



Supporting Users of Archives for Open Dialogue in Digital Humanities

Pia Borlund, Nils Pharo and Ying-Hsang Liu

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 10, 2022

Supporting Users of Archives for Open Dialogue in Digital Humanities

Pia Borlund, Nils Pharo, and Ying-Hsang Liu (Oslo Metropolitan University, Norway)
{piabor, nilsp, [yinghsan](mailto:yinghsan@oslomet.no)}@oslomet.no}

Introduction

This paper reports work in progress and presents preliminary empirical findings from a European collaborative research project in digital humanities with references to user access to digital cultural heritage archives in the context of post-colonial history. The overall objective of the project is to study colonial audio-visual heritage across the archives involved and open a dialogue between the archives and a variety of users. Participation at this conference is a first-class instance to discuss the results with the research community, receive feedback, and mature the future presentation of results to relevant parties.

This study aims to identify the users of the archives and to understand the users' information needs (Ingwersen, 2000) and their information search behaviour (Wilson, 1999) through a mixed-method approach of questionnaire and interview (Morse, 2016). The archives in question are the Sound & Vision (the national audio and video archive of The Netherlands), INA (the national audio and video archive of France), and the Pitt Rivers Museum, Oxford, UK.

In general, there exist relatively few archive user studies (Rhee, 2015, p. 37). Hence, the present study is such a contribution. Further, Rhee argues that archive user studies should be conducted with “greater scientific rigour, precision, validity, and reliability” and not by the archivists themselves (Rhee, 2015, p. 37). The replies to the questionnaire inform about who the users are and their motivation and purpose for visiting the archives, which is complemented with in-depth interviews with six scholars that provide insights about the scholars' use of audio-visual archives for the purpose of postcolonial research.

The remaining paper is organized as follows: The following section introduces related work. Hereafter the test design and methodology of the mixed-method approach of questionnaire and interview are detailed. Next, the preliminary findings from the questionnaire and interviews are outlined. The final section addresses the ongoing work as part of the research project.

Related Work

Given our objective is to identify the users of the three archives in question, it is relevant to investigate who the archive users are in previous, related research. Here we see that heavy users of digitised archives include historians, genealogists and family historians (Chang & Lee, 1999; Darby & Clough, 2013; Wang, 2014), and further historians have used electronic means to locate primary materials (Tibbo, 2003). Media professionals have sought short pieces of audio-visual materials from an audio-visual archive (Huurnink et al., 2010). Another group of archive users includes documentary film directors and activists (Buchanan & Bastian, 2015; Fersters, 2019). Even though recent studies have focused on the impact of digital resources on the workflow of scholars (Green & Lampron, 2017; Kamposiori et al., 2019), from user perspectives, we still have a limited understanding of the similarities and differences among the different groups of archive users when they engage with these resources.

From the perspective of information seeking and retrieval, research has focused on the search behaviour of user search behaviour in archives or the use of finding aids. A large-scale study of search logs in the National Library of France has revealed user intents, user search behaviour and content interesting to users (Sumikawa et al., 2019). Singh and Anand (2017) were concerned with the design of search tasks for studying user search behaviour in archives. In a typical interactive information retrieval study applying a controlled user experiment, Freund and Toms (2016) found that task types affect the user search behaviour when users (genealogists and historians) are interacting with finding aids in print and digital forms.

Overall, since these studies have been designed to study user search behaviour to improve user interface design, these findings have revealed the search strategies of users of archive users, and how the functionalities of search and browsing interfaces can be developed to accommodate the interaction patterns.

Test design and methodology

Questionnaire

To bridge the archives and the users for a dialogue on information access, we conducted an online survey study by using a questionnaire as a data collection technique. This allowed us to specify the information needs of the different user groups who access the digital archives and their search strategies. The questionnaire was prepared in English, French, and Dutch to accommodate the linguistic differences between our partner organisations and their users.

