Facilitating Extended Reality in Museums through a Web-Based Application

Theodor Norill Kvalvaag
Sindre Stolberg Fardal

Master's thesis in Software Engineering at

Department of Computer science, Electrical
engineering and Mathematical sciences,
Western Norway University of Applied Sciences

Department of Informatics,
University of Bergen

June 2023

Western Norway
University of
Applied Sciences

Abstract

As technology continues to advance at an unprecedented pace, the need for museums to stay updated with the digital transformation of their exhibitions becomes crucial. However, typically, only around 5% of museums' budget is allocated towards digital and communication initiatives. Therefore, it raises the question of how museums can leverage the potential of emerging technologies like Extended Reality (XR) within their existing resource framework. One approach is the development of tools to facilitate the integration of such technology, but the process and specific requirements for such a tool need to be addressed.

This project was guided by the objective of developing an administration tool for museums to assist in the facilitation of multiple head-mounted displays and extended reality applications. Principles and guidelines from User Experience and User Interface Design were employed, resulting in the development of a web application with a focus on ease of use.

The evaluation of the developed web application involved two groups, each consisting of five participants, including museum employees and students. The user tests were assessed using measures such as the NASA Task Load Index (NASA-TLX), Concurrent Think Aloud (CTA), and interviews. The participants exhibited a positive response to the simplistic design, highlighting the potential of the web application for regular museum employees without specific technical expertise. Although the number of evaluation rounds was limited, the developed web application serves as a proof of concept, demonstrating how an administration tool can be created to facilitate the use of XR technology in museums.

By making MuseumXR an open-source platform, it can serve as a valuable resource for future research projects or act as a foundation for the development of systems aiming to streamline the integration of XR in museums or similar institutions.

Acknowledgements

We would like to thank our supervisors at Western Norway University of Applied Sciences (HVL), Harald Soleim and Atle Birger Geitung for their continuous guidance and support and for their invaluable feedback throughout our work on this thesis.

Thanks to our external supervisor from Spello, Audun Klyve Gulbrandsen for providing us with the option to explore the domain of Extended Reality technology in museums and letting us form our own thesis. He has contributed significantly to this thesis with his support and expertise throughout the development process, and it has been a joy having him on the team.

We would also like to thank the University Museum of Bergen for their collaboration to make this project possible. We are grateful for their willingness to part take in this project and let us use their museum during this project.

Thanks to Bodil Moss for giving advice on academic writing and thanks to Malik Beglerovic for giving us advice on searching in research databases.

Thanks to TekLab for letting us participate and present our master's thesis at the international seminar ”Reclaiming the Metaverse” at Media City Bergen.

And lastly, thanks to all the individuals who willingly participated in our user tests, as their involvement enabled us to thoroughly evaluate and refine the solution.

Contents

List of Figures

Acronyms

API Application Programming Interface.
AR Augmented reality.

CMS Content Management System.
CTA Concurrent Think Aloud.

DSR Design Science Research.

HLS HTTP Live Streaming.
HMD Head Mounted Display.

MR Mixed reality.
MRTK Mixed Reality Toolkit.
MVP Minimum Viable Product.

NASA-TLX NASA Task Load Index.
NN/g Nielsen Norman Group.

PoV Point of View.

RTA Retrospective Think Aloud.

SDK Software Development Kit.
SWAT Subjective Workload Assessment Technique.

UI User interface.
UX User experience.

VR Virtual reality.

WDP Windows Device Portal.

WP Workload Profile.

XR Extended reality.

Chapter 1

Introduction

This chapter introduces the motivation and context for this thesis, followed by the goal, research question and methodology.

1.1 Motivation

Museums play a crucial role in preserving, protecting, and promoting a local community or nation's collective memory, knowledge, and history [1]. They serve as a community and a venue for learning about the history and heritage of a specific area. With the rapid advancement of technology in contemporary society, it is essential for museums to remain current with the digitalization of their exhibitions [2, 3, 4, 5]. Museums that lack digital tools may be characterized by an abundance of static content, in which the medium used to convey information could be enhanced to improve visitor learning outcomes and interest [6, 7, 8].

Extended reality (XR) is an umbrella term encompassing Virtual reality (VR), Augmented reality (AR), and Mixed reality (MR) technologies. These technologies leverage computer-generated visuals and interactive scenarios to enhance or substitute real-world environments, thereby creating immersive and adaptive experiences for users. This technology can be employed to create immersive and captivating interactive experiences, thereby offering increased value to visitors [9, 7, 10]. However, museums may have varying requirements depending on what type of museum it is, the content being exhibited, or the purpose of their digital extension. The Louvre Museum in Paris has produced a digital replica of the renowned Mona Lisa painting, which allows individuals to experience it in a more personal manner from home using VR [11]. Similarly, the National Museum of Natural History, also located in Paris, features MR exhibitions augmented with virtual 3D models and animations, resulting in an immersive experience in which virtual elements blend with the real world [12]. These distinct technologies serve different functions: VR enables individuals to view the Mona Lisa from any location and at any time, while the MR employed at the Natural History Museum provides an immersive experience without excluding the real world. Illustrations can be seen in Figure 1.1.

(a) Mona Lisa VR experience from the Louvre museum in Paris [11]

(b) MR experience from the National Museum of Natural History in Paris [12]

Figure 1.1: VR and MR experience

In a joint effort with an external supervisor from Spello, a keen interest was expressed in exploring the potential of incorporating this type of technology in museum environments. Spello, a firm dedicated to the creation of XR solutions for education and learning, has been an active participant in several educational and cultural heritage initiatives, ranging from sensor-based XR applications to fully immersive virtual museums. The initial proposal put forth by Spello envisioned the creation of a Unity app for XR devices, with the intention of exploring innovative uses of sensor technology. Alongside this, the development of an associated Content Management System (CMS) to manage virtual elements was also suggested. However, this proposal served only as a guideline, and articulating an independent idea for the thesis within the domain was encouraged.

To investigate the potential within this domain, a collaboration with a museum was necessary and an inquiry was initiated with the University Museum of Bergen. They expressed their interest in the exploration of XR technology in the museum and were willing to participate in this project. The University Museum of Bergen is a multidisciplinary research institution located in Bergen comprising natural and cultural history. Through dialogue with the museum, an initial decision was made to enhance one of their natural history exhibitions, specifically, one related to two 8000 year old human skulls that were discovered in the local area. Spatial constraints within the layout of the exhibition room compelled the museum to adopt a less-than-ideal exhibition display, resulting in an unintuitive presentation. By utilizing the potential of XR, these limitations could be effectively addressed and the exhibition experience significantly enhanced.

While technological advancements confer numerous benefits, the incorporation of such innovations into pre-existing business models often manifests challenges. A study conducted by the International Council of Museums including almost 1600 responses from museums from 107 countries illuminates the fact that 75% of museums lack personnel designated for digital activities. Moreover, over 40% of museums earmark less than 5% of their overall budget for digital and communicative initiatives [13]. It is regrettable that strategizing around digital integration does not rank highly on the priority list for museum planners [14, 15, 16, 17, 18]. When deliberating over the deployment of an extended reality experience, various factors must be considered. A large proportion of museum visitors may not have prior exposure to the requisite device and might require guidance. If the device is a Head Mounted Display (HMD), assisting users might be challenging due to the inability to view their experiences. This could necessitate the allocation of trained staff and time, resources that may not be readily accessible to many museums.

1.2 Problem description

In the initial stages, neither the requirements nor functionalities were explicitly delineated, thereby necessitating a careful alignment of the museum's needs with the capabilities of the technology as a critical step in the development cycle. One of the requirements that emerged from these discussions was that if the museum were to integrate novel technology, it should be done without compromising the social and physical aspects of the museum experience.

While constructing an application in VR is feasible, it alienates users from the physical exhibition as the immersive nature of VR may cause a detachment from the real world. Conversely, MR and AR allow users to maintain a connection to the physical environment, affording them the opportunity to view the physical exhibit while receiving supplemental information. This integration unlocks natural and intuitive interactions among the 3D human, computer, and environmental dimensions[19]. Given this, and considering the resources available for this project, a decision was made to utilize the Hololens 2, a MR HMD.

The process of narrowing down the idea for the thesis to explore innovative enhancements for the museum's skull exhibition proved to be time-consuming. Thorough research was undertaken to investigate diverse approaches for augmenting the engagement level of museum exhibitions, and the utilization of gamification was explored as one potential method. Further research as well as conversations with the representatives from the museum was done discussing different ideas with gamification in mind.

During this phase of the project, challenges pertaining to the implementation and administration of the intended idea emerged. During a meeting involving a representative from the museum industry with experience in innovative technology from several other museums in Norway, valuable insight into an occurring problem pertaining to the neglect of the museum staff's role in administering new technology arose. The trend seemed to be that often when new technology was introduced in museums, proper systems and routines for administering it were disregarded. The findings done by the International Council of Museums show that museums on average use less than 5% of their budget on digital initiatives could be one of the reasons for this [13].

Through deliberations with the external supervisor and subsequent discussions with the representatives from the museum, a change in direction for the project aimed at facilitating the implementation and administration was concluded.

1.3 Goal

The goal of this study is to design and develop a tool, coined MuseumXR, that facilitates the implementation and operation of XR technology in museums, with a particular emphasis on ease of use. MuseumXR is a Minimum Viable Product (MVP) web application that can be used by museum employees to manage both HMDs and virtual exhibitions. The proposed system not only aims to showcase the practical implementation and operation of new technology but also serves as a suggested design for such a system. To be able to perform user tests and evaluate the MuseumXR platform, a simple MR application was developed.

The thesis also seeks to shed light on the inherent challenges related to the administration of digital exhibitions within the museum context. A key aspect is the identification of difficulties arising from the creation of custom administration functionalities for each individual digital exhibition. Without a consistent and universally accepted approach for digital artists to create these exhibitions, the development of a solution to administer them can become cumbersome and inefficient. Artists creating digital content for museums might need to create custom solutions for each museum, rendering the value of a tool in streamlining and managing virtual exhibitions limited. This problem is further exacerbated by the fact that museums, already struggling with resource allocation for digital activities, may be deterred from investing in new technology that lacks a standardized framework. By establishing a standard or a set of guidelines for digital artists working on virtual exhibitions, the integration of extended realities in museums could become easier and more accessible, ultimately enabling museums to better leverage the potential of XR experiences.

This leads to the following research question for this thesis:

- How can an administration tool be developed for museums to aid in the facilitation of multiple head-mounted displays and extended reality applications?

1.4 Scope

The scope of this thesis is confined to the functionality of MuseumXR, specifically targeting the personnel envisaged as users of the system, and the evaluation of the system as an administration tool. The investigation and development of a standard or conformity pertaining to the administration of digital exhibitions are recognized as significant elements, yet they lie beyond the scope of this thesis.

1.5 Methodology

In order to address the research question, a thorough examination of guidelines and principles pertaining to User experience (UX) and User interface (UI) design was undertaken with a focus on limiting the required workload on museum staff. Additionally, an exploration of the development process and an in-depth analysis of the museum industry context were conducted. The research conducted in these domains formed the basis for the development of the web application and provided guiding principles throughout the duration of the project.

Design Science Research (DSR) methodology [20] was employed in this study due to its aptitude for gauging the efficacy of a system within a museum context, utilizing knowledge established through academic research and the contextual environment for the development of an artefact. Both qualitative and quantitative assessment methodologies were employed in this research. Case studies, encompassing semi-structured interviews and surveys, were carried out in a bid to ensure comprehensive data collection and analysis.

The evaluation of MuseumXR employed the NASA Task Load Index (NASA-TLX) [21], which served as an evaluation methodology designed to ascertain the workload associated with the use of the web application. To facilitate real-time feedback, the think-aloud protocol [22] Concurrent Think Aloud (CTA) technique was implemented during field testing. Continuous observation was maintained, and a semi-structured interview was subsequently administered. This approach ensured a comprehensive evaluation, integrating both quantitative and qualitative aspects to derive meaningful insights.

The web application serving as the administration tool was created with Flutter Web[23] for the frontend solution and Firebase [24] for the backend solution.

The MR application for this project was constructed utilizing the Mixed Reality Toolkit offered by Microsoft [25], incorporated within Unity [26]. It is worth noting that it was not ready for deployment and use as its principal purpose was to serve as a prototype to demonstrate prospective capabilities rather than a final product.

Chapter 2

Background

In this chapter, an overview of the current state of museums and their use of technology is provided, referencing related studies in the field. Also, a clear explanation of the key technologies and terms used in this project is included.

2.1 Current state of museums and technology

In recent years, museums have been increasingly exploring and adopting new technologies to enhance visitor experiences, improve accessibility, and expand their reach [27]. However, the integration of new technology in museums has not been without its challenges. As mentioned in chapter 1, a big portion of museums lacks dedicated staff for digital activities and allocates only a small fraction of their budget for digital and communication initiatives [13]. It can also be stated that the degree of usability and accessibility is frequently disregarded from the standpoint of staff members at cultural institutions, who are mandated to oversee interactive installations [28]. As it stands, research has done comparatively very little towards the evaluation and analysis of user experience pertaining to the tools [29] supplied for multimedia content configuration within the cultural landscape as well as numerous others.

To address these challenges, some researchers and practitioners have called for greater collaboration among museums, technology providers, and academia to develop best practices, share resources, and build capacity in the sector [30]. This collaborative approach has led to the emergence of research projects and consortia such as the European Union-funded CHESS project, which aims to create personalized, adaptive, and engaging museum experiences using AR and other digital technologies [31]. As the landscape of technology continues to evolve, museums must adapt and invest in building their digital capabilities to stay relevant, meet the expectations of increasingly tech-savvy visitors, and fulfil their mission of preserving and disseminating cultural heritage in the digital age [32].

2.2 Related work

2.2.1 XR in museums

Research indicates that there is a gap between the enhancement of exhibitions in museums and the administration and operation of such XR implementations. Research done in 2022 by Silva et al. into the use of XR in museums shows that from the comparatively small number of publications they were able to gather for analysis, the use of XR in museums is still in its early phases [33]. What they do find however is that most of the studies done within the XR field in museums pertain to the experience the visitor is engaged in in terms of immersion and meaningful content. Margetis et al. and Banfi et al. research show that XR is successful in achieving more engaging content that attracts and encourages revisits from visitors [34], [35]. The gap between curators in the museum and the new technologies is something Silva et al. highlight and Galdieri and Carrozzino suggest new categories of professionals with insight into both domains are needed to fill this gap [30].

2.2.2 XR management systems in museums

Galdieri and Carrozzino's research [30] focuses on the development of a tool aimed at empowering curators to independently create virtual exhibits that accurately simulate the real environment of their exhibitions. Although not directly aligned with the scope of this thesis, their study shares common ground with regard to enhancing the museum experience, addressing challenges encountered during the implementation of new technology in contexts where limited technological advancements have been previously utilized. They highlight the importance of creating tools that are user-friendly enough for non-IT-experts to be able to use. Further difficulties pertaining to the limited technical skills of the curators were also discussed and underlines the fact that many employees' lack of IT competence shows the importance of developing a system that is accessible and easy enough for non-IT personnel to use.

A recent study by Saviano et al. recognises that the issues staff of cultural institutions face are often overlooked when engaging with this type of technology[28]. The research paper was published through the "IEEE Conference on Virtual Reality and 3d User Interfaces Abstracts and Workshops (VRW)" in late March 2023 [36] and was not available for most of the timeline of this project. It has therefore not laid any foundation for this project's thesis but is discussed as it is highly relevant to the underlying topic. One of the aspects this research looks at is developing a CMS accessible to museum staff without IT expertise. The MiRA (Environmental Mixed Reality) project as presented by Saviano et al. and illustrated in Figure 2.1 is at the time of writing this thesis under development and is a collaboration between five multidisciplinary work units from three universities and an enterprise. MiRA is a project that seeks to create a CMS that can be used in a variety of museums, but the research paper only covers the academic basis for creating such a CMS and seeks to understand the underlying problems and challenges tied to it. It should be noted that despite the relevance, this project is more of an advanced tool to control and configure XR installations specifically aimed for visor-less experiences.

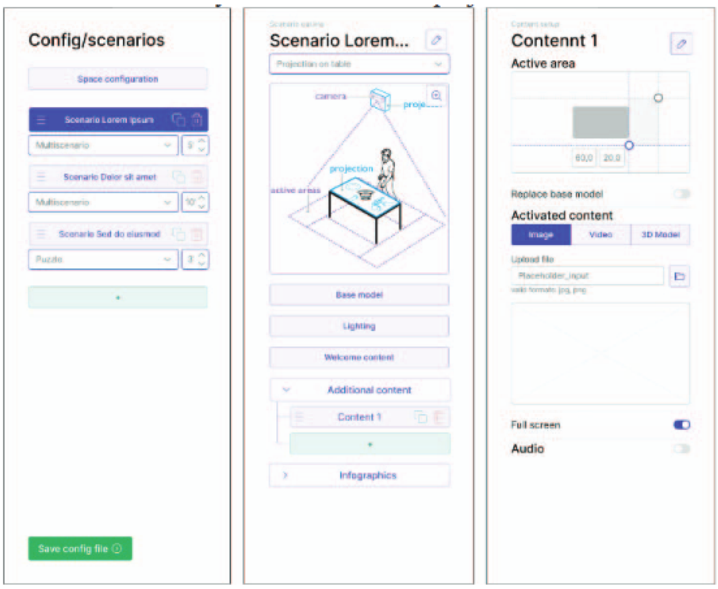


Figure 2.1: Three of the four columns composing the interface of the proposed MiRA CMS [28]

In light of the absence of product creation and conducted field tests, their project refrains from offering a definitive conclusion, instead reiterating the recurrent challenges and issues associated with personnel in cultural institutions.

2.2.3 Similar products

ClassVR

ClassVR [37] is a VR solution that provides hardware, software and content for educational purposes. They offer their own standalone virtual headsets together with a teacher portal that makes it possible for teachers to manage the content the students are experiencing as well as guide students through these experiences. This is a solution custom-made for education in schools which limits its possibilities for cross-domain. It offers ready made lesson plans and guides serving as a uniformity/standard to help teachers integrate VR and AR into their lessons. As will be discussed in chapter 6, to introduce technology like this in domains such as education, uniformity is needed, something ClassVR solves with its teacher portal depicted in Figure 2.2.

This applies to several other cross-domains where new technology is introduced. One of the goals of museums, like educational institutions, is to educate people. If this can be done in a seemingly successful way with ClassVR in education, it highlights the possibilities and use of systems/platforms like this in other areas as well, like a museum.

ClassVR has done several case studies where they have interviewed schools that have taken this product into use. One of these case studies is from Westhaven School in North Somerset, England [38]. For Westhaven, one of the important

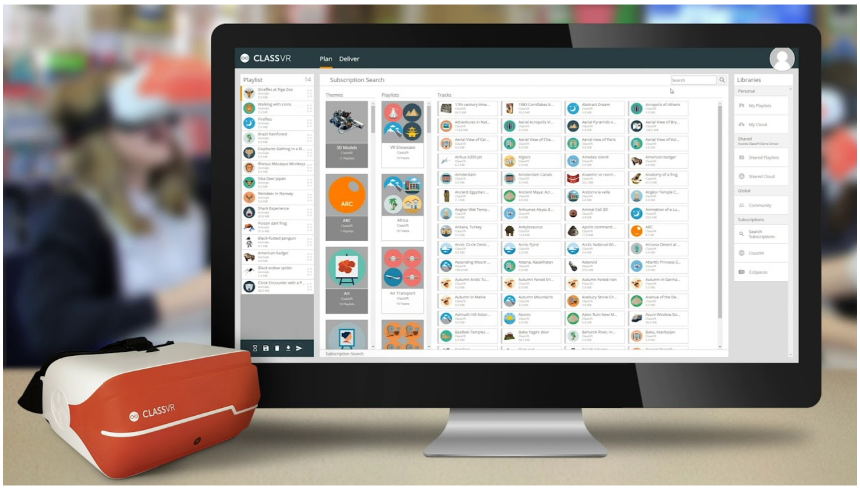


Figure 2.2: ClassVR teacher portal [37]

factors when choosing new technology was ensuring it can be easily integrated into lessons by teachers. The Network Manager said “I could see the technology was optimised for a classroom environment and that the management interface would be a great fit for us. We can easily navigate the portal and find content. ClassVR is so simple, so easy for us to manage, so user friendly.” [38].

The key lesson from many of these case studies is the importance of having a central system for managing and controlling content, especially in a situation like a classroom. Even though a museum’s situation is not exactly the same, there are still many similarities that highlight the value of having a complete system that can handle both the hardware and software for XR.

HMD management

Tools for mainly managing HMDs also exist, two examples being ArborXR and ManageXR. They both offer very good management of the HMDs, but little to no software related management. More specifically, the tools are limited to having a connection to the HMDs, so it is for example possible to install and deploy applications, but unlike ClassVR, they cannot manage content inside the applications.

There exist additional tools designed to manage HMDs, such as ArborXR [39] and ManageXR [40]. The commonality observed is that while these tools facilitate effective management of HMDs, they exhibit minimal to no capacity for software-related management. Specifically, these tools demonstrate the ability to establish connectivity with the HMDs, enabling application installation and deployment. However, in contrast to ClassVR, these tools lack the functionality to manage content within the applications. In addition, ArborXR and ManageXR are more advanced tools targeting companies seeking to manage a large number of HMDs, that might require designated personnel.

Comparing these to the proposed solution in this thesis, MuseumXR has a smaller scale HMD management solution, whilst also offering a software management interface for applications to the HMDs.

2.3 Game engine

Game engines are important in the creation of interactive 2D and 3D content, including MR applications, by offering an extensive set of tools within a developmental environment conducive to the crafting of engaging and immersive experiences. Among the game engines employed in MR development, Unity [26] and Unreal Engine [41] are recognized as two of the more prevalent choices [42]. Both engines provide robust MR support and a wide spectrum of capabilities. Nevertheless, Unity was selected for this project due to the researcher's greater familiarity with its interface and functionality.

Unity had, as of 2020, about 60% of the virtual and augmented reality content [26, 43]. This includes about 90% of all emerging augmented reality platforms including the Microsoft HoloLens and "dominates the virtual reality business", according to Fortune [44]. Its powerful and versatile game engine has gained widespread popularity among developers for its comprehensive set of MR-specific features [45]. MRTK for Unity won the Auggie Award for best developer tool 2021 [46]. As indicated by itch.io [47], Unity boasts a substantial and engaged community of developers, which can provide valuable assistance for troubleshooting and sharing best practices.

2.4 Extended Reality

XR is a term that encompasses various technologies that extend or enhance the way we perceive and interact with the world around us [48]. There are three main subcategories of XR: VR, AR, and MR.

VR is a fully immersive experience that replaces the user's real-world environment with a computer-generated one. VR typically requires a HMD or other device to view and interact with the virtual world and it is commonly used for gaming, entertainment, and training purposes [48].

AR is a technology that enhances the user's real-world environment by overlaying digital information on top of it. AR can be experienced through a variety of devices, such as smartphones, tablets, or specialized HMDs and is commonly used for education, training, and advertising purposes [48].

MR is a blend of VR and AR, combining aspects of both. In MR, the real and virtual worlds are seamlessly integrated, creating a more natural and intuitive experience and can be experienced through specialized HMDs that track the user's movements and adjust the virtual elements accordingly [49].

In particular, MR has great potential for use in museums [50]. By combining the physical artefacts and exhibits with digital information and interactive elements, MR can enhance the visitor's understanding and engagement with the content [51]. MR can be used to create virtual tours of historical sites, bring exhibits to life with animations and sound effects, or provide additional context and background information.

MR is the interaction between the extremities shown in Figure 2.3.

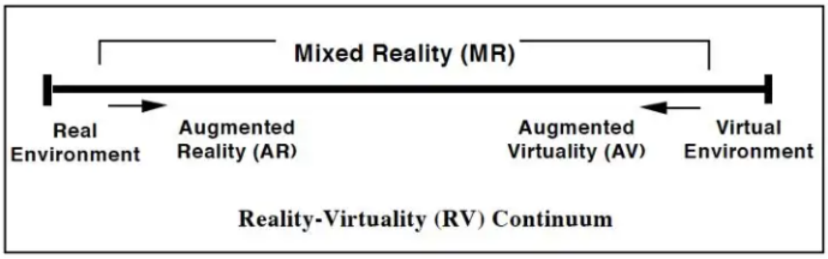


Figure 2.3: Reality-Virtuality continuum [52]

Mixed Reality Display Types MR displays can be classified into seven types according to Milgram [53] in order to help people better understand the concept.

1. Monitor based (non-immersive) video displays. Displaying video of the real world overlaid with digital images
2. A HMD showing video. The same as type 1, but with a head-mounted display
3. Optical see-through HMD. Virtual images are superimposed over real images on a see-through display.
4. Video see-through HMD. The same as type 3, but with virtual graphics superimposed on the video of the real world.
5. Monitor based AV system. Displaying 3D graphics on a monitor with video superimposed.
6. Immersive or partially immersive AV. Visualizing 3D graphics with video superimposed on them in an immersive display.
7. Partially immersive AV systems. Additional real-object interactions are possible with AV systems, such as interacting with one's own hand.

