



Image searching in an open photograph archive: search tactics and faced barriers in historical research

Elina Late¹ · Hille Ruotsalainen¹ · Sanna Kumpulainen¹

Received: 23 February 2023 / Revised: 22 November 2023 / Accepted: 17 December 2023
© The Author(s) 2024

Abstract

During the last decades, cultural heritage collections have been digitized, for example, for the use of academic scholars. However, earlier studies have mainly focused on the use of textual materials. Thus, little is known about how digitized photographs are used and searched in digital humanities. The aim of this paper is to investigate the applied search tactics and perceived barriers when looking for historical photographs from a digital image archive for research and writing tasks. The case archive of this study contains approximately 160,000 historical wartime photographs that are openly available. The study is based on a qualitative interview and demonstration data of 15 expert users of the image collection searching photographs for research and writing tasks. Critical incident questions yielded a total of 37 detailed real-life search examples and 158 expressed barriers to searching. Results show that expert users apply and combine different tactics (keywords, filtering and browsing) for image searching, and rarely using one tactic only is enough. During searching users face various barriers, most of them focusing on keyword searching due to the shortcomings of image metadata. Barriers were mostly in the context of the collection and tools. Although scholars have benefited from the efforts put into digitizing cultural heritage collections, providing digitized content openly online is not enough if there are no sufficient means for accessing the content. Automatic annotation methods are one option for creating metadata to improve the findability of the images. However, a better understanding of human information interaction with image data is needed to better support digitalization in the humanities in this respect.

Keywords Digital humanities · Digital images · Image archives · Image search · Image retrieval

1 Introduction

During the last decades, considerable investments have been made to enhance digitalization and open access to data in the humanities. In consequence, cultural heritage collections have been digitized for the use of academic scholars and others interested in the contents. Digitization of archival materials and online access have not only offered greater availability for the use of such collections but also created

new research possibilities for the digital humanities (DH). For example, the use of digital data includes quantitative approaches accompanied by more traditional methods such as close reading [34]. Open access to digital content has also changed the traditional role of the archives from being mere protectors and preservers of records toward data providers [43].

Historical photographs are one example of cultural heritage collections that have been digitized widely. In the fields of art history and visual studies, a tradition for theoretical considerations of images as data and shaping of visual scholarship exists [16, 27, 42]. Nevertheless, there is a dearth of user-centered studies, focusing on how historical digitized photographs are utilized and searched in DH research. Most of the earlier studies have focused on textual collections such as books or newspapers [2, 15]. Yet, photographs are important primary sources in the history domain, and they are used, for example, for knowledge creation [3, 9].

✉ Elina Late
elina.late@tuni.fi
Hille Ruotsalainen
hille.ruotsalainen@tuni.fi
Sanna Kumpulainen
sanna.kumpulainen@tuni.fi

¹ Faculty of Information Technology and Communication Sciences, Tampere University, Kalevantie 4, 33014 Tampere, Finland

It is commonly agreed that images are difficult to find as their searching is mostly reliant on textual descriptions [60]. Creating metadata for images may be challenging as the reading of an image depends on the viewer, resulting in potential divergent readings of the same image. However, images are always born and represented in a specific context that influences their reading [27] and the searchers' interpretations of the images may change throughout the search process [12]. Searching for historical photographs becomes even more challenging due to the frequently incomplete or historically contextual descriptions associated with these images [51]. This makes image searching in a historical context a challenging research field that imposes requirements beyond the mere availability of digitized materials. Merely opening the image collections is not enough if the contents cannot be found, accessed, interoperated, and reused, as advocated by the FAIR principles [59].

However, studies focusing on historical image searching practices from digital archives are rare. It is not yet known how scholars manage to find historical images and what barriers they face when doing so. This information is vital in improving and supporting the usability of such collections [6]. This necessitates qualitative research on image searching to gain a deeper understanding of how people search for images in their real-life tasks and interpret their experiences [10, 39]. The information needs are derived from the context of human behavior, which in the present study is the scholarly work of history researchers. The context shapes how meanings are interpreted within the images.

This study seeks to address this research gap by studying the search tactics employed and barriers perceived when finding images from a digital image archive containing historical wartime photographs. This collection was not originally intended for research purposes but rather to provide illustrations for the propaganda organization that operated during the Second World War in Finland. After its digitization in 2013, the collection has been a popular source of image data for, e.g., genealogists, other hobbyists, and history researchers in Finland. The study is based on qualitative interview and demonstration data collected from expert users of the collection who search photographs for research and writing tasks.

Our research questions are:

RQ1. What search tactics are used for historical image searching?

RQ2. What barriers do image seekers experience when searching for the images?

The article will first provide a background on image uses in DH research, search tactics used, and perceived barriers to image searching. We continue by describing the research methods, followed by results and discussion.

2 Background

2.1 Use of images in digital humanities

Over the past decade the digitalization of research materials and tools for humanities has heavily influenced scholars' ways of working [21, 34, 57]. Although some scholars might still favor print over digital formats [20], many see digital collections as essential for conducting research [54]. While previous research has focused, for example, on the production and needs for image metadata [8, 35, 46], there is no comprehensive understanding of how humanities scholars are able to access digitized images for their research purposes.

Yet, digital images are important primary sources (i.e., research material) in history research. According to Chasanoff [9], photographs provide a valuable historical reference for verification, documentation, or corroboration. Historians use photographs for historical reasoning showing and learning "what things looked like then" [9]. Historians place significant importance on the trustworthiness associated with reputable institutions like archives and the origin of the photographs when utilizing these images. They desire original descriptive information, such as captions, keywords, subject headings, original medium, and the size of the images [9]. The contextual information about the creator of the metadata is also crucial for humanities scholars [46]. In some cases, digital surrogates cannot replace the original paper photographs, but historians can use the digital collections as a tool for finding the photographs they need [9]. This applies also to digitized textual materials [14, 54].

Beaudoin [3] studied image use among archaeologists, architects, art historians and artists. She discovered that images were used for various purposes such as knowledge creation, conceptual modeling, inspiration, cognitive recall, critical thinking, communication, emotion, engagement, marketing, proof, social connection, translation, and trust. In her study, it was observed that the image use differed among user groups. Specifically, those in archaeology and art history used images most often for knowledge creation of their lecture presentations, as well as for their research and subsequent publications. Indeed, Fidel's seminal study analyzed image use as a data pole and an object pole [17] and later McCay-Peet and Toms studied image use for illustration and for information [40]. Conniss et al. [13] categorized the use of images as knowledge construction into four distinct categories: information processing, information dissemination, learning, and ideation. Other image needs that have been identified in the research include entertainment, aesthetic appreciation, engagement, inspiration, and social interactions [10].

