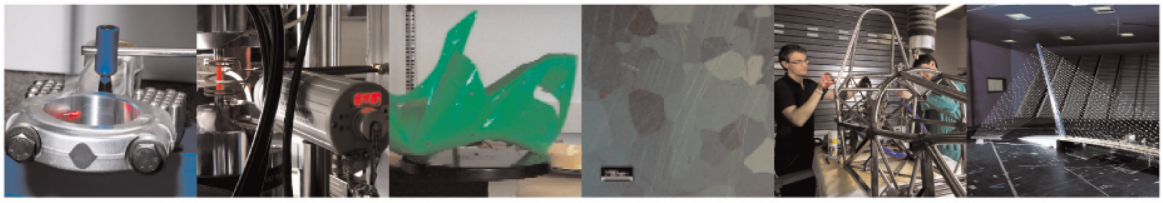

Integrating quantitative and qualitative analysis to evaluate digital applications in museums

Malik, Umair Shafqat;Micoli, Laura Loredana;Caruso, Giandomenico;Guidi, Gabriele

This is a post-peer-review, pre-copyedit version of an article published in Journal of Cultural Heritage. The final authenticated version is available online at:
https://doi.org/10.1016/j.culher.2023.06.002

Integrating Quantitative and Qualitative Analysis to Evaluate Digital Applications in Museums

Umair Shafqat Malik[a], Laura Loredana Micoli[a], Giandomenico Caruso[a], Gabriele Guidi[b]

[a] Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy, umairshafqat.malik@polimi.it

[b] Virtual World Heritage Lab, Indiana University, Bloomington (IN), USA, gabguidi@iu.edu

Abstract

Analyzing the effectiveness of a digital museum application is generally a complex process. This is due to the interaction between a digital tool capable of a limited number of narration paths and a human being whose multifaceted reactions are uniquely affected by many cultural, psychological, and emotional elements. The classical approaches in the literature can be divided into two main areas: qualitative methods, usually based on interviews at the end of the experience for assessing the cognitive effect of the digital interaction, and quantitative approaches, generally more oriented with the assessment of the subjects' emotional reactions. Among the quantitative methods, we can mention those dealing with measurements of physiological parameters during the experience, such as ECG, EEG, temperature, etc. This paper proposes a novel approach based on integrating a qualitative process using a two-stage interview, with the quantitative analysis of the user's keystrokes during his digital interaction with the application. The results show how the quantitative component can decouple the interview answers from the interviewer's possible interferences, simultaneously delivering: i) a more transparent assessment of the user's learning

experience and ii) mapping of the application's least effective sections. This provides crucial input for improving the design of digital applications for museums.

Keywords

Digital application for museums. Learning experience. Digital interaction assessment. Qualitative analysis. Quantitative analysis.

1. Introduction

In the last few decades, digital applications have grown exponentially in all areas of science, production, and daily life. The field of Cultural Heritage is no exception in this sense, both as it relates to the study, diagnostics, and restoration of art as well as its exhibition.

Museums increasingly use multimedia applications and technological devices to accomplish traditional goals, such as classifying, archiving, collecting, and exhibiting (Parry, 2013), and new practices to capture visitors' attention and enhance the global comprehension of cultural aspects.

Various technologies classified as Mixed Reality (MR) allow the visitor to explore exhibition contents moving from the real world to the virtual one with different degrees of immersivity (Milgram et al., 1995), improving the cultural learning opportunities (Bekele & Champion, 2019; Petrelli, 2019). Currently, thanks to multiple devices and technologies operable both inside and outside the museum, it is possible to integrate complementary information into the exhibition, delivered through engaging narratives.

Since the beginning of the 2000s, the concept of Intangible Cultural Heritage (ICH), made explicit by the 2003 UNESCO convention (UNESCO, 2003), has commonly become part of the debate related to the safeguarding and dissemination of culture. Subsequently, the dichotomy between tangible and intangible was overcome, by considering that any cultural asset represents an entity

that carries various meanings linked to the socio-cultural, economic, and religious contexts within which it was created and used. The semantic relationships between different objects can therefore represent important keys for defining museum collections and appropriate exhibition narratives (Carboni & de Luca, 2016; Kirshenblatt-Gimblett, 2004). The interaction between a user and the cultural material mediated by digital technology becomes particularly effective for collections composed of physical items strictly interlinked by many semantic connections also associated to non-explicit ICH components, also known as Complex Museum Collection (Micoli et al., 2020). Examples of this kind can be ancient objects related to religious rituals whose meaning is very difficult to be explained and understood if they are observed out of context (Capurro & Lupo, 2016).

In these cases, it would not be easy to communicate the meaning of both of single items and of the whole collection to the public with traditional exhibition displays. Digital approaches and tools, in several forms, represent an effective support for museum scope to reveal the embedded complexity connecting items. For example, virtual models, multimedia supports, and hyperlinks can be useful to deliver specific information not explicitly discernable by the items alone and to realize narrative contents that can more effectively clarify the articulated connections among the objects to visitors. However, a crucial point in the design of such application is represented by their actual effectiveness in term of educational content, interest for the user, ease of use, and accessibility.

In this work we present a novel mixed-methods analysis approach to evaluate visitors' interaction with a digital application. The approach involves, first, interviewing the visitors before and after the experience, which provides feedback about the amount of information gathered by each visitor through the experience according to their own assesment. Though useful, such opinion is generally

affected by the visitors' personal background, their state of mind during the visit, possible bias about the topic, their attitude to comply with the interviewer's expectations, etc.

To mitigate a user's biases, this method has been coupled with a quantitative analysis of the user's keystrokes during the digital interaction with the museum application. Each interview can be associated with data about the individual interactions between the user and the application under test. Such analysis allows to critically interpret the answers to the interviews, providing a clearer view of their interaction.

2. Research Aim

This research aims to propose a novel mixed-methods methodology to analyze the effectiveness of interactive digital applications in museums regarding users' learning experience. Such methodology can be applied to testing the outcome of interactive digital applications in museums based on narrative and 2D/3D multimedia content.

The evaluations of digital applications are typically carried out with questionaries/interviews to sample end users, but this could mislead the analysis due to the lack of objectivity in qualitative user data. The authenticity of qualitative data can be affected by multiple human factors such as multiple biases depending on users' personal background, the will to make the interviewer happy, possible embarrassment for not knowing the topic, and other typically human issues, which can significantly affect the evaluation and the outcome can be completely different from reality. The approach proposed in this article uses a mixed-methods analysis, where the interviews are connected to the corresponding sequence of keystrokes associated with each user during their navigation in the interactive museum application. The results reveal actual user behaviors that richly complement the mere interview-based qualitative analysis. We think this could significantly

contribute to defining metrics for museum application effectiveness, providing good guidance for designers.