A purposive sampling technique was used to survey the users of digital archives (Kelly, 2009). A blurb with a link to the English questionnaire was hosted on the pages of the photograph and film collections of the Pitt Rivers Museum website. The English version was also distributed to the users of digital archives through our contacts at the Sound & Vision archive, the EUscreen Foundation, as well as digital humanities researchers through our social networks. The French and Dutch versions were distributed to our contacts of digital humanities scholars in France and the Netherlands, respectively.

Interviews

Six semi-structured interviews were conducted with scholars using the three archives, two from each of the archives. The interviews were conducted online and were audio-recorded and transcribed by one of the authors. Interviews lasted between 25 minutes and one hour. The questions asked in the interviews concerned the scholars' information search behaviours, their information needs, their search strategies, and what factors make their searches succeed or fail.

Interviewees were recruited via the archives' websites and by directly approaching users of the archives. The interviews were analysed using the thematic analysis (Braun & Clarke, 2006), where two of the authors independently reviewed and coded the interviews before discussing their analysis.

Results

Questionnaire overview

We received a total of 46 responses from November 2021 to April 2022. The responses came from various countries, including France, the UK, Netherlands, the US, Canada, Saudi Arabia, Germany, Norway, Serbia, Slovenia, Romania, Spain, Kenya, China and Indonesia (n = 41). It is expected that our partner organisations and contacts represented the top 3 countries (France, UK and the Netherlands). Respondents aged from 19 to 77, with an average of 47 years old (n = 41, mean = 46.7, median = 49.0). The gender distribution was 28 female

(62.2%), 15 male (33.3%), and 2 prefer not to say (4.4%) (n = 45). Respondents came from a diverse group of users, with archivist, historian, media professional, student and university lecturer represented between 10% to 17% of responses. Activist represented only 4.3% while other user groups accounted for 26.1%. The other user groups include communication specialist, information professional (excluding media) and researchers who work in an institution or independently in the fields of anthropology, music, and media.

Regarding the user’s capacity of visit, the top response was job (20, 43.5%), followed by study (15, 32.6%) and interest (11, 23.9%). Some respondents specified their specific topics of interest, mostly in history, and there were notable capacities of visit to the digital archives, such as a mix of work and interest (social activist), “finding photos related to my community” and “restitution and repatriation”.

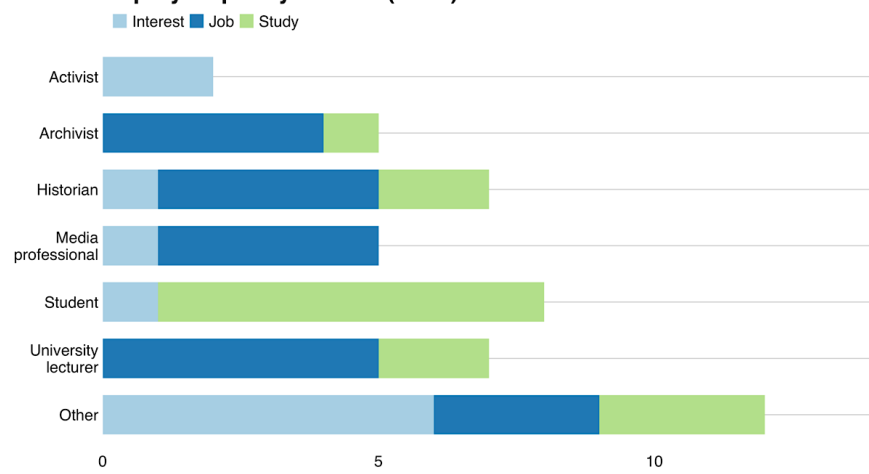
Most respondents had some (14, 30.4%) or a lot (13, 28.3%) of familiarity with the collection, while there were still a few with none (9, 19.6%), little (4, 8.7%) or a bit (6, 13.0%) familiarity. About a quarter of their visits were the first time (11, 24.4%), and the other categories included a few times (13 28.9%), several times (7, 15.6%), frequent (11, 24.4%) and always (3, 6.7%). As such, most respondents were frequent visitors who had a high level of familiarity with the collection.

Questionnaire – user groups

Since this study is concerned with the information needs and search strategies of different user groups of digital archives, this section reports the results by user groups.