2.2 Image searching

Although images are visual data, they are mostly searched using text-based queries rather than by images. Cho et al. [10] and Westman [58] offer extensive literature reviews on image information behavior including image searching. Generally, keyword searching and browsing are the central tactics in image searching and the choice of the method that is applied depends on the image need and functionalities provided by the search system [13].

Studies have shown that keyword searching is the prior tactic for image searching within and beyond the history domain [32, 38, 41]. Users tend to favor very short queries, such as isolated terms or simple expressions [11, 38, 49]. Choi and Rasmussen [12] focused on keyword searching and analyzed the subject contents of queries in a digital image archive of American history. The number of search keys used by the participants varied between 1 and 15, the average being 4.87. Users having a general/abstract request used more search keys compared with those having a specific, generic/nameable, or subjective request. Participants rated the date, title, and subject descriptions as the most important factors representing images. Identified keyword categories were the names of the kind of person, things, geographical names, kind of event, action, condition, individual names, and the date or period of time.

In addition, browsing is an important search tactic and is often accompanied by keyword searching [9, 29, 32, 37, 38]. Browsing can be used either for finding the images or for getting familiarized with the images to be able to perform keyword searching [13]. The study by Göker et al. [22] showed that creative professionals used keyword searching for targeted searching but in the case of broader searches, they used browsing by categories. Browsing can also be used to avoid the unintended exclusion of important images originating from too narrow queries [9, 37]. Further, browsing has shown to be an attractive tactic for those users who have little knowledge about the domain or collection [19]. According to Münster et al. [44], art historians may use digital images for serendipitous discovery when looking for inspiration at the beginning of research processes. Also, Matusiak [38] found that browsing was more likely to be exploited by those users who were less confident with their digital search skills. The confident users were more likely to use keyword searching as a search tactic. However, the affordances that are provided by the search interface influence user's browsing behavior, and sorting images into categories and providing thumbnails of images encourages browsing [24, 50].

2.3 Barriers to image searching

Barriers are something that restrict or hinder the search process and cause negative affections [31, 52]. To date, only

a handful of studies have concentrated on the analysis of barriers to image searching, and only a few of them have positioned themselves in the context of searching historical images. These studies focus mainly on art history. Thus, there is clear lack of studies in this respect.

Beaudoin and Brady [5] studied the image use by archaeologists, architects, art historians, and artists. They discovered problems related to discoverability, copyright, size, and quality. Despite the large digitization projects over the past years, another study showed that art historians were suffering from the lack of open-access visual materials [44]. However, the availability of materials varied between geographical locations and research specialties. Other barriers recognized in the study were related to the quality of metadata, the resolution of the images, and the indication of rights of usage. Art historians perceived the interfaces often as hard to use and browsing too time-consuming.

More generally, literature review by Cho et al. [10] revealed obstacles in image searching that were related to semantic problems, content-based issues, technical limitations, issues of aboutness, inclusivity issues, search skills, and cognitive overload. Semantic problems that were seen as the most important relate to the terminology or language used in image retrieval systems. Text-based searching requires metadata that has many shortcomings starting from the mismatch between the contents and the textual descriptions, or user interpretation and cost [28]. Clearly, images always contain more information than can reasonably be included in textual descriptions. Vocabularies and ontologies that are developed for the systems can quickly become outdated as the user needs are contextual and dynamic. Thus, they require constant maintenance and upkeep [56]. Terminology for describing images may not always meet the needs of different users, such as professionals vs. non-specialists [11]. Content-based problems are related to the identification of image attributes [32]. Some studies have identified technical barriers such as long load times, size, format, and resolution of images [24, 25].

Beyond image retrieval, Kumpulainen and Late [31] studied the context of barriers to information interaction faced by academic historians who were using digitized historical newspapers. They provide an analysis frame for barriers occurring in four contexts: collection, task, tools, and socio-organization. In their study, barriers related to searching and selecting activity appeared mostly in the contexts of the collection (e.g., OCR data) and tools (tools were not available or they were too complicated).

3 Finnish wartime image collection

This study investigates the use of a unique digital collection of Finnish wartime photographs containing around 160.000



Fig. 1 The photograph archive interface. <http://sa-kuva.fi>

photographs from the Second World War during 1939–1945. The collection is provided by the Finnish Defence Forces and is also available in print format. The images in the collection portray life on the home front, events and operations at the front, the war industry, leisure time at the front, damages in bombings, and the evacuation of Finnish Karelia. The photographs were often used for wartime propaganda, and they were mainly taken by wartime Information Company photographers. Most of the photographs are in black and white, and a small share of the material consists of color photographs or video recordings. The collection was published online in 2013 and is openly available for all users (<http://sa-kuva.fi/>).

The digitized photographs can be accessed via an online search interface (see Fig. 1) including keyword search, advanced search (Boolean operators), and browsing. Filtering options available include a pull-down menu for the predefined stages of the war (Winter War, Continuation War, Lapland War), temporal searching based on specific date information, and selecting videos and images in color. In addition, users can select “those without dates” to include images lacking date metadata in their results. The result page includes 15 thumbnail images for browsing. The users can click a thumbnail to open a larger image, access related metadata and load the image. At this stage, users can also submit additional information about the image for the archive. Archive description, guidelines, and terms of use are provided.

The search is based on textual machine-readable metadata of the images. Metadata is based on information that was mostly created during wartime by the photographers who were instructed to provide the name of the photographer, the

location, and the subject or event in the image. This information was manually entered into the digital archive to form the metadata for the images. However, metadata is partly missing because of the chaotic times during the photographing. The photographers did not always have a chance to write any kind of description, at least not in detail. Metadata also contains spelling errors and other mistakes concerning, for example, the date and location. The metadata has never been edited or proofread. Metadata is mainly in Finnish, some in Swedish.

4 Research data and methods

The research data were collected by semi-structured in-depth interviews and complementing demonstrations about how images were searched from the digital archive. The data consist of 15 interviews collected during five months from November 2021 to April 2022. All interviewees were active users of the archive who regularly searched images for the purposes of research and writing tasks. The selection of the interviewees proceeded partly through the contacts of the research team, partly through contacting organizations where the archive was known to be used and partly by snowballing method (each interviewee was asked whether there was someone else (colleague, etc.) suitable for the interview). Among the interviewees there were two scholars who had previously worked in the Finnish Defence Forces’ Photograph Archive before the collection was digitized (in 2013). The profile of interviewees is presented in Table 1.