Type 3 MR displays, like the HoloLens developed by Microsoft, let virtual information appear on top of the real world; meanwhile, type 1 MR displays show virtual characters interacting with real people via a camera feed.

2.5 MR device

To experience MR, an HMD is typically required. HMDs are devices that are worn on the head and provide a display for the user to view the virtual elements. The Hololens 2, which will be used for this project, features advanced sensors and tracking technology, including a depth sensor, an IR camera, and a six-degree-of-freedom motion controller [54]. These allow for more natural and intuitive interaction with the virtual elements. It also includes a built-in computer, allowing for more powerful and sophisticated applications to run directly on the device instead of being connected to an external computer. With its 6DoF motion controller, the Hololens 2 is good in terms of tracking and control,

Figure 2.4: Microsoft HoloLens 2 [55]

enabling accurate and organic interactions with virtual objects. The device is shown in Figure 2.4.

2.6 Front-end

Flutter Web [23] emerged as an advantageous choice for the front-end framework for this project's website. Despite the existence of alternative web development frameworks, such as React [56], Angular [57], and Vue.js [58], to name a few, Flutter Web was selected due to its superior support for UI customization and its capacity to generate high-quality, responsive interfaces [23].

The extensive array of programmable widgets and features offered by Flutter Web significantly contributed to its suitability for this project. The straightforward integration of buttons, text fields, and layouts, courtesy of pre-designed elements, built in libraries, permits the creation of a cohesive, refined UI that is aesthetically pleasing and user-friendly. Furthermore, Flutter Web facilitates the use of a single codebase for a potential mobile app, thus ensuring a consistent user experience across multiple devices.

Compared to other web development frameworks, Flutter Web offers a polished experience to create UI for both mobile and web. With other solutions such as React, Angular, or Vue.js, separate codebases for mobile and web would be required, which can be time-consuming and may require different skillsets.

An additional advantage of Flutter Web is its adherence to a reactive programming approach, simplifying the management of user interactions and the real-time updating of the UI. This feature is essential for this project, given the website's role in managing MR equipment and digital exhibitions. Ensuring a user-friendly and responsive UI is therefore important.

In summary, Flutter Web was selected as the preferred framework for crafting the UI of the proposed web application. This choice was motivated by Flutter's broad array of customizable widgets and tools, its capacity for constructing high-quality, adaptive interfaces, and its cross-platform capabilities. Moreover, its reactive programming model, alongside its swift development cycle featuring the "hot-reload" function that facilitates real-time code modification, adds to its appeal. These qualities render Flutter an optimal selection for creating a visually engaging, user-friendly, and efficient UI suitable for the target user group: museum employees. Furthermore, the potential for using the same codebase for a mobile application ensures a consistent user experience across various platforms, which is a notable advantage offered by Flutter.

2.7 Backend

The backend solution used for this project was Firebase [24]. Firebase is a Backend-as-a-Service (BaaS) [19] platform that provides a variety of tools and services for mobile and web application development. It is developed and maintained by Google, and it offers several features such as real-time databases, authentication, hosting, cloud storage, and more. Firebase is also a fully managed platform, which means that it takes care of server maintenance, allowing developers to focus on building their applications.

The primary reason for selecting Firebase was its extensive suite of services, which, when leveraged, could alleviate the necessity of building these features from the ground up. Additionally, Firebase integrates with Flutter Web, thereby simplifying the process of incorporating backend functionality into the project. With a robust set of documentation and substantial community support, Firebase provides a repository of resources for resolving potential issues.

It deserves mention that the choice of Firebase not only proves advantageous for this specific project but also holds potential benefits for future research, whether these involve the continuation of this project or the derivation of inspiration from it. Given its wide adoption within the industry and its purpose-built features aimed at addressing common challenges faced by developers, Firebase presents an attractive option for subsequent projects.

As will be elaborated upon in chapter 4, Firebase was accompanied by certain challenges during development, particularly concerning communication with the Unity application. Consequently, a Python Flask server was established to mediate the communication between the Unity application and Firebase. The choice of Python Flask as the server choice was dictated by the developers' familiarity with Python and the consideration that the remaining time was insufficient for the development of a more refined server solution.

As illustrated in Figure 2.5, the client communicates directly with Firebase backend services through the web application, eliminating the need for any middleware between the service and the web application. Because of this, almost all queries and other related backend services are written directly in the client application. As for the connection between Unity and Firebase, the traditional setup is more representative.

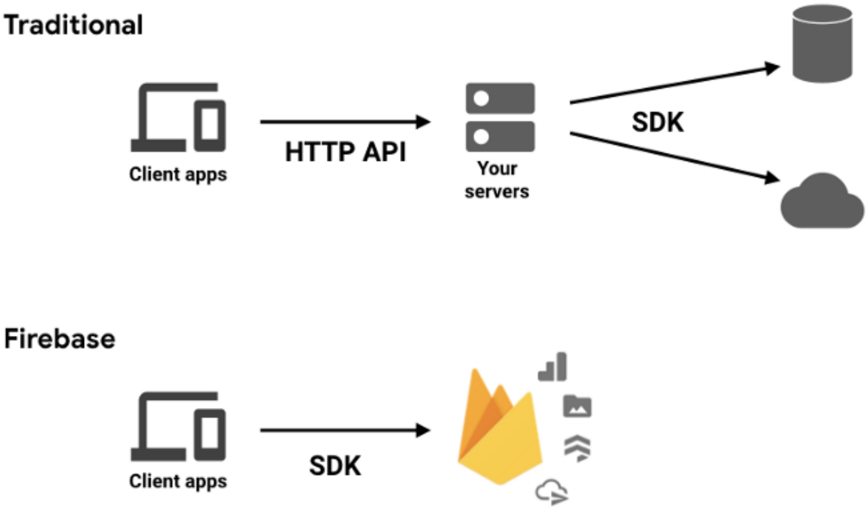


Figure 2.5: Illustration of Firebase vs. traditional backend

Chapter 3

Research Methodology

To answer the research question posed in this thesis, a research methodology is needed. The research method chosen for this project is a revised version of Design Science Research [20]. In this chapter an overview of the Design Science Paradigm is given, why it was chosen for this particular project as well as how it was applied. An overview of the development methodology and evaluation methodology is also covered.

3.1 Design Science Research

The DSR paradigm is an approach to research that emphasizes the creation and evaluation of artefacts to solve practical problems. Its aim is to generate novel and valuable evaluations that establish new concepts, techniques, technical competencies, and products to successfully analyse, design, implement, manage, and utilise information systems [20]. It pushes us to strive for new and efficient ways of accomplishing tasks related to information systems.

To evaluate if the DSR paradigm is the appropriate approach to a project, Hevner et al. [20] propose a framework to define the boundaries for understanding information systems research. The diagram presented in Figure 3.1 showcases the connections that exist between a company's Business Strategy and its information technology strategy as well as the connection between organizational infrastructure and information systems infrastructure.

> The effective transition of strategy into infrastructure requires extensive design activity on both sides of the figure organizational design to create an effective organizational infrastructure and information systems design to create an effective information system infrastructure [20].

Organizational design involves creating a framework that enables the company to achieve its goals efficiently, while information systems design involves creating a framework for utilizing technology to accomplish the company's objectives. Careful planning and design are required in both aspects to ensure a proper

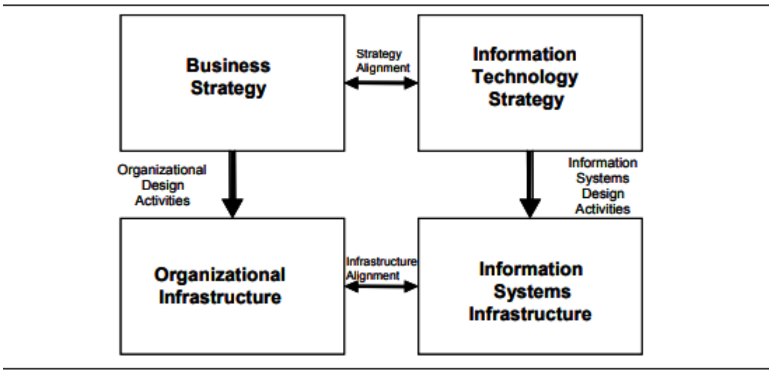


Figure 3.1: Organizational Design and Information Systems Design Activities [20]

alignment with the company's overall strategy and to support its operations effectively.

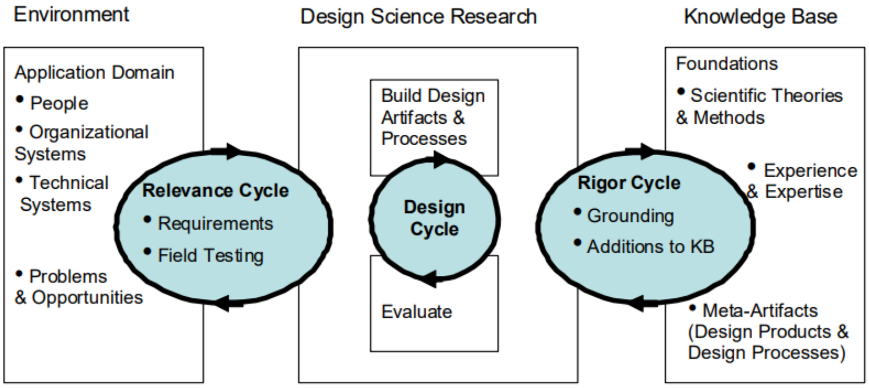


Figure 3.2: Information Systems Research Framework [59]

Hevner et al. [59] also offer a framework that shows the three different cycles present in DSR. Figure 3.2 shows the DSR paradigm defined as three cycles linking the three contextual environments present in a research project. The relevance cycle begins with identifying opportunities and problems in an application context. This must include both requirements for research as well as acceptance criteria for evaluation. The output generated from the DSR must be tested in the application domain to determine whether additional iterations are needed, be it the performance of the artefact or more fundamental problems with the input requirements.

DSR utilizes a broad range of resources, including scientific theories, engineering methods, domain-specific expertise and experiences, and pre-existing artefacts within the relevant field of application. The rigour cycle ensures that the designs produced are research contributions and not routine designs.

The internal design cycle is crucial in the DSR process and it involves iterative activities of constructing, evaluating, and improving the artefact. It is the core of the process and although it depends on input from the relevance and rigour cycles, its partial autonomy is important to underline.

3.2 Guidelines

Hevner et al. [20] offer 7 guidelines for how to approach Design Science as a research method. These guidelines can be seen in Figure 3.3, and states that an artefact must be created innovatively and purposefully for a specific problem domain, from guidelines 1 and 2. The artefact must be rigorously defined, novel, coherent and internally consistent, and thoroughly evaluated for its utility, from guidelines 3, 4 and 5. An effective solution should be found within the problem space from guideline 6, and guideline 7 states that the results must be effectively communicated to both technically as well as managerial-oriented audiences.

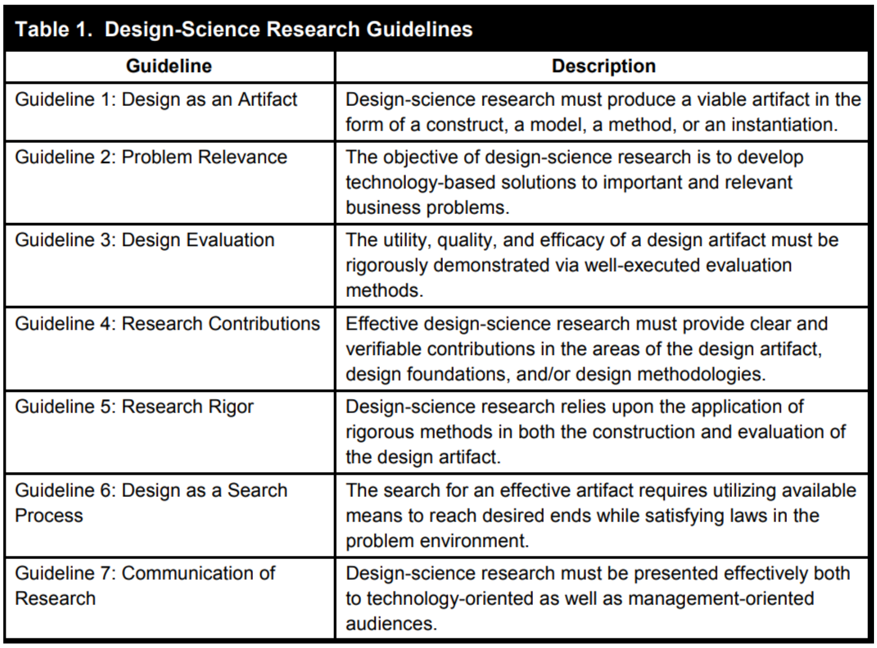


Table 1. Design-Science Research Guidelines

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Figure 3.3: Design Science research guidelines [20]

The design-science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts.

3.3 Applied to our project

As explained in chapter 1, this thesis introduces MuseumXR, an MVP based on relevant research, showcasing how a platform can be created to facilitate the implementation of XR technology in a museum. Design Sciences' goal to extend the boundaries of human and organizational capabilities by creating new and innovative artefacts [20], therefore fits well with the idea of this project. The MuseumXR website is not yet fully scalable and does not entirely represent a realistic environment and therefore it may not align entirely with the

DSR paradigm. However, it serves as a suitable tool for a smaller-scale research project aimed at exploring its potential. This limitation is important to acknowledge and may impact the generalizability of the study's findings. This is further discussed in chapter 6. Nonetheless, the research conducted serves as a valuable contribution to the field and provides insights into the feasibility of implementing similar tools in the field of XR in museums.

3.3.1 Application of the Design Science Research guidelines in the project

Guideline 1: Design as an Artifact

MuseumXR is developed within the framework of the DSR paradigm, and represents an instantiation artefact in accordance with Guideline 1, as depicted in Figure 3.3. The platform is designed to enable more efficient integration of XR technologies in museum settings, by providing museum employees with a tool to administer equipment and software. Although the platform was not fully optimized for use in a real-world environment during the project timeline, the objective was to demonstrate the feasibility of such a system for the museum.

Guideline 2: Problem Relevance

Several problems that exist today in the integration and operation of XR technology in museums are addressed by the proposed solution. The creation of a website aims to facilitate a resource-efficient and easy way of handling XR technology, which is achieved by reducing the amount of manual processing and technical skills required. It is worth noting that the University Museum of Bergen had not previously used any XR technology before this study. The proposed solution is addressing the limitations that the museum faces due to a lack of resources and expertise by offering a website that manages most of the tasks needed for implementing and operating XR technology.

Guideline 3: Design Evaluation

Evaluation plays a crucial role in DSR, as stipulated in guideline 3 by Hevner et al. depicted in Figure 3.3, which emphasizes the need to rigorously demonstrate the utility, quality, and efficacy of a design artefact through well-executed evaluation methods. Such methods provide vital feedback to the construction phase regarding the quality of the design process and the design product under development. In this project, multiple evaluation methods were employed to assess the MuseumXR platform. To perform qualitative analysis, interviews and CTA evaluation methods were utilized to obtain the test subject's opinions regarding the suitability of MuseumXR as a solution in a work environment. To obtain statistical feedback on the usability of the MuseumXR website, a quantitative analysis of the NASA-TLX results was utilized. However, it is important to note that the second round of prototype testing was done with students unaffiliated with the museum, which may be deemed limited in the context of DSR as the participants may not have had the same level of contextual understanding as someone working at the museum. Although the understanding of the context of the museum may be missing from these students' evaluation of the website, a discussion can be had about to what degree this impact the results. This is further discussed in chapter 5.

Guideline 4: Research Contributions

The research project presented in this thesis aims to make a contribution to the field of XR in museums by addressing the problems related to the implementation and management of XR in museums. The web application developed in this project serves as an MVP and is not yet fully optimised and lacks some desired functionality. While the tests conducted with the two small groups have shown promising results, as outlined in chapter 5, the solution has not yet been validated in a real-world setting with visitors as the test had to be done in experimental circumstances, due to the tasks being scenario based. The research project's contribution lies in its potential to enhance the implementation and administration of XR in museums, and it represents a significant and novel step towards achieving that goal. By open sourcing the codebase for this project it allows others to take inspiration or contribute beyond what was achieved throughout this project.

Guideline 5: Research Rigor

The present study's prototype was developed through an iterative design process involving museum employees and an external supervisor with insight into implementing XR in museums from several other projects, as per the DSR paradigm. The process entailed several rounds of conversations with the collaborators to elicit their expertise, requirements and expectations regarding the envisioned solution. These requirements were incorporated into the design and development process of the prototype. Two rounds of user testing were conducted, one with museum employees and one with students, to assess the functionality and usability of the prototype. However, due to time constraints, it was not feasible to conduct any additional rounds of testing to refine and improve the prototype further. Nonetheless, the design and development process of the prototype adhered to the principles of rigour and relevance, which are essential for both design-science and behavioral-science research. The artefact's rigour was maintained through established guidelines for website development discussed in chapter 4 and appropriate evaluation methods discussed in section 3.5. The development process also involved the incorporation of collaboration partners' requirements and expectations into the design process and the assessment of the prototype's applicability and generalizability, ensuring relevance to the intended user base.

Guideline 6: Design as a Search Process

In this project, a website for managing MR HMDs and digital exhibitions in a museum environment was developed, underpinned by Jakob Nielsen's heuristic principles for UX design. The guideline Design as a Search Process implies a process of navigating through a space of possible design solutions, guided by constraints such as time, resources, and specific project requirements. However, the project's unique contextual factors and time limitations prevented the comprehensive application of this heuristic search process. These factors served as the uncontrollable 'laws' that dictated the direction of the design journey.

The resulting website was a 'satisficing' solution - a term derived from design science literature indicating a solution that is sufficient and satisfactory for the current needs, albeit with the understanding that future refinements may be necessary. This solution was reached by navigating the design space under the given constraints, embodying the essence of "Design as a Search Process".

Following field testing, the website yielded constructive feedback, yet a more thorough investigation is warranted. The goal is to better ascertain the degree to which the heuristic search process was deployed, and its impact on the user experience.

Despite the progress made, the current state of the website marks a point in the ongoing journey of design science, not a final destination. The website will be subjected to further iterations, each guided by the principle of "Design as a Search Process", refining its functionality and user experience in response to additional research and feedback. In this way, the website's evolution embodies the iterative, search-driven nature of design science.

Guideline 7: Communication of Research

This project has effectively communicated the design-science research to both technical and managerial audiences, focusing on the construction, use, and the potential benefits of the developed artefact. The research findings and comprehensive details of the project are encapsulated in this thesis. Further, the code for the prototype has been made open-source, fostering transparency and promoting future innovation and research.

3.3.2 Design Science Research cycles

Figure 3.4 showcases how we have applied the Design Science paradigm in our project. In the relevance cycle, the primary focus was to identify the practical problem domain and to understand the specific needs and requirements of the museum. Extensive engagement with museum employees and external supervisors, as well as academic research, was conducted to establish the necessity for an XR implementation platform. Initially, the aim of the project was to enhance a single museum exhibition with digital elements through mixed reality. However, after careful deliberation and consultations, it became apparent that museums faced challenges in implementing and managing XR experiences partly due to a lack of dedicated personnel and technical expertise. Consequently, the project's objective was redefined to develop a user-friendly platform that would simplify the implementation and management of XR experiences for museum employees.

The rigour cycle involved drawing upon existing scientific knowledge, theories, and methods to inform the design of the XR implementation platform. A thorough literature review was conducted on XR technologies, museum experiences, UI and UX design to ensure that the prototype was built upon a solid theoretical foundation. Heuristics for effective UX design were employed, and a comprehensive search for similar solutions was undertaken. Although it can not be definitely concluded, no existing solutions catering specifically to museum use cases were found, thus indicating an innovative nature of the proposed platform.

The design cycle focused on the development, evaluation, and refinement of the prototype through an agile approach. The project employed tasks, user stories, sprints, and iterations to systematically improve the design. While the evaluation of the prototype involved qualitative and quantitative methods during two user tests with museum employees and students, it is acknowledged that a more rigorous evaluation process with more tests and refinements would have been preferable for better alignment with DSR principles. Despite this limitation, the project demonstrates the practical application of DSR methodology in designing an innovative XR implementation platform tailored to the unique needs of museums.

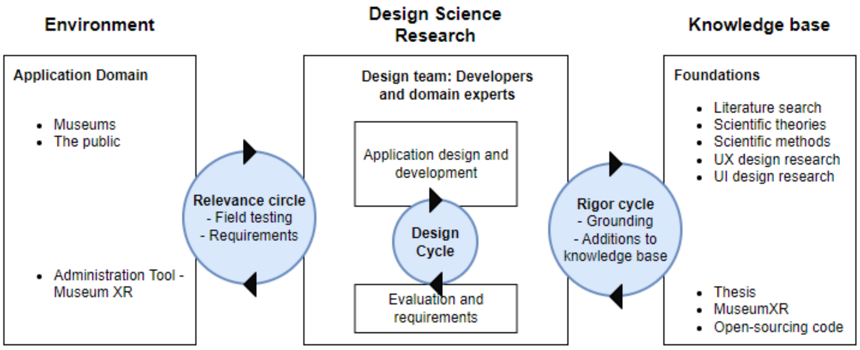


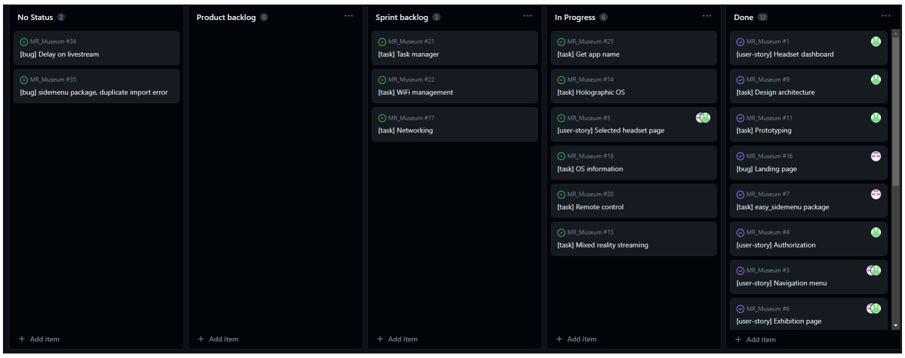
Figure 3.4: Information Systems Research Framework [59]

3.4 Development methodology

An Agile-inspired approach was employed throughout the project, aligning with the principles of DSR. While the consistency of sprint durations was not always maintained due to varying focus on development during different periods, the flexibility and adaptability of the Agile methodology allowed us to address changing requirements and emerging insights. This approach emphasized iterative development, continuous evaluation of the solution, and close collaboration with the museum, domain experts and external supervisors, ensuring the development went in a relevant and correct direction.

The Agile process was characterized by a series of sprint iterations managed with GitHub projects illustrated in Figure 3.5. Each sprint usually lasted between 1-3 weeks, which involved planning, development, testing, and review stages, and as in DSR, this fostered an environment of continuous learning and knowledge creation [60]. This iterative process facilitated the refinement of the solution and enabled us to better address the problem at hand through incremental improvements, despite the varying lengths of sprints.

Throughout the project, adherence to the chosen development methodology was challenged during certain time periods. Establishing consistent communication with contacts at the University Museum of Bergen proved to be difficult, which further complicated the collaboration with the museum. Additionally, unanticipated challenges specific to the HoloLens extended the duration of certain sprints beyond expectation. Limited documentation on the HoloLens Application Pro-



No Status 2

MR_Museum #34
[bug] Delay on livestream

MR_Museum #35
[bug] sidemenu package, duplicate import error

Product backlog 0

Sprint backlog 3

MR_Museum #21
[task] Task manager

MR_Museum #22
[task] WiFi management

MR_Museum #17
[task] Networking

In Progress 6

MR_Museum #25
[task] Get app name

MR_Museum #14
[task] Holographic OS

MR_Museum #5
[user-story] Selected headset page

MR_Museum #18
[task] OS information

MR_Museum #20
[task] Remote control

MR_Museum #15
[task] Mixed reality streaming

Done 12

MR_Museum #1
[user-story] Headset dashboard

MR_Museum #9
[task] Design architecture

MR_Museum #11
[task] Prototyping

MR_Museum #16
[bug] Landing page

MR_Museum #7
[task] easy_sidemenu package

MR_Museum #4
[user-story] Authorization

MR_Museum #3
[user-story] Navigation menu

MR_Museum #6
[user-story] Exhibition page

Figure 3.5: Example from one sprint in GitHub projects

gramming Interface (API) and a lack of expertise in this area contributed to these difficulties. However, during phases primarily focused on front-end development, the methodology yielded positive results when collaboration with all relevant partners was successfully conducted.