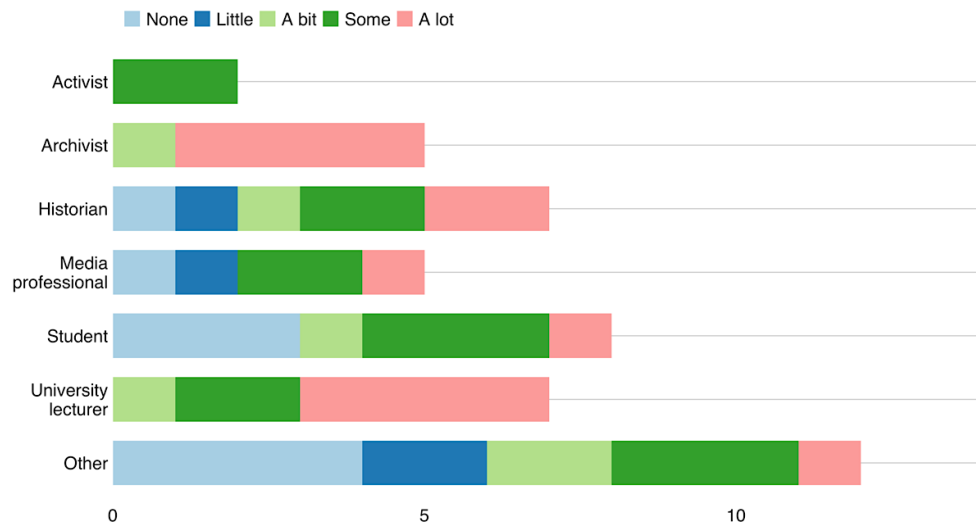
Most user groups visited the digital archive in the capacity of job, except for students and activists. The capacity of visits for activists was interest, which was shared by most other user groups, except by archivists and university lecturers. The capacity of visits of study was also shared by most user groups, except by activists and media professionals. The user groups of other and historians had more diverse capacities of visits than other groups. It is worth noting that other capacities of visit were related to the characteristics of the digital archives regarding indigenous culture and colonial history, including finding photos related to my community” and “restitution and repatriation”. Overall, the results show diverse user groups of digital archives who visited the collection in the capacity of job, study and interest, as well as the specific contexts of indigenous culture and colonial history.

User Group by Capacity of Visit (n=46)



All user groups had at least some familiarity with the collection. The levels of familiarity of some and a lot showed in almost all user groups. The category of none familiarity with the collection represented all user groups, except for activist, archivist and university lecturer. Overall, the results reveal that all user groups were quite familiar with the collection and there were still cases where users were not familiar with the collection at all.

User Group by Familiarity with Collection (n=46)



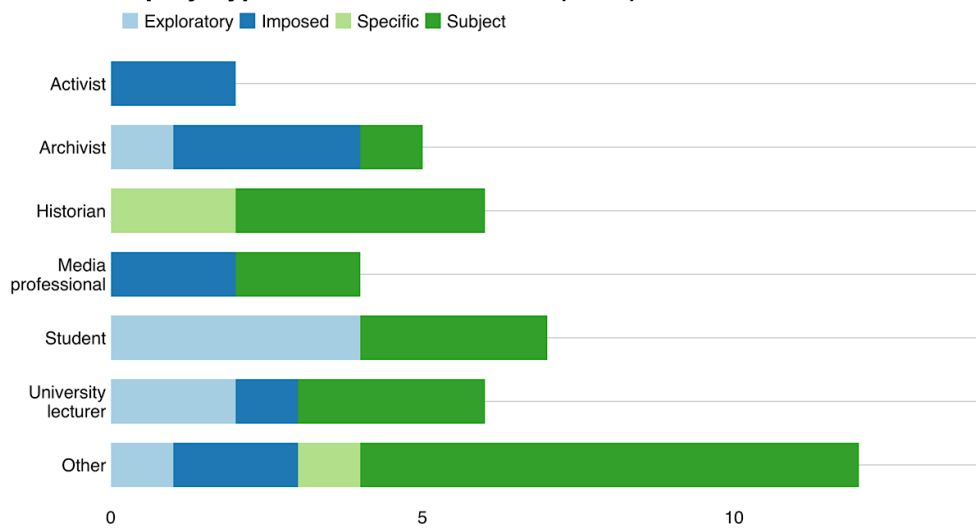
Questionnaire – information needs

Regarding user information needs, we characterised the types of search questions by the specificity of search topics and ranked them by frequency:

- Subject (to search for topics of interest with identifiable topics): "Visual representations in anthropological films of the 20s and 30s";
- Imposed (to search the digital archive on behalf of others): "Information required by the broadcaster";
- Exploratory (to explore topics of interest without clearly identified search questions): "at this stage, I just want to see what photographs are in the collection";
- Specific (to search for clearly identified topics of interest, i.e., known-item search): "The Victoria, British Columbia, photographer Richard Maynard (1832–1907) and any photos in your holdings".

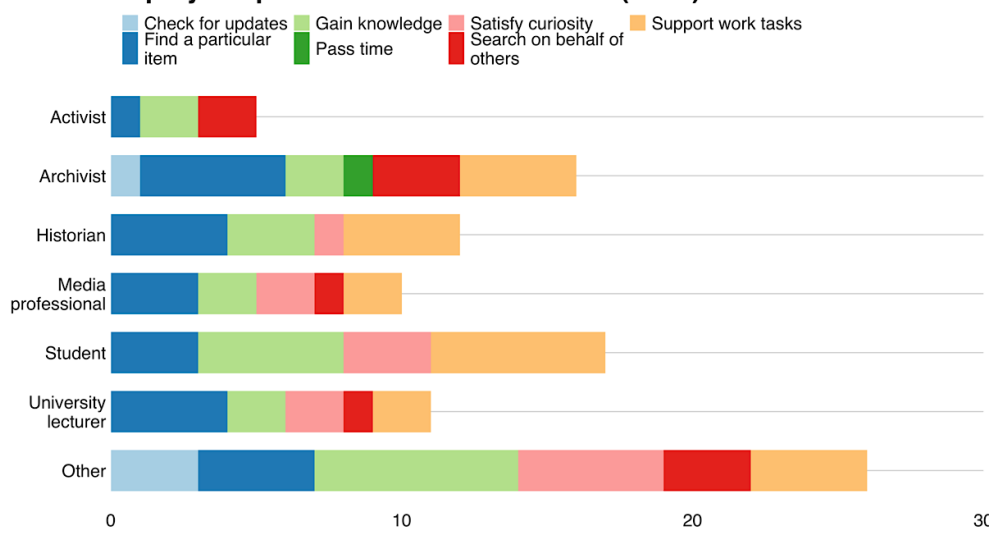
Searching the archive on behalf of others (i.e., imposed search) appeared in most user groups, except for historians and students. As expected, subject searching was important for searching the digital archives. Overall, the distribution of types of search questions by user groups demonstrates the different levels of information needs when they search for digital archives.

User Group by Types of Search Question (n=42)



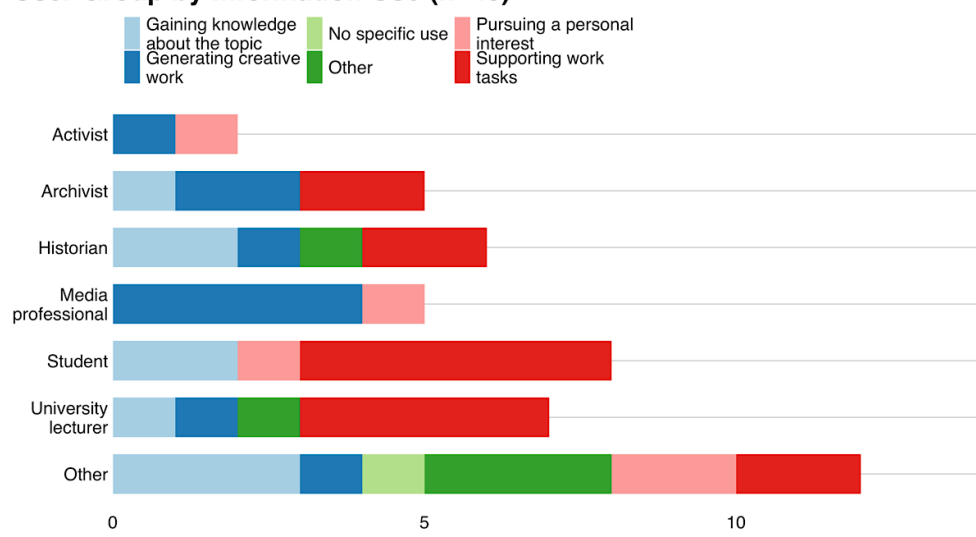
The question of why they are seeking this information is concerned with the motivations of information seeking. In this study, respondents were primarily trying to find a particular item, gain knowledge, and support work tasks. ‘To satisfy curiosity’ appeared in most user groups, except for activists and archivists.