Interviews were done online using video conference tool Zoom, and all interviews were video-recorded. Interviews

Table 1 Profile of the interviewees

	Age range and the number of interviews
Age	26–59 years (mean 44, median 46)
Work organization	University (9), library/museum (2), private company (4)
Discipline	Archeology (5), history (8), art history (1), ethnology (1)
Profession	Academic scholar (11), professional writer (4)

were in Finnish, and data were collected until saturated. The video files were transcribed full for analysis. The average length of one interview was 37 min, and the interviews' audio data run, in total, 9 h and 10 min. Before the interviews, informed consents were collected from the interviewees. Part of the interview data has been used in previous study [35].

Interview questions included background information such as current status, research field and age. The following questions applied a variation of the critical incident technique [18] where interviewees were asked to describe the ways they had used the image collection by searching, selecting, and saving images from the archive for their recent task. This way we were able to collect descriptions of critical incidents with the collection. The interview guide is included in the Appendix A. However, interviews did not necessarily follow the order of the questions in the guide, but the guide was used as a checklist to keep track of the interview. The interviewees were free to talk about their user experiences in any order they wanted. In the demonstrations, interviewees were also asked to recall a recent or typical search topic and demonstrate how they searched the images. They were encouraged to think aloud and describe their searches. During the demonstrations, interviewees also brought up difficulties they had faced when searching. This way we were able to collect insights beyond the interview guide and the participants were able to articulate better their needs concerning the photograph archive. In addition, the demonstrations helped the interviewees to recall their concrete working practices and the barriers they faced during searching. Participants shared their screens during the online demonstration and their screens were video-recorded. The voice recordings were transcribed into text.

4.1 Data analysis

Researcher triangulation was used to increase the validity of the study. The data were collected by one scholar and initially analyzed by another. Finally, the coding and the interpretation of the results were discussed with the research team to reach consensus. Content analyses were conducted using Atlas.ti software and Microsoft Excel. The content analyses consisted of the iterative readings of the interview transcripts, open coding, and selective coding [55]. The open coding focused on instances in the data describing information about the search tactics and barriers to searching. Quotations from

the data were entered into two Excel spreadsheets. The first spreadsheet contained the descriptions of the applied search tactics that were further coded. We identified a total of 37 search descriptions from the data. Based on the coding, seven combinations of search tactics were identified. By search tactic we mean either searching by keywords, filtering, or browsing. The descriptions of the use of different tactics and their combinations were based on the interview data. The second spreadsheet contained the descriptions of barriers to searching. By a barrier, we mean the difficulties, obstacles, or frustrations expressed by the interviewees during their image search process. A total of 158 search-related barriers were traced. The search tactic the barrier was related to was coded in the spreadsheet. Next, the context of the barriers was coded according to the model by Kumpulainen and Late [31]. This model was chosen because it is particularly suitable for categorizing barriers within the context of DH research and digitized collections. To study the relationship between the search tactic and the context of the barrier, the variables were cross-tabulated. Quotations were selected from the interviews to illustrate the search tactics and barriers faced during searching. Quotations were loosely translated from Finnish to English.

5 Findings

5.1 Search tactics

Interviewees searched the images from the collection for research and writing purposes. Interviewees saw the collection as an important source for images, and its digitization offered new possibilities for historical research.

When these images were published on the web, it was a huge cultural investment, it basically revolutionized in many ways the wartime history research, in the same way as publishing the wartime journals [online]. [P15]

Interviewees searched images for illustration and information. Illustration concerned searching images for books, articles, social media updates, or presentations. When searching images for information, users' goal was to collect research data, fact-check, create teaching materials and assignments and get help with historical reasoning. However,

Table 2 Tactics applied for image searching and number of cases in the interview data

Tactics	N. of cases
I. Keyword (Image ID)	3
II. Advanced keyword + filtering + browsing	11
III. Keyword + filtering + browsing	8
IV. Advanced keyword + browsing	6
V. Keyword + browsing	3
VI. Filtering + browsing	5
VII. Browsing	1

in many cases, the needs overlapped. For example, interviewees searching images for research data often also used the images for illustration. Similarly, searching images for illustration often included searching for information, as the interviewees learned about the topics of interest and about the collection during the process. During the searching, images not necessarily relevant to the topic were found, but they were interesting for other tasks.

Through these images, you'll get an impression of the scenery and the events. Although I did not use the image directly [for visualization] it helped me to figure out what it was like. For example, what it was like for the soldiers located on the battlefield. [P6]

Interviewees searched either for a specific image from the collection or one or a set of images related to the topic. When searching images for research data, interviewees aimed to find all the images of the topic of interest. The interviewees used and combined three different search tactics, namely keyword searching, filtering, and browsing for image retrieval. Their prior knowledge about their information needs and about the image they were looking for influenced the selection of search tactics. From the interview data we identified seven different combinations of search tactics, which are presented in Table 2. In case I users were able to trace the specific image with a simple keyword, while in the most complex situation (case VII) browsing was the only search tactic available. Between these two extremes, the users selected and combined different tactics: keyword, filtering and browsing (case II, III), keyword and browsing (case IV, V), and filtering and browsing (case VI). Most commonly users combined all three tactics to search images (cases II, III). All tactics and combinations were used for searching images for both illustration and information.

In the simplest case (case I), the users knew the image they were looking for and they already had specific information about it. In this case, they were able to use, for example, a photo ID number as a keyword and locate the image easily without having the need to apply any other tactics. Photo ID numbers were found to be a very useful way to track images

and users collected ID numbers from already found images to be able to trace them again. Occasionally, users asked for the photo ID numbers from the authors who had used the images in their publications. Photo ID numbers were also used to get on track with other images about the same topic in the collection.

Historians' detective work goes like this. You'll get a grip of one image that has an ID number, then you follow the numbers, image by image. This way you can get images related to the same topic. [P15]

However, typically the situation was more complex, and the less information the users had the more they needed to combine browsing with other tactics to locate the image or set of images they were searching for. Their information need was either too abstract or they had no information about the metadata related to the image in the collection. Typically, the information need itself was broad and abstract (such as "the feelings of soldiers") and there was no simple way to approach solving it.

In most cases, users were able to make keyword searches (cases II, III, IV, V). Users searched with keywords using, for example, named entities (e.g., names of people, locations, or buildings), objects (e.g., animals, vehicles), temporal keywords (e.g., summer), roles (e.g., soldier, child), activities (e.g., skiing) or events (e.g., Christmas). They also searched with non-visual attributes such as the name of the photographer.