3. State of the Art

For the past three decades, the interactive museum experience model (Dierking & Falk, 1992) has been a reference point for researchers in museum experience and visitor studies. This model simplifies our understanding of museum experience by considering it "an interaction that occurs in the personal context of the visitors, the physical context they encounter and the social context they experience" for all types of museums and all types of visitors. Similarly, other authors have argued to evaluate the museum experience as a "fit" between visitors' meaning-making and the methods used by museums. Visitors' meaning-making combines visitors' agenda, personal context, and how they create meaning for themselves, rather than passively receiving the content offered by a museum exhibition (Rounds, 1999; Silverman, 1995). Subsequently, drawing from literature on tourism, museum, and visitor studies, (Packer & Ballantyne, 2016) characterized the visitor experience as: "(i) inherently personal and subjective; (ii) responsive to the affordances of external or staged activities, settings, or events; (iii) bounded in time and space and (iv) significant to the visitor (i.e., having an impact on the visitor that makes it noticeably different from everyday life)."

Prior visitor theories and museum experience models gave rise to a new relationship between visitor needs and museum goals in contemporary museum practices. Museums are increasingly working to understand and respond to their visitors' behavior to stay "relevant" (Simon, 2016). Therefore, museum practices such as audience engagement (Mcintyre, 2014; Visser et al., 2013), participatory experience (Simon, 2010), collaborative design with visitors and interdisciplinary

professionals (Avram et al., 2019; Ciolfi et al., 2015; Vavoula & Mason, 2017), and personalization (Ardissono et al., 2012; Raptis et al., 2019) are gaining popularity in recent years.

Furthermore, the literature on such contemporary museum practices highlights the potential of digital technologies in improving the visitor experience and helping to accomplish museums' goals. Consequently, museums are increasingly adapting technological applications to involve visitors. In 2016, a New Media Consortium Horizon report highlighted the short-term, mid-term, and long-term impacts of emerging technologies on education and interpretation in museums worldwide (Freeman et al., 2016). The report reveals the importance of emerging technologies to further museums' interpretation goals and enhance their visitor experiences by developing effective digital strategies. Still, relatively few studies investigate the way visitors interact with digital museum exhibitions and its effects on their experience.

The early visitor studies demonstrated the short and long-term effects of visitors' agenda, motivations, the context of the visit, and interaction in the museum on their learning outcomes (Ellenbogen, 2003; Falk et al., 2004; Packer & Ballantyne, 2002; Stevenson, 1991). These studies adopted quantitative and qualitative methods to evaluate the learning outcomes. While quantitative data in these studies were collected through simple questionnaires and time-tracking visits, diverse qualitative methods were also introduced in these early studies. Such qualitative methods include ethnographic techniques to understand the culture of learning environments (Brewer, 2000), content analysis to identify code and categorize patterns or themes in qualitative data (Bernard & Ryan, 1998), personal meaning mapping (PMM) to measure the effects of a specific educational experience on visitors' personal, conceptual, attitudinal, and emotional understanding (Falk et al., 1998; Falk & Dierking, 2003; van Winkle & Falk, 2015) and the comparison of data collected in various contexts of interaction.

Further studies explored visitors' perceptions of specific types of interactions, the role of interactivity in visitor experience, and visitor behavior analysis for designing interactions. Drawing from past studies, the researchers employed both qualitative and quantitative methods for the evaluation of visitor experiences for different purposes, such as: the perception of digital and physical reproductions and their effects on the museum learning (Lindgren-Streicher & Reich, 2007); interface design evaluation (Hornecker, 2008); motivation and behavior study of online visitors (Fantoni et al., 2012; Skov & Ingwersen, 2014); evaluation of visitor experience at onsite exhibitions by visualizing visitor flows; preferences in the physical spaces (Strohmaier et al., 2015); usability evaluation of virtual tour applications by employing the System Usability Scale (SUS)[1] (Othman et al., 2021); multidisciplinary psychosomatic assessment for user experience design evaluation (Bonacchi et al., 2018).

Evolving from past studies, the user experience evaluation based on mixed methods, which this article will illustrate, represents a new approach to evaluating visitor interaction with digital applications in museums by connecting the users' keystrokes with the qualitative analysis of their interviews.

4. Methodology

Starting from the analysis of a specific interactive digital experience, the methodology is designed to verify the effectiveness of the narratives for digital applications in museums and to assess the learning outcomes of such applications. We propose a mixed methods approach to collect and evaluate data on visitors' interaction with interactive applications installed alongside the physical artifacts to explain their meanings. Three types of data collection are proposed: (a) data collection

[1] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

concerning visitors' interaction with the application by recording the keystroke logs; (b) data collection concerning learning outcomes generated by interviewing the visitors at the entrance and exit of the exhibition before and after their visit and (c) the collection of visitor demographic data to create visitor profiles with a questionnaire.

4.1. Data collection

Although the specific content analyzed and the exact types of interactive applications in past studies were not the same as those proposed in this study, the literature review provided useful insights into data collection and research methodologies for visitor studies. The literature suggests the necessity of collecting both quantitative and qualitative data for thoroughly evaluating visitors' experiences and learning from their interactions within a museum context. Several approaches have been employed in the past for quantitative data collection, including interaction sequence tracking, visitor time logging, user study surveying, and using questionnaires. On the other hand, the approaches used for qualitative data collection in past museum visitor studies have included PMM, structured/semi-structured/open-ended interviews, follow-up phone interviews after the visits, observations and handwritten notes by the researcher, and open conversations with the visitors. Furthermore, collecting the demographic data to create visitor profiles is considered important to analyze the collected data considering the individual context of the visitor. This can be done with either a questionnaire or a baseline interview. For an in-depth analysis and learning aspects of visitors' interaction, the literature recommends data collection before, during, and after the interaction and also not informing the visitors about the purpose of the study so as to collect unbiased opinions.

Most of recommendations concerning data collection have been covered in this study's methodology by carefully designing a protocol of semi-structured interviews, interaction logging,

open-ended and structured questionnaires. This data can then be integrated with the physical observations taken by the researcher during the visitors' interaction with the application.

The research methodology proposed in this article is structured according to the following protocol:

1. The participants are divided into two groups based on their background: (a) expert visitors are those who are already familiar with the subject and have already experienced other similar digital applications; (b) inexperienced visitors are those who had either none or very few experiences of the subject. Moreover, visitors are divided based on how they interacted with the application (i.e., sole visitors and group visitors).

2. The visitors' keystrokes are recorded while using the digital application. For each interaction, an automatic log entry can be generated to record the time and type of interaction for each visitor, who can be identified by using the recorded time stamps.

3. The previously identified visitors are interviewed before and after interacting with the digital application, as suggested by the principles of PMM (Falk & Dierking, 2003). The demographic information of the visitor (age, sex, profession, interests, etc.) is collected through structured and open-ended questionnaires at the end of the experience to assign a visitor profile to each interview.

