Larson, K., & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent for information retrieval. Proceedings of CHI 1998: 25-32.

Lewis, C., Polson, P., Wharton, C., & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. Proceedings of CHI 1990: 235-242.

Makri, S., Blandford, A. & Cox, A.L. (2011). This is what I'm doing and why: Methodological reflections on a naturalistic think-aloud study of interactive information behavior. Information Processing & Management 47: 336–348.

McDonald, S., Edwards, H.M., & Zhao, T. (2012). Exploring think-aloud methods in usability testing: An international survey. IEEE Transactions on Professional Communication 55(1): 2-19.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. Cognitive Science 10: 151–177.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Molich, R. and Nielsen, J. (1990). Improving a human-computer dialogue. Communications of the ACM: 338-348.

Nørgaard, M. & Hornbæk, K. (2006). What do usability evaluators do in practice?: An explorative study of think-aloud testing. Proceedings of the 6th Conference on Designing Interactive Systems: 209–218.

Olmstead-Hawala, E.L., Murphy, E.D., Hawala, S., & Ashenfelder, K.T. (2010). Think-aloud protocols: Analyzing three different think-aloud protocols with counts of verbalized frustrations in a usability study of an information-rich web site. Proceedings of the 2010 IEEE International Professional Communication Conference: 60-66.

Polson, P., Rieman, J., & Wharton, C. (1992). Usability inspection methods: Rationale and examples. University of Colorado, Boulder Technical Report 92-07.

Shi, Q. (2008). A field study of the relationship and communication between Chinese evaluators and users in thinking aloud usability tests. Proceedings of the 5th Nordic Conference on Human-Computer Interactions: Building Bridges: 344–352.

Spencer, R. (2000). The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. Proceedings of CHI 2000: 353-359.

Wharton, C. Rieman, J.. Lewis, C. and Polson, P. (1994) The Cognitive Walkthrough Method: A Practitioner's Guide. In J. Nielsen and R. Mack (eds.) Usability Inspection Methods (New York: Wiley) 105-140.

Websites

Live Music Census Dashboard and Maps: https://www.openmuse.eu/

Finno-Ugric Sampo: http://135.181.91.51:3006/en

Annexes

Annex I: Concurrent think-aloud test protocol

a. Description of the protocol

Think-aloud methods are a well-established method of studying "task-based cognitive processes" in general, and human-computer interactions in particular (McDonald et al. 2012, pg. 3).

In the "classic" concurrent think-aloud (CTA) method, based on Ericsson and Simon's work (1993), a single user undertakes a set of target tasks, while an evaluator observes and records them. Ericsson and Simon recommend taking standardised measures to maximise the utility and neutrality of the narrative data elicited: these include offering warm-up / practice opportunities; keeping written scripts as neutral as possible; keeping evaluator probes as general as possible (e.g., "keep talking"); and generally minimising interaction between the evaluator and the user.

The protocol begins with a warm-up during which the think-aloud concept is concisely explained, and the participant is given the chance to practice thinking aloud while completing a simple task (e.g., searching for a credible information source on Google). The participant is then given a selection of written scripts describing tasks to perform on the OpenMusE websites. The scripts are kept as neutral as possible. During the test itself, the evaluators are instructed to minimise their interaction with the participants, whenever possible restricting their probes to Ericsson and Simon's Level 1 and 2 prompts and restrained "acknowledgement tokens". Level 3 prompts are only used if absolutely necessary to the successful completion of the test.

- Levels 1 and 2: reminders to "keep talking" and acknowledgement tokens (e.g., "um-hmm")
- Level 3: questions about specific actions or the thoughts behind them (e.g., "why did you click on that purple tab?")

The test protocol is designed to take ca. 20-25 minutes. The participants are also asked to conduct a ca. 5–10-minute exit interview, for a total length-of-interview of ca. 30 to maximum 40 minutes. The exit interview begins with a few open questions about overall usability and ends with the administration of the system usability scale (SUS) questionnaire, which consists of nine general usability questions scored using a five-point Likert-type scale, as well as a socio-demographic questionnaire. After completing these questionnaires, the participants are given the option of sharing their SUS scores with the evaluator and explaining why they scored particular SUS items the way that they did; however, this is not mandatory.

The test will be conducted online using screen-sharing and will be video recorded.