Regarding the tools employed in the development process, as previously discussed in chapter 2, the web application was created using Flutter v.3.7. Meanwhile, the XR application was developed utilizing the Unity game engine, specifically version 2021.3.15f1. Firebase was used as a cloud backend server and the microservice backend servers were developed with Python using Flask v.2.3.

Other tools and software used in the project:

- Visual Studio Code [61] as IDE for programming
- Uizard [62] high-fidelity prototype drawings
- draw.io [63] for figures and visualizing models
- Figma [64] for illustrations and drawings
- Git and GitHub [65] for version control
- GitHub Projects [66] for sprints and tasks
- Asana [67] for planning and tasks related to writing
- Miro [68] for planning

All tools are visualized in Figure 3.6

3.5 Evaluation methodology

In choosing an evaluation method for this research project, several conditions had to be considered. The research question and goal, the characteristics of the population being studied, the context of the use, and the resources available for conducting the evaluation. An evaluation of the psychometric properties of the evaluation method was also important to ensure that the results obtained from the evaluation were meaningful and accurate.

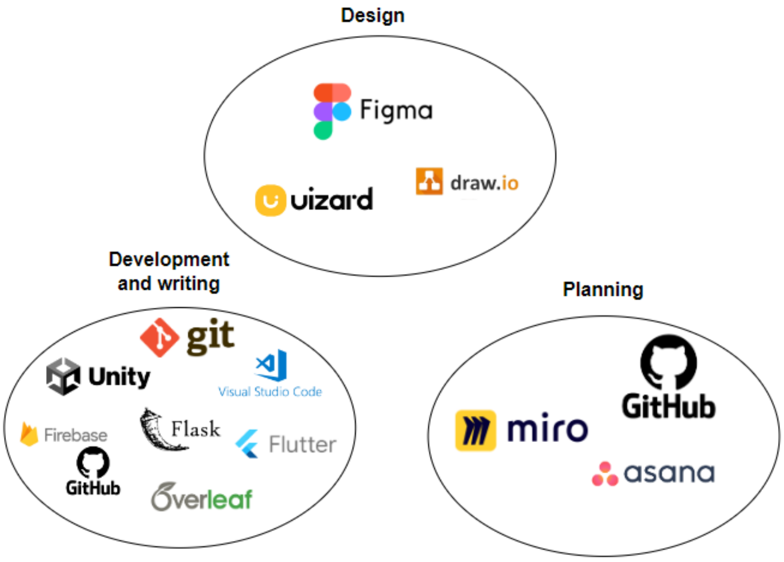


Figure 3.6: Overview of the tools used

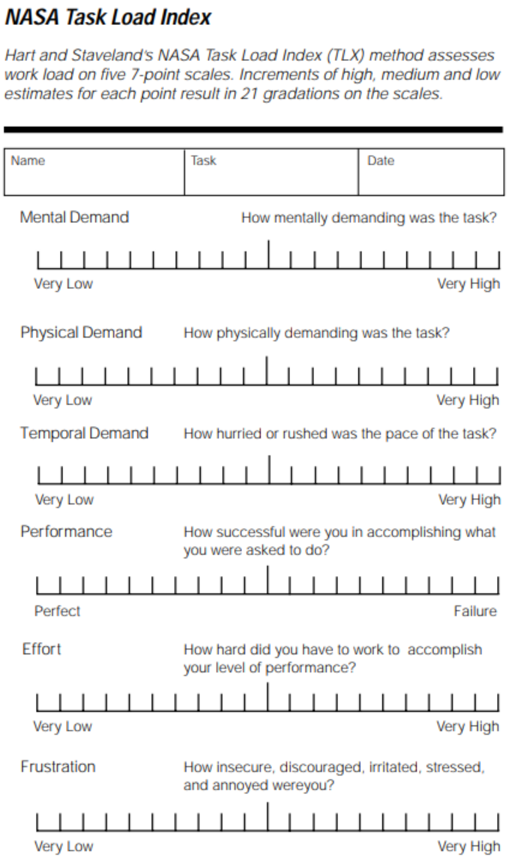
3.5.1 Criteria for the evaluation

Evaluating a web-based application can be done in many ways all depending on what the focus of the evaluation is. From the research question in chapter 1 "How can an administration tool be developed for museums to aid in the facilitation of multiple head-mounted displays and extended reality applications?", the aim is to evaluate how/if the solution proposed in this project can help museums implement XR to their exhibitions. To determine the criteria for success for an evaluation of this kind the context of the museum has to be taken into consideration. The important factors for a museum context need to be established in order to know what type of evaluation method to choose. Some of these factors were established in the initial iterations of communication with the museum representatives. These included a lack of technical expertise and practical factors such as how the layout of the museum would affect the implementation. The museum's layout spans multiple floors, posing a challenge due to the presence of only one station, namely the reception desk, available at the entrance for employee use. Being able to access the application on different devices such as an iPad or phone would therefore be necessary for this museum and other museums in the same situation. The problem of insufficient technical expertise was identified not only within the context of this museum but was also observed by certain representatives from the museum in their experiences with the implementation and operation of technology in other museums. Based on insights gained through communication with these representatives, it appears that having dedicated IT staff is not a common practice among museums in Norway. Hence, while some factors may be specific to a particular museum, there is a justifiable basis to infer that other museums face similar circumstances. This would require the application to be not only user-friendly for individuals lacking technical expertise but also necessitate a limitation on the level of the workload from museum employees regarding the utilization of such a solution.

3.5.2 NASA-TLX

From considering the different factors, the evaluation method chosen for this project was the NASA-TLX. NASA-TLX is a well-established and widely used method for evaluating the workload associated with performing a task. It was initially developed to assess the workload of pilots but has since been adapted and used in a range of other contexts, such as website design evaluation [69].

The NASA-TLX consists of six subscales, each of which evaluates a different aspect of a task's workload as depicted in Figure 3.7. These subscales include mental demand, physical demand, temporal demand, performance, effort, and frustration. Not all of these aspects may apply to using a web application in a museum setting and the NASA-TLX lets you weigh these aspects pair-wise before performing the actual evaluation of the tasks you are given. The pair-wise comparison involves presenting participants with a series of choices, where they must indicate their preference or judgment between two aspects at a time. This is repeated for all possible pairs of aspects and gives a nuanced insight into the relative significance of the different workload dimensions.



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 3.7: NASA-TLX Scoresheet [70]

Hart demonstrates that the NASA-TLX can be modified to eliminate the weighting process, known as Raw TLX (RTLX), which exhibits comparable sensitivity to the original NASA-TLX [69]. This finding aligns with the assertion made by Virtanen et al. "This Raw-TLX is an effective and straightforward approach when the importance of dimensions is roughly equal." [71], albeit under the condition of approximately equal importance among the dimensions. Consequently, if any differences exist among the various aspects, it becomes necessary to incorporate the weighting process in the evaluation. The examination of individual subscales in addition to the overall workload serves to highlight the evaluative strength of this method: the diagnostic value the component subscales possess in the assessment.

The NASA-TLX has been shown to be a reliable and valid measure of task workload across a range of contexts. However, it is important to note that the NASA-TLX is a self-report measure and is therefore subject to some limitations. For example, participants may not accurately perceive or report their workload, or may not understand the subscales. Hart and Staveland put forth a persuasive argument, as cited in their doctoral thesis, indicating that the subjective perception of workload offers a more appropriate means of assessment in contrast to the pursuit of objective measures like heart rate, which may exhibit substantial variations based on the characteristics of the task at hand [72].

To calculate the subscale scores, the participant rates each subscale on a 20-point scale. The points are labelled from 0-20 and are anchored by descriptive terms such as "low" and "high". The participant selects the point on the scale that best represents the perceived workload for each subscale. For example, on the mental demand subscale, the participant might select the point on the scale labelled "high" to indicate that the mental demands of the task were high.

The subscale scores are then weighted according to their relative importance in the task. The weighting is determined by the participant performing the pairwise comparison between the aspects which results in a ranking of the aspects between zero and five. The weights are then multiplied by the subscale scores and summed to obtain a weighted subscale score. The overall workload score is obtained by summing the weighted subscale scores and dividing by 15. The division by 15 is to normalize the scores to a 0-100 scale.

3.5.3 Alternatives to NASA-TLX

Several other evaluation methods were considered to evaluate the workload of the tasks. Subjective Workload Assessment Technique (SWAT) and Workload Profile (WP) are two other methods that evaluate workload. SWAT [73] employs a subjective rating approach utilizing three levels, namely low, medium, and high, for evaluating workload across three dimensions: time load, mental effort load, and psychological stress load.

Tsang and Velazquez proposed the multidimensional instrument known as WP [74]. Drawing from Wickens' multiple resource model [75], the WP aims to integrate the benefits of secondary task performance-based procedures, characterized by high diagnosticity, with subjective techniques that offer high subject acceptability while requiring minimal implementation and avoiding intrusiveness. Despite its potential, Tsang and Velazquez acknowledge the need for further comprehensive research to explore the properties and efficacy of the Workload Profile technique.

Both the SWAT and the WP offer valuable insights into individuals' cognitive demands. While SWAT and WP have demonstrated utility, the NASA-TLX emerged as a preferred choice for this project due to its comprehensive and validated nature. NASA-TLX captures workload across multiple dimensions, including mental, physical, and temporal demands, offering a robust and standardized approach for workload assessment. The participant's weightings presented in chapter 5 also substantiate this as some aspects were more important than others and therefore gave a more nuanced evaluation. Its established reliability and wide applicability made NASA-TLX a preferred option when requiring accurate workload evaluation.

3.5.4 Concurrent Think Aloud

The Thinking Aloud evaluation method encompasses several approaches, including the CTA and Retrospective Think Aloud (RTA) methods. CTA involves participants verbalizing their thoughts and actions during the test, while RTA is conducted retrospectively, with users reflecting on the tasks either from memory or by reviewing recorded videos. Bowers and Snyder's study found that RTA participants tended to offer suggestions and explanations, while CTA participants primarily provided descriptions of their actions [76]. Similarly, Haak et al. discovered that "These results indicate that the CTA method is a more faithful representative of a strictly task-oriented usability test, while the RTA method is likely to yield a broader gamut of user reactions.", referring to the results in her paper on retrospective versus concurrent think aloud protocols on testing the usability of an online library catalogue [77].

In terms of task performance, Alhadreti and Mayhew found no significant differences between RTA and CTA groups in their study, both in task performance and subjective ratings of the system in their investigation into the different types of think aloud protocols [78]. Bowers and Snyder also reported "No performance differences between subjects using concurrent and retrospective protocols were found" [76]. Since the performance disparities appear to be negligible and considering the project's aim to evaluate website usability in the context of XR in the museum, the CTA method was deemed the most suitable approach.

CTA holds significant importance in user experience research as a widely utilized method for understanding user interactions with products, systems, or interfaces. By capturing participants' thoughts and actions in real-time, CTA provides valuable insights into users' cognitive processes and the challenges they face during task execution [79]. Additionally, this method is particularly effective in uncovering unreported issues such as confusion or misunderstandings, leading Nielsen to assert that "Thinking aloud may be the most unique and valuable method in usability engineering" [80].

3.5.5 Walkthrough of the user tests

To evaluate the MuseumXR website, scenarios were made to simulate possible real-world situations which also encapsulated the functionality of the website. The participants were first given an overview of the project including information about the collaborating parties, motivation for the project and other miscellaneous information about the concepts of XR, the hardware used and so on. They then got to play around with a test application in the Hololens to get a better feel for how it worked. After they felt comfortable enough to continue they were presented with the overall plan for the user test and the different evaluation methods which were going to be used during and following the test, such as the NASA-TLX, CTA and a short interview. They were then given a small survey with a few questions to map their experience of using various technological devices and programs to possibly give a better understanding of their answers in the evaluation of the website.

The main tasks for the user test were divided into four, with each task being associated with a scenario intended to emulate a plausible museum visitor experience. Before each of the tasks, the scenarios were first performed to the participants to set the scene. They were subsequently encouraged to explore by themselves to find the solution to the task connected to that scenario. The participants could ask questions during the evaluations if they met any difficulties or had any questions. An emphasis was also put throughout the evaluation on reminding the participants to think aloud during the tests, as Ericsson and Simon describe the importance of this in their prescribed guidelines [81].

Relatively quickly after the participants were done with the tasks, the NASA-TLX was explained to them in greater detail and they were able to read the instructions themselves before being presented with the questionnaire. After performing the NASA-TLX questionnaire, a brief semi-structured interview was performed to get some more subjective insight from the participants about the project as well as the evaluation itself.

Chapter 4

Design and Solution

In the following chapter, decisions made during the development of the web application are detailed together with key principles adhered to during the design process. This is succeeded by an architectural overview of all the components incorporated in the application. Finally, a discussion concerning potential alternatives instead of Hololens for HMDs is presented.

4.1 Design choices

In the initial design phase of the web application, several factors were considered to ensure it would be easy to use and user-friendly. This was important as this tool is to be used by museum employees that do not necessarily have any technical skills. The importance of this was confirmed in early dialogue iterations with representatives from the museum. While considering these factors, maintaining best practices for designing administration tool webpages was also kept in mind.

First, simple wireframes were developed to outline a simple layout and structure of the application as seen in Figure 4.1. Once a clear understanding of the design goals was present, we transitioned to using Uizard [62], a prototype tool that allows rapid creation of high-fidelity prototypes; see Figure 4.2. Testing different ideas, experimenting with different layouts, and getting valuable feedback early in the development were made possible due to Uizard. This also ensured a thorough requirement and expectation validation before actually starting to work on the solution.

The design process underpinning this project was largely informed by guidelines established by the Nielsen Norman Group (NN/g) [82], an internationally renowned entity in the realm of research-based user experience [83]. This organization disseminates a wide array of articles, principles, and laws centred around UX and UI design.

One such principle is "Jakob's Law of the Internet User Experience" [84], which proposes that users assign their time to other websites and, as such, become familiarized with the dominant design standards and conventions. Consequently,

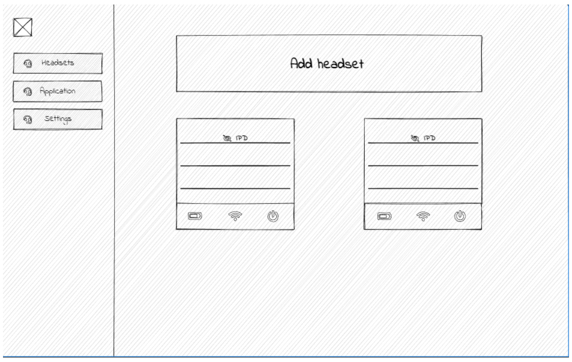


Figure 4.1: Wireframe of HMD dashboard

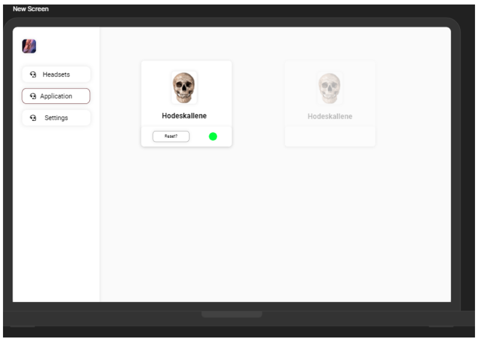


Figure 4.2: High fidelity prototype of application dashboard

upon visiting a site, users anticipate a similar user experience to that of other websites. Given this understanding, a key consideration in the design process for this solution was to maintain intuitiveness and ease of use.

In alignment with the utilization of NASA-TLX for assessing workload, specific design principles were adopted with the aim of mitigating the cognitive workload that a user might encounter when visiting the website. For instance, "Miller's Law" suggests that the human brain can retain an average of 7 +- 2 objects in short-term memory [85]. To lessen the number of possible objects to remember, users are supported by the constant display of relevant information, such as the details of the HMD they are currently managing, and the accessible navigation menu from any point within the app.

"Fitt's Law" was another guiding principle adopted during the design process. This law posits that the time taken by an individual to move a pointer, such as a mouse cursor, to a target area is determined by the ratio of the distance to the target and the size of the target [86]. This law brings to attention a few crucial factors such as the actual distance between elements and the concept of the prime and magic pixels. The prime pixel represents the origin of the user's action, for instance, the location of a button that has been clicked, whereas the magic pixels are those farthest away from the prime pixel. By conscientiously avoiding the magic pixels, and considering the overall distance and the size of UI elements, the design process is aligned with the spirit of Fitt's Law, thereby ensuring an efficient and user-friendly experience.

In terms of specific guidelines or heuristics used when designing, Jakob Nielsen's heuristics [87] were used actively, as this is a widely recognized and proven set of principles to use when designing systems [88]. Jakob Nielsen is part of the NN/g and his heuristics are a set of general principles for evaluating the user experience of websites and web-based applications. They are based on Nielsen's extensive research and experience in the field of usability engineering [89] and are intended to help designers and developers create websites that are easy to use and understand. Numerous studies have demonstrated the effectiveness of these heuristics in identifying usability issues and enhancing user experience [90].

Although MuseumXR is an MVP or a proof of concept rather than a fully-fledged product, following these heuristics proved to be very beneficial, particularly due to the evaluation and testing throughout the development process and when doing the NASA-TLX evaluation. By following these principles, we ensured a strong foundation for usability, even in the early stages.

In addition, the decision aligns with the chosen research method, namely DSR. The heuristics serve as guidelines in creating a better artefact that addresses real-world problems and contributes to the knowledge base in the field, which is at the core of DSR [91]. Recent research supports the continued relevance and importance of usability heuristics in modern web development [92]. In the subsequent subsection, we will provide an explanation of each heuristic, as well as elaborate on how these principles were applied in the project.

4.1.1 Application overview

In the following two figures, an overview of the application is presented. In Figure 4.3 the HMD overview is shown with one device connected. Whilst in Figure 4.4 an overview of the digital exhibitions is shown where both of the exhibitions are activated.

4.1.2 Jakob Nielsen's heuristics

Visibility of system status

> *"The system should always keep users informed about what is going on, through appropriate feedback within a reasonable time [87]."*

To ensure good system status visibility to our application users, we ensured that continuous feedback on the user's actions was provided. Examples include progress indicators when content is loaded, easy-to-understand error messages and confirmation dialogues after performing actions as depicted in Figure 4.5.

Match between system and the real world

> *"The system should speak the user's language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order [87]."*

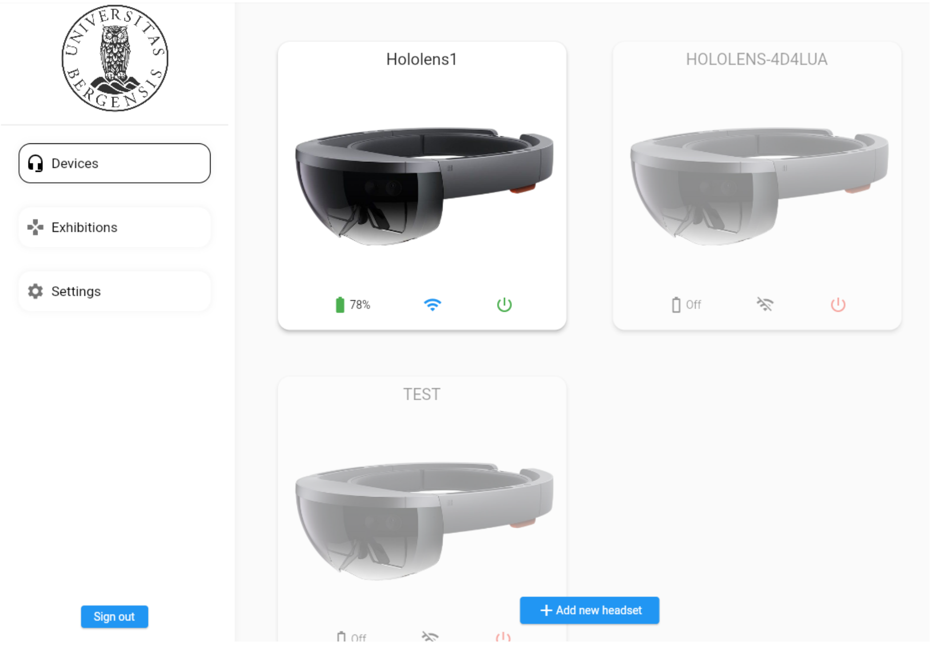


Figure 4.3: The HMD overview in MuseumXR

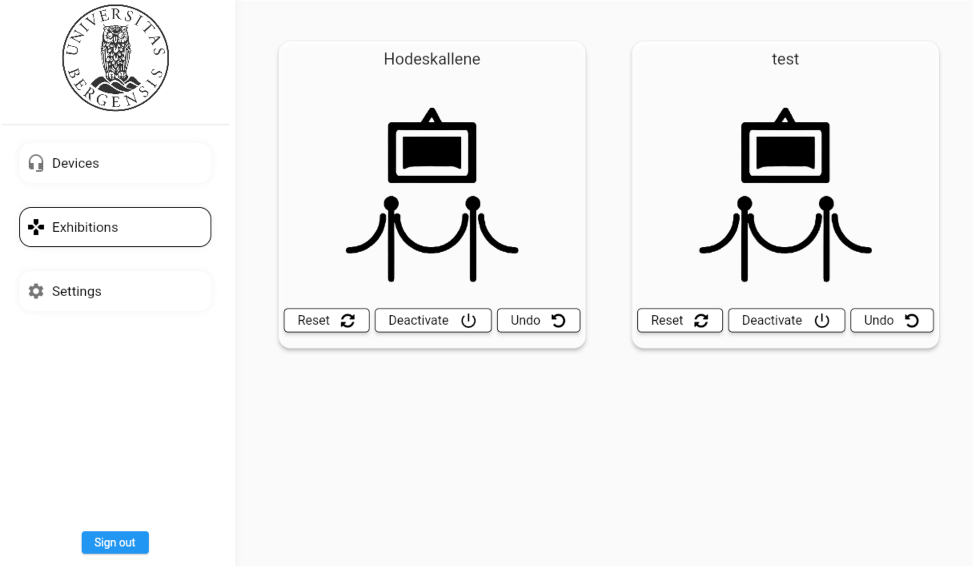
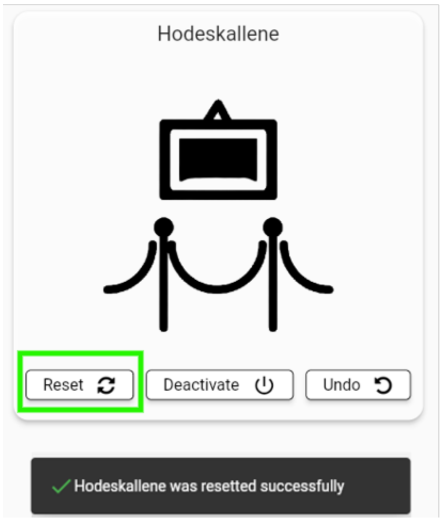


Figure 4.4: The exhibition overview in MuseumXR

The choice of icons, terminology, and additional interface components was attempted to cater to museum employees, ensuring that the interface aligns with user cognition.



✓ Hodeskallene was resetted successfully

Figure 4.5: Example of feedback given after pressing the "Reset" button on the exhibition overview page

User control and freedom

> *"Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo [87]".*

To mitigate the possibility of users experiencing disorientation or confinement within the application, the principal navigation menu has been designed to maintain constant visibility, irrespective of the user's ongoing activity. As depicted in the flowchart in Figure 4.6, the application's primary functions can be accessed through minimal steps.

MuseumXR has been developed as an MVP. Thus, for specific actions requiring significant development time to establish a functional reversal mechanism, alternative 'undo' buttons were introduced as a practical solution seen in Figure 4.7. Although not functional for all actions, these buttons were designed to elicit valuable feedback during the testing phase. As the testing process was consistently supervised, it was feasible to emulate the undo functionality without revealing its inoperative nature to the user.

Consistency and standards

> *"Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions [87]."*

Utilizing Flutter as the Software Development Kit (SDK) for the development of the web application facilitated the maintenance of a consistent design style throughout the layout and various components or widgets within the application. This can be primarily attributed to Flutter's built-in design packages, which automatically style the incorporated components.