4.2. Evaluation

Among several methods for data evaluation, the most coherent methods with the type of data collection are proposed for this study and adapted to evaluate the visitors' interaction and learning outcomes. The quantitative data collected from interaction logging and a part of the questionnaire can be analyzed through summary statistics for central tendency and visualizations. Qualitative

data collected through interviews and physical observations can be analyzed through the content analysis (Drisko & Maschi, 2016) of pre-interview and post-interview responses of the visitors.

5. Experimental study

The mixed-methods approach proposed in the previous section has been tested using as a case study a digital application about the ancient Egyptian collection of the Milan Civic Archaeological Museum, developed within the framework of the PERVIVAL project (http://www.pervival.polimi.it/). That digital application aims to link the exhibited physical objects with their inherent immaterial elements to create a synthesis that reveals the appropriate meaning of a specific object either for its physical significance or its multiple connections related to the ancient Egyptian funerary rituals.

To communicate this archaeological knowledge in the form of an engaging narrative (Roussou, 2001), an interdisciplinary team (Katifori et al., 2018) worked iteratively (Manson, 2015) on a subset of grave goods from the museum collection. The objects were selected, digitized, and used in the virtual application to create rich visual content supported by reliable historical references.

The virtual interactive application, developed with the platform Unity3D and presented through a touch-screen kiosk (Burmistrov, 2015), is based on a tree map noting the three main conditions for achieving eternal life according to the beliefs of the ancient Egyptians (Andrews, 1994; Lise, 1979; Miatello, 2018; Pinch, 2004; Redford, 2005; Scalf & Lowry, 2017; Stewart, 1995): 1) Preserve the intact body through mummification and the placing of the remains within a coffin, a sarcophagus, and an appropriate tomb; 2) Feed the deceased with real or symbolically represented food; 3) Face ordeals while the spirit moves into the afterlife with the help of rituals, amulets and coffin decorations, this being described in the papyrus of the “Book of the Dead” that was frequently included with the burial goods.

All the digital elements and pages belonging to one of the main conditions is distinguishable thanks to the graphic language adopted based on color and logos. The navigation of the app could be either guided for non-expert users, following the predefined path suggested by the authors, or free, for more expert users, with the possibility to choose a single object/piece of content to explore, creating in this way personalized paths into the digital content.

The application has been hierarchically structured into four levels corresponding to a further branching of the contents and different interaction modalities ((Micoli et al., 2020)), the resulting levels being: I) synthetic description of the main section; II) single or multiple items descriptions through different media (text, images, animations); III) animation of 3D digitization of single or multiple items, enriched with interactive tools (replay, magnifying glass, hotspot point); IV) details of specific item represented with different media such as videos or drawings.

The navigation can be done with different approaches according to the user's expertise and time available, following predefined paths or choosing the content to be explored. In duration terms, the two extreme ways to navigate correspond to the minimum and maximum required time: 3 minutes are necessary for a synthetic overview of the whole material of the three main themes at level I; twenty to twenty-five minutes are necessary to completely navigate all of the contents across all levels.

The data on visitors' experience and learning was collected based on their interaction with the PERVIVAL application installed at the Civic Archaeological Museum within the exhibition "Sotto il Cielo di Nut" (Fig 1.a).

(a) (b)

Figure 1: PERVIVAL installation at the exhibition "Sotto il Cielo di Nut" in the Civic Archeological Museum of Milan. (a) A group of visitors (research participants) interacting with the application; (b) Interview with users of the application at the entrance/exit of the exhibition.

A total of 50 visitors (the research participants) who interacted with the PERVIVAL application were interviewed. Before the interaction, the visitors were asked questions about their knowledge of the content that was being delivered by the application. The same questions were repeated after the interaction, and the answers were connected with the log information to analyze the learning aspects of the interaction. This activity provided several insights into the importance of designing the interaction based on the type of content (3D or otherwise), selection of narratives, and interrelation between the content type and narrative. The data from interviews was integrated with the interaction logging, questionnaires, and physical observations the researchers took during the visitors' interaction with the application (fig. 2).

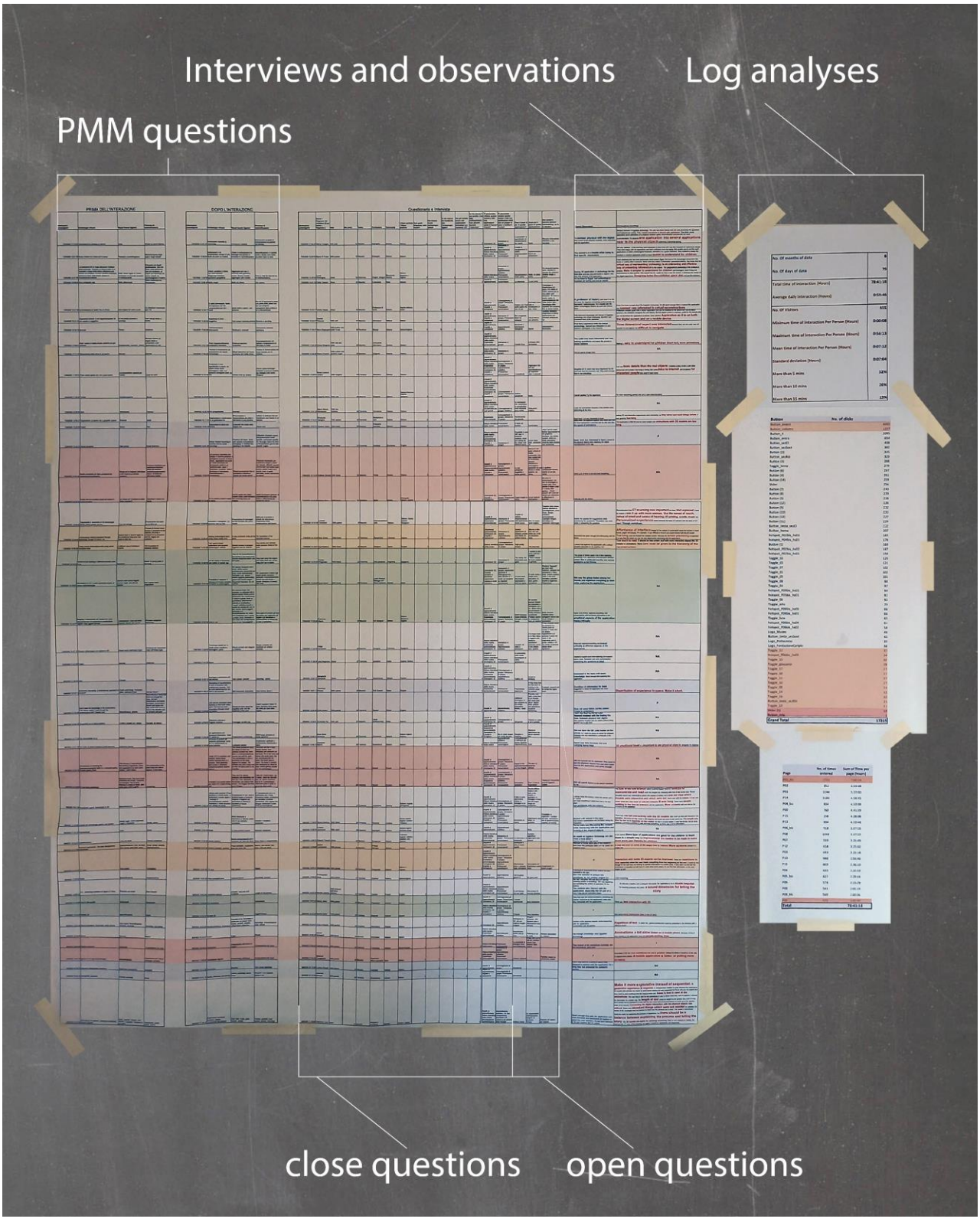

Figure 2: Different types of data integration to evaluate the user experience in the reported example.