Users needed to be creative and use their imagination in selecting the keywords if they did not have prior knowledge of the possible metadata. They needed background information about the collection and its provenance, and information about the historical context of the images. They were trying to figure out how the image may have been described in the original image caption within the historical context. Thus, the keywords used in searching did not necessarily represent the information needs directly, but the user's conjecture of the words used in the image captions.

A bread or a coffee or a hymnal or a tombstone are examples of those that are included in the descriptions. But if you think about things like bravery, or fear or sadness or joy, that are very contextual and qualitative concepts [...] In this type of material that was strictly guided by the propaganda organization [...] every sentence was looked at very closely. But if you search for an image about bravery or joy you need to go around and look from the viewpoint of a certain event or location. We know that at a certain time and in a certain place something meaningful for the war happened. Can we find images that were taken after this event that represent joy or sadness? Or can we find images of sadness by just using the keyword "funerals"? [P1]

Finding the right keywords took a lot of time and nerves from the users but during the process, they gained experience of the collection that aided the searching.

It was about the tenth keyword that started to produce the content we wanted to have. [...] You need to learn how the engine works...and what which keywords to use. [P14]

In addition to using their own imagination, users located suitable keywords from other sources such as historical books, and original physical photo envelopes, they asked for information from their colleagues, and from already found images. Many users were experienced in using the advanced retrieval features in keyword searching, such as Boolean operators (AND, OR, NOT) and truncation (cases II, IV). These were helpful in including synonyms, and language or spelling variants in the search. Although the data were mainly in Finnish, some captions were written in Swedish, which the users needed to consider.

I've experienced that it's easiest to find images from there if you're creative in including synonyms and use truncation at the right places. [P3]

Filtering options provided by the system were also typically used (cases II, III, VI). The most used filter was "include those without dates", which broadened the search to those images that lacked metadata about the date. If the users were interested in images from a certain timeframe or had knowledge about the time of the image, they were able to filter the results by date or by the stage of the war (Winter war, Continuation war, Lapland war). They could also filter the images by color (color vs. black and white) and by content type (still vs. video), but these options were not often used. Indeed, none of the interviewees used video materials from the collection.

Browsing was included in most cases (cases II, III, IV, V, VI, VII). Typically, users needed to browse hundreds of images from the search results. The less specific the search, the more they needed to browse. During browsing the interviewees checked the additional information related to the images from the metadata. Although browsing was usually combined with some other search tactic, in some cases there were no other tactics available. One of the interviewees had browsed the entire collection (case VII). Furthermore, some interviewees browsed through the whole search result from the first to the last image. However, when there were hundreds or even thousands of images to browse, interviewees might start browsing the images randomly trying to find the right image by change.

5.2 Barriers to image searching

Although the interviewees were in general very pleased with the collection, all agreed that searching the collection was challenging. In total 158 expressions of barriers related to searching the images from the collection were identified from the data. Most of the barriers ($n = 102$, 65%) were related to keyword searching and the rest to filtering ($n = 29$, 18%) and browsing ($n = 17$, 11%). Some barriers ($n = 10$, 6%) were related to more general problems in searching.

Barriers were categorized according to their context to collection, tools, socio-organizational or task levels. Most of the barriers were in the context of collection ($n = 94$, 60%) and tools ($n = 54$, 34%). Only a few barriers were related to the socio-organizational context ($n = 6$, 4%) or the context of the task ($n = 4$, 3%). To study the contextual barriers across search tactics, the search tactic and the context of the barrier were cross-tabulated (see Fig. 2).

5.3 Barriers to keyword searching

Barriers related to keyword searching were mostly in the contexts of collection ($n = 70$, 69%) and tools ($n = 23$, 23%). Some barriers were also in the socio-organizational ($n = 6$, 6%) and task context ($n = 3$, 3%).

In the context of the collection barriers to keyword searching were mainly related to the metadata of the images. Metadata barriers were caused by the lack of incompleteness or wrong/inaccurate metadata. Part of the images in the collection had no metadata or metadata was incomplete concerning, for example, the date or the place of the photograph was taken.

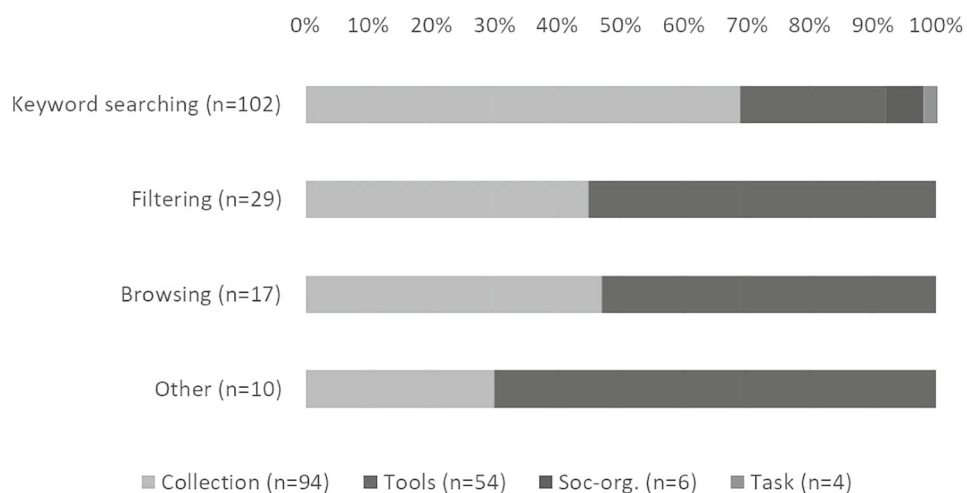
As far as I remember 10 percent or 15 percent of the photographs have no captions, so those cannot be searched by any means, so they are left out by everyone. It is a pity. [P1]

Metadata was found also incomplete, and in many cases, photographs in one film roll contained the same metadata.

The image captions do not include the names of the persons. The whole film roll, or I think there are many of those, have only the caption "Mannerheim's return visit to Germany". That is all. [P7]

Metadata was often too abstract and not specific enough to meet the users' interests. Users also struggled with the randomness of metadata. Some images included additional information, for example, about the photographer or the location, but many did not. In addition, there was variation in the locations, as some were described at the level of regions, and others at the level of towns or even villages. Wrong or inaccurate metadata also caused barriers. Especially the dates in

Fig. 2 The share of barriers across different contexts in different search tactics



the metadata were often wrong or the names of locations or objects in the images (such as vehicles) were wrongly named.