Moreover, the design adhered to the F-Pattern concept proposed by the NN/g

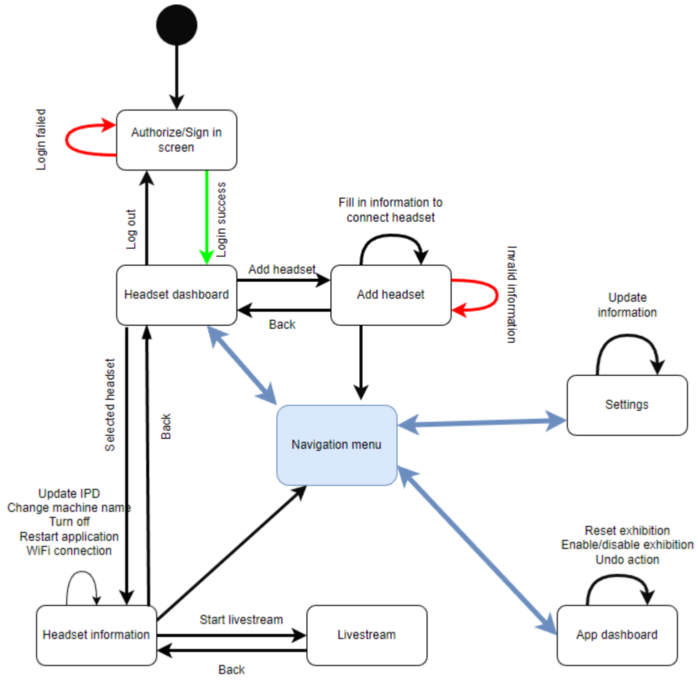


Figure 4.6: Flowchart of the web application

Figure 4.7: Illustration of the undo button

[93], delineating a common method of scanning web content. According to this concept, users initially read the upper horizontal section of a web page, followed by another horizontal scan just below the first, still situated in the upper portion of the page. Subsequently, a vertical scan of the left side is conducted [93]. Consequently, the primary navigation menu was strategically positioned on the

left side to capture users' attention during the initial scanning phase of the page. At the same time, the main content was placed in the upper central area of the webpage. This design approach is illustrated in Figure 4.8.

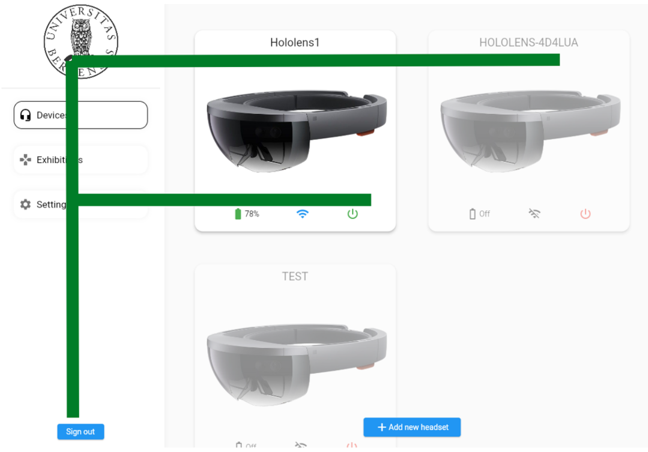


Figure 4.8: F-pattern illustration

Error prevention

> *"Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action [87]."*

Minimizing errors was emphasized; however, as previously mentioned, errors were still present, given that this version primarily serves as a proof of concept rather than a final solution. The predominant sources of errors were predominantly attributable to connection-related issues involving one or more background microservices. To mitigate this, the status of the microservices was meticulously monitored during testing and evaluation, ensuring that potential errors in the feedback received would exert minimal influence on the results. In addition, straightforward yet effective measures, such as type validation on all input fields, were implemented to prevent errors.

Recognition rather than recall

> *"Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate [87]."*

Given the application's limited number of elements and options, extensive modifications were not required to adhere to this concept. Nevertheless, simple colour codes were applied to specific actions to facilitate recognition. Moreover, the HMD name was prominently displayed on all pages related to or associated with it, ensuring its visibility.

Flexibility and efficiency of use

> *"Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions [87]."*

Considering that the application primarily functions as an MVP, most features were not designed to accommodate acceleration for more advanced users. However, the architecture, which will be discussed in greater detail in the architecture section, has been structured in a manner that is amenable to future adaptations, potentially incorporating shortcuts. This approach enables, for instance, the implementation of role-based access controls, ensuring that museums can restrict specific features and options solely to qualified roles, should they wish to do so.

The necessity of this feature was highlighted during the most recent testing sessions with museum employees, wherein inquiries emerged regarding adding HMDs. Given that this procedure is slightly more advanced and likely to occur infrequently, it may be prudent to reserve this feature for individuals possessing greater responsibility or technical expertise.

Aesthetic and minimalist design

> *"Interfaces should not contain information that is irrelevant or rarely needed. Every extra unit of information in an interface competes with the relevant units of information and diminishes their relative visibility [87]."*

We strived to create a minimalistic and aesthetically pleasing design for our web application. As already mentioned, Flutter makes this easy with its built-in design components. To reduce visual noise, we utilized whitespace effectively and adopted a clean and modern design with few colours and options.

4.1.3 Key design elements and usability

As mentioned in the preceding section about User control and Freedom, specific components exhibit greater prominence than others. The navigation menu, for instance, is one such component that, as illustrated in the flowchart in Figure 4.6, is accessible from virtually any location within the application. This feature enables users to navigate the app and efficiently perform their desired actions.

4.1.4 Functionality

"The goal is not so much to support the design of more powerful authoring tools as it is to design tools that meet the needs of realistic user audiences." [94].

As emphasized by Murray in his discourse on Theory-based Approaches to Authoring Tool Usability, it is posited that the scope of functionalities in a tool should be constricted to the specific necessities of its intended user group. Rather than focusing on incorporating an extensive range of functionalities with an objective of creating an advanced tool, it is crucial to curate the tool to align with the user's needs. In the case of MuseumXR, the proposed users are identified as museum employees. Hence, the development emphasis was placed on generating a solution offering essential functionalities tailored to their museum-related tasks.

During the initial phases of functional planning, hypothetical scenarios were developed to embody potential situations that might arise in an ordinary workday. The intention was to assimilate functions that could efficiently manage these situations.

In the course of creating these scenarios, it became apparent that the ability to monitor visitors' experiences within the HMD was important and thus the implementation of a livestream option was done. Given that museums would typically operate with multiple HMDs in a real-world scenario, the functionality to provide an overview of all HMDs, displaying current battery level, power status, and WiFi status, was identified as beneficial and subsequently integrated.

Despite the daily use relevance of the aforementioned functionalities, the ability to add new HMDs via the website was incorporated. While not deemed vital for regular use, its absence would necessitate manual data input on the backend server, an operation deemed complex for the intended users. The addition of new HMDs via the MuseumXR website, though simplified compared to the backend server procedure, still requires certain technical knowledge such as finding the HMD's IP address. Given sufficient exposure and experience with the technology and MuseumXR, this would be a relatively simple task.

Functionality relating to administering the digital content of the XR application was developed to showcase simple yet useful ways the employees could manage content in XR applications, depicted in Figure 4.9. Resetting the position of the digital elements and enabling/disabling the digital elements were implemented. The reset functionality gives the employees the option to help visitors reset certain elements within the application if the elements for example were to become unmanageable, whilst the disabling/enabling functionality serves as a way for the employees to decide which digital exhibitions should be present. An application for the HMD pertaining to the physical skull exhibition was developed.

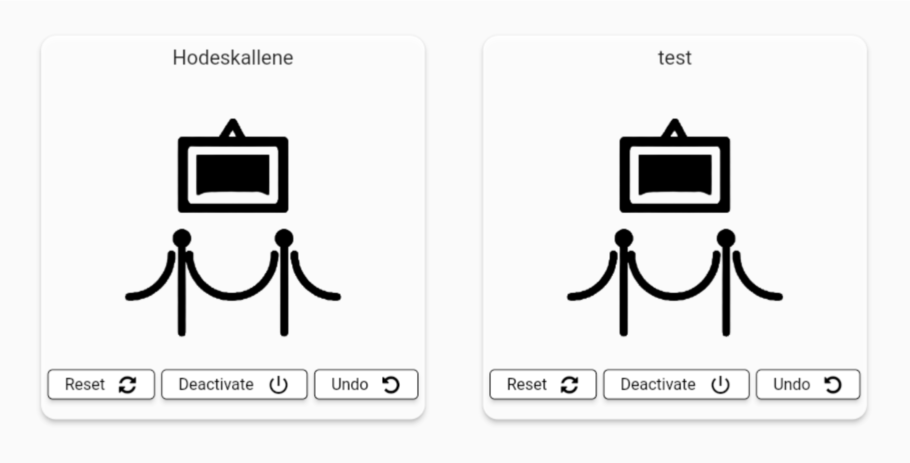


Figure 4.9: Illustration of functionality to administer digital content

This application contains a table at a fixed location and an interactable skull placed on top of it. An accompanied script is attached to the skull that connects the relevant parameters to MuseumXR. The size, position and scale of the skull were connected to the reset functionality in MuseumXR meaning that if the Reset button is pressed on the website, size, position and scale reset to their original values resulting in the skull being placed on top of the table again in the application. The enable/disable functionality in MuseumXR deactivates, making the skull disappear, or activates, making it appear again.

The focus of the functionality developed in relation to the administration of the HMD application's digital content was to exemplify straightforward yet effective methods by which employees could manage content within XR applications. The implemented capabilities included resetting the position of digital elements and enabling or disabling these elements. The reset feature provides employees with the capacity to assist visitors by restoring particular objects within the application to their original positions, for instance, if the objects were to become unmanageable. The disable/enable functionality allows employees to dictate which digital exhibitions should be visible or not to the visitors.

A corresponding application for the HMD was created in Unity, specifically for the physical skull exhibition at the museum. This application encompasses a stationary 3D model of a table with an interactive 3D model of a skull positioned atop it. An affiliated script is connected to the skull, linking relevant parameters to MuseumXR. The skull's size, position, and scale were linked to the reset function in MuseumXR, resulting in the resetting of these elements to their original values if the Reset button is pressed on the website. Consequently, the skull is restored to its initial position atop the table within the application. The enable/disable functionality within MuseumXR allows for the activation or deactivation of the skull, causing it to appear or disappear, respectively. A picture of the 3D skull is displayed in Figure 4.10.

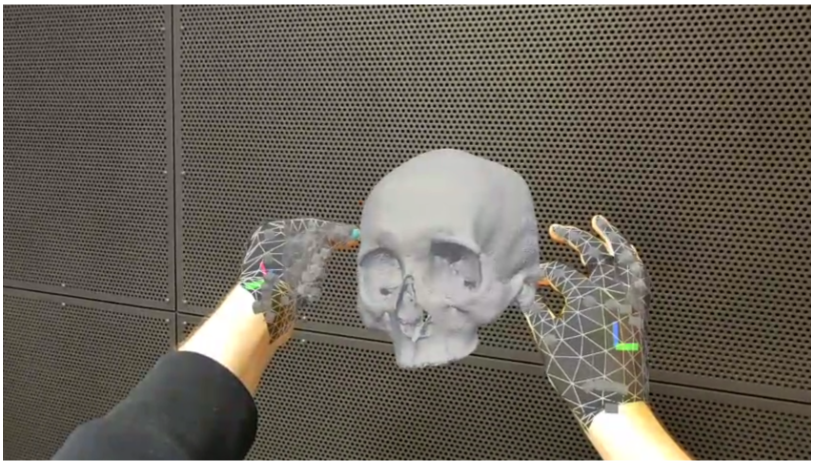


Figure 4.10: The skull from the Unity application

4.2 Architecture

The overall architecture of the application consists of several key components that can be broadly divided into two main parts: the front end and the back end. The front end serves as the user interface for the web application. It provides an intuitive means for the end user to manage the HMDs and virtual exhibition elements. The backend is responsible for handling data input, communicating between all the different components, and maintaining the overall state of the application. In the following subsections, each aspect of the architecture will be discussed in greater detail, starting with the front end, followed by the back end, security, and scalability.

4.2.1 Frontend

The decision to utilize Flutter as the frontend framework was a significant factor in shaping the overall frontend architecture. As will be discussed in the [sec:backend]backend section, it affected the backend quite a bit. Choosing Flutter had many benefits, one being that Flutter is a cross-platform UI toolkit designed to allow code reuse across multiple systems. It also runs in a Virtual Machine (VM) when developing, offering stateful "hot reloads" without recompiling. This provides fluent and effective workflows. Flutter apps are also compiled directly to machine code or JavaScript if targeting the web, resulting in better compatibility due to being native to the targeted system [23].

The application's frontend structure was carefully organized to ensure ease of navigation, maintainability, and scalability and has the following file structure:

- 'assets': This directory stores all static files, mostly images for this project.
- 'models': All classes are defined here to ensure consistency throughout the application. Examples are classes representing the different API responses from the Hololens.
- 'screens': All the different screens within the application.
- 'services': A directory containing files for handling logic related to different services, such as connection to the hololens and all database-related logic.
- 'widgets': A collection of custom-made reusable widgets.

As highlighted in chapter 2, Flutter incorporates built-in libraries to ensure consistency and functionality. However, it is essential to note that many high-level features are implemented as packages, including some from third parties. One such example is HTTP protocol support. Although Flutter is a cross-platform UI toolkit, it does not necessarily guarantee that packages are equally cross-platform. This posed a challenge for this project, as it relied on packages to handle MP4 chunks from the Hololens to facilitate a live stream of the Hololens display within the web application.

After extensive troubleshooting, it was determined that Flutter Web was not ideally suited for this task. Two primary concerns arose in addressing this issue: first, the format needed to be supported by both Flutter and the web browser, and second, the video had to be displayed as a live stream rather than a standalone video. A method for manipulating the received data was required to tackle these challenges. Flutter offers an HTML package called dart:html that enables the manipulation and parsing of HTML data outside the browser [95]. Nonetheless, this approach resulted in either unsupported video formats in the browser or an inability to initiate video playback before closing the connection. As a result, it was concluded that although the format was supported, it was incompatible with the live-stream context.

Due to the scarcity of online resources to address this issue, the decision was made to utilize an intermediate server to convert the live-stream data to HTTP Live Streaming (HLS) before accessing it from the web app. Further details regarding this solution will be elaborated upon in the backend subsection.

4.2.2 Backend

The initial architecture markedly differed from the final version. Initially, the backend comprised solely of a cloud database component and various APIs responsible for managing the communication between all the components, as depicted in Figure 4.11. This strategy aimed to minimize the required services and components by employing Firestore from Firebase in conjunction with the web app and Unity app. All logic would be implemented within either the web app or the Unity app. Nonetheless, this approach proved unfeasible due to several technical limitations, prompting a revision of the architecture to the configuration illustrated in Figure 4.12.

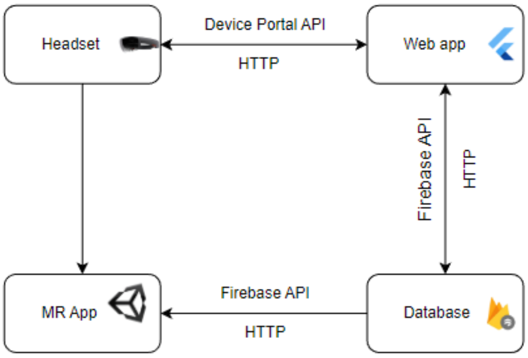


Figure 4.11: Initial architecture of the system

Firestore from Firebase was selected as the service for managing database services. The primary reasons are integration with Flutter, scalability, and ease of use. However, implementation proved more difficult than anticipated. As illustrated in Figures 4.12 and 4.11, both the web app and the Unity app necessitate communication with the database for the system to function effectively. Firebase documentation shows it offers robust support for Flutter, with all libraries being compatible [24]. At first glance, Unity also appeared to be supported. Although support is provided for numerous Unity platforms, not all are covered.

Initially, it was believed that Hololens would be supported due to its Windows compatibility, which purportedly supported the required package - Cloud Fire-

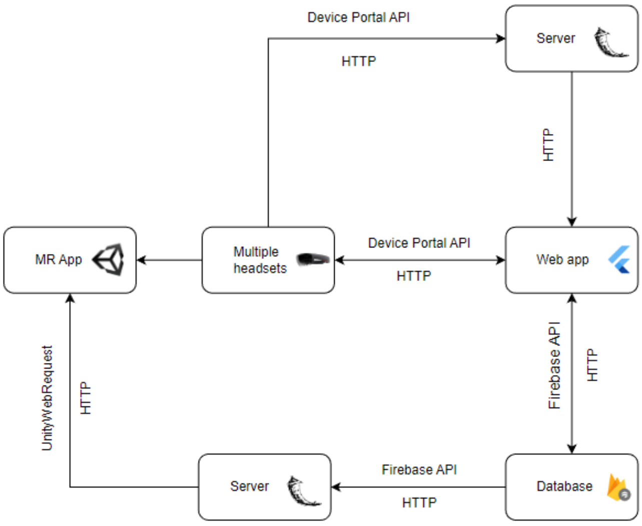


Figure 4.12: Final architecture of the system

store [96]. Regrettably, after extensive troubleshooting and research, it was concluded that Hololens was not supported. This discovery was made after considerable time had been invested in the project. Consequently, a dilemma arose: either change the server solution and modify the completed client-side code to connect to Firestore from Flutter or devise a solution enabling the Hololens application to connect to Firestore.

As most of the database connection occurred between the web application and the code had already been written, the decision was made to create an additional intermediate server acting as a microservice between the Unity app and Firebase to manage all communication.

Unity application

In the Unity application, each object is associated with an individual script responsible for managing communication with the microservice that handles interactions with the database. These scripts contain variables with initial values, and in scenarios such as resetting the position, the script monitors changes in the database. Once a change related to the reset variable is identified in the database, the reset function is activated. Presently, these variables need to be manually inserted in both the Unity script and the web page. However, as discussed in chapter 8, there is potential for enhancing this process by allowing the web page to automatically retrieve any variable that a developer has implemented in the script for modification. A schematic representation of the existing Unity-related architecture is presented in the subsequent subsection, which explains the operation of the microservice that manages database communication.

Microservice to handle Unity and database communication

The intermediate server that handles the communication between the Firebase database and the Unity app is built with Python and the Flask framework. This service ensures an efficient connection between the database and the Unity app. By using the UnityWebRequest API, HTTP requests can be made from the Unity app to the server, which again uses the Firebase API to fetch updates from the database. The architecture is currently designed in a way that new variables can easily be added to expand the features available. All that is needed is to add the variables in the Firestore document in the same location as the other variables that can be changed and make sure to include an initial value in the C script, and include logic based on what kind of value it is.

The following sequence explains an example scenario of how the server would interact with the rest of the system. See Figure 4.13 for an illustration of the points listed.

1. The end user wants to reset the position of an exhibition at the museum and presses the button "Reset"
2. The Firebase API is then utilized to update the document containing the exhibition
3. The server is listening to changes and detects a change in the document
4. Server then has the updated values in JSON format to the Unity application
5. The updated values are then received in the Unity application, the script in the Unity application serializes the values and reads that the position reset button was pressed, and runs the function "resetPosition()".
6. Position is then changed in the application

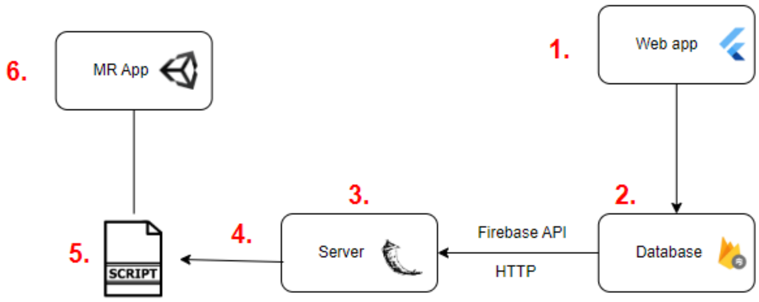


Figure 4.13: Workflow when a user presses "Reset exhibition"

Microservice for live stream possibilities

The intermediate server that handles the communication between the livestream from the Hololens 2 and the website is also built with Python and the Flask framework. This microservice was included due to the incompatibilities with Flutter and the stream from the Hololens. Its purpose is to send a GET request to get the live stream from Hololens, convert it to HLS, and make it available on a URL that can be accessed through Flutter. The solution serves its purpose, but it is not ideal due to the increased delay on the live stream. Approximately five seconds of delay between what the Hololens user sees and the displayed content in the web app was the case with this solution. If the call were made directly from a web app, it is likely that the delay would be comparatively smaller.

4.2.3 Scalability

Scalability plays a crucial role in any architecture solution, as it influences the overall performance and capacity of the system to handle additional users and workload. Using Cloud Firestore as the backend database has several advantages in terms of scalability, such as its ability to handle real-time updates and automatically scale after the number of users [97]. However, as a cloud service offered by Google, the cost escalates with the number of users [98]. In addition, as was realized during the development of this project, its API is not supported for Unity Hololens applications, which is a major downside for this specific use case. In hindsight, using something other than Firestore would probably be the better option due to the cloud aspect, and the main benefit of easily incorporating it disappeared, and a local server was needed anyways.

4.2.4 Security

Implementing security measures was essential in light of integrating various services and a database within this solution. The adoption of Firebase as the database service significantly streamlined security, thanks to its robust rule system that efficiently prevents unauthorized user access [99]. Cloud Firestore security rules allow for granular control over access to documents and collections based on specific conditions such as the user's authentication state or other document properties [100]. These rules are defined using a flexible, JSON-like syntax and are evaluated on every request, ensuring that only authorized users can read or write data.

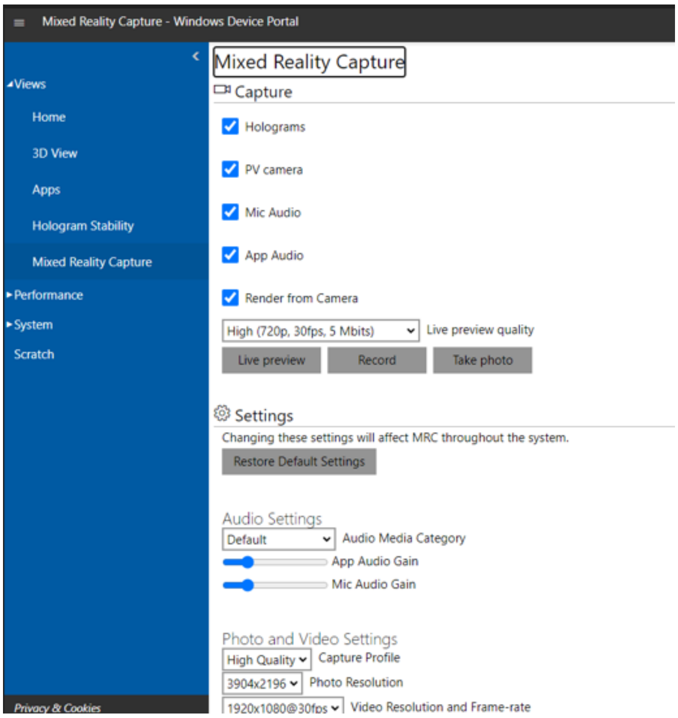
In addition to the Firestore rule system, Firebase Authentication played a crucial role in enhancing the solution's security. Behind the scenes, Firebase Authentication relies on secure protocols like OAuth 2.0 and OpenID Connect to handle user credentials, tokens, and sessions [24]. When a user successfully authenticates, Firebase generates a unique JSON Web Token (JWT) for them, which is then used to securely identify the user in subsequent requests to Firestore or other Firebase services [39].

By combining the power of Firebase's authorization solution with its sophisticated rule system for Cloud Firestore, the creation and management of application users were greatly simplified, while ensuring robust security for the underlying data. It also makes it easy to create user-based access control later by assigning roles to each user and modifying the rules.

4.3 Key features and functionalities

4.3.1 Device Portal

The Microsoft HoloLens 2 device was used in this project, and the Device Portal API reference was utilized to access the device's settings and configuration options, facilitating the development process. The Windows Device Portal (WDP) [101] for HoloLens makes the management and configuration of the device available through a website, using either Wi-Fi or USB connections. A web server operates on the HoloLens, which can be accessed through a web browser on a personal computer. Mixed Reality Streaming presents a live stream from the HoloLens to display the user's Point of View (PoV) as seen through the device. Power settings are also available, allowing for the assessment of the system's low power state and current battery levels. Remote control and task management enable the shutting down or restarting of the target device, as well as the initiation and termination of specific applications. Accessing information through a website, rather than wearing the device and navigating its settings, is made possible by the network and OS information settings found within the WDP. Data pertaining to IP configuration and general OS details, such as machine name, can be retrieved. Although all the functionality related to HMDs in MuseumXR can be achieved from the WDP, it is limited to an overview of one HMD at a time and contains a lot of options and functionality unnecessary for the target audience for MuseumXR. A snippet of WDP is displayed in Figure 4.14



Mixed Reality Capture

Capture

☑ Holograms

☑ PV camera

☑ Mic Audio

☑ App Audio

☑ Render from Camera

High (720p, 30fps, 5 Mbits) Live preview quality

Settings

Changing these settings will affect MRC throughout the system.

Audio Settings

Default Audio Media Category

App Audio Gain

Mic Audio Gain

Photo and Video Settings

High Quality Capture Profile

3904x2196 Photo Resolution

1920x1080@30fps Video Resolution and Frame-rate

Figure 4.14: Snippet of WDP

4.4 HMD APIs

The current solution was developed primarily for the HoloLens as this was the only MR HMD available for this project. However, to ensure versatility if museums want to include other HMDs, it is crucial to support other MR HMDs apart from the HoloLens. Different HMDs will likely have unique APIs or SDKs, necessitating specific code for each HMD, which complicates making the web app compatible with most devices.