User Group by Purpose of Information Search (n=97)



And how are they going to use this information? The most frequent answers are for supporting work tasks, generating creative work, and gaining knowledge about the topic. Supporting work tasks was shared by all user groups, except for activists and media professionals. Generating creative work was shared by all user groups, except for students. Other use of information includes ‘Learning more about my heritage’, ‘It depends on what I find’, and ‘To disseminate research’.

User Group by Information Use (n=45)



Overall, these findings suggest a diverse group of users of archives with different information needs and information use. Aside from finding a particular item to support work tasks, it can be found in photos related to the community to learn about one's heritage, or it can be an exercise of restitution and repatriation. This opens a dialogue for the polyvocal interpretations of colonial heritage collections.

The interviews

The interviews are conducted with six scholars (participant 1-6 (P1-P6)) who search for audio-visual material as part of their research on post-colonial cultural heritage. The interviewees do all have years of search experience and experience with archives, and their foci of research are varied which is reflected in how they search and what they search for.

P1 and P3 both search the INA in France. P1 studies video content documenting how North African immigrants' memories are portrayed online, concerning the Algerian war. Whereas P3 investigates the colonial past of Marseille by looking at the representation, the speeches, and the public use of the colonial past of the city of Marseilles in the media (television, radio, and the web).

P2 and P6 search the Sound & Vision archive in the Netherlands with P2 focusing on the content of the archives, and how the archive records' metadata represent views and opinions on past times. P6 is interested in Indonesian Gamelan music and cultural performances as well as images of Suriname and Indonesia, and how music, sound, and images are used as part of news reports.

The scholars P4 and P5 both search the Pitt Rivers Museum's film and photography collections and work on a joint project where they study three prominent and quite different British anthropologists: Beatrice Black Cold, Ursula Graham Bauer, and Frederick Chapman Spencer. The focus is on the Western gaze as depicted via the British anthropologists and their meeting with colonized people in the 1920s-1940s.

Information needs

The scholars' information needs are categorised according to the three different types of information needs devised by Ingwersen (e.g., 2000). At an overall level, the three types of information needs are the verificative information need, the muddled topical information need, and the conscious topical information need. In short, the verificative information need is about the verification of information and fact-finding. The remaining two types of information needs are both of topical nature and differ with respect to prior knowledge of the user concerning the topic in question. The muddled topical information need represents an information need where the user knows little or nothing about the search topic and therefore would like to learn more, and the conscious topical information need is an information need where the user is familiar with the topic and would like to update or further explore the topic.

The scholars' information needs are dominated by being of the type conscious topical information need pointing to their familiarity with their research topics, but all three types of information needs appear. An illustrative example of a conscious topical information need is seen with P2 who investigates how people talk: "It's either looking at a historical question so, how did we talk? Uh, you know, how did the Netherlands talk about a subject? For example, right? [...] Also, things like word choices in the metadata. So not just the archival record itself, but literally, what the metadata tells us about the institution, about the people that were labelling the document itself". Another example is from P4 and is seen by how the western gaze of the anthropologists is identified by studying the same films over and over again: "I tend to see the films many times because every time I look at them something new comes up".