During analysis by SINUS, the CTA transcripts will be coded using the same problem categories utilised in the CWW:

- User (U): The problem is due to the user's experience and knowledge.
- Hidden (H): The interface gives no indications that the function is available or how it should be used.
- Text and icon, including headings/link labels (T): Placement, appearance and content can easily be misinterpreted or misunderstood.
    - Unfamiliar text/icon: the user does not know what the label means
    - Confusable text/icon: the label could easily be confused with another label

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

    - Goal-specific competing text/icon: the similarity of an incorrect label to the goal is greater than the similarity of the correct label to the goal
    - Other
- Sequence (S): Functions and operations have to be performed in an unnatural sequence.
- Feedback (F): The interface gives unclear indications of what the user is doing or has done.
- Other (O)

The following codes identifying specific attitudes toward the interface will also be used in the transcript analysis (Hornbaek 2006):

- Annoyance
- Anxiety
- Complexity
- Control
- Engagement
- Flexibility
- Fun
- Intuitive
- Learnability
- Liking
- Physical discomfort
- Want to use again

Finally, the following quantitative performance metrics will be measured per task:

- Effectiveness: tasks completed
- Effectiveness: errors in each task
- Effectiveness: tasks with errors
- Effectiveness: task error intensity (qualitative assessment by evaluator, 1-10 scale)
- Efficiency: task time
- Efficiency: unnecessary actions (i.e., deviations from shortest path to task completion)
- Satisfaction: proportion of users complaining during the CTA

b. Introduction

- Introduction of the interviewer, topic and duration of the interview.
- Language rule: Interview in English (even with German speakers), "First name"/"last name"
- Necessity of recording
- Explicit reference to anonymity of the analysis: no names are mentioned in the report
- We want to get your feedback on the usability of these websites. We are testing the website, not the interviewee; If you cannot complete a task, that reflects on the website, not on you.
- There are no wrong and no right answers, speak without filters
- Obtain OK

Interviewer instruction: **start recording!**

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

c. Warm-up task

Interviewer: this test will use a method called "think-aloud", in which you complete a few tasks while narrating what you are doing and thinking. I'll show you an example of how this is done and then ask you to try it yourself. As an example, I'll give myself the task of finding the current unemployment rate in Germany. Now I will start the task while thinking aloud.

> *I'm opening my web browser and navigating to google.com, the search engine I usually use. Now I'm searching "unemployment rate Germany". Here I see the AI reply, and it gives me a couple of sources. I'm not sure if I can trust them, so I will keep looking. The next website that shows up, is by the German statistics office. I trust this source, so I will take a look. Here it says 3,9% and the last data is from July 2025, so very recent. Task completed.*

Do you have any questions about the think-aloud procedure that I just demonstrated?

Now I'd like you to try. Please find the number of people who claimed asylum in [your country] in 2025.

d. User tasks

Interviewer instruction: ask the user to share their screen, open their web browser, and navigate to https://www.openmuse.eu/. Send the link in the chat.

Interviewer: We will take a look at the OpenMusE website today and perform a couple of simple tasks. Before we start the tasks, what is your very first impression of the website?

Interviewer instruction: ask the user to complete each task. DO NOT describe the proper use flow – allow the user to figure it out. While the user is completing the task, try to speak as little as possible – however, if the user forgets to keep speaking aloud, remind them.

If the user appears to have completed a task, confirm verbally that the task is complete and move to the next. If the user cannot complete a task, wait until the user has reached a point of frustration and then ask if the user would like to move on.

https://www.openmuse.eu/		
Task	**Description**	**Scenario (use flow)**
1	*Please find the Live Music Map of Helsinki and count the number of bars in this city.*	*Select "Live Music Census" → click "Live Music Maps" → Select "Helsinki → Select "bars" → Find the number of bars*
2	*Please, locate the bar "Filmriss" in Mannheim and visit its Instagram page.*	*Select "Live Music Census" → click "Dashboard" → Type "Filmriss" → click on the venue → click on the Instagram link*
3	*Please find out the gender distribution of live music audiences in Lviv.*	*Select "Live Music Census" → click "Dashboard" → Select "Lviv" → Select all genders → scroll down the audience survey results to find the gender distribution*
4	*Please compare live music venues in Vilnius and Helsinki: Which top three factors affect the respective venues most negatively?*	*Select "Live Music Census" → click "Dashboard" → Select "Vilnius" and "Helsinki" → scroll down the venue survey results to find the challenges and compare.*

Interviewer instruction: tell the user that we will look at another section, but that we would like to ask them some short questions before that.

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Interviewer: thank you for completing the tasks! We'd like to ask a couple of follow-up questions before moving on to the next section.

1. After using the website for a while, what was your impression of it in general?
2. Overall, how easy or difficult was it to complete the tasks? Why?
3. After completing the tasks, had your impression of the website changed?
4. Compared to current websites on the internet: does the website offer anything new and valuable, or does it mostly duplicate what is already out there?
5. Overall, do you think you and your colleagues would use the website and its content in the course of your work? Why or why not?