5.1. Application Logging

To analyze the visitor interaction with the application and to verify the effectiveness of the narratives designed for the application, the log information was recorded. The log created an entry for each interaction on the screen. Each log entry included the following data:

- time and date of the interaction
- name of the page on which the interaction occurred
- type of interaction (tapping a button, moving a slider, zoom in/zoom out)

Such data were collected for 73 days over 5 months (6 working days for 2 months, only weekends for 3 months). In this period, 4,532 visitors[2] entered the exhibition hall where the PERVIVAL application is installed. The log data was then analyzed for:

- minimum, maximum, and average time of interaction per person
- most and least interacted pages
- most and least used buttons
- the visitors' path of interaction

To calculate the interaction time per person, it was necessary to identify any singular person from the available log information. Even if there was a button on each page to finish the interaction and return to the homepage, most visitors did not return to the homepage after finishing their visit. Furthermore, several passing visitors just tapped the screen without having real interaction with the application. Therefore, the log data was filtered first to remove such anomalies.

[2] The data on total number of visitors of the exhibition was provides by the civic archeological museum of Milan.

An example of how the interaction time has been calculated for a particular visitor by filtering misrepresenting entities is presented in figure 3. The time of real interaction of visitors with the application was calculated by highlighting all the log entries containing the cover page. The time of such entries was then subtracted from the time of the entry before the next interaction on the cover page. The entries in which the interaction at the cover page was repeated or very near the preceding entry (less than 5 seconds of time difference) were ignored.

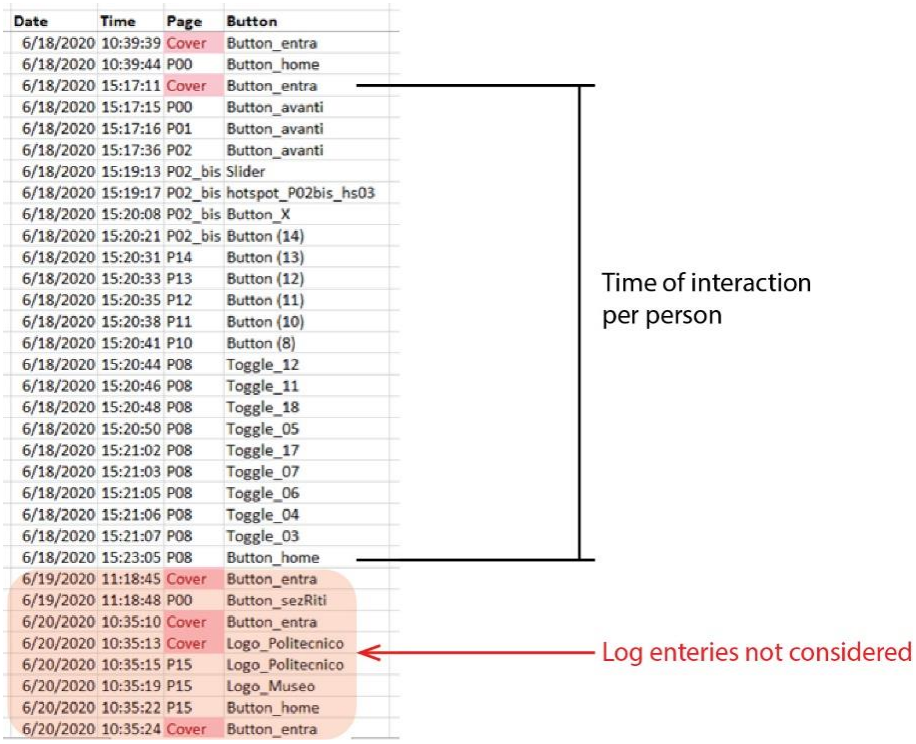

Figure 3: An example of the log entries to calculate the interaction time for each particular visitor. The highlighted area is an example of false entries.

After identifying the real interactions for each visitor who interacted with the application, it emerged that out of 4,532 visitors who visited the exhibition, a total of 639 (14%) interacted with the application, and 77.5 hours of interaction were recorded. Subsequently, the minimum maximum and average interaction time per person was calculated (Table 1). Furthermore, the

visitors were divided into four groups based on the total interaction time per person: more than 3, 5, 10, and 15 minutes, respectively. The percentage of these groups among the sample data is also reported in the following table. While most visitors experienced a synthetic overview of the application, the data clearly shows that the time expected to explore the application in detail did not correspond to the time that a visitor would actually spend on average with the application.

Table 1: Log data of the analyzed museum application

No. of days of data	73
Total time of interaction (Hours)	77:34:44
Average daily interaction (Hours)	1:03:46
No. Of visitors who interacted	639
Minimum time of interaction Per Person (Hours)	0:00:08
Maximum time of interaction Per Person (Hours)	0:56:13
Mean time of interaction Per Person (Hours)	0:07:17
Standard deviation of interaction Per Person (Hours)	0:07:07
More than 3 mins	65%
More than 5 mins	53%
More than 10 mins	26%
More than 15 mins	13%

The log data was also analyzed to understand the visitors' navigation and general interests in different parts of the application and the most common types of interaction. A Sunburst Diagram, usually used to visualize hierarchical data, can also be used to identify user navigation sequences for various applications or websites (Rodden, 2014). Such diagram is used here to visualize the interaction data of the visitors with the application (Figure 4). The visualization is based on the total number of interactions with a specific button on each page of the application within the recorded log data. The first hierarchical level represents the pages of the application where the interaction occurred, and the second hierarchical level represents different buttons on each page.

The width of each segment represents the number of interactions where wider segments are those with more interactions and vice versa.