An illustration of the differences in APIs between MR HMDs can be seen by comparing Magic Leap [102] with the HoloLens. Magic Leap possesses a set of APIs accessible through the Lumin SDK [103], whereas the HoloLens APIs adhere to WDP API. Although the Lumin SDK and WDP provide analogous features, the procedures for setting up, retrieving, and processing the information vary between the two platforms. Consequently, these differences emphasize the importance of accommodating the unique requirements of each HMD's APIs and SDKs in developing a universally compatible solution.

Given that MuseumXR accesses and updates specific variables within the Unity application, establishing a standard would greatly benefit developers by simplifying the process of incorporating support for this project. This is also in line with other standards becoming more and more popular, such as OpenXR, a royalty-free, open standard that provides high-performance access to AR and VR. They seek to simplify AR/VR software development, enabling applications to reach a wider array of hardware platforms without having to port or re-write their code and subsequently allowing platform vendors supporting OpenXR access to more applications [104]. The same principles can be applied to this project, having a standard would allow both existing and new applications more easily compatible with MuseumXR.

Due to the project's time constraints, an SDK was not developed to assist with this process. Ideally, digital artists working on museum projects would utilize a Unity plugin to do this together with the OpenXR standard, making the application available for most HMDs and also supporting integration with administrative solutions like this project's web application.

Chapter 5

Results

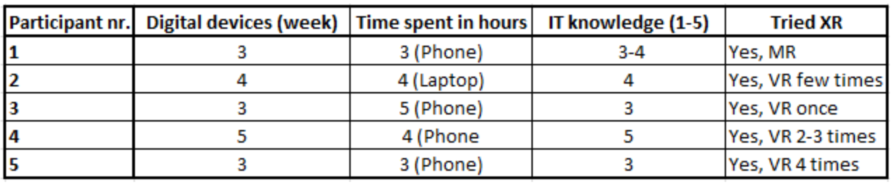
This chapter presents the results from the two user evaluations conducted in this project, the improvement between the two rounds and an overview of the total results from different perspectives.

5.1 First round of evaluations

The first user evaluation was conducted in early April 2023, with five participants working at the University Museum of Bergen. Among these participants, four were students with part-time employment at museum the and one served as a curator.

5.1.1 Technology assessment questionnaire

As detailed in section 3.5, all participants were asked to complete a brief questionnaire designed to assess their relationship to everyday technological devices and their experience with using API equipment. A simplified version of the results is presented in Figure 5.1. The "IT knowledge" column in Figure 5.1 is a reflection of the participant's own assessment of their IT knowledge. The results showed that the majority of the participants had a reasonable level of exposure to new technology and had tried XR at some point in the past.



Participant nr.	Digital devices (week)	Time spent in hours	IT knowledge (1-5)	Tried XR
1	3	3 (Phone)	3-4	Yes, MR
2	4	4 (Laptop)	4	Yes, VR few times
3	3	5 (Phone)	3	Yes, VR once
4	5	4 (Phone	5	Yes, VR 2-3 times
5	3	3 (Phone)	3	Yes, VR 4 times

Figure 5.1: Simplified results from the technology assessment questionnaire

The completion of the questionnaire was followed by inquiries made to participants about any IT systems frequently used in their daily work at the museum. Findings from these inquiries suggested that such use was fairly minimal. Nevertheless, a system did exist on local computers, enabling control over various functions such as lights and curtains. Multiple participants additionally noted that this system was primarily employed once or twice daily for the activation or deactivation of features. These observations further suggest that the operation of technical systems by museum employees is infrequent and that there appears to be no stipulation for the employees to possess specific technical knowledge for their roles at the museum.

5.1.2 NASA-TLX

Following the completion of the technology assessment questionnaire and a brief demonstration of the HoloLens device, participants engaged in tasks designed for the NASA-TLX assessment. After completing all the tasks, participants were directed to rank the significance of the NASA-TLX workload aspects in order of their contribution to the overall workload. The weighting of the workload aspects was carried out via a pairwise comparison, where participants were asked to identify the workload they believed exerted a more significant impact on the overall demand. The decision to utilize a single pairwise comparison across all tasks was made due to the similarity in the nature of the tasks. This process involved comparing all the workloads against each other and incrementing the weight of a workload by 1 each time it was selected over another. Given that there are six workloads, this results in a total of 15 comparisons. For instance, if a participant considers the mental workload to be more significant than all the other workloads, the mental workload would receive a final weight of 5, as it was selected over other options five times. The respective weights for each participant are illustrated in Figure 5.2, while the average weights for all participants on each dimension are displayed in Figure 5.3. As observable from the figure, there is substantial variability among the participants with regard to the perceived importance of each workload category. Notably, while some participants deemed the performance workload as the most critical, others regarded it as the least significant, with the exception of the universally low-rated physical demand. This contrast could be attributed to varying interpretations of the individual workload categories.

In the subsequent subsections for the different tasks, all presented figures display the adjusted rating of the participant's responses for each specific workload. This entails adjusting the data according to each participant's weight on each workload, followed by calculating the average by comparing all the answers. More specifically, this is calculated using the following formulas where R_i represents the raw rating of the i-th participant and W_i is the weight of the specific workload for the i-th participant. A_i represents the adjusted rating.

$$A_i = R_i * W_i$$

Owing to the implemented weighting system, the spectrum of plausible numerical outcomes extends from 0 to 500. Yet, the maximum value observed across all tasks has been identified as 180, prompting the adjustment of the graphical representation to cease at a limit of 200. Furthermore, it should be noted that the precise numerical quantities depicted within these graphs are mainly intended for comparative purposes within the context of this thesis. As discussed in section Summarised results, these quantities are subjected to a process of

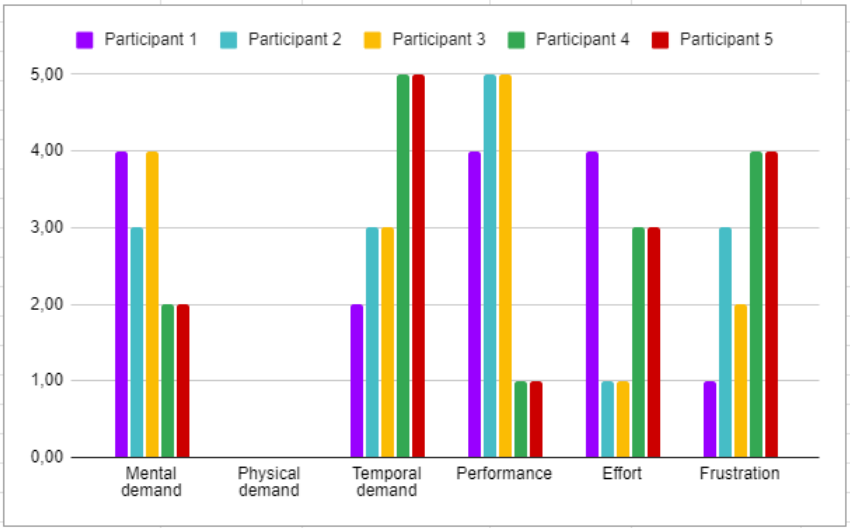


Figure 5.2: Weights for each participant

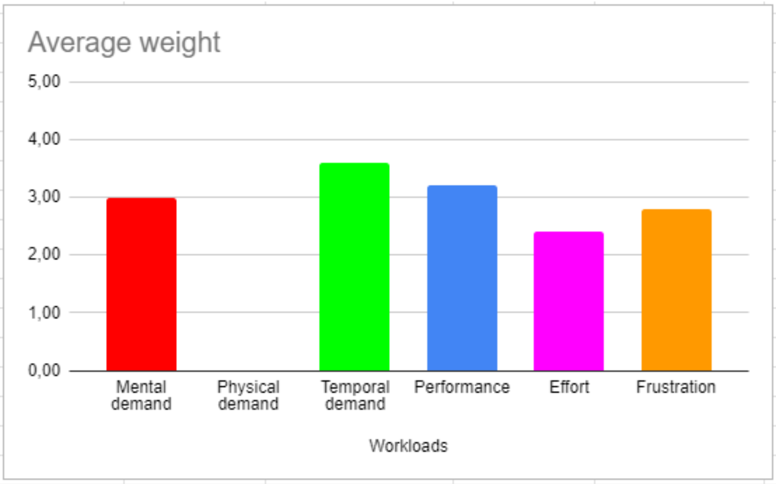


Figure 5.3: Average weights

normalization and compared to findings from external studies for a more comprehensive comparative analysis.

Creating a semi-realistic environment

Considering the experimental nature of the functionality of the MuseumXR website and the user tests, we tried to create a semi-realistic environment and scenario for each task. for both rounds of the user evaluation, a room was set up with a table and a computer, where the MuseumXR website was accessible, to simulate the reception desk at a museum. While one of the researchers acted as a visitor in the museum. The participant was placed at the desk and had access to the MuseumXR website while the researcher acting as the visitor of the museum stood on the floor in front of the desk. The scenario was then explained and the participant was thereafter encouraged to act, using the website, based on the scenario given.

As only the first scenario is going to happen at the reception desk in a real situation, the environment created was not perfect, but due to constraints of the MuseumXR availability only being on the computer at the time of the user tests, as well as the time available for conducting the tests, all of the scenarios was done like this.

Task 1

In the first scenario, we simulated a situation where two visitors wanted to experience the new MR exhibition and the availability of two HMDs needed to be assessed. As part of simulating a possible scenario that could happen in the museum, it was ensured that one HMD was turned off, while the other was on, with both devices having a full battery. The task for the museum employee in this scenario would be to check the power and battery status of the HMDs via the website. The majority of participants found this task to be relatively undemanding as illustrated in figure 5.4, which is likely attributable to the immediate display of HMD statuses since the initial screen when opening MuseumXR is the HMD overview. However, the think-aloud protocol used during the evaluation provided valuable insight into possible confusion points. For instance, two participants expressed uncertainty regarding whether the HMD with a "0%" power status was entirely depleted or simply turned off. Additionally, the card display of the offline HMD was clickable on the MuseumXR website. This was not an intended feature and caused some confusion for the participants.

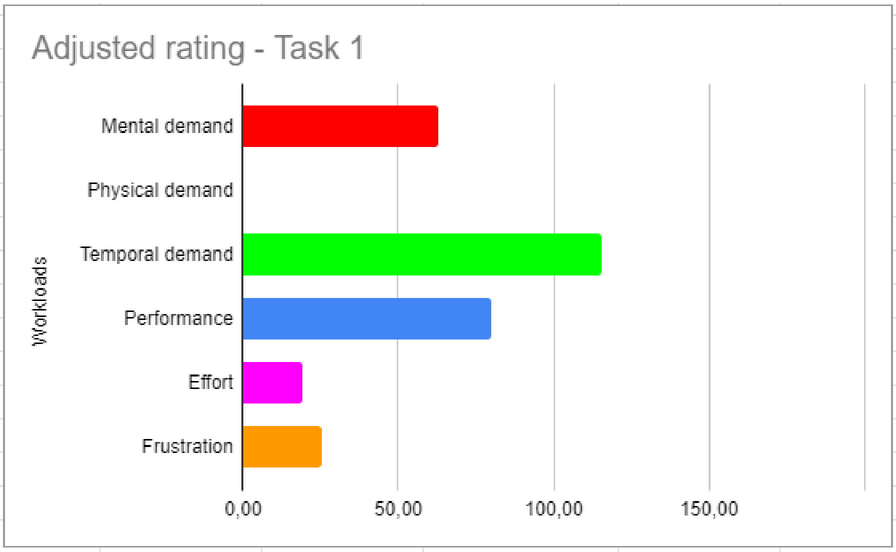


Figure 5.4: Average results from Task 1

Task 2

In the second task, participants were presented with a scenario in which a visitor had altered the scale and position of an object, making it difficult to restore it to its original state and size. The task's objective was to encourage participants to examine the HMD's live stream to fully comprehend the issue and subsequently reset the position of that element in the application. The participants' approaches to this task varied; some chose to reset the exhibition immediately, while others navigated within the application before viewing the live stream and then resetting the exhibition. One participant attempted to reset the application on the HMD itself, which would have been effective in a single-user session but was not the optimal solution in a multiplayer context, where resetting the application would not typically resolve the issue.

As depicted in Figure 5.5, the second task proved more demanding than the first. This is likely due to the less straightforward nature of the task and the requirement to navigate within the application. During this task, it was also discovered that the terminology used, specifically "application", led to confusion among participants. They did not necessarily associate the term with the page that could modify the exhibition's behaviour. Further discussion revealed that multiple participants struggled a bit to understand the different roles of the application page and the HMD page. Two participants explicitly expressed the belief that all changes, whether related to the exhibition or the HMD, would be possible from the HMD overview. One of them said during the task "Ok, so since the user is not seeing the expected behaviour I find it logical to click on their HMD and see what option lies there.".

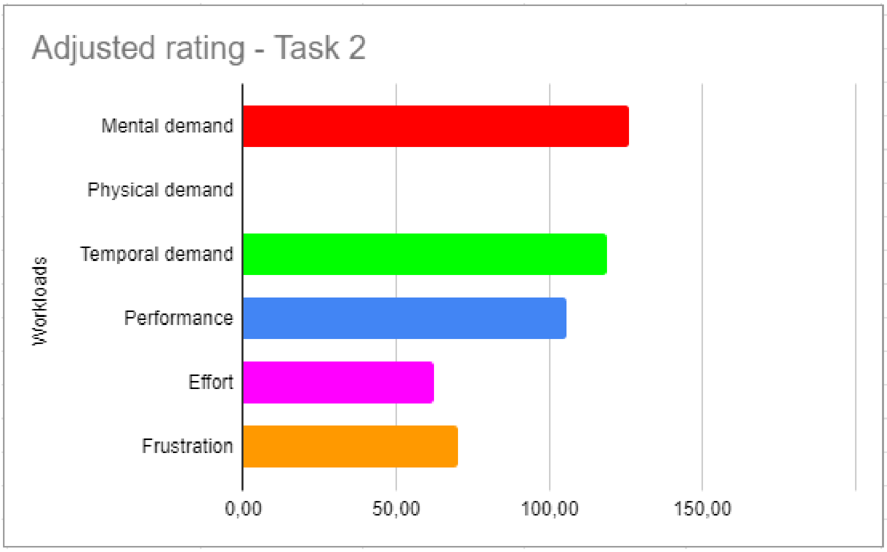


Figure 5.5: Average results from Task 2

Task 3

In the third task, participants faced another scenario, wherein the HMD user had accessed some settings or navigated out of the intended application. Similar to the previous task, the objective was to first identify the issue before executing any reset commands. This time, however, the focus was on restarting the application on the HMD. The approach to this task varied among participants. One participant, who had reset the Unity application in the previous task, quickly realized that this task was better suited for that action and proceeded accordingly, achieving the ideal solution. Others had to ask questions and view the live stream to determine the appropriate course of action. Since the web application's specific functionality was never explicitly stated by the researchers, it was initially a bit unclear to the participants where and what to look for. Several participants also expressed uncertainty about the consequences of restarting the Unity application, indicating a hesitation to perform certain actions in fear of doing something wrong.

As shown in Figure 5.6, this task's results indicate that it was more demanding than the other two tasks. Temporal demand was particularly high, which some participants elaborated on. One participant stated, "Since the visitor mentioned that they were on some settings presumably for the HMD, I felt that I needed to fix it quickly before the user could end up accidentally changing settings that I am not aware of how to fix." Frustration levels were also notably higher for this task compared to the others, which aligns with the previously mentioned response from one of the participants as well as the following comment made by another participant during the task: "Right now, I feel a bit stressed because I don't know for sure how to fix this, even though I can see it on the live stream, I don't know what happened or how it came to this."

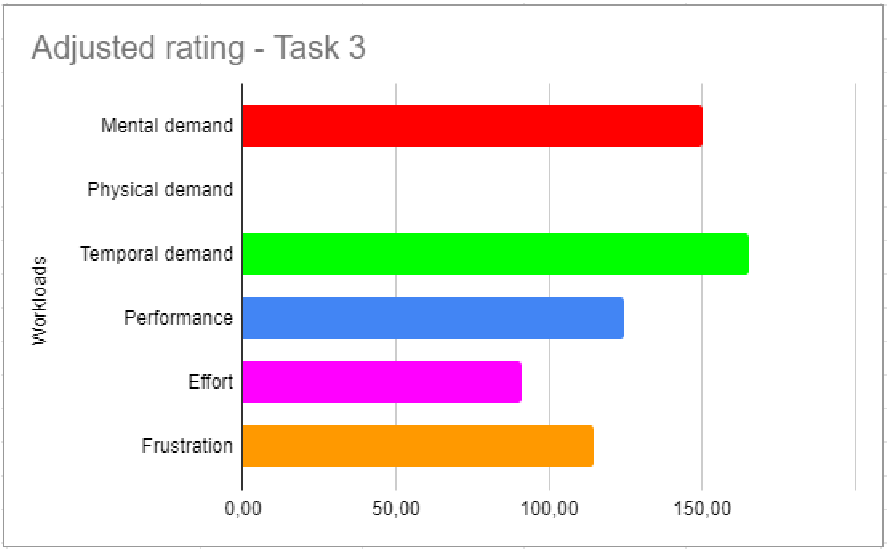


Figure 5.6: Average results from Task 3

Task 4

In the final task, the participants were instructed to deactivate the skull exhibition. As depicted in Figure 5.7, all participants deemed this task notably less demanding and easier than the preceding tasks. Temporal demand emerged as the most significant factor, rating 50. While the scores for temporal demand, performance, and effort were similar, the weight assigned to temporal demand resulted in a higher rating. Participants mostly attributed the ease of this task to its straightforward nature and their prior knowledge from previous tasks about the location of the reset button for the exhibition.

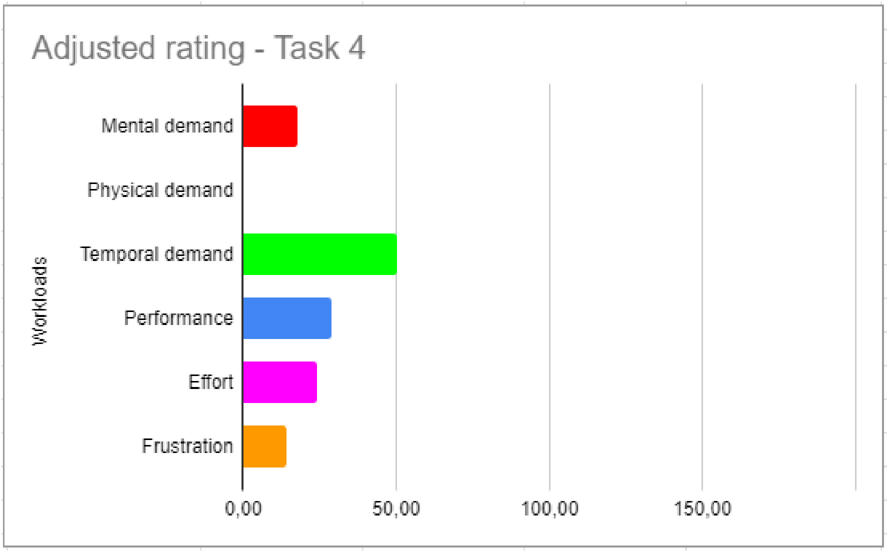
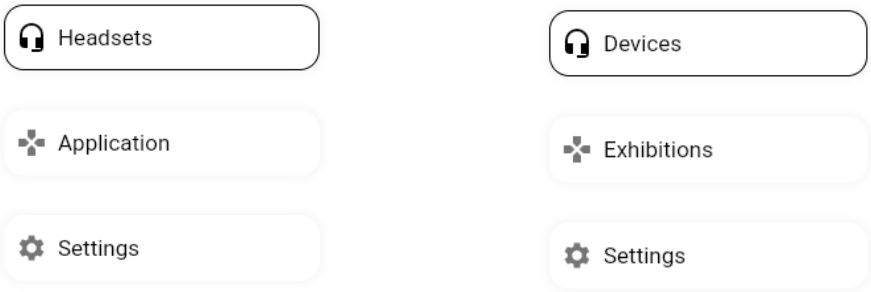


Figure 5.7: Average results from Task 4

5.2 Improvements before the second round of evaluations

On the basis of the flaws discovered during the first evaluation, multiple improvements were made. Ideally, more improvements would have been done, but due to the limited time before our next evaluation, some had to be dropped. As two of the participants mentioned that it was unclear what ”Application” referred to, illustrated in Figure 5.8, changes were made to the wording, which included changing ”Application” to ”Exhibitions” to make it more clear what the button redirects you to. Changing the displayed information on battery levels was also made, making sure to display ”off” instead of ”0%” when the HMD was powered off to avoid any confusion related to if it was off and out of battery, or just off. These adjustments are illustrated in Figure 5.9 and align well with the heuristic discussed in chapter 4, ”Match between system and the real world” which states ”The system should speak the user’s language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order [87].”.



(a) Before the changes (b) After the changes

Figure 5.8: Language changes between the rounds

(a) Before the changes (b) After the changes

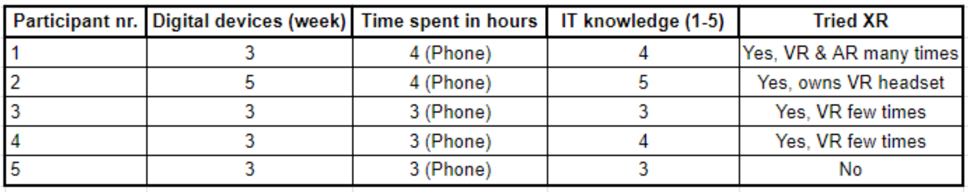
Figure 5.9: Changes made to HMD information

5.3 Second round of evaluations

The second round of evaluations involved a new group of five participants, mirroring the procedural approach employed in the initial round. However, the demographic in this round shifted to encompass students unaffiliated with the museum, a divergence from the ideal scenario. Communication hurdles impeded the ability to secure a second group of employees from the University Museum of Bergen to participate in the user tests. Despite this divergence, it merits highlighting that the majority of the museum's employees that would use MuseumXR in a real scenario are students belonging to different institutes. As four out of the five participants in the first group were students working part-time at the museum, the drawback of using a group outside of the museum was deemed minor. The results received from the first test group were very similar to each other where they first explained what kind of system they used at the museum as well as who regularly used them. The question within the technology assessment questionnaire pertaining specifically to the museum was therefore unlikely to yield incremental insights if posed to an extended sample of museum personnel.

5.3.1 Technology assessment questionnaire

The result from the technology assessment questionnaire in the second group are displayed in Figure 5.10. As depicted by the figure, the results are relatively similar to the first group.



Participant nr.	Digital devices (week)	Time spent in hours	IT knowledge (1-5)	Tried XR
1	3	4 (Phone)	4	Yes, VR & AR many times
2	5	4 (Phone)	5	Yes, owns VR headset
3	3	3 (Phone)	3	Yes, VR few times
4	3	3 (Phone)	4	Yes, VR few times
5	3	3 (Phone)	3	No

Figure 5.10: Simplified results from the technology assessment questionnaire

5.3.2 NASA-TLX

The outcome of the pairwise comparison is presented in figure 5.11 and 5.12. Similar to the first group, the second group demonstrated a wide range of responses.

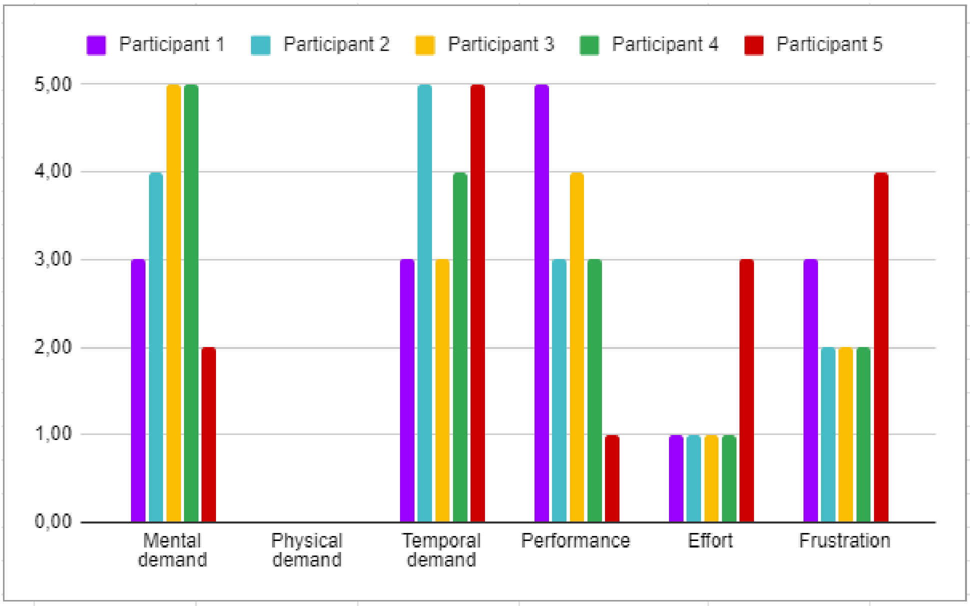


Figure 5.11: Weights for each participant

Results for the second group

The outcomes obtained from the NASA-TLX for the second group were similar to those of the first group for certain tasks. The first task had a difference primarily in the performance rating, which saw a reduction from 87 to 64. Additionally, a minor shift was observed in the workload dynamics between Effort and Frustration. The results are depicted in Figure 5.13. Throughout the execution of this task, all participants resolved the presented challenge in an easy manner, providing negligible to no significant insights during the think-aloud process.