The following example of an information need from P6 can be categorised as either a muddled topical or conscious topical information need depending on the level of prior knowledge of the topic in question of P6: "And then you know, what I'm looking for sort of the different ways that music is used, different kinds of music and the effect that it might have on viewers to sort of shape the narratives of, of the news and of history, looking at them as historical objects as well". P6 provides a clearcut example of a muddled information need pointing out how they search for coffee and chocolate to learn what is to be found in the archive: "we've been looking to see what there is about coffee and chocolate. Because those are colonial products that very often have colonial images attached to them and they're quite literally".

P4 and P5 both demonstrate verificative information needs for P4 it is seen by how the search for photos is used to document the identified western gaze in the films: "But that's not my aim. My aim is not, you know, is really to find some photographs that relate to the same kind of Western gaze that we have seen in the films". Whereas P5 is very much oriented towards know item search, hence categorised as verificative information need, which is illustrated by the following excerpt: "We decided the anthropologists beforehand. So what I did is just, I looked for a clear, with a simple control F search, the name of the anthropologists and, you know, I, so [...] to identify the movies I was interested in, and then I was looking at the movies".

Search strategy

The scholars' search strategies have some similarities, but they also adapt collection-specific strategies. In general, keyword searching is used by all scholars, often they plan which keywords to use, e.g., P6 states "I guess, I started out just thinking about keywords that might bring up colonial contexts or problematic, or sort of controversial aspects of history. So, a lot of that has to do with bringing in my own knowledge of Dutch colonial history and what may be some of those points of interest or controversy. I've also used some resources that other

people have compiled about lists of controversial terms in the sort of cultural heritage context.” P1 explains “So, I write a list keyword list and I also for each research I make. Uh, I note all the keywords I have used and all the results I have [...] I have to write something like 10 keywords with Boolean operators”.

P4 and P5 both analyse a small collection of films, and state that they know which films to look at, illustrated by the following quote from P4: “I normally go straight into the film collection because that's what I'm looking at the moment, so I don't go into the website and you know, I've got the link already set up”. Both, however, compile terms from the film metadata to use for searching the collection of still images in the archive: “you can get to the archive the photographic archive via the knowledge that you have acquired” (P4).

The scholars searching large archives often perform explorative, trial and error-based searching. P2 explains it in the following way “Usually, I try to start with a really broad query and see what gets me, and again, it's that sort of narrowing down. So then what I'll do is usually start with a broad query look into one or two documents that are clearly relevant to what I'm looking for and [...] I will notice some words that leap out that are really relevant that I wouldn't have known to use as a search term and then I use those” and P2 continues “in other cases where I've just been curious to see what there is, and that's sort of a keyword search a lot, and it's usually a matter of trying out several different keywords and seeing what shows up”.

Search challenges

The scholars meet different kinds of challenges when trying to find information in the archives. Some are related to the complexity of the search system. This is illustrated by two examples from P1 and P5: “Yeah, uh, first thing I have to say is that web archives are very specific. To find information in Web archive is a methodological challenge and uh. It's a kind of research question [in its own right] in a way” (P1). P5 comments on the difficulties in performing location-based searching: “I believe, there are several like location or geographically related fields that are not necessarily clear because some with dropdown closed options, like you have continent and then country”.

Another type of challenge pointed out by P5 is the very ‘time-consuming process’ of interacting with films, which is also made note of by P4, who says “I tend to see the films many times because every time I look at them something new comes up”. P2 has come up with a strategy to cope with this challenge:

So you actually bypass the archival metadata and go straight into the transcript itself. Which is great 'cause you can pinpoint literally specific instances of people saying something. So if there's a term of phrase you want, you can go straight to the moment in the broadcast where they're using that phrase. Because one of the peculiarities of working with audio-visual material specifically is that it's extremely time-consuming to actually sit down and watch and or listen to. You know, sometimes it'll be an hour-long document, and if you're looking for something this big. So one of my search strategies is to avoid that at all costs. And well, not to avoid it, but avoid it until I absolutely have to do it. (P2)

The need for help

To overcome the challenges, many scholars seek help from archivists, e.g., P3: “I have to talk to the documentalist in Marseille because the principal office of the INA is in Paris, but I am working in Marseille”. P2 has been working with the archivists and “discovered that there are

very few people actually at the institution, who have a really good overview of what's in the collection. There are one or two people". P1 also collaborates with archivists, stating "I am also working on different projects and in collaboration with web archivists. [...] Sometimes they also share with me data and metadata. For instance, I have metadata of 3000 videos related to the Maghrebi struggle during the 2000 and they had sent me the database in an Excel file, so it's another way to search inside".