Interviewer instruction: ask the user to navigate to http://135.181.91.51:3006/en, the data sharing space that will soon be integrated into our general website. Send the link in the chat.

Interviewer: now we'll move to the data sharing space. Before we start the tasks, what is your very first impression of this section?

Interviewer instruction: ask the user to complete each task in the same manner as before.

http://135.181.91.51:3006/en		
Task	**Description**	**Scenario (use flow)**
5	*Please find out what this website is about.*	*Click on "Info" → Select "About this page"*
6	*Please find a song in Komi and open it on Spotify.*	*Select "Music" → Click on "Language" → select "Komi" → chose a song → click on "Spotify Track ID"*
7	*Please find out more information about a song, any song, by Hilda Griva.*	*Select "Music" → Click on "Performer" → select "Hilda Griva" → chose a song → Click on the title*
8	*Please export a list of songs from Talsi.*	*Select "Music" → Click on "Location" → select "Talsi" → Click on "Export"*
9	*Please find the Journal Article on the Revival of Finno-Ugric Studies and open or download it*	*Select "Publications" → Click on "Type of Publication" → select "Journal Article" → Click on "Revival …" → Scroll to URL → Open the link*

e. Exit interview

Interviewer instruction: inform the user that we will end the test with a few open questions, followed by a short questionnaire.

Interviewer: thank you for completing the tasks! We'd like to ask a couple of follow-up questions.

6. After using the website for a while, what was your impression of it in general?
7. Overall, how easy or difficult was it to complete the tasks? Why?
8. After completing the tasks, had your impression of the website changed?
9. Compared to current websites on the internet: does the website offer anything new and valuable, or does it mostly duplicate what is already out there?
10. Overall, do you think you and your colleagues would use the website and its content in the course of your work? Why or why not?

Interviewer: finally, we'll go through two quick questionnaires together.

f. System usability scale questionnaire

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Table XX. System usability scale

	Item	Strongly disagree (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Strongly agree (5)
1	I think that I would like to use this [project] frequently.					
2	I found the [project] unnecessarily complex					
3	I thought the [project] was easy to use.					
4	I think that I would need the support of a technical person to be able to use this [project].					
5	I found the various functions in this [project] were well integrated.					
6	I thought there was too much inconsistency in this [project].					
7	I imagine that most people would learn to use this [project] very quickly.					
8	I found the [project] very cumbersome to use.					
9	I felt very confident using the [project].					
10	I needed to learn a lot of things before I could get going with this [project].					

The SUS scoring calculation is based on a specific framework as follows:

a. $\Sigma_{\text{odd numbered questions}} - 5 = X$
b. $\Sigma_{\text{even numbered questions}} - 5 = Y$
c. $(X+Y) * 2.5 = Z$ (where max Z=100)

g. Socio-demographic questionnaire

1. What type of organisation do you work for?
    a. Governmental (public organisations working in one country)
    b. Intergovernmental (public organisations working across more than one country e.g. EU organisations)
    c. Non-governmental (faith-based)
    d. Non-governmental (non-faith based)
    e. Private company
    f. Freelancer
    g. Other (please specify in the next question)
2. On what level does your organisation mostly work?
    a. International

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

   b. Federal/national
   c. State/prefectural
   d. Local/communal
3. In what sector does your organisation primarily operate?
   a. City planning
   b. Music performance/art
   c. Music management
   d. Advocacy in culture
   e. Marketing
   f. Concert promotion
   g. Education/Academia in music-related fields
   h. Researcher
   i. Radio or similar, looking for local music
   j. Librarians
   k. Export of music
   l. (Local) politics
4. About how many years have you been working in your current field?
   a. 1-5 years or less
   b. 6-10 years
   c. More than 10 years
5. What is your gender identity?
   a. Male
   b. Female
   c. Other
   d. Choose not to answer
6. How old are you?
   a. 18-19
   b. 20-29
   c. 30-39
   d. 40-49
   e. 50-59
   f. 60-69
   g. 70 or above
   h. Choose not to answer
7. What is the highest level of education you have completed?
   a. Primary education
   b. Secondary education
   c. Vocational training or professional certificate
   d. Bachelor's degree
   e. Master's degree
   f. Doctorate

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295

Annex II: Test data

Please see the usability analysis in the file OpenMusE_usability_analysis, downloadable via this secure link: https://cloud.sinus-institut.de/public/cb6cef

© 2023 OpenMusE | HORIZON-CL2-2022-HERITAGE-01-05 | Grant Agreement No. 101095295