In relation to the second and third tasks, a considerable decrease in effort levels was noted. The participants did not provide specific commentary to clarify this

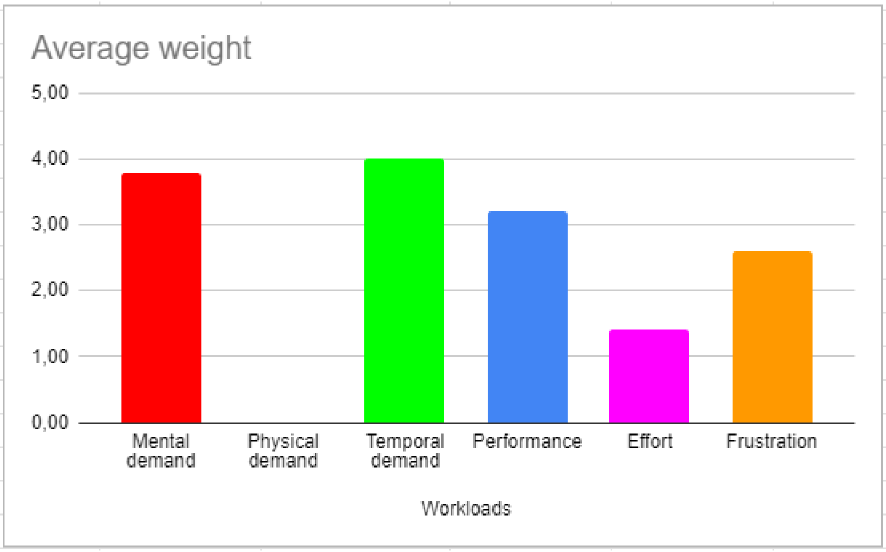


Figure 5.12: Average weights

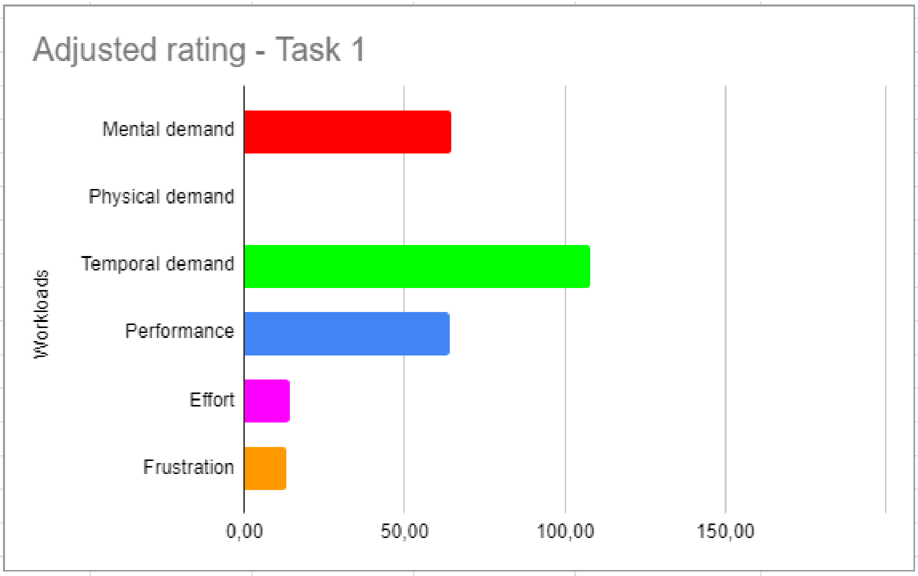
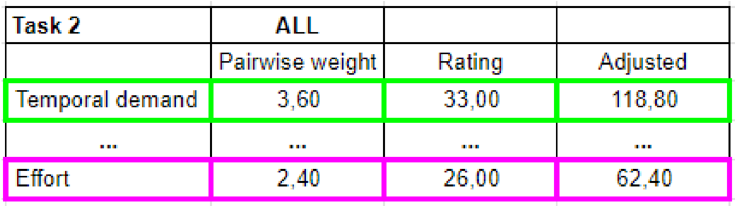


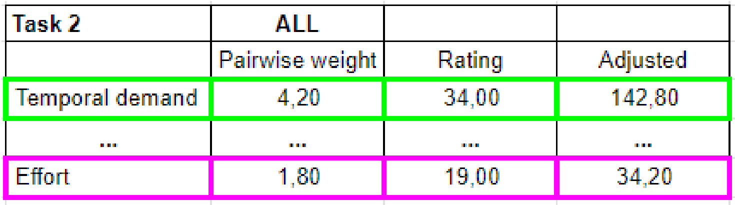
Figure 5.13: Average results from task 1

variation. A comparison of weight distribution between the first and second group shows that the second group assigned the effort workload aspect 0.6 points less than the first, partially accounting for the difference in the adjusted ratings of this aspect. Nonetheless, a large proportion of the difference arises from the variation in ratings. The initial rating on effort in Group 1 was 26 for task 2 and 38 for task 3, whereas Group 2 rated it 19 and 24, respectively. This depiction can be observed in Figure 5.14.



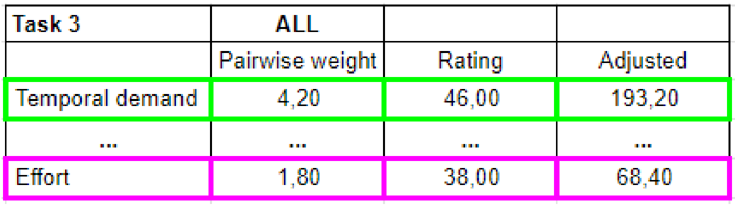
Task 2	**ALL**		
	Pairwise weight	Rating	Adjusted
Temporal demand	3,60	33,00	118,80
...	...	...	...
Effort	2,40	26,00	62,40

(a) Task 2, Group 1



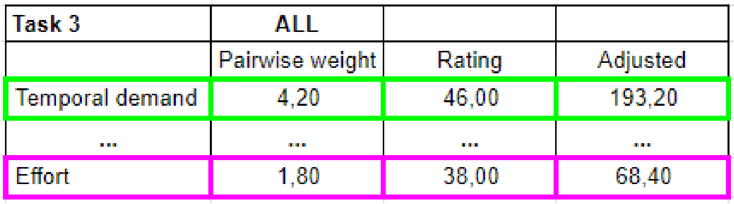
Task 2	**ALL**		
	Pairwise weight	Rating	Adjusted
Temporal demand	4,20	34,00	142,80
...	...	...	...
Effort	1,80	19,00	34,20

(b) Task 2, Group 2



Task 3	**ALL**		
	Pairwise weight	Rating	Adjusted
Temporal demand	4,20	46,00	193,20
...	...	...	...
Effort	1,80	38,00	68,40

(c) Task 3, Group 1



Task 3	**ALL**		
	Pairwise weight	Rating	Adjusted
Temporal demand	4,20	46,00	193,20
...	...	...	...
Effort	1,80	38,00	68,40

(d) Task 3, Group 2

Figure 5.14: Comparison of task 1 and 2

A weight difference of 0.6 was also observed in Temporal Demand, but the ratings between the groups were nearly identical. Group 1 assigned a rating of 33 on Temporal Demand for task 2 and 46 for task 3, whereas Group 2 gave it 34 and 43, respectively. Thus, it can be concluded that the weight difference is solely responsible for the differences in adjusted ratings of Temporal Demand between the two groups. The compared numbers are shown in Figure 5.14, whilst the graphical representation of the rating results for the second group is depicted in Figure 5.15 and 5.16.

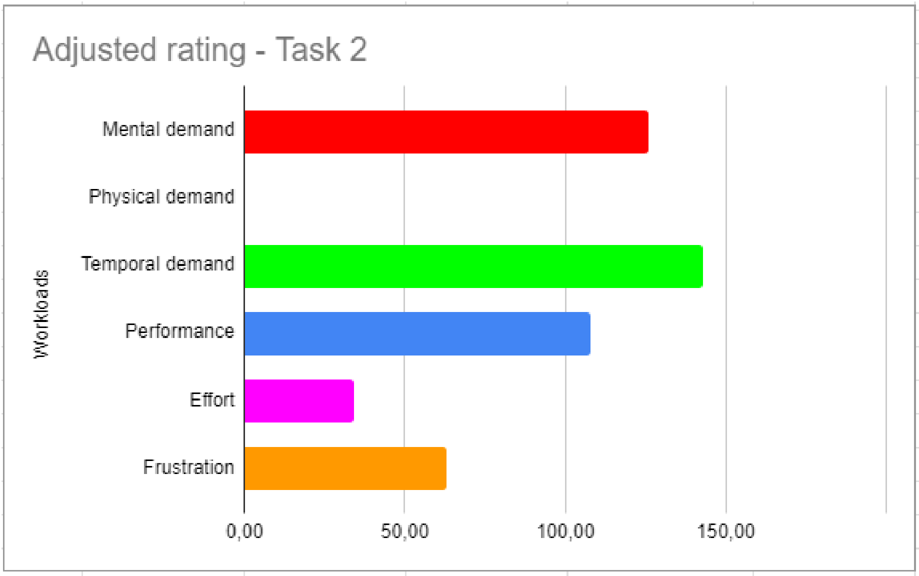


Figure 5.15: Average results from task 2

During the participants' think-aloud process, several points emerged during these two tasks. One was that sideways scrolling between the different pages of the navigation menu was possible, but resulted in errors. This was not an anticipated feature by the developers but was included in a third-party Flutter package that was used to create the navigation menu, natively supporting sideways scrolling. It was also suggested that despite the presence of feedback for the users for most actions performed within the website, some of the feedback was not clear enough and could have been slightly enhanced.

Regarding the final task, similar to the first group, all participants completed it without encountering any difficulties. Many participants stated that this task was considerably straightforward, owing to their acquired familiarity with the system from the preceding tasks. The results are illustrated in Figure 5.17.

5.4 Summarised results

The MuseumXR website solution underwent testing by a total of 10 individuals, segmented into two groups: One consisting of one curator at the University Museum of Bergen and four students with part-time employment at the museum, and the other consisting of five students unaffiliated with the museum.

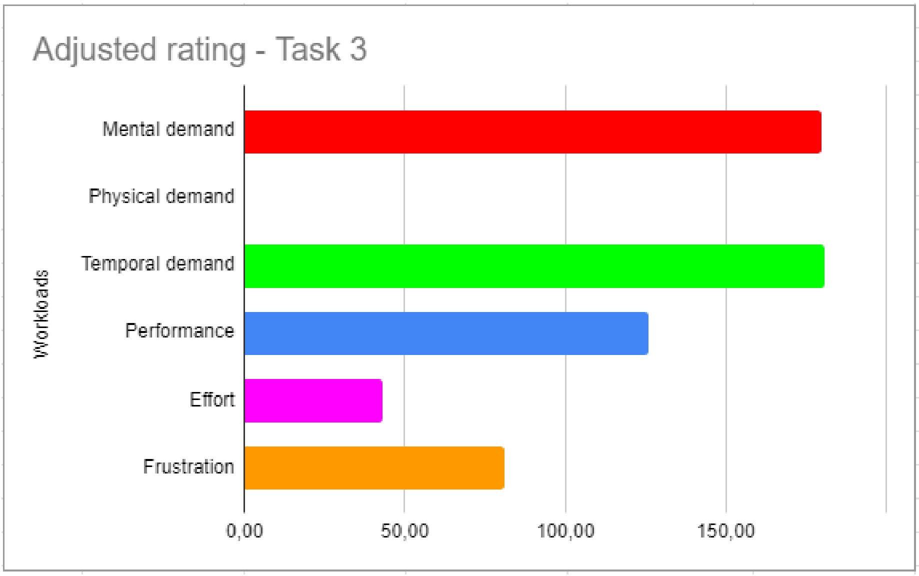


Figure 5.16: Average results from task 3

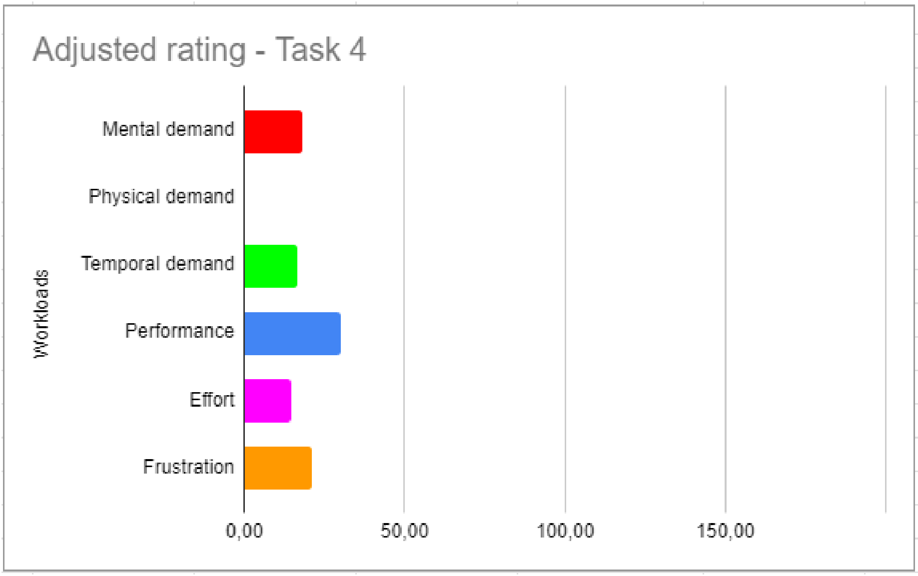


Figure 5.17: Average results from task 4

Summarizing the findings, we can refer to Figure 5.18 to see the overall workload for each task. The scores for the overall workload are presented as the sum of all the ratings divided by 15 to normalize the weighted score since the sum of all the weights always adds up to 15. The following formula is used, where d is dimension and OW overall workload :

$$OW_i = \frac{\sum_{d=1}^{m}(W_i, d * R_i, d)}{15}$$

Despite the variation in weights, the final outcome, as illustrated by Figure 5.18, demonstrates notable similarity across both groups and too little information to draw any conclusion. However, if we look at Figure 5.19 which displays all the different workloads with their average score rating across all tasks, we can see the noticeable difference in mental demand, effort and frustration. While it is difficult to conclude why these aspects are different, the changes discussed in this chapter including improvements in language usage and concise information could be a contributing factor as to why effort and frustration are lower, especially since these problems were explicitly mentioned by the first group as frustrating.

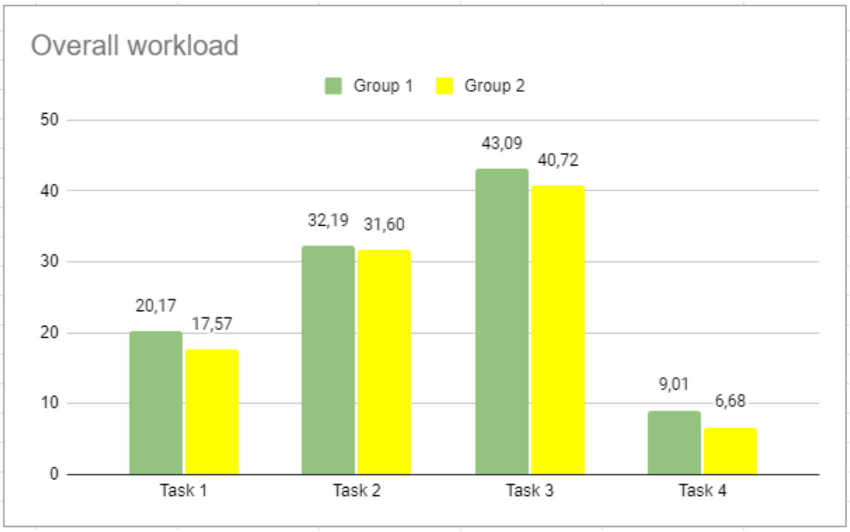


Figure 5.18: Overall workload results from both groups

The official guidelines for NASA-TLX [21] do not provide specific instructions on interpreting the results in terms of positive or negative outcomes based on the numerical results. The results of a weighted NASA-TLX evaluation display the relative contribution of each dimension to the overall workload. Because NASA-TLX is used in a variety of sectors with vastly different tasks, comparing the numerical results against other studies can therefore give an inaccurate interpretation of the results. However, a comprehensive meta-analysis conducted in 2015 by Grier, which scrutinized over 1000 NASA-TLX scores, sought to enhance the interpretability of these scores [105]. The study's findings emphasize the importance of the task type in determining the positive or negative aspects of a score. Correspondingly, the paper offers a table showcasing the Cumulative Frequency Distributions based on Task Type. A selection of examples from this table, demonstrating the range of outcomes, is presented in Figure 5.20. Of all the task types in the study, "Computer Activities" is the most correlated to this paper, and as depicted by the figure, it has a range between 7,46 and 78.

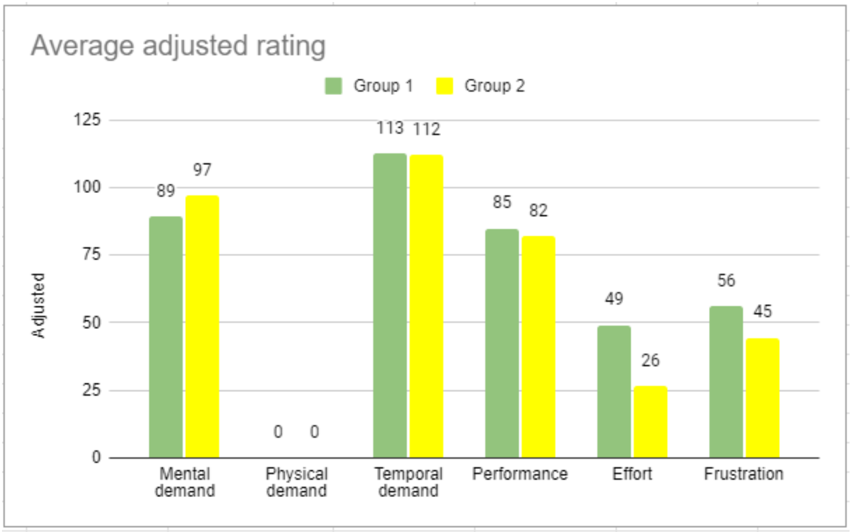


Figure 5.19: Average adjusted rating on all workloads for both groups

Cumulative Frequency Distributions of NASA-TLX					
Task type	Min	25%	50%	75%	Max
Computer Activities	7.46	20.99	54.00	60.00	78.00
Classification	8.00	30.15	46.00	51.20	84.30
Physical Activites	40.83	50.98	62.00	71.83	75.19

Figure 5.20: Cumulative Frequency Distributions of NASA-TLX [105]

The 50% mark is at 54, and all tasks were well within this threshold as well, indicating that the tasks have an overall workload better than 50% of the scores in the study within the task group ”Computer Activities”.

5.5 Interview

Following the completion of the NASA-TLX evaluation, the groups were subjected to a series of questions, conducted in an interview-like format, with the objective of delving deeper into their perspectives on MuseumXR and their experiences with the NASA-TLX evaluation. During the open-ended questions regarding their experience, the majority of participants conveyed their perception of the technology as intriguing and beneficial for museum applications. However, their reflections were predominantly concentrated on the general usage of XR in museums, with less emphasis on MuseumXR. Several participants offered innovative ideas and perspectives on potential XR applications.

While participants were provided with an overview of the necessary steps for task completion without the utilization of MuseumXR, they expressed difficulties in envisaging an alternative approach without prior experience in resolving the presented problems without the tool. Nevertheless, each participant who addressed this issue concurred on the necessity for a straightforward yet effective method of XR administration for its successful implementation.

Another aspect emphasised during the interview was the application of guided tours. Given the University Museum of Bergen's current lack of such tours, this feature has not been extensively addressed in this thesis. Nonetheless, as posited by the participant discussing this topic, guided tours might represent one of the most advantageous applications of MuseumXR. This approach could facilitate a highly customizable and dynamic experience, wherein a guide maintains control over various parameters of the digital exhibitions at all times.

A participant further suggested that MuseumXR could assist museums in addressing accessibility challenges. The possibility of adjustable parameters, which could be adjusted from group to group, would enable the content to be tailored to specific needs, such as accommodating the viewing angle for individuals using wheelchairs.

The platform's simplicity was emphasized by two participants, noting how this aspect facilitated a comprehensive understanding of the functionalities, making it accessible to individuals lacking specific IT skills or experience. This was further articulated by one participant, drawing a comparison with the technology currently employed at the museum, where reliance on external IT professionals was common in instances of troubleshooting beyond basic restarts. Such dependency was largely attributed to the limited technical proficiency among the museum personnel and a certain apprehension of causing further issues, prompting a general reluctance to attempt any solutions.

Despite being introduced as an MVP with limited functionalities compared to a finished product, the simplicity of the website was appreciated. It was indicated, given adequate initial guidance, that most museum employees could effectively operate the system. This sentiment was echoed by several participants, who felt that the test session itself served as a good introduction to the system, enabling them to navigate it without encountering immediate difficulties. These observations align with the NASA-TLX results for task 4 in the user tests, where the overall workload was considerably reduced mainly due to increased familiarity with the system, developed over the course of preceding tasks.

Striking a balance between a system that is user-friendly enough for the majority to use, while retaining sufficient functionality to offer value, is a challenging task. However, participant feedback suggests that the website was not overly complicated for routine users.

As for the feedback on the testing procedure, the overall impression was positive, with some participants recognizing the thorough and professional approach applied in the test scenarios. However, a few participants struggled with fully understanding the distinct aspects of the NASA-TLX, resulting in considerable time dedicated to its introduction and explanation. The distinction between the 'frustration' and 'mental demand' aspects, in particular, was difficult for some to delineate clearly.

Chapter 6

Discussion

In this chapter, a further discussion of the results presented in the previous chapter is undertaken. Additionally, attention is given to the evaluation methodology, the selection of tasks for evaluation, and the identification of limitations.

6.1 Museum

The objective of this research project was to develop a platform to simplify the implementation and operation of XR technology within the museum context. Domain experts from the University of Bergen Museum provided valuable insights into the current state of technology utilisation across a spectrum of Norwegian museums, revealing a pronounced deficiency in resources and tools dedicated to the administration of emerging technology. A significant imbalance has been observed between resources allocated to the implementation of new technology and those dedicated to its subsequent management. It was also expressed that museums often exhibit a tendency to prioritize visitor experience, inadvertently neglecting the needs of the employees responsible for the technology's administration. This, coupled with the limited resources typically allocated to technology within museums [13], culminates in a less than ideal foundation for technology integration in museum environments. While the solution advanced in this study may not directly address the core issue of resource distribution within museums, it attempts to show museums how administration tools can be implemented without the need to reorganize or invest in designated personnel. Developing a solution that is effective in these regards also potentially encourages museums to allocate more resources towards technology advancements.

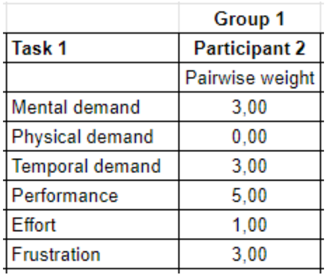
6.2 Design

During the development phase, several heuristics and principles from NN/g [83] were adhered to. These heuristics and principles were utilized as guiding frameworks throughout the development process. Despite the developers' adherence to the proposed heuristics during the creation process, the user evaluation indicated that further improvements could still be made to better align with these principles. One such instance involved the language employed within MuseumXR. Efforts were made to utilize understandable language, yet feedback pertaining to the terms "application" and "headsets" in the navigation menu suggested that these attempts were not entirely successful.

6.3 Evaluation methodology

6.3.1 Weighting in NASA-TLX

With respect to the weighting applied by all participants, the principle behind using weights aims to garner feedback more tailored to the participant's experience. Nonetheless, some results suggest a lack of coherence concerning the choice of which weight has a more significant impact. As an illustration, a logical discrepancy was noted in Participant 2's weights displayed in Figure 6.1, resulting in equal weight scores between some of the aspects. This could be attributed to various factors, however, given the 15 different comparisons involved, it is plausible that the weighting procedure may not have received sufficient attention. Nevertheless, assuming participants gave this process due consideration, it is unlikely that the outcomes would have been materially altered to an extent of significance.