The big problem is of course the metadata...it is based on the wartime captions and those have been collected during that time. The interests were very different compared with what they are if you study cultural or wartime history today. So, the interest of the image descriptions does not meet today's research interests. [P3]

In addition, the language used in the metadata caused barriers to keyword searching. Image captions were written in natural language, so the users needed to consider inflectional forms, synonyms, variants, and typos when selecting the keywords. Captions included abbreviations that were not necessarily known by the users. Captions were written mostly in Finnish, but some contained captions in Swedish that needed to be considered when searching.

Barriers in the context of tools were most often related to expectations derived from more developed search systems such as Google. Users expected the system to support their keyword searching by automatically broadening the search, taking typos and spelling variants into account and to give recommendations.

Now we all probably use Google, and you know how it works, it gives recommendations. Let's say you search with the word "sauna" and then Google can suggest using the word "ladle" or "heater". [P7]

Although Boolean operators were often used and found helpful, some interviewees saw these as outdated. However, some desired more advanced search options where they could focus their search on certain or even various metadata labels. Other barriers in the context of tools were related to a lack of knowledge about the search system, lack of, or misleading guidelines, and searching skills. For example, interviewees

described situations where they did not understand how they got the results. This caused a feeling of insecurity for the users as they did not know how the system worked.

If you search by "dog", you'll also get those where it is not mentioned in the caption. There is a dog in the image, but it is not in the caption. There is some system behind this, but it has not been described in the guidelines. [P9]

Socio-organizational barriers to keyword searching were related to the lack of historical knowledge of the collection and its metadata. The language of the captions was historical and contained words and names of places that were not used in contemporary language or that had changed during the time. Captions also contained, for example, abbreviations and propaganda jargon used during the wartime and negative expressions were avoided in the captions. Because of the specific nature and the wartime conditions during which the captions for the photographs were created, users had problems imagining what the metadata could contain. They needed knowledge about the historical context and the historical language to be able to formulate the keyword queries for searching.

Images with swastikas? Extremely delicate topic in every possible way, which is guaranteed not to be found in the caption. [P1]

Barriers in the context of the task were related to the time the interviewees had for searching with different keywords. Users were forced to delimit their searches in certain keywords although they knew that all images were not found because of the lack of time.

5.4 Barriers to filtering

Barriers related to filtering were in the context of tools ($n = 16, 55\%$) and collection ($n = 13, 45\%$). Barriers in the context of tools related to lack of filtering options, technical issues in filtering (it did not work), and how the options were provided in the interface. Because of the commonly shared experience that the filtering did not work properly, many interviewees held back from using the filters.

Basically, I don't use at all the filters on time spans or dates, it doesn't just...I feel it just mixed it up, there is something wrong with the code. [P3]

Barriers in the context of the collection were related to metadata issues of the collection. Because the filters were based on the metadata (e.g., the dates of the images), filters gave wrong results or left images outside the search. Since many images lacked metadata about the date, users needed to select "search from those without date" to include all images in their search. This option was available for the Finnish interface only. It was a major barrier for especially inexperienced users but caused incomplete results every time the users forgot to select the button.

At first, I did not get it. It is a bit hidden in the top corner [the option to include images without a date]. You would though it comes automatically. I can imagine that many users miss this and then those without dates are never found. [P7]

5.5 Barriers to browsing

Barriers to browsing were related either to the context of tools ($n = 9, 53\%$) or the collection ($n = 8, 47\%$). In the context of tools, barriers were related to the interface and how results were provided. Interviewees found the interface problematic for browsing since the thumbnail images were small and could not be quickly browsed when looking for details. Interviewees were also insecure about how the result list was organized, and they wished abilities for organizing the results, for example, according to the date of the image to help the browsing.

The order of the results is a bit fuzzy for me. I want them in chronological order, it would help to follow what happened. Now it is somehow random...in a way it has been categorized according to the photo series. I can see there is photo one, photo two...how they were in the film roll, but sometimes it gets mixed. [P5]

In the context of the collection, barriers were related to the number of images to browse. Interviewees also reported that they often lose track when browsing several images.

He said that there is this image but the only way to find it is by browsing. We have some idea, that it was taken during the Continuation War, but it is still four years, so there is a lot to browse. [P2]

5.6 Other barriers to searching

Interview data contained some barriers ($n = 10$) related to searching that could not be categorized into keyword searching, filtering, or browsing but were more general by nature. These barriers were in the context of tools and the collection. In the context of tools, barriers were related to possibilities provided by the interface to navigate and edit their search. Interviewees using the images as research data were also helpless in evaluating the completeness of their search results as the interface gave no means for this. Purely technical barriers included the low capacity of the website and the fact that the interface was optimized for desktop use and was not fully functional when used with a tablet or mobile phone. In the context of the collection, sudden changes in the collection caused frustration when images were removed from the database.

I remember it was in that location and I went through all of these. I know where it was, on this last page. And now it is not here, so it has been removed from the gallery. Probably because it was a sensitive topic. So...this is what happened. [P2]

Interviewees also saw the lack of resources for developing the collection as a barrier. Users were able to send error messages and new information about the photographs to the archive, but there were not enough personnel to integrate this information into the collection. Because of the lack of metadata and the large number of images in the collection, searching in general was found time-consuming and difficult.

Often it is very slow to find even one specific image from the big collection. [P9]

6 Discussion

Digitization of image collections has changed the work of academic historians who are using images in their research and writing tasks. Digital images are available for online use, yet their searching is not as straightforward as one might think. Our analysis based on qualitative interview and demonstration data indicates that searching for digitized images requires advanced searching skills, knowledge about the collection and its origins, and about the historical events. Numerous identified barriers to image searching indicate the complexity of digital scholarship. Therefore, it is easy to understand why, for example, Mussell [45] calls

for historically reflexive media literacy skills to understand the influence of digitization on the archival contents.

Interviewed scholars used images for illustration and information, which is similar to findings by McCay-Peet and Toms [40]. However, we did not find any differences in search behavior between the purposes of use. The purposes for use, illustration, and information were overlapping and intertwined, and they were impossible to differentiate in this kind of real-life setting.