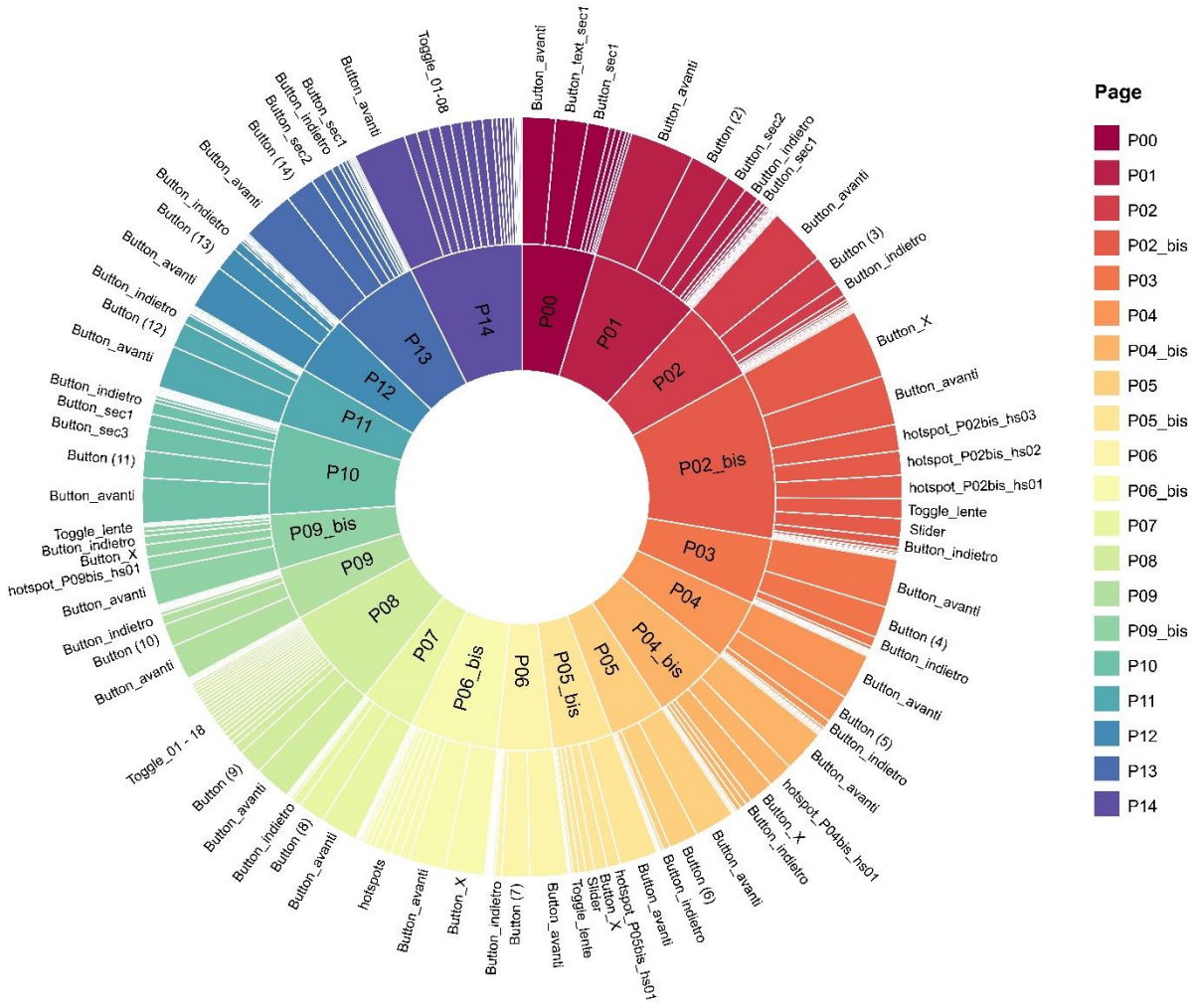

Figure 4: Sunburst diagram of visitor interaction at different pages of PERVIVAL application. The first hierarchical level represents the pages and buttons are at the second level. The width of each piece represents the number of interactions.

It can be seen from the visualization that most of the interactions occurred on the page "P02_bis", relating to the mummy. On this specific page, the visitors explored all the contents in detail which is represented by the width of several segments in the second hierarchical level representing

buttons. Overall, the most used buttons were “Button_avanti” (forward) and “Button_indietro” (back). It implies that, apart from the page “P02_bis”, most of the visitors took a synthetic overview of the whole application without exploring each page in great detail. This observation also corresponds to the average time of interaction per person calculated i.e. 3-5 minutes for most of the visitors.

To confirm this observation and to demonstrate the most common paths of interaction, the log data was analyzed to draw a cumulative overview of the Interaction Sequences (ISs) associated with each of the app users. Each recognizable IS was assigned to any identified visitor (a total of 639) in the log data. Each interaction starts at the “cover page,” and the only path from that page is to page P00 (start page). Any single user had different levels of interaction with the app, spending more or less time in front of the screen depending on their interest. For a suggestive cumulative representation of such data, only the first 31 interactions attributed to each of the 639 users were considered and plotted in the heatmap shown in figure 5.

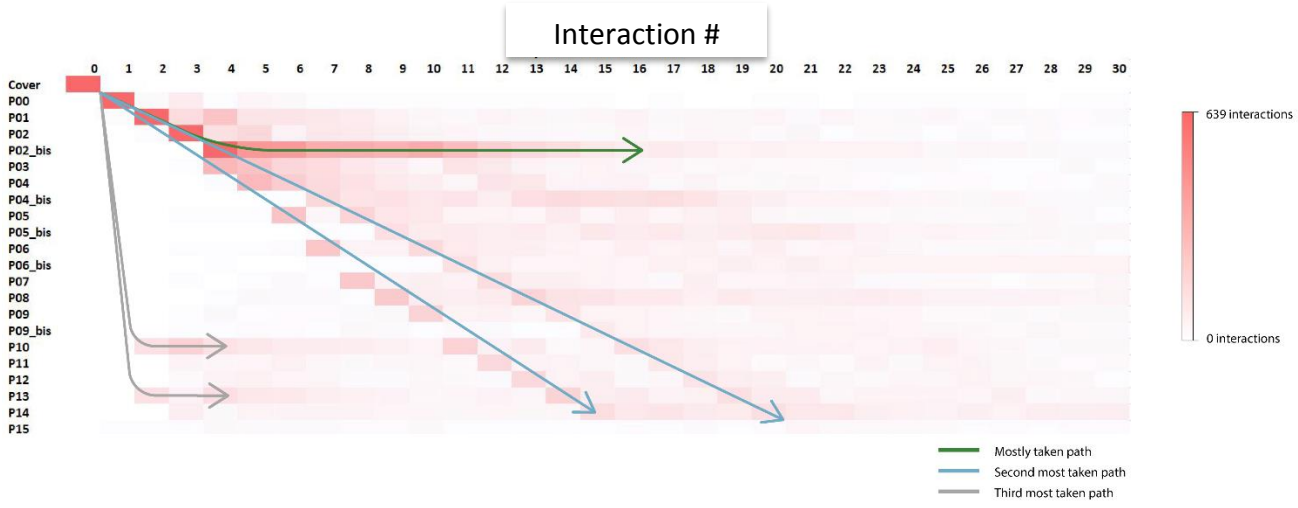

Figure 5: Heatmap of an Interaction Sequences of all visitors interacting with the museum application used for this test. The most used paths are represented by arrows.