	Group 1
Task 1	**Participant 2**
	Pairwise weight
Mental demand	3,00
Physical demand	0,00
Temporal demand	3,00
Performance	5,00
Effort	1,00
Frustration	3,00

Figure 6.1: Weights from Participant 2 in Group 1

6.3.2 Rounds of evaluations

Given that only two evaluation rounds of the MuseumXR website were conducted with NASA-TLX, it is not straightforward to extract concrete conclusions from these findings. It would have been ideal to have more time between the evaluation rounds, thereby affording additional opportunities for improvements based on the feedback obtained. More importantly, increasing the number of testing rounds would have potentially provided more robust and reliable data.

However, it's important to note that even with these limitations, the results garnered do provide an initial indication. They offer valuable insight into how the workload is perceived when users interact with MuseumXR. Though not exhaustive, these findings serve as a useful starting point for understanding user experiences, pointing towards areas of potential improvement and refinement in future iterations of the tool.

Think-aloud

As the participants were encouraged to perform the think-aloud method during the user tests, this might have affected the workload scores of the NASA-TLX evaluation.

Firstly, the added cognitive demand of verbalizing thoughts and actions can increase the perceived mental workload. This might have led to higher mental demand scores on the NASA-TLX scale. The need to articulate thoughts can distract from the task at hand, leading to a subjective feeling of increased workload.

Secondly, the think-aloud protocol might also increase temporal demand – another dimension of the NASA-TLX. Participants might feel that they need to complete tasks more slowly to fully articulate their thought process, which could increase the perceived time pressure.

Additionally, the nature of the think-aloud process could lead to increased effort and frustration scores. If users struggle to articulate their thoughts, or if they become aware of their struggles through the act of verbalizing them, this could contribute to a sense of greater effort and frustration.

Lastly, the think-aloud protocol might affect performance, the self-evaluation of how successful participants think they are at accomplishing the tasks. If participants made errors that they only realized upon verbalizing their thought processes, they may perceive their performance to be worse.

In light of these potential impacts on the NASA-TLX scores, it was crucial to consider the trade-offs when deciding to use the think-aloud protocol in combination with the NASA-TLX. Despite the potential for inflating workload scores, there were key benefits to this combination.

The first benefit was the complementary nature of these methods in terms of data richness. The NASA-TLX primarily provided quantitative scores across six workload dimensions but offered little qualitative insight into the reasons behind these scores. By contrast, the think-aloud method offered qualitative data that illuminated the user's thought process during task execution. Participants that verbalized their thoughts shed light on why certain dimensions of the NASA-TLX were rated highly. For instance, in the third task of the user tests, several workload aspect scores, including frustration and temporal demand, were observed to be significantly higher than in other tasks. The causes of these increases were inferred from the verbal expressions provided by the participants.

Lastly, this combination of methods was valuable in terms of providing actionable feedback for UX and UI improvements. By using the NASA-TLX, we can quantify areas of high perceived workload, but without the context, it can be challenging to address these issues effectively. The think-aloud method offered this context by providing insights into why the users were struggling. This helped the designers pinpoint specific areas for improvement.

6.4 Task choice

Regarding the selection of tasks for the NASA-TLX evaluation, the focus was to incorporate activities that represented a comprehensive range of the web application's functionality. The nature of the tasks was carefully designed to simulate realistic scenarios, ensuring a comprehensive measure of all workload dimensions encompassed by the NASA-TLX framework. For instance, in Task 2, the objective was to reset an exhibition. During this task, participants were introduced to a scenario wherein one of the evaluators posed as a museum visitor experiencing issues with resizing the digital exhibition. However, the problem was not explicitly stated; participants were only informed that the visitor was having difficulty comprehending the state of the digital exhibition. If tasks were made overly explicit, such as "Reset the skull exhibition", they would not accurately replicate real-life scenarios.

Task 3 adopted a similar approach to Task 2, while Task 4 exhibited significantly different ratings compared to the others. As briefly discussed in Chapter 5, this disparity could be attributed to the simplistic nature of Task 4, which was more straightforward than the others. The task required participants to disable a digital exhibition, and unlike the previous tasks, the solution was not left to the participant's discretion. Moreover, the disable button's proximity to the reset button meant that participants, having remembered its location from previous tasks, could execute this task with relative ease.

While these tasks do not encompass all possible scenarios, they effectively measure the core functionality of the MuseumXR website while at the same time mimicking possible situations occurring in the museum.

Further deliberation on the implications of task design regarding NASA-TLX scores reveals that the disparity in task complexity and the level of guidance provided could have significantly swayed the workload scores. Task 4's straightforward nature possibly led to diminished overall workload scores, thereby making this task seem less demanding relative to the others. These observations stimulate a broader discussion on how task design influences the user's perceived workload and how this, in turn, might shape the overall system evaluation. On the other hand, the variation in task difficulty also ensures a balanced representation of the workload associated with the workload as a whole and not only difficult tasks.

From this standpoint, the task selection process emerges as a critical component in NASA-TLX evaluations. A balance must be achieved between formulating tasks that reflect real-world scenarios and those that explicitly test certain features or functionalities. Future evaluations with a more functionality-enriched system could potentially profit from a more diverse task selection, covering a wider array of use cases, to ensure a comprehensive system assessment.

Moreover, the sequence in which tasks were presented merits attention. Participants might have grown increasingly familiar with the system and more proficient at task execution as they advanced through the tasks, thereby potentially influencing the workload scores. Future studies should consider randomizing task orders to mitigate this.

Lastly, individual differences among participants could have affected their perceived workload. Different users may exhibit varying levels of familiarity with similar systems, diverse learning rates, or unique interpretations of the task instructions. A questionnaire was conducted initially, revealing a moderate level of familiarity with IT systems. However, the incorporation of recognition and consideration of these individual differences could prove to be a valuable extension to this study. Furthermore, such an approach may contribute to the refinement of the NASA-TLX method for future evaluations.

6.5 Number of participants

In research conducted by Landauer and Nielsen [106], it was demonstrated that the number of usability issues identified in a usability test involving 'n' users can be represented by the formula:

$$N(1 - (1 - L)^n)$$

In this equation, 'N' refers to the complete set of usability issues inherent in the design, while 'L' signifies the fraction of usability issues that are detected during the testing of a single user. On average, across numerous projects evaluated by NN/g [106], the typical value of 'L' was found to be approximately 31%. Plotting this curve for L equal to 31% yielded the subsequent result:

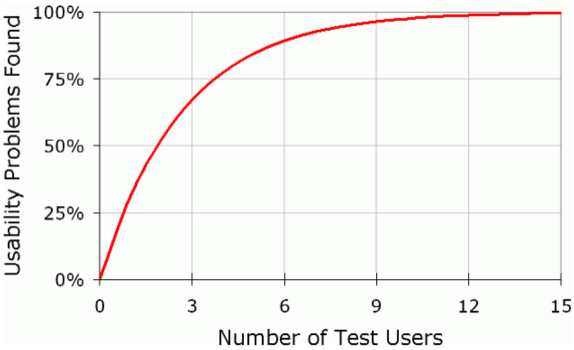


Figure 6.2: Usability Problems Found [106]

As indicated by the graph, approximately 85% of usability issues are detected with just five participants. Consequently, Landauer and Nielsen advocate for conducting multiple rounds of evaluations with five participants each, rather than fewer rounds with larger groups. Guided by this advice, five individuals were chosen for each evaluation round in this study. However, considering NASA-TLX's characteristics as a subjective, multidimensional assessment tool, it could have been advantageous to incorporate a larger number of participants to gather a more extensive range of perspectives. Consequently, the graph depicted in Figure 6.2 might not wholly represent the characteristics of NASA-TLX.

An interesting inference can be drawn from the results, as portrayed in Figure 6.3, highlighting a remarkable correlation between the two groups. This correspondence may subtly imply that the number of participants engaged in this study was sufficient to secure meaningful results.

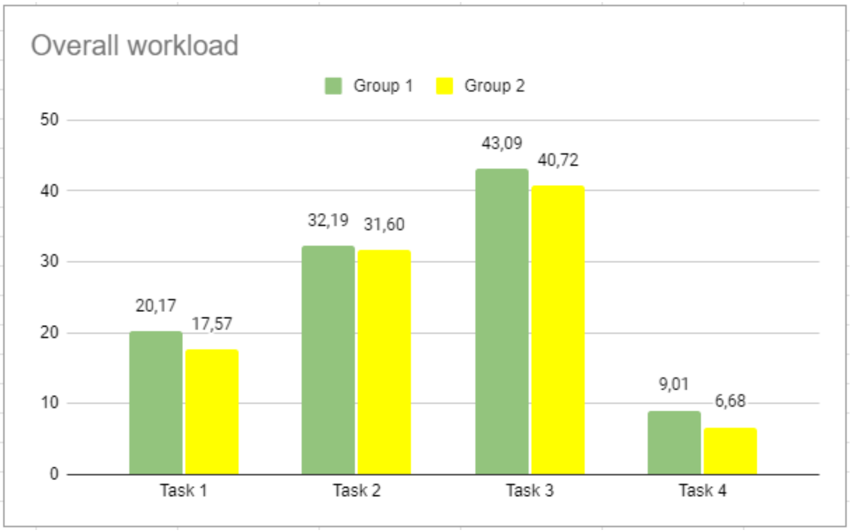


Figure 6.3: Overall workload results from both groups

6.6 Limitations

6.6.1 Experimental environment

The decision to conduct user tests within an experimental environment was driven by the need to simulate potential challenges or scenarios that participants may encounter. By artificially creating these situations, the aim was to elicit feedback on the system's performance and identify any issues that may not have naturally arisen in an actual working environment within a reasonable time frame. By subsequently analyzing the qualitative data gathered from interviews, we were able to compensate for some of the limitations in replicating authentic working conditions. This approach allowed us to assess the system's performance and obtain valuable feedback that may not have been possible solely through quantitative measurements. Although the experimental environment may have presented certain constraints, the insights gained from the interviews provided substantial evidence of the platform's positive reception and its potential for effective implementation in a practical context.

As the user tests were performed within this experimental environment, it is important to acknowledge that the researchers acted as the museum visitors using the HMDs, which potentially shielded the test participants from other situational factors or contextual aspects related to the museum experience. Consequently, the quantitative results obtained from the analysis of the NASA-TLX questionnaire, assessing workload, may possess a degree of inaccuracy.

6.6.2 XR Content standard

Throughout the development of the administration tool, a challenge regarding the involvement of artists and their role within the context of XR in museums has been identified. The proposed solution incorporates functionality that enables the management of digital exhibitions through the website's exhibition page. However, the absence of a standardized approach or framework for implementing these exhibitions introduces issues of expense and impracticality. While beyond the immediate scope of this project, it is crucial to consider the artist's ability to align their exhibitions with the functionalities offered by the website, as the effectiveness of the administration tool relies on this compatibility.

To achieve ideal integration, further research and development of a standard or conformity mechanism is required. Although the concept was acknowledged during the project, it was determined to be beyond the project's scope. Nonetheless, it is important to recognize that the absence of a standard or conformity mechanism constitutes a limitation in achieving a more comprehensive solution.

Chapter 7

Conclusion

This thesis delineates the methodology adopted for the development of an application, the underlying design principles, and the rationale behind the decisions made during the process. The primary objective was to devise an application capable of mitigating the technical impediments encountered by museums in integrating XR into their exhibitions, thereby simplifying the administration of XR hardware and software for personnel lacking IT expertise.

The research question in this thesis was:

- How can an administration tool be developed for museums to aid in the facilitation of multiple head-mounted displays and extended reality applications?

The answer to this research question is considered answered in Chapter 4, which provides insights into the design of the application, and in Chapter 3, which outlines the developmental process. Furthermore, the evaluations' outcomes reflect the viability of this solution. It is important to note that while the proposed method offers a viable approach to designing such an application, it does not necessarily imply that it is the exclusive or optimal solution.

The application underwent evaluation by two distinct groups: one comprised of museum personnel and the other comprised of students from the University of Bergen. While a larger and more representative sample size from the museum environment might have yielded more comprehensive and definitive insights, the positive outcomes of this study indicate that the application is a feasible tool for its intended use.

Chapter 8

Future Work

8.1 Other HMDs

The market consists of other MR HMDs that could potentially integrate with this solution. The application of a modular design paradigm facilitates the implementation of functionalities for alternative HMDs. This proposition is contingent on these HMDs having API support that mirrors that of the Hololens. In the event that this condition is not met, it might necessitate the creation of a design that supports different layouts based on the HMD.

8.2 Module based design

An enhancement that would render benefits to any forthcoming solution involves augmenting the system's modularity concerning the parameters accessible via the application dashboard.

Adjusting one exhibition's content based on the visitors in terms of their age group, the purpose of the visit or just experimenting with different ways of showing an exhibition, open up a lot of customization and tailor-made content possibilities that the museum employees could administer themselves without having to involve IT personnel.

Presently, parameter integration within the application is conducted manually, leading to a marked deficiency in customization options. It is likely that different museums have varying requirements for the parameters they wish to adjust.

For instance, a structural modification of the application could be implemented to autonomously retrieve parameters from the database, thereby establishing a design that accommodates an unlimited quantity of adjustable parameters. This proposed reconfiguration could ideally coincide with the database's automated retrieval of parameters from Unity scripts. The fusion of these elements would result in the application dashboard updating reflexively in response to modifications in scripts, thereby greatly increasing the adaptability and customizability of the system. However, this must be done with care, as the simplicity of the dashboard for museum employees is a fundamental aspect of the project.

By maintaining a user-centred approach to the design and development of the MuseumXR tool, a comprehensive solution for managing and enhancing the museum experience with XR technologies could be achieved, enabling museums to leverage such technology effectively.

8.3 Future framework or SDK

In relation to the previous subsection, future research would greatly benefit from the development of a standard framework or a SDK for Unity, designed to facilitate the administration of parameters from the application dashboard. This tool would be particularly advantageous for developers and artists crafting content for museums, providing a streamlined interface for parameter management. The proposed framework or SDK could significantly reduce the current reliance on manual parameter addition within the application, resulting in a more user-friendly and efficient method of customization. Moreover, the proposed solution harmonizes with the current trend in the XR field of embracing open-source technologies such as OpenXR [104]. The creation of a standard framework or SDK would enable greater accessibility and transparency, while also fostering community-driven improvements and innovations.

Bibliography

[1] Neil MacGregor and Michael Morpurgo. “Museums Matter.” en. In: (2015), p. 3. URL: `shorturl.at/PUY16`.

[2] Dai-In Han, Timothy Jung, and Alex Gibson. “Dublin AR: Implementing Augmented Reality (AR) in Tourism.” In: *Information and Communication Technologies in Tourism 2014*. Journal Abbreviation: Information and Communication Technologies in Tourism 2014. Jan. 2014, pp. 511–523. ISBN: 978-3-319-03972-5. DOI: `10.1007/978-3-319-03973-2_37`.

[3] Anne-Cecilie Haugstvedt and John Krogstie. “Mobile augmented reality for cultural heritage: A technology acceptance study.” In: Nov. 2012, pp. 247–255. ISBN: 978-1-4673-4660-3. DOI: `10.1109/ISMAR.2012.6402563`.

[4] Federica Palumbo, Gandolfo Dominici, and Gianpaolo Basile. “Designing a mobile app for museums according to the drivers of visitor satisfaction.” In: June 2013.

[5] *The Future of Technology in Museums*. en-US. June 2020. URL: `https://www.cambridgeseven.com/about/news/the-future-of-technology-in-museums/` (visited on 09/07/2022).

[6] Wei Xu et al. “Effects of technology application on museum learning: a meta-analysis of 42 studies published between 2011 and 2021.” In: *Interactive Learning Environments* 0.0 (Sept. 2021). Publisher: Routledge _eprint: https://doi.org/10.1080/10494820.2021.1976803, pp. 1–16. ISSN: 1049-4820. DOI: `10.1080/10494820.2021.1976803`. URL: `https://doi.org/10.1080/10494820.2021.1976803` (visited on 10/16/2022).

[7] M. Claudia tom Dieck and Timothy Hyungsoo Jung. “Value of augmented reality at cultural heritage sites: A stakeholder approach.” en. In: *Journal of Destination Marketing & Management*. Special edition on Digital Destinations 6.2 (June 2017), pp. 110–117. ISSN: 2212-571X. DOI: `10.1016/j.jdmm.2017.03.002`. URL: `https://www.sciencedirect.com/science/article/pii/S2212571X16300774` (visited on 02/24/2023).

[8] Timothy Hyungsoo Jung and M. Claudia tom Dieck. “Augmented reality, virtual reality and 3D printing for the co-creation of value for the visitor experience at cultural heritage places.” In: *Journal of Place Management and Development* 10.2 (Jan. 2017). Publisher: Emerald Publishing Limited, pp. 140–151. ISSN: 1753-8335. DOI: `10.1108/JPMD-07-2016-0045`. URL: `https://doi.org/10.1108/JPMD-07-2016-0045` (visited on 02/24/2023).

[9] Eleanor Cranmer et al. “Understanding the Acceptance of Augmented Reality at an Organisational Level: The Case of Geevor Tin Mine Mu-

seum." en. In: *Information and Communication Technologies in Tourism 2016*. Ed. by Alessandro Inversini and Roland Schegg. Cham: Springer International Publishing, 2016, pp. 637–650. ISBN: 978-3-319-28230-5 978-3-319-28231-2. DOI: 10.1007/978-3-319-28231-2_46. URL: http://link.springer.com/10.1007/978-3-319-28231-2_46 (visited on 02/24/2023).

[10] Alexandra Bec et al. "Management of immersive heritage tourism experiences: A conceptual model." en. In: *Tourism Management* 72 (June 2019), pp. 117–120. ISSN: 0261-5177. DOI: 10.1016/j.tourman.2018.10.033. URL: https://www.sciencedirect.com/science/article/pii/S0261517718302681 (visited on 02/24/2023).

[11] *The Mona Lisa in virtual reality in your own home.* en. URL: https://www.louvre.fr/en/what-s-on/life-at-the-museum/the-mona-lisa-in-virtual-reality-in-your-own-home (visited on 10/15/2022).

[12] *Revivre, les animaux disparus en réalité augmentée.* en. URL: https://www.jardindesplantesdeparis.fr/fr/programme/galeries-jardins-zoo-bibliotheques/revivre-animaux-disparus-realite-augmentee-4077 (visited on 10/15/2022).

[13] *Museums, museum professionals and COVID-19: ICOM and UNESCO release their full reports.* en-US. URL: https://icom.museum/en/news/museums-museum-professionals-and-covid-19-survey-results/ (visited on 10/28/2022).

[14] Sofia Pescarin et al. "Archeovirtual 2011: An evaluation approach to virtual museums." In: *2012 18th International Conference on Virtual Systems and Multimedia*. Sept. 2012, pp. 25–32. DOI: 10.1109/VSMM.2012.6365903.

[15] Larissa Neuburger, Julia Beck, and Roman Dr. Egger. "Chapter 9 The 'Phygital' Tourist Experience: The Use of Augmented and Virtual Reality in Destination Marketing." In: Nov. 2018, pp. 183–202. ISBN: 978-1-78756-292-9. DOI: 10.1108/978-1-78756-291-220181009.

[16] Fotis Kitsios, Eleni Champipi, and Evangelos Grigoroudis. "Cultural and Creative Industries Innovation Strategies for New Service Development Using MCDA." In: Jan. 2016, pp. 69–84. ISBN: 978-3-319-33001-3. DOI: 10.1007/978-3-319-33003-7_4.

[17] Fotis Kitsios, Eleni Champipi, and Evangelos Grigoroudis. *New service development: strategy and innovation process in cultural and creative industries using MCDA*. June 2015.

[18] Annarita Sorrentino et al. "Does digital technology improve the visitor experience? A comparative study in the museum context." In: *International Journal of Tourism Policy* 10 (Jan. 2020). DOI: 10.1504/IJTP.2020.10029192.

[19] *What is mixed reality? - Mixed Reality — Microsoft Learn.* URL: https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality (visited on 05/31/2023).

[20] Alan Hevner et al. "Design Science in Information Systems Research." In: *Management Information Systems Quarterly* 28 (Mar. 2004), p. 75.

[21] *TLX @ NASA Ames - Home.* URL: https://humansystems.arc.nasa.gov/groups/TLX/ (visited on 06/01/2023).

[22] World Leaders in Research-Based User Experience. *Thinking Aloud: The #1 Usability Tool*. en. URL: `https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/` (visited on 06/01/2023).

[23] *Flutter architectural overview*. en. URL: `https://docs.flutter.dev/resources/architectural-overview` (visited on 05/01/2023).

[24] *Firebase*. en. URL: `https://docs.flutter.dev/data-and-backend/firebase` (visited on 05/01/2023).

[25] MaxWang-MS. *Mixed Reality Toolkit 3 Developer Documentation - MRTK3*. en-us. URL: `https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/` (visited on 09/21/2022).

[26] Unity Technologies. *Unity Real-Time Development Platform — 3D, 2D VR & AR Engine*. en. URL: `https://unity.com/` (visited on 09/21/2022).

[27] Jim Richardson. *How Are Museums Putting the AR into Arts and Culture*. en. Mar. 2021. URL: `https://www.museumnext.com/article/how-are-museums-putting-the-ar-into-arts-and-culture/` (visited on 05/19/2023).

[28] Miriam Saviano, Viktor Malakuczi, and Lorenzo Imbesi. “Visor-less XR in museums. A Content Management System for Immersive Installations.” In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Mar. 2023, pp. 551–556. DOI: `10.1109/VRW58643.2023.00122`.

[29] Charlie Hargood and Daniel Green. “The Authoring Tool Evaluation Problem.” en. In: *The Authoring Problem*. Ed. by Charlie Hargood et al. Series Title: Human–Computer Interaction Series. Cham: Springer International Publishing, 2022, pp. 303–320. ISBN: 978-3-031-05213-2 978-3-031-05214-9. DOI: `10.1007/978-3-031-05214-9_19`. URL: `https://link.springer.com/10.1007/978-3-031-05214-9_19` (visited on 05/19/2023).

[30] Riccardo Galdieri and Marcello Carrozzino. “Towards dedicated software tools to assist the creation of Virtual Exhibits.” en. In: *SCIRES-IT - SCIentific RESearch and Information Technology* 7.1 (Aug. 2017). Number: 1, pp. 17–28. ISSN: 2239-4303. DOI: `10.2423/i22394303v7n1p17`. URL: `http://www.sciresit.it/article/view/12650` (visited on 05/26/2023).

[31] *CHESS - The CHESS project*. URL: `https://chessexperience.eu/` (visited on 04/10/2023).

[32] Kjetil Sandvik. “Ross Parry (ed.): Museums in a Digital Age. London & New York: Routledge. 2010.” In: *MedieKultur: Journal of media and communication research* 27 (June 2011), p. 8. DOI: `10.7146/mediekultur.v27i50.5245`.

[33] Manuel Silva and Luís Teixeira. “eXtended Reality (XR) Experiences in Museums for Cultural Heritage: A Systematic Review.” en. In: *Intelligent Technologies for Interactive Entertainment*. Ed. by Zhihan Lv and Houbing Song. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Cham: Springer International Publishing, 2022, pp. 58–79. ISBN: 978-3-030-99188-3. DOI: `10.1007/978-3-030-99188-3_5`.

[34] George Margetis et al. “X-Reality Museums: Unifying the Virtual and Real World Towards Realistic Virtual Museums.” en. In: *Applied Sciences* 11.1 (Jan. 2021). Number: 1 Publisher: Multidisciplinary Digital Publish-

ing Institute, p. 338. ISSN: 2076-3417. DOI: `10.3390/app11010338`. URL: `https://www.mdpi.com/2076-3417/11/1/338` (visited on 05/19/2023).

[35] Fabrizio Banfi, Raffaella Brumana, and Chiara Stanga. "Extended reality and informative models for the architectural heritage: from scan-to-BIM process to virtual and augmented reality." en. In: *Virtual Archaeology Review* 10.21 (July 2019). Number: 21, pp. 14–30. ISSN: 1989-9947. DOI: `10.4995/var.2019.11923`. URL: `https://polipapers.upv.es/index.php/var/article/view/11923` (visited on 05/19/2023).

[36] *IEEE Xplore - Conference Table of Contents*. URL: `https://ieeexplore.ieee.org/xpl/conhome/10108528/proceeding` (visited on 05/19/2023).

[37] *Norsk Skole Virtuell Virkelighet*. no-no. URL: `https://www.classvr.com/no/` (visited on 05/26/2023).

[38] *Westhaven School (Student Age 4-18)*. en-gb. URL: `https://www.classvr.com/case-studies/westhaven-school-student-age-4-18/` (visited on 05/31/2023).

[39] *Manage AR and VR Devices - VR Device Management — ArborXR*. en-US. URL: `https://arborxr.com/` (visited on 05/31/2023).

[40] *ManageXR — XR Device Management For Oculus, Pico, HTC Vive*. URL: `https://www.managexr.com/` (visited on 05/31/2023).