Our results show that expert users apply and combine various search tactics, namely keyword searching, filtering, and browsing when searching images for research and writing tasks. The results support the earlier findings showing the importance of keyword searching in image retrieval [32, 38, 41]. However, according to our findings keyword searching alone is an appropriate tactic only when the searcher has specific knowledge about the specific image (s)he was looking for (e.g., image ID). Thus, keyword searching is most often combined with filtering and/or browsing. In addition, our data show that the keywords used in searching do not necessarily represent the information needs directly, but the user's conjecture of the words used in the image captions. Thus, studies focusing only on the keywords used for image retrieval, without any context of searching, do not necessarily reveal the actual information needs of the users.

Filtering is a search tactic that can never be used alone but it is always combined with another tactic. In many studies, filtering is considered as a part of browsing or as a category search [22, 38]. In this study, filtering was analyzed as a separate tactic. However, the ability to filter is always dependent on the functionalities provided by the system and its interface. In the case of this study, users applied mostly temporal filtering. The use of filters based on image contents (color, image type) was not observed in the data as the users were more interested in images on a conceptual level.

Browsing was included in most cases when searching for images, but most often browsing was used in combination with other tactics, as alone it is time-consuming. However, in extreme cases, historians may be willing to browse thousands of images if no other tactic is supported. Earlier studies have shown the varying reasons for browsing, such as serendipity and broadening searches [9, 22, 37, 44]. Matusiak [38] argued that heavy browsing was related also to the lack of computer skills. However, our findings based on interviews and demonstrations with expert users show that browsing was almost always part of the searching process and depended on the information the user already had about the image and the collection, not necessarily on their computer skills. In addition, all images are not available for searching by keywords (in the case there is no metadata) and the only possible tactic to find them is to browse. Also, the interface features may influence users browsing behavior [49].

Historical images that were originally not intended for research purposes, may be difficult to obtain in research. Analysis related to the perceived barriers to image searching showed that there were barriers to all search tactics. However, most of them focused on keyword searching. Most barriers were in the context of collection and tools. Results are in line with the study about using historical newspapers, where searching-related barriers were also mostly in the contexts of collection and tools [31]. Barriers in the context of task and socio-organization are rarely related to searching activity, but more often to planning and reporting activities. Clearly, the main problem with the use of the case archive was the inaccurate metadata or the lack of metadata. Our case archive is unique in the sense that it contains historical photographs with metadata produced during wartime. Although this type of metadata carries historical value and evidence, it causes many difficulties for searching the contents. The image captions work as a key to the collection but its shortcomings (lack of incompleteness, inaccuracy) restricted the use of the collection.

Similar problems are most likely faced by many digitized cultural heritage collections that have limited resources for developing the collection and the services. However, putting efforts into creating expert vocabularies may be disappointing as studies have shown problems with the controlled ontologies for image archives as they may become quickly outdated or do not meet the needs of the users [56]. Studies have even indicated an ontological gap between researchers and information specialists in their perceptions and needs for image metadata [46, 47]. Indeed, data are often reused differently than what was originally intended [7]. User tagging and content-based image retrieval (CBIR) have offered new possibilities for creating metadata and supporting image searching with a lower cost and higher density. However, the study by Beaudoin [4] showed that CBIR was found helpful by users interested in formal characteristics such as color, shape, composition, and texture of the images, while users interested in known-items, themes, or specific locations did not find similar advantages. Archaeologists and art historians taking part in the study by Beaudoin [4] were not interested in CBIR systems because they wanted to rely on the textual retrieval of images. Their finding underlines the importance of user studies in system development, and this necessitates the study of real-life user needs.

When there are sparse or lacking metadata as in the present case, automatic annotation methods to create metadata would be an option to improve the findability of the images [35]. Since users want to search with textual queries the only option is to provide them textual annotations or descriptions. CBIR has advanced lately, and it is already possible to recognize people, objects, events, and landscapes from images. The novel methods allow the recognition of photographic arrangements such as distance between objects, or camera

and objects, or even to recognize the main characters in the images [36, 53]. If the users are willing to use images as search input, using reverse image search can be implemented. If one interesting photograph is found, similar can be easily retrieved with this kind of search tool. Automatic methods have also been used for producing image captions with short descriptions [8].

When images are used as research data, scholars need information, for example, about the aboutness of data, characteristics of data, metadata, and secondary information about data [30]. Therefore, Huvila [26] calls for paradata (data on the processes of its creation, curation, and use) for digitally shared research data. In the case of a historical photograph archive that contains, for example, propaganda materials, paradata is essential for using and analyzing the images. In addition, scholars may need paradata concerning the digitization process of the contents (what was digitized and what was not). Yet, the situation is not ideal as Hansson and Dahlgren [23] argue that research data archives do not support describing images at the item level but provide metadata on the dataset level and often focus on metadata about publications. Important questions for scholars using the images as research data include: how the collection and the captions were created, what mistakes/missing information there is in the data, what type of images there are in the collection, are there gaps in the number of photographs in certain times, and how did the censorship affect the collection?

Digitization has offered new possibilities for historical research and has changed the way research is done. Due to the digitization, archival materials are easily accessed online, and new research paradigms are emerged through the development and application of new (often computational) methods. Research activities such as tool development and data preparation have become new research practices in the digital humanities [21]. Yet, these developments have mainly focused on textual materials. There are already examples of DH projects utilizing computational methods such as computer vision for image analysis [15, 33] and using novel methods have been discussed in the literature [1, 48]. However, future research will show the possibilities of digital and multimodal scholarship for taking advantage of large digital image collections.

This study does not come without any limitations. Because our data collection focused on one specific image collection, our results may not be generalizable beyond similar collections. For example, the content of the collection, the quality of provided metadata, and the search interface influence search behavior and perceived barriers. Therefore, every collection creates a unique user experience. However, qualitative case studies like the one at hand, are necessary for gathering in-depth information about user behavior. Additionally, we analyzed only professional use of the collection as our interviewees searched the images for their research and

writing tasks. It must be noted that image collections are used for various tasks and by various users creating other types of information needs and search behaviors. Thus, future studies should broaden their scope beyond academic users.

7 Conclusions

Scholars in the digital humanities have benefited from the efforts put into the digitization of cultural heritage collections, such as historical photographs. However, providing digitized content openly online is not enough if there are no sufficient means for accessing the content. This study analyzed image search tactics and barriers to searching in research and writing tasks in the context of a historical image archive. Results showed that expert users apply and combine different tactics for image searching and rarely using only one tactic is enough. During searching users face various barriers, most of them focusing on keyword searching. Barriers were mostly in the context of the collection and tools. Especially, the shortcomings of metadata related to the images caused problems in searching. Although image searching has developed during the last years, the developments have not always reached the digitized cultural heritage collections.