Each interaction, indexed from 0 to 30, is represented on the x-axis in the heatmap, while y-axis represents specific pages of the application. The darkness of each cell of the map, is proportional to the number of total interactions associated with any page/interaction index combination. The fact that all users start from the cover page at step 0 and navigate to P00 at step 1 determines the two dark cells in the top left corner of that diagram. While the navigation in the app's contents is branched in different optional paths, the cumulative analysis of the recorded keystrokes associated with all the 639 recorder users allows making graphically evident the preferred paths. Such a reading has been further highlighted by the colored arrows drawn in Fig. 5. It can be seen that the most chosen path by visitors was the interactions and exploration in detail of page "P02_bis" (the path represented by green arrow). The second most taken path was that of a synthetic overview of all the pages represented by blue arrows. Some of the visitors also took an alternative path of directly reaching the 2nd and 3rd page among the "level 1" pages selection from "P00". These paths are represented by grey arrows in the heatmap. The most common paths taken by the visitors explain how most visitors interacted with the application for an average time of interaction of 3-5 minutes (Table 1).Table 1

5.2. Personal meaning mapping (PMM)

To understand the visitors' background and the manner by which they learned from the digital experience, we used the PMM approach. The specific museum application tested here was designed to create interest and understanding among the visitors about the Egyptian funerary collections through digital reconstructions and narratives. Therefore, before interacting with the application, the participants were asked to share their knowledge of three relevant keywords: *virtual archeology*, *Egyptian funerary rituals*, and *mummification process*. After the interaction, the participants were asked to re-evaluate their answers from before the interaction and decide if

they wanted to change their thoughts about these keywords. This helped to determine whether the visitor conceptions of these same three words changed after the interactive learning experience.

Interacting with the application significantly changed the meaning of the three keywords among visitors. They became more familiar with the subject topic, and in particular there was a substantial change in their perception of the word "digital" in the context of archeology. Most of the participants changed their answers after the visit, from hardly knowing anything to producing an entire explanation of a phenomenon or naming objects to explain the process. Only 15% of the participants did not significantly change their answers after the interaction. This included two types of visitors: expert visitors who were already familiar with the subject or visitors who only spent a superficial amount of time exploring the application (data verified from the log).

Virtual archeology and its applications were understood better by the visitors after interacting with the application under test. Before the interaction, most visitors related virtual archelogy to merely 3D models. After the visit, virtual archeology applications were considered a shift from traditional passive museum learning to an active form of learning. The post-visit responses of the participants revealed that experiencing the application had changed their view of virtual archaeology from simply the use of 3D digitization to instead mentioning specific qualities, such as it being: "modern," "educational," "futuristic," "virtual interaction," and "a simulation of past." The visitors defined this as an interesting way to understand, visualize, and teach the history of archeological artifacts in detail. Additionally. the importance of different types of 3D scanning and survey for documenting objects and also the reconstruction of archeological finds was mentioned several times in the interviews.

Similarly, after the interaction the understanding of material and immaterial connections improved significantly among the visitors. They were able to explain the meaning of material items and of

rituals connected to them with great details in post-interview PMM. Such learnings from the interactive experience also directly corresponded to the individual log information, gathered through a precise tracking of login/logout time connected with the entry/exit interview time. For example, two of the visitors who explained in detail the rituals outspread in the exhibition spent 15 and 23 minutes interacting with the application, respectively, double and more than triple the average interaction time.

5.3. Questionnaire and interviews

The participants were also asked to fill out a questionnaire at the end of the visit. Out of 50 visitors of the exhibition who gave their consent to participate in the research, 52% were female, and 48% were male. While 29 visitors participated in this research alone, 21 were in a group of at least two persons (friends, family, or partners). The age of participants ranged from 20 years to 62 years, with most of them less than 35 years old (62%). As the application under study was monolingual, most participants were Italian (74% of the participants). All the participants could understand the Italian language except three, who were accompanied by a translator. Therefore, the questionnaire and interviews were conducted in the Italian language.

To understand the background of the participants, they were presented with five closed questions, with a range of 1 to 5, 1 containing the least weight and 5 containing the most weight. The overview of the responses provided in Table 2 shows that most participants were frequent visitors of museums in general and archeological museums specifically. Even though there were some experts on Egyptian archeology and digital applications for archeology among the participants, most were less familiar/unfamiliar with the subject and type of application under study. Nevertheless, most of the participants felt comfortable answering the questions.

Table 2: Background of the research participants on a 5 steps scale (1 = low; 5 = high).

n=50 **52% female; 48% male**	**Min value**	**Max value**	**Mean**	**Standard deviation**
Frequency of visiting museums	2	5	3.54	0.98
Frequency of visiting archeological museum	1	5	3.42	1.13
Familiarity with Egyptian archeology	1	5	2.84	1.03
Familiarity with digital archeology	1	5	2.64	1.13
Ease at answering question	1	5	4.62	0.77

As anticipated, half of the participants in this research did not feel comfortable answering the questions verbally to the researcher. Therefore, the interview questions were also provided in the form of the second and third part of the questionnaire with 2 closed and 3 open questions, respectively, to understand the visitors' opinions and assess different aspects of interaction within the application.

The two closed questions in the second part were presented as multiple-choice checkboxes to understand the most effective aspects of their interactive experience with the application when they were alone or in a group. In these questions, the selection of multiple options was allowed. The overview of the responses in Table 3 demonstrates that while alone, the participants appreciated the learning aspects and interactivity through 3D animations. As expressed during the interviews and open-ended questions, the results of closed questions also clearly show that the experience with the application could not be easily personalized for every type of visitor and improvements in terms of interface and affordance could be made. On the other hand, when asked about the interaction with the application in a group, most participants found it useful for group engagement and fostering conversations.

Table 3: The effectiveness of different aspects of the digital application as voted by the research participants.

When alone		
	No. of participants	**% of total**
Learning aspects	30	60%
Interactivity	29	58%
Personalization	17	34%
3D animations	26	52%
Look and feel	13	26%

When in Group		
	No. of participants	**% of total**
Group engagement	26	52%
Group interaction	26	52%
Collective learning	21	42%
Fostering conversations	25	50%

The third part of the questionnaire consisted of three open questions inviting the participants to describe merits, demerits, and suggestions about their experience.

Most participants found the application synthetic, simple, and rich in educational information. In terms of visual impact, the interactive 3D graphics, animations, different layers of details, high-quality images, and informative text were valued by the visitors. The participants were enthusiastic to see the 3D reconstructions. In particular, the expert visitors verified the validity of the digital reconstructions to reconstruct different situations and render the details that can't be seen in the physical objects. The interactive nature of the experience was appreciated as the visitors could choose their own pace to personalize navigation and selection for exploring the whole collection or parts of it. For the visitors who were experiencing a digital application for archeology in a museum for the first time, it was an innovative way for museum learning, reconstructing the past, and understanding history.