[41] *Unreal Engine — The most powerful real-time 3D creation tool*. en-US. URL: `https://www.unrealengine.com/en-US` (visited on 01/25/2023).

[42] Marcus Toftedahl Blogger September 30 and 2019. *Which are the most commonly used Game Engines?* en. Section: production. Sept. 2019. URL: `https://www.gamedeveloper.com/production/which-are-the-most-commonly-used-game-engines-` (visited on 01/25/2023).

[43] *Learn to create AR and VR apps with Unity*. en. URL: `https://blog.unity.com/manufacturing/learn-to-create-ar-and-vr-apps-with-unity` (visited on 11/22/2022).

[44] *This company dominates the virtual reality business, and it's not named Oculus*. en. URL: `https://fortune.com/2015/03/19/unity-virtual-reality/` (visited on 11/22/2022).

[45] qianw211. *Choosing your engine - Mixed Reality*. en-us. June 2022. URL: `https://learn.microsoft.com/en-us/windows/mixed-reality/develop/choosing-an-engine` (visited on 01/25/2023).

[46] *Auggie Awards Winners - most recognized AR VR Industry Awards*. URL: `https://www.awexr.com/usa-2021/auggie-winners` (visited on 01/25/2023).

[47] *Most used Engines*. en. URL: `https://itch.io/game-development/engines/most-projects` (visited on 01/25/2023).

[48] *Top 40 XR/AR Use Cases / Applications with Examples in 2023*. en-US. URL: `https://research.aimultiple.com/ar-use-cases/` (visited on 01/25/2023).

[49] qianw211. *What is mixed reality? - Mixed Reality*. en-us. URL: `https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality` (visited on 10/15/2022).

[50] Stella Sylaiou et al. "Leveraging Mixed Reality Technologies to Enhance Museum Visitor Experiences." In: Aug. 2018. DOI: `10.1109/IS.2018.8710530`.

[51] Nikolche Vasilevski and James Birt. "Analysing construction student experiences of mobile mixed reality enhanced learning in virtual and aug-

mented reality environments.” en. In: *Research in Learning Technology* 28 (Jan. 2020). ISSN: 2156-7077. DOI: `10.25304/rlt.v28.2329`. URL: `https://journal.alt.ac.uk/index.php/rlt/article/view/2329` (visited on 01/25/2023).

[52] *Figure 1: Definition of Mixed Reality (MR) (Milgram & Colquhoun, 1999).* en. URL: `https://www.researchgate.net/figure/Definition-of-Mixed-Reality-MR-Milgram-Colquhoun-1999_fig3_235093883` (visited on 01/25/2023).

[53] Paul Milgram et al. “Augmented reality: a class of displays on the reality-virtuality continuum.” In: *Telemanipulator and Telepresence Technologies*. Vol. 2351. SPIE, Dec. 1995, pp. 282–292. DOI: `10.1117/12.197321`. URL: `https://www.spiedigitallibrary.org/conference-proceedings-of-spie/2351/0000/Augmented-reality--a-class-of-displays-on-the-reality/10.1117/12.197321.full` (visited on 11/22/2022).

[54] *HoloLens 2—Overview, Features, and Specs — Microsoft HoloLens.* en-us. URL: `https://www.microsoft.com/en-us/hololens/hardware` (visited on 01/25/2023).

[55] *Microsoft HoloLens 2 (NJX-00014 for bedrift — Atea eShop.* no-NO. URL: `https://www.atea.no/eshop/product/microsoft-hololens-2/?prodid=4831563` (visited on 01/25/2023).

[56] *React – A JavaScript library for building user interfaces.* en. URL: `https://reactjs.org/` (visited on 01/25/2023).

[57] *Angular.* URL: `https://angular.io/` (visited on 01/25/2023).

[58] *Vue.js - The Progressive JavaScript Framework — Vue.js.* URL: `https://vuejs.org/` (visited on 01/25/2023).

[59] Alan Hevner. “A Three Cycle View of Design Science Research.” In: *Scandinavian Journal of Information Systems* 19 (Jan. 2007).

[60] Shirley Gregor and Alan Hevner. “Positioning and Presenting Design Science Research for Maximum Impact.” In: *MIS Quarterly* 37 (June 2013), pp. 337–356. DOI: `10.25300/MISQ/2013/37.2.01`.

[61] *Visual Studio Code - Code Editing. Redefined.* en. URL: `https://code.visualstudio.com/` (visited on 05/01/2023).

[62] *Uizard — App, Web, & UI Design Made Easy — Powered By AI.* en. URL: `https://uizard.io/` (visited on 04/17/2023).

[63] *diagrams.net.* URL: `https://app.diagrams.net/` (visited on 05/01/2023).

[64] *Figma: the collaborative interface design tool.* en. URL: `https://www.figma.com/` (visited on 05/01/2023).

[65] *Build software better, together.* en. URL: `https://github.com` (visited on 05/01/2023).

[66] *About Projects.* en. URL: `https://ghdocs-prod.azurewebsites.net/en/issues/planning-and-tracking-with-projects/learning-about-projects/about-projects` (visited on 05/01/2023).

[67] Asana. *Manage your team’s work, projects, & tasks online • Asana.* en. URL: `https://asana.com/` (visited on 05/01/2023).

[68] *The Visual Collaboration Platform for Every Team — Miro.* en-US. URL: `https://miro.com/` (visited on 05/31/2023).

[69] Sandra Hart. “Nasa-task load index (Nasa-TLX); 20 years later.” In: vol. 50. Oct. 2006. DOI: `10.1177/154193120605000909`.

[70] Jay Shively. “Workload Measurement in Human Autonomy Teaming: How and Why ?” en. In: ().

[71] Kai Virtanen et al. "Weight watchers: NASA-TLX weights revisited." In: *Theoretical Issues in Ergonomics Science* 23.6 (Nov. 2022). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/1463922X.2021.2000667, pp. 725–748. ISSN: 1463-922X. DOI: `10.1080/1463922X.2021.2000667`. URL: `https://doi.org/10.1080/1463922X.2021.2000667` (visited on 05/11/2023).

[72] Jeff Sauro PhD. *10 Things to Know about the NASA TLX – MeasuringU*. en-US. URL: `https://measuringu.com/nasa-tlx/` (visited on 05/04/2023).

[73] Gary B. Reid and Thomas E. Nygren. "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload." en. In: *Advances in Psychology*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Human Mental Workload. North-Holland, Jan. 1988, pp. 185–218. DOI: `10.1016/S0166-4115(08)62387-0`. URL: `https://www.sciencedirect.com/science/article/pii/S0166411508623870` (visited on 05/11/2023).

[74] Pamela Tsang and VELMA VELAZQUEZ. "Diagnosticity and multidimensional subjective workload ratings." In: *Ergonomics* 39 (Apr. 1996), pp. 358–81. DOI: `10.1080/00140139608964470`.

[75] Wickens C. D. "Information processing, decision-making, and cognition." In: *Handbook of Human Factors* (1987). Publisher: John Wiley & Sons, pp. 72–107. URL: `https://cir.nii.ac.jp/crid/1572261549731782016` (visited on 05/11/2023).

[76] Victoria A. Bowers and Harry L. Snyder. "Concurrent versus Retrospective Verbal Protocol for Comparing Window Usability." en. In: *Proceedings of the Human Factors Society Annual Meeting* 34.17 (Oct. 1990), pp. 1270–1274. ISSN: 0163-5182. DOI: `10.1177/154193129003401720`. URL: `http://journals.sagepub.com/doi/10.1177/154193129003401720` (visited on 05/08/2023).

[77] Maaike van den Haak, Menno De Jong, and Peter Jan Schellens. "Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue." In: *Behaviour & Information Technology* 22.5 (Sept. 2003). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0044929031000, pp. 339–351. ISSN: 0144-929X. DOI: `10.1080/0044929031000`. URL: `https://doi.org/10.1080/0044929031000` (visited on 05/08/2023).

[78] Obead Alhadreti and Pam Mayhew. "Rethinking Thinking Aloud: A Comparison of Three Think-Aloud Protocols." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: Association for Computing Machinery, Apr. 2018, pp. 1–12. ISBN: 978-1-4503-5620-6. DOI: `10.1145/3173574.3173618`. URL: `https://dl.acm.org/doi/10.1145/3173574.3173618` (visited on 05/08/2023).

[79] Weipeng Chen et al. "Automated comprehensive evaluation approach for user interface satisfaction based on concurrent think-aloud method." en. In: *Universal Access in the Information Society* 17.3 (Aug. 2018), pp. 635–647. ISSN: 1615-5297. DOI: `10.1007/s10209-018-0610-z`. URL: `https://doi.org/10.1007/s10209-018-0610-z` (visited on 05/08/2023).

[80] Jakob Nielsen. *Usability Engineering*. en. Google-Books-ID: 95As2OF67f0C. Morgan Kaufmann, Sept. 1994. ISBN: 978-0-12-518406-9.

[81] K. Anders Ericsson and Herbert A. Simon. *Protocol Analysis: Verbal Reports as Data.* en. Apr. 1993. DOI: `10.7551/mitpress/5657.001.0001`. URL: `https://direct.mit.edu/books/book/4763/Protocol-AnalysisVerbal-Reports-as-Data` (visited on 05/08/2023).

[82] World Leaders in Research-Based User Experience. *Nielsen Norman Group: UX Training, Consulting, & Research.* en. URL: `https://www.nngroup.com/` (visited on 06/01/2023).

[83] World Leaders in Research-Based User Experience. *About Nielsen Norman Group: UX Training, Consulting, & Research.* en. URL: `https://www.nngroup.com/about/` (visited on 06/01/2023).

[84] World Leaders in Research-Based User Experience. *End of Web Design.* en. URL: `https://www.nngroup.com/articles/end-of-web-design/` (visited on 06/01/2023).

[85] World Leaders in Research-Based User Experience. *How Chunking Helps Content Processing.* en. URL: `https://www.nngroup.com/articles/chunking/` (visited on 06/01/2023).

[86] World Leaders in Research-Based User Experience. *Fitts's Law and Its Applications in UX.* en. URL: `https://www.nngroup.com/articles/fitts-law/` (visited on 06/01/2023).

[87] World Leaders in Research-Based User Experience. *10 Usability Heuristics for User Interface Design.* en. URL: `https://www.nngroup.com/articles/ten-usability-heuristics/` (visited on 04/17/2023).

[88] Jakob Nielsen and Rolf Molich. "Heuristic evaluation of user interfaces." en. In: *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90.* Seattle, Washington, United States: ACM Press, 1990, pp. 249–256. ISBN: 978-0-201-50932-8. DOI: `10.1145/97243.97281`. URL: `http://portal.acm.org/citation.cfm?doid=97243.97281` (visited on 04/24/2023).

[89] Robin Jeffries et al. "User Interface Evaluation in the Real World: A Comparison of Four Techniques." en. In: ().

[90] Ebba Hvannberg et al. "Heuristic evaluation: Comparing ways of finding and reporting usability problems." In: *Interacting with Computers* 19 (Mar. 2007), p. 225. DOI: `10.1016/j.intcom.2006.10.001`.

[91] Jan vom Brocke, Alan Hevner, and Alexander Maedche. "Introduction to Design Science Research." In: Sept. 2020, pp. 1–13. ISBN: 978-3-030-46780-7. DOI: `10.1007/978-3-030-46781-4_1`.

[92] Obead Alhadreti and Pam Mayhew. "To intervene or not to intervene: An investigation of three think-aloud protocols in usability testing." In: *Journal of Usability Studies.* Vol. 12. May 2017, pp. 111–132. URL: `https://uxpajournal.org/intervene-think-aloud-protocols-usability-testing/`.

[93] World Leaders in Research-Based User Experience. *F-Shaped Pattern For Reading Web Content (original eyetracking research).* en. URL: `https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content-discovered/` (visited on 04/17/2023).

[94] Tom Murray. "Coordinating the Complexity of Tools, Tasks, and Users: On Theory-based Approaches to Authoring Tool Usability." In: *International Journal of Artificial Intelligence in Education* 26 (Nov. 2015). DOI: `10.1007/s40593-015-0076-6`.

[95] *html — Dart Package*. en-us. URL: https://pub.dev/packages/html (visited on 05/01/2023).

[96] *Add Firebase to your Unity project — Firebase for Unity*. en. URL: https://firebase.google.com/docs/unity/setup (visited on 05/01/2023).

[97] *Firestore*. en. URL: https://firebase.google.com/docs/firestore (visited on 05/01/2023).

[98] *Understand Cloud Firestore billing — Firebase*. en. URL: https://firebase.google.com/docs/firestore/pricing (visited on 05/01/2023).

[99] *Get started with Cloud Firestore Security Rules*. en. URL: https://firebase.google.com/docs/firestore/security/get-started (visited on 04/27/2023).

[100] *Structuring Cloud Firestore Security Rules*. en. URL: https://firebase.google.com/docs/firestore/security/rules-structure (visited on 04/27/2023).

[101] Karl-Bridge-Microsoft. *Windows Device Portal overview - UWP applications*. en-us. Oct. 2022. URL: https://learn.microsoft.com/en-us/windows/uwp/debug-test-perf/device-portal (visited on 06/01/2023).

[102] *Home*. en-us. URL: https://www.magicleap.com/en-us/ (visited on 06/01/2023).

[103] *Developer Portal — Magic Leap*. URL: https://ml1-developer.magicleap.com/learn/reference/public/v0.26.0/UnityAPI/index.html (visited on 06/01/2023).

[104] *OpenXR - High-performance access to AR and VR —collectively known as XR— platforms and devices*. en. Section: API. Dec. 2016. URL: https://www.khronos.org/openxr/ (visited on 05/01/2023).

[105] Rebecca A. Grier. "How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores." en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59.1 (Sept. 2015). Publisher: SAGE Publications Inc, pp. 1727–1731. ISSN: 2169-5067. DOI: 10.1177/1541931215591373. URL: https://doi.org/10.1177/1541931215591373 (visited on 06/01/2023).

[106] World Leaders in Research-Based User Experience. *Why You Only Need to Test with 5 Users*. en. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (visited on 05/31/2023).

[../main.tex]subfiles

Appendix A

Technical assessment questionnaire

- How many digital devices do you use in a typical week?
- Which device do you use the most? Approximately how many hours per day/week do you spend using it?
- Throughout a regular weekday (including work, study, and leisure time), what digital tools/services do you utilize?
- On a scale of 1 to 5, how would you rate your IT competence?
- Are you familiar with VR/AR/MR headsets, and have you used them before? If yes, how frequently?
- Does your museum have any systems/programs/procedures related to digital technologies? If so, please specify.

Appendix B

Interview Questions

- Do you have any thoughts or reflections on the experience you have just had?
- How do you imagine it would be to use such a solution on a regular basis? Imagine that there are a couple of exhibitions at your museum that have virtual associated elements.
- Do you envision this as a product simple enough for employees without IT expertise to be able to use?
- Do you have any thoughts on whether one option would be more stressful than the other when it comes to the device for the admin portal? For example, the counter desk versus being out among the exhibits. Which device would you choose?
- Are there any additional functionalities you would like to have? If so, what would they be?

Appendix C

NASA-TLX Results

	Group 1		
Task 1	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	10,00	40,00
Physical demand	0,00	5,00	0,00
Temporal demand	2,00	45,00	90,00
Performance	4,00	35,00	140,00
Effort	4,00	10,00	40,00
Frustration	1,00	15,00	15,00
Weighted rating:			21,67

	Group 1		
Task 2	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	10,00	40,00
Physical demand	0,00	5,00	0,00
Temporal demand	2,00	15,00	30,00
Performance	4,00	20,00	80,00
Effort	4,00	10,00	40,00
Frustration	1,00	10,00	10,00
Weighted rating:			13,33

	Group 1		
Task 3	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	15,00	60,00
Physical demand	0,00	5,00	0,00
Temporal demand	2,00	15,00	30,00
Performance	4,00	60,00	240,00
Effort	4,00	20,00	80,00
Frustration	1,00	50,00	50,00
Weighted rating:			30,67

	Group 1		
Task 4	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	10,00	40,00
Physical demand	0,00	5,00	0,00
Temporal demand	2,00	0,00	0,00
Performance	4,00	15,00	60,00
Effort	4,00	0,00	0,00
Frustration	1,00	0,00	0,00
Weighted rating:			6,67

	Group 1		
Task 1	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	15,00	45,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	10,00	30,00
Performance	5,00	20,00	100,00
Effort	1,00	5,00	5,00
Frustration	3,00	5,00	15,00
Weighted rating:			13,00

	Group 1		
Task 2	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	55,00	165,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	25,00	75,00
Performance	5,00	50,00	250,00
Effort	1,00	40,00	40,00
Frustration	3,00	40,00	120,00
Weighted rating:			43,33

	Group 1		
Task 3	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	75,00	225,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	55,00	165,00
Performance	5,00	50,00	250,00
Effort	1,00	60,00	60,00
Frustration	3,00	40,00	120,00
Weighted rating:			54,67

	Group 1		
Task 4	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	0,00	0,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	15,00	45,00
Performance	5,00	10,00	50,00
Effort	1,00	25,00	25,00
Frustration	3,00	5,00	15,00
Weighted rating:			9,00

	Group 1		
Task 1	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	10,00	40,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	20,00	60,00
Performance	5,00	30,00	150,00
Effort	1,00	10,00	10,00
Frustration	2,00	5,00	10,00
Weighted rating:			18,00

	Group 1		
Task 2	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	50,00	200,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	40,00	120,00
Performance	5,00	50,00	250,00
Effort	1,00	20,00	20,00
Frustration	2,00	25,00	50,00
Weighted rating:			42,67

	Group 1		
Task 3	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	40,00	160,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	55,00	165,00
Performance	5,00	50,00	250,00
Effort	1,00	25,00	25,00
Frustration	2,00	40,00	80,00
Weighted rating:			45,33

	Group 1		
Task 4	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	5,00	20,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	5,00	15,00
Performance	1,00	10,00	10,00
Effort	1,00	0,00	0,00
Frustration	2,00	5,00	10,00
Weighted rating:			3,67

	Group 1		
Task 1	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	40,00	80,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	50,00	250,00
Performance	1,00	20,00	20,00
Effort	3,00	5,00	15,00
Frustration	4,00	10,00	40,00
Weighted rating:			27,00

	Group 1		
Task 2	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	40,00	80,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	25,00	125,00
Performance	1,00	35,00	35,00
Effort	3,00	20,00	60,00
Frustration	4,00	15,00	60,00
Weighted rating:			24,00

	Group 1		
Task 3	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	45,00	90,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	40,00	200,00
Performance	1,00	25,00	25,00
Effort	3,00	25,00	75,00
Frustration	4,00	35,00	140,00
Weighted rating:			35,33

	Group 1		
Task 4	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	5,00	10,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	10,00	50,00
Performance	1,00	15,00	15,00
Effort	3,00	10,00	30,00
Frustration	4,00	5,00	20,00
Weighted rating:			8,33

	Group 1		
Task 1	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	30	60
Physical demand	0,00	40	0
Temporal demand	5,00	35	175
Performance	1,00	20	20
Effort	3,00	10	30
Frustration	4,00	10	40
Weighted rating:			21,66666667

	Group 1		
Task 2	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	55	110
Physical demand	0,00	10	0
Temporal demand	5,00	60	300
Performance	1,00	10	10
Effort	3,00	40	120
Frustration	4,00	35	140
Weighted rating:			45,33333333

	Group 1		
Task 3	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	75	150
Physical demand	0,00	20	0
Temporal demand	5,00	65	325
Performance	1,00	10	10
Effort	3,00	60	180
Frustration	4,00	40	160
Weighted rating:			55

	Group 1		
Task 4	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	10	20
Physical demand	0,00	20	0
Temporal demand	5,00	40	200
Performance	1,00	10	10
Effort	3,00	15	45
Frustration	4,00	10	40
Weighted rating:			21

	Group 2		
Task 1	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	5,00	15,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	25,00	75,00
Performance	5,00	15,00	75,00
Effort	1,00	5,00	5,00
Frustration	3,00	5,00	15,00
Weighted rating:			12,33

	Group 2		
Task 2	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	40,00	120,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	20,00	60,00
Performance	5,00	40,00	200,00
Effort	1,00	25,00	25,00
Frustration	3,00	30,00	90,00
Weighted rating:			33,00

	Group 2		
Task 3	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	70,00	210,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	30,00	90,00
Performance	5,00	50,00	250,00
Effort	1,00	25,00	25,00
Frustration	3,00	30,00	90,00
Weighted rating:			44,33

	Group 2		
Task 4	**Participant 1**		
	Pairwise weight	Rating	Adjusted
Mental demand	3,00	10,00	30,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	0,00	0,00
Performance	5,00	10,00	50,00
Effort	1,00	10,00	10,00
Frustration	3,00	0,00	0,00
Weighted rating:			6,00

	Group 2		
Task 1	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	15,00	60,00
Physical demand	0,00	10,00	0,00
Temporal demand	5,00	10,00	50,00
Performance	3,00	5,00	15,00
Effort	1,00	5,00	5,00
Frustration	2,00	5,00	10,00
Weighted rating:			9,33

	Group 2		
Task 2	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	35,00	140,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	25,00	125,00
Performance	3,00	50,00	150,00
Effort	1,00	15,00	15,00
Frustration	2,00	15,00	30,00
Weighted rating:			30,67

	Group 2		
Task 3	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	65,00	260,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	50,00	250,00
Performance	3,00	50,00	150,00
Effort	1,00	25,00	25,00
Frustration	2,00	35,00	70,00
Weighted rating:			50,33

	Group 2		
Task 4	**Participant 2**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	0,00	0,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	10,00	50,00
Performance	3,00	5,00	15,00
Effort	1,00	5,00	5,00
Frustration	2,00	5,00	10,00
Weighted rating:			5,33

	Group 2		
Task 1	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	5,00	20,00	100,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	20,00	60,00
Performance	4,00	30,00	120,00
Effort	1,00	15,00	15,00
Frustration	2,00	5,00	10,00
Weighted rating:			20,33

	Group 2		
Task 2	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	50,00	200,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	50,00	150,00
Performance	5,00	50,00	250,00
Effort	1,00	25,00	25,00
Frustration	2,00	25,00	50,00
Weighted rating:			45,00

	Group 2		
Task 3	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	50,00	200,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	50,00	150,00
Performance	5,00	50,00	250,00
Effort	1,00	30,00	30,00
Frustration	2,00	10,00	20,00
Weighted rating:			43,33

	Group 2		
Task 4	**Participant 3**		
	Pairwise weight	Rating	Adjusted
Mental demand	4,00	5,00	20,00
Physical demand	0,00	0,00	0,00
Temporal demand	3,00	10,00	30,00
Performance	5,00	10,00	50,00
Effort	1,00	5,00	5,00
Frustration	2,00	5,00	10,00
Weighted rating:			7,67

	Group 2		
Task 1	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	5,00	35,00	175,00
Physical demand	0,00	10,00	0,00
Temporal demand	4,00	40,00	160,00
Performance	3,00	40,00	120,00
Effort	1,00	5,00	5,00
Frustration	2,00	5,00	10,00
Weighted rating:			31,33

	Group 2		
Task 2	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	40,00	80,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	30,00	150,00
Performance	1,00	30,00	30,00
Effort	3,00	10,00	30,00
Frustration	4,00	20,00	80,00
Weighted rating:			24,67

	Group 2		
Task 3	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	55,00	110,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	30,00	150,00
Performance	1,00	50,00	50,00
Effort	3,00	20,00	60,00
Frustration	4,00	20,00	80,00
Weighted rating:			30,00

	Group 2		
Task 4	**Participant 4**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	5,00	10,00
Physical demand	0,00	0,00	0,00
Temporal demand	5,00	0,00	0,00
Performance	1,00	15,00	15,00
Effort	3,00	10,00	30,00
Frustration	4,00	15,00	60,00
Weighted rating:			7,67

	Group 2		
Task 1	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	10	20
Physical demand	0,00	15	0
Temporal demand	5,00	40	200
Performance	1,00	10	10
Effort	3,00	20	60
Frustration	4,00	5	20
Weighted rating:			20,66666667

	Group 2		
Task 2	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	45	90
Physical demand	0,00	0	0
Temporal demand	5,00	45	225
Performance	1,00	10	10
Effort	3,00	20	60
Frustration	4,00	15	60
Weighted rating:			29,66666667

	Group 2		
Task 3	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	60	120
Physical demand	0,00	0	0
Temporal demand	5,00	55	275
Performance	1,00	10	10
Effort	3,00	20	60
Frustration	4,00	40	160
Weighted rating:			41,66666667

	Group 2		
Task 4	**Participant 5**		
	Pairwise weight	Rating	Adjusted
Mental demand	2,00	10	20
Physical demand	0,00	0	0
Temporal demand	5,00	0	0
Performance	1,00	10	10
Effort	3,00	10	30
Frustration	4,00	10	40
Weighted rating:			6,666666667

Appendix D

Source Code

Link to source code: https://github.com/theodornk/MuseumXR