In addition to P1, several scholars use supportive software to handle the archive data. P2 mentions the TV guide from BBC – "BBC Genome" and Clariah's¹ Media suite. The Media suite and Excel are also used by the other Dutch scholar (P6) to interact with the archive content.

What to do next...

To report work in progress and present the preliminary empirical findings from this European collaborative research project at the final conference of the international network Digitization and the Future of Archives: Digital archives, Big Data and Memory represent the initial steps in the process of coding and analysing the collected data as well as disseminating the results. Still, more analyses are to be carried out in that five in-depth interviews remain to be analysed. Further, the two sets of data, the questionnaire and the interviews are to be coordinated to complete the portrait of the archive users and their information needs in the context of colonial audio-visual heritage. The findings are to be presented to the research consortium in the early autumn of 2022 providing the opportunity for a dialogue between the archives of Sound & Vision, INA, and the Pitt Rivers Museum and the users. Finally, the results are to be reported in writing and submitted to peer-reviewed journals and conferences hereby making the case of a recent study of archive users.

Funding

This work is funded in part by The Research Council of Norway – JPICULTURE-Cultural heritage and global change (Project number: 323808).

References

- Buchanan, A., & Bastian, M. (2015). Activating the archive: Rethinking the role of traditional archives for local activist projects. *Archival Science*, 15(4), 429–451.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Chang, S.-J. L., & Lee, Y.-Y. (1999). Exploring information use of historical records and archives: From the user's point of view. *University Library Quarterly*, 3(3), 62–76.
- Darby, P., & Clough, P. (2013). Investigating the information-seeking behaviour of genealogists and family historians. *Journal of Information Science*, 39(1), 73–84.
- Fersters, I. (2019). *Making an archive come to life: A comparative case-study of three documentary film directors as archival users*. Unpublished master thesis, Lund University.
- Freund, L., & Toms, E. G. (2016). Interacting with archival finding aids. *Journal of the Association for Information Science and Technology*, 67(4), 994–1008.

¹ Dutch Digital Humanities research infrastructure: <https://www.clariah.nl/>

- Green, H. E., & Lampron, P. (2017). User engagement with digital archives for research and teaching: A case study of Emblematica Online. *Portal: Libraries & the Academy*, 17(4), 759–775.
- Huurnink, B., Hollink, L., Van Den Heuvel, W., & De Rijke, M. (2010). Search behavior of media professionals at an audiovisual archive: A transaction log analysis. *Journal of the American Society for Information Science and Technology*, 61(6), 1180-1197.
- Ingwersen, P. (2000). Users in context. In *European Summer School on Information Retrieval* (pp. 157-178). Springer, Berlin, Heidelberg.
- Kamposiori, C., Warwick, C., & Mahony, S. (2019). The impact of digitization and digital resource design on the scholarly workflow in art history. *International Journal for Digital Art History*, 4(4), 3.11-3.27.
- Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1–2), 1-224.
- Morse, J. M. (2016). *Mixed method design: Principles and procedures*. Routledge.
- Rhee, H. L. (2015). Reflections on archival user studies. *Reference and User Services Quarterly*, 54(4), 29-42.
- Singh, J., & Anand, A. (2017). Designing search tasks for archive search. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 361-364.
- Sumikawa, Y., Jatowt, A., Doucet, A., & Moreux, J. P. (2019). Large scale analysis of semantic and temporal aspects in cultural heritage collection's search. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 77–86.
- Tibbo, H. R. (2003). Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *American Archivist*, 66(1), 9–50.
- Wang, L.-C. (2014). A study of archives users' information behavior in context. Unpublished doctoral dissertation, National Taiwan University.
- Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249-270