The humanities field has not been at the forefront of the use of digital technologies. Nevertheless, the transformation toward using digital methods in the field has been enabled by digital collections. The digital library field has been widely interested in textual data, but image materials need also to be considered since the importance of images has expanded along with, e.g., social media images. Therefore, when developing future research infrastructure and digital libraries for the digital humanities, it is important also to include image data. Further, to overcome the difficulties in accessing the information, identifying the barriers to the use of the collection is necessitated. This yields a better understanding of human information needs and may suggest information system designs beyond data-centric methods. Our paper indicated several practical implications for supporting different image search tactics. The most important points of development relate to improving the metadata by exploiting different ways of including CBIR and collaborative methods such as user-tagging.

Acknowledgements This work was supported by the Intelligent Society Profiling Action funded by Academy of Finland and by Academy of Finland grant number 351247.

Funding Open access funding provided by TampereUniversity (including Tampere University Hospital).

Declarations

Conflict of interest Only adult participants with informed consents were interviewed in the research. Therefore, no ethical review was required according to The Ethics Committee of the Tampere region.

Ethical approval All protocols were conducted according to their guideline: "Ethical review on non-medical research in the field of the human sciences" (<https://www.tuni.fi/en/research/responsible-science-and-research/ethical-reviews-in-human-sciences>).

Informed consent Informed consent was obtained from all participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A Interview guide

1. Background information

- Age/year of birth
- Education
- Academic position/place of work
- Municipality of residence

2. The context of image use

- For what kind of activities and tasks do you use the image archive for?
- Are there any phases in the tasks? Can you describe these phases?
- At what phase specifically do you use these images? [How?]

3. CI: If you think about the last time you used the collection...

- What kind of searches did you make?
- How did you choose the images from the results?
- What did you look at from the images?
- Was it easy for you to find the images you needed? In what way easy/hard?
- Did you save the images or information about the images? Why? How?

4. Information sharing

- Do you share information about the collection with colleagues or others interested in the same topic?
- What kind of information do you share?
- How, in which forum?

5. Suggestions for improvement and comments

- Did you have any difficulties using the image archive? What kind difficulties did you face?
- What do you think about the instructions and the smoothness of the search in the image archive?

6. "Search engine of the perfect world"

- What would be the perfect world searching engine like? What could you do with it?

7. Demonstration

- Think about a specific image or images that you have searched from the collection recently.
- Can you show me how did you find the images?
- What problems did you face when searching?

8. Ending

- How was the interview?
- Can you think of anything else related to this topic?
- Do you want to name some other person you know is using the image collection?

Thank you!

References

1. Azar, M., Cox, G., Impett, L.: Introduction: ways of machine seeing. *AI Soc* **36**, 1093–1104 (2021). <https://doi.org/10.1007/s00146-020-01124-6>
2. Bartalesi, V., Pratelli, N., Lenzi, E.: Linking different scientific digital humanities in digital humanities: the imago case study. *Int. J. Digit. Libr.* **23**, 1–15 (2022). <https://doi.org/10.1007/s00799-022-00331-4>
3. Beaudoin, J.E.: A framework of image use among archaeologists, architects, art historians and artists. *J. Doc.* **70**(1), 119–147 (2014). <https://doi.org/10.1108/JD-12-2012-0157>
4. Beaudoin, J.E.: Content-based image retrieval methods and professional image users. *J. Am. Soc. Inf. Sci.* **67**(2), 350–365 (2016). <https://doi.org/10.1002/asi.23387>
5. Beaudoin, J.E., Brady, J.E.: Finding visual information: a study of image resources used by archaeologists, architects, art historians, and artists. *Art Doc. J. Art Libr. Soc. North Am.* **30**(2), 24–36 (2011)
6. Borgman, C.L.: Designing digital libraries for usability. In: Bishop, A.P., Van House, N.A., Battenfield, B.P. (eds.) *Digital Library Use: Social Practice in Design and Evaluation*, pp. 85–118. MIT Press, Cambridge Mass (2003)
7. Borgman, C.L., Darch, P.T., Sands, A.E., et al.: Knowledge infrastructures in science: data, diversity, and digital libraries. *Int. J. Digit. Libr.* **16**, 207–227 (2015). <https://doi.org/10.1007/s00799-015-0157-z>
8. Cetinic, E.: Towards generating and evaluating iconographic image captions of artworks. *J. Imag.* **7**(8), 123 (2021). <https://doi.org/10.3390/jimaging7080123>
9. Chassanoff, A.M.: Historians' experiences using digitized archival photographs as evidence. *Am. Arch.* **81**(1), 135–164 (2018). <https://doi.org/10.17723/0360-9081-81.1.135>