On the other hand, the participants found that the pace of the animations was slower than expected. Exploration of an object in detail was only possible after the animations finished. It was not

possible to re-launch the animations after a certain point during the course of the animations. Moreover, although though the long text narratives were appreciated by the expert visitors, the inexperienced visitors found them distracting and confusing. They instead suggested adding links that could open the text information on a separate page or even connect them to the internet for further exploration. The 3D animations and interactive tools were preferred by non-expert visitors; they recommended creating additional elements of multimedia interaction, such as: audio narratives, sounds, and customized creation tools.

For adults, it was easy to navigate and understand, but all participants with children asserted that the interface should be made simpler to understand for children and teenagers. Additionally, it was less intuitive for a non-expert user to navigate through different levels of content. For example, most visitors did not realize that they could switch through different sections, therefore they were forced to follow the predefined path. They suggested improving the affordance of the tools and animated tooltips to guide them through the navigation.

Sometimes, visitors found it difficult to associate application contents with the physical objects. The most discussed reason was the screen's position with respect to the physical objects, which made it difficult for visitors to look at the objects while using the application. For this reason, it was suggested to change the immobile experience into an itinerant experience by using multiple screens positioned in proximity of each object. Another suggestion was to introduce augmented reality, which could make it truly immersive and engaging thus, passing from a passive to an active form of media.

6. Discussion and Conclusions

This article introduces a novel methodology for testing the outcome of interactive digital applications in museums based on narrative and 2D/3D multimedia content. Such a test is typically carried out with questionaries to sample end-users, but this has provoked criticism. The lack of objectivity that may arise from people under interrogation due to various biases depending on their personal background e.g.: the will to make the interviewer happy, possible embarrassment for not knowing the topic among others, can significantly affect the evaluation. It is a typical situation when the measurement process affects the measurand in such a destructive way that the outcome may significantly diverge from reality.

The evaluation of the visitor experience presented here is based on a mixed-mode methodology, grounded in quantitative and qualitative data collected from the visitors who interacted with their digital experience in the museum. The purpose is to provide feedback about the effectiveness of a specific museum application, filtering out the end-users' biases which possibly affected interviews and questionnaire answers.

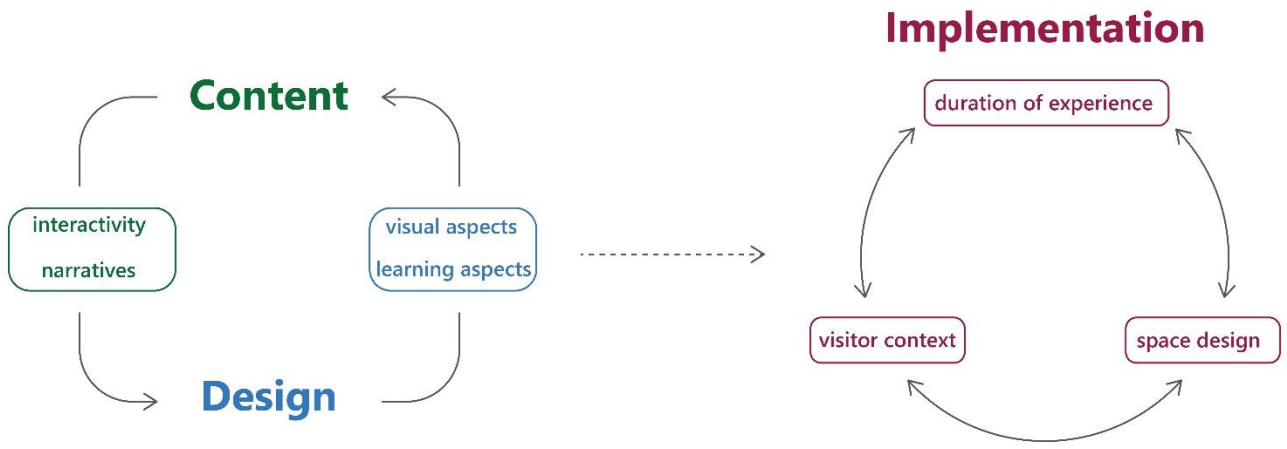

Figure 6: Evaluation of visitor experience based on visitors' interaction.

As exemplified in figure 6, the user experience was evaluated based on the analysis of visitors' navigation through interaction logging, their satisfaction with the experience, and most repeated improvements commonly suggested during the interviews.

In conclusion, the results presented in this research demonstrate how a mixed-mode analysis, where the interviews are 1:1 connected to the corresponding sequence of keystrokes associated with each user during their navigation in the interactive museum application, may reveal actual user behaviors that richly complement the simpler interview-based analysis.

The mixed-mode analysis approach used in this research appears to offer valuable insights into user behavior, which can be used to enhance the design and effectiveness of museum applications. The methodology may also be applicable to other areas of User eXperince (UX), User Interface (UI), and Human Computer Interaction (HCI) design, potentially providing a useful tool for designers to better understand user needs and behaviors. This will thus extend the applicability of this research from museums to much broader contexts such as consumer product marketing.

Defining metrics for app effectiveness is crucial for improving the overall user experience, and we think that the findings of this research could prove to be a valuable resource for designers looking to create engaging and intuitive applications.

7. References

Andrews, C. (1994). *Amulets of ancient Egypt*. British Museum Press.

Ardissono, L., Kuflik, T., & Petrelli, D. (2012). Personalization in cultural heritage: The road travelled and the one ahead. *User Modeling and User-Adapted Interaction*, *22*(1–2), 73–99. https://doi.org/10.1007/s11257-011-9104-x

Avram, G., Ciolfi, L., Spedale, S., Roberts, D., & Petrelli, D. (2019). Co-design goes large. *Interactions*, *26*(5), 59–63. https://doi.org/10.1145/3348793

Bekele, M. K., & Champion, E. (2019). A comparison of immersive realities and interaction methods: Cultural learning in virtual heritage. *Frontiers in Robotics and AI*, *6*, 91.

Bernard, H. R., & Ryan, G. (1998). Text analysis. In *Handbook of methods in cultural anthropology* (Vol. 613).

Bonacchi, A., Ferrari, A., Camilleri, L., & Chiesi, F. (2018). Multidisciplinary Psychosomatic Assessment for UX Design Evaluation. *Studies in Digital Heritage*, *2*(1), 106–120. https://doi.org/10.14434/sdh.v2i1.25652

Brewer, J. (2000). *Ethnography*. McGraw-Hill Education (UK).

Capurro, R., & Lupo, E. (2016). *Designing Multivocal Museums. Intercultural Practices at Museo Diocesano, Milano*. Politecnico di Milano.

Carboni, N., & de Luca, L. (2016). Towards a conceptual foundation for documenting tangible and intangible elements of a cultural object. *Digital Applications in Archaeology and Cultural Heritage*, *3*(4), 108–116. https://doi.org/10.1016/j.daach.2016.11.001

Ciolfi, L., Petrelli, D., Mcdermott, F., & Avram, G. (2015). Co-design to Empower Cultural Heritage Professionals as Technology Designers. In *Co-design to Empower Cultural Heritage Professionals as Technology Designers*. https://doi.org/10.1007/978-3-319-13018-7

Dierking, L. D., & Falk, J. H. (1992). Redefining the museum experience: The interactive experience model. *Visitor Studies*, *4*(1), 173–176.

Drisko, J. W., & Maschi, T. (2016). *Content analysis*. Pocket Guide to Social Work Re.

Ellenbogen, K. M. (2003). Museums in family life: An ethnographic case study. In *Learning conversations in museums* (pp. 92–112). Routledge.

Falk, J. H., & Dierking, L. D. (2003). Personal meaning mapping. *Museums and Creativity: A Study into the Role of Museums in Design Education*, 10–18.

Falk, J. H., Moussouri, T., & Coulson, D. (1998). The Effect of Visitors ' Agendas on Museum Learning. *Curator: The Museum Journal*, *41*(2), 107–120.

Falk, J. H., Scott, C., Dierking, L., Rennie, L., & Jones, M. C. (2004). Interactives and Visitor Learning. *Curator: The Museum Journal*, *47*(2), 171–198.

Fantoni, S. F., Stein, R., & Bowman, G. (2012). Exploring the relationship between visitor motivation and engagement in online museum audiences. *Museums and the Web*.

Freeman, A., Adams Becker, S., Cummins, M., McKelroy, E., Giesinger, C., & Yuhnke, B. (2016). NMC Horizon Report: 2016 Museum Edition. In *New Media Consortium*. https://doi.org/10.1111/j.2151-6952.2010.00064.x

Hornecker, E. (2008). "I don't understand it either, but it is cool"—Visitor interactions with a multi-touch table in a museum. *2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*, 113–120.

Kirshenblatt-Gimblett, B. (2004). Intangible Heritage as Metacultural Production. *Museum International*, *56*(1–2), 52–65. https://doi.org/10.1111/j.1350-0775.2004.00458.x

Lindgren-Streicher, A., & Reich, C. A. (2007). Visitor usage of digital and physical artifacts in two museum programs. *Visitor Studies*, *10*(2), 152–167.

Lise, G. (1979). *Museo archeologico: Raccolta egizia*. Electa.

Mcintyre, A. (2014). *The Spectrum of Audience Engagement: A relationship policy matrix*.

Miatello, L. (2018). Examining texts and decoration of Peftjauauiaset's coffins in Milan. *Égypte Nilotique et Méditerranéenne*, *11*, 41–133.

Micoli, L. L., Caruso, G., & Guidi, G. (2020). Design of digital interaction for complex museum collections. *Multimodal Technologies and Interaction*, *4*(2), 1–20.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies*, *2351*, 282–292.

Othman, M. K., Nogoibaeva, A., Leong, L. S., & Barawi, M. H. (2021). Usability evaluation of a virtual reality smartphone app for a living museum. *Universal Access in the Information Society*. https://doi.org/10.1007/s10209-021-00820-4

Packer, J., & Ballantyne, R. (2002). Motivational Factors and the Visitor Experience: A Comparison of Three Sites. *Curator*, *45*(3), 183–198.

Packer, J., & Ballantyne, R. (2016). Conceptualizing the Visitor Experience: A Review of Literature and Development of a Multifaceted Model. *Visitor Studies*, *19*(2), 128–143. https://doi.org/10.1080/10645578.2016.1144023

Parry, R. (2013). *Museums in a digital age*. Routledge.

Petrelli, D. (2019). Making virtual reconstructions part of the visit: An exploratory study. *Digital Applications in Archaeology and Cultural Heritage*, *15*, e00123. https://doi.org/10.1016/j.daach.2019.e00123

Pinch, G. (2004). *Egyptian Mythology: A Guide to the Gods, Goddesses, and Traditions of ancient Egypt*. Oxford University Press.

Raptis, G. E., Fidas, C., Katsini, C., & Avouris, N. (2019). A cognition-centered personalization framework for cultural-heritage content. *User Modeling and User-Adapted Interaction*. https://doi.org/10.1007/s11257-019-09226-7

Redford, D. B. (2005). *The Oxford encyclopedia of ancient Egypt*. Oxford University Press.

Rodden, K. (2014). Applying a sunburst visualization to summarize user navigation sequences. *IEEE Computer Graphics and Applications*, *34*(5), 36–40. https://doi.org/10.1109/MCG.2014.63

Rounds, J. (1999). Meaning making: A new paradigm for museum exhibits. *Exhibitionist*, *18*(2), 5–8.

Scalf, F., & Lowry, K. B. (2017). *Book of the dead: Becoming God in ancient Egypt*. The Oriental Institute of the University of Chicago.

Silverman, L. H. (1995). Visitor Meaning-Making in Museums for a New Age. *Curator: The Museum Journal*, *38*(3), 161–170. https://doi.org/10.1111/j.2151-6952.1995.tb01052.x

Simon, N. (2010). *The participatory museum*. Museum 2.0.

Simon, N. (2016). *The art of relevance*. Museum 2.0.

Skov, M., & Ingwersen, P. (2014). Library & Information Science Research Museum Web search behavior of special interest visitors. *Library and Information Science Research*, *36*(2), 91–98. https://doi.org/10.1016/j.lisr.2013.11.004

Stevenson, J. (1991). The long-term impact of interactive exhibits. *International Journal of Science Education*, *13*(5), 521–531. https://doi.org/10.1080/0950069910130503

Stewart, H. M. (1995). *Egyptian shabtis*. Shire.

Strohmaier, R., Sprung, G., Nischelwitzer, A., & Schadenbauer, S. (2015). Using visitor-flow visualization to improve visitor experience in museums and exhibitions. *MW2015: Museums and the Web 2015, September*, 1–6.

UNESCO. (2003). *Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO. https://ich.unesco.org/en/convention

van Winkle, C. M., & Falk, J. H. (2015). Personal meaning mapping at festivals: A useful tool for a challenging context. *Event Management*, *19*(1), 143–150.

Vavoula, G., & Mason, M. (2017). Digital exhibition design: Boundary crossing, Intermediary Design Deliverables and processes of consent. *Museum Management and Curatorship*, *32*(3), 251–271. https://doi.org/10.1080/09647775.2017.1282323

Visser, J., Richardson, J., Lawrence, R., Fox, L., Dixon, M., Mckenzie, B., Margarone, C., Allen-Greil, D., Baruzzi, D., Appel, H. H., Melchionna, L., Kelly, L., Joseph, M., Zietzke, O., Desaive, P.-Y., Guerra, R., Dodge, R., Bausson, S., Boddeker, S., … Butler, T. (2013). *Digital Engagement in Culture, Heritage and the Arts*. 67.