10. Cho, H., Pham, M.T., Leonard, K.N., et al.: A systematic literature review on image information needs and behaviors. *J. Doc.* **78**(2), 207–227 (2022). <https://doi.org/10.1108/JD-10-2020-0172>
11. Choi, Y., Hsieh-Yee, I.: Finding images in an online public access catalogue: Analysis of user queries, subject headings, and description notes. *Can. J. Inf. Libr. Sci.* **34**(3), 271–295 (2010). <https://doi.org/10.1353/ils.2010.0004>
12. Choi, Y., Rasmussen, E.M.: Searching for images: The analysis of users' queries for image retrieval in american history. *J. Am. Soc. Inform. Sci. Technol.* **54**(6), 498–511 (2003). <https://doi.org/10.1002/asi.10237>
13. Conniss, L., Ashford, J., Graham, M.E.: Information seeking behaviour in image retrieval: Visor i final report. *Art Libr. J.* **27**(2), 46–47 (2002)
14. Conway, P.: Digital transformations and the archival nature of surrogates. *Arch. Sci.* **15**(1), 51–69 (2015). <https://doi.org/10.1007/s10502-014-9219-z>
15. Dee, H.M., Hughes, L.M., Roderick, G.L., et al.: Visual digital humanities: using image data to derive approximate metadata. In: Foster, R. (ed.) *Managing Digital Cultural Objects: Analysis, Discovery and Retrieval*, pp. 89–110. ALA Neal-Schuman, Chicago (2016)
16. Elkins, J.: *Visual Studies: A Skeptical Introduction*. Routledge (2003)
17. Fidel, R.: The image retrieval task: implications for the design and evaluation of image databases. *New Rev. Hypermed. Multimed.* **3**(1), 181–199 (1997). <https://doi.org/10.1080/13614569708914689>
18. Flanagan, J.C.: The critical incident technique. *Psychol. Bull.* **51**(4), 327 (1954). <https://doi.org/10.1037/h0061470>
19. Frost, C.O., Taylor, B., Noakes, A., et al.: Browse and search patterns in a digital image database. *Inf. Retr.* **1**(4), 287–313 (2000). <https://doi.org/10.1023/A:1009979200555>
20. Geddes, K.O., Czapor, S.R., Labahn, G.: *The shallows: How the internet is changing the way we think, read and remember*. Atlantic Books Ltd (2010)
21. Given, L.M., Willson, R.: Information technology and the humanities scholar: Documenting digital research practices. *J. Am. Soc. Inf. Sci.* **69**(6), 807–819 (2018). <https://doi.org/10.1002/asi.24008>
22. Göker, A., Butterworth, R., MacFarlane, A., et al.: Expeditions through image jungles: the commercial use of image libraries in an online environment. *J. Doc.* **72**(1), 5–23 (2016). <https://doi.org/10.1108/JD-01-2014-0019>
23. Hansson, K., Näslund Dahlgren, A.: Open research data repositories: Practices, norms, and metadata for sharing images. *J. Am. Soc. Inf. Sci.* **73**(2), 303–316 (2022). <https://doi.org/10.1002/asi.24571>
24. Hastings, S.K.: Evaluation of image retrieval systems: Role of user feedback. *Libr. Trends* **48**(2), 438–452 (1999)
25. Huang, K., Kelly, D.: The daily image information needs and seeking behavior of Chinese undergraduate students. *Coll. Res. Libr.* **74**(3), 243–261 (2013). <https://doi.org/10.5860/crl-331>
26. Huvila, I.: Improving the usefulness of research data with better Paradata. *Open Inform. Sci.* **6**(1), 28–48 (2022). <https://doi.org/10.1515/opsis-2022-0129>
27. Jordanova, L.J.: *The Look of the Past: Visual and Material Evidence in Historical Practice*. Cambridge University Press (2012)
28. Jörgensen, C.: Attributes of images in describing tasks. *Inform. Process. Manag.* **34**(2–3), 161–174 (1998). [https://doi.org/10.1016/S0306-4573\(97\)00077-0](https://doi.org/10.1016/S0306-4573(97)00077-0)
29. Jörgensen, C., Jörgensen, P.: Image querying by image professionals. *J. Am. Soc. Inform. Sci. Technol.* **56**(12), 1346–1359 (2005). <https://doi.org/10.1002/asi.20229>
30. Korkeamäki, L., Keskitalo, H., Kumulainen, S.: Task information types related to data gathering in media studies. *J. Doc.* **78**(7), 528–545 (2022). <https://doi.org/10.1108/JD-04-2022-0082>
31. Kumpulainen, S., Late, E.: Struggling with digitized historical newspapers: contextual barriers to information interaction in history research activities. *J. Am. Soc. Inf. Sci.* **73**(7), 1012–1024 (2022). <https://doi.org/10.1002/asi.24608>
32. Kumpulainen, S., Ruotsalainen, H.: Searching wartime photograph archive for serious leisure purposes. In: *Linking Theory and Practice of Digital Libraries: 26th International Conference on Theory and Practice of Digital Libraries, TPDL 2022, Padua, Italy, September 20–23, 2022, Proceedings*. Springer, pp. 81–92, (2022) https://doi.org/10.1007/978-3-031-16802-4_7
33. Lang, S., Ommer, B.: Transforming Information Into Knowledge: How Computational Methods Reshape Art History. *DHQ Digital Human. Quarterly.* **1**;15(3) (2021)
34. Late, E., Kumpulainen, S.: Interacting with digitised historical newspapers: understanding the use of digital surrogates as primary sources. *J. Doc.* **78**(7), 106–124 (2022). <https://doi.org/10.1108/JD-04-2021-0078>
35. Late, E., Ruotsalainen, H., Kumpulainen, S.: In a perfect world: exploring the desires and realities for digitized historical image archives. *Proceed. Assoc. Inform. Sci. Technol.* **60**(1), 244–254 (2023). <https://doi.org/10.1002/prat.2.785>
36. Late, E., Ruotsalainen, H., Seker, M., et al.: From textual to visual image searching: User experience of advanced image search tool. In: *booktitle=International Conference on Theory and Practice of Digital Libraries*. Springer, pp. 277–283, (2023) https://doi.org/10.1007/978-3-031-43849-3_24
37. Markkula, M., Sormunen, E.: End-user searching challenges indexing practices in the digital newspaper photo archive. *Inf. Retr.* **1**(4), 259–285 (2000). <https://doi.org/10.1023/A:1009995816485>
38. Matusiak, K.K.: Information seeking behavior in digital image collections: A cognitive approach. *J. Acad. Librariansh.* **32**(5), 479–488 (2006). <https://doi.org/10.1016/j.acalib.2006.05.009>
39. Matusiak, K.K.: Studying information behavior of image users: An overview of research methodology in list literature, 2004–2015. *Library Inform. Sci. Res.* **39**(1), 53–60 (2017). <https://doi.org/10.1016/j.lisr.2017.01.008>
40. McCay-Peet, L., Toms, E.: Image use within the work task model: images as information and illustration. *J. Am. Soc. Inform. Sci. Technol.* **60**(12), 2416–2429 (2009). <https://doi.org/10.1002/asi.21202>
41. Ménard, E., Khashman, N.: Image retrieval behaviours: users are leading the way to a new bilingual search interface. *Library Hi Tech.* **32**(1), 50–68 (2014). <https://doi.org/10.1016/j.acalib.2006.05.009>
42. Mitchell, W.T.: *Iconology: Image, Text, Ideology*. University of Chicago Press (2013)
43. Mizruchi, S.L.: *Libraries and archives in the digital age*. Palgrave Macmillan (2020). <https://doi.org/10.1007/978-3-030-33373-7>
44. Münster, S., Kamposiori, C., Friedrichs, K., et al.: Image libraries and their scholarly use in the field of art and architectural history. *Int. J. Digit. Libr.* **19**(4), 367–383 (2018). <https://doi.org/10.1007/s00799-018-0250-1>
45. Mussell, J.: *The Nineteenth-Century Press in the Digital Age*. Palgrave Macmillan, Basingstoke, Hampshire (2012)
46. Näslund Dahlgren, A.: Image metadata. from information management to interpretative practice. *Museum Manag. Curator.* (2022). <https://doi.org/10.1080/09647775.2022.2073562>
47. Näslund Dahlgren, A., Hansson, K.: What an image is: the ontological gap between researchers and information specialists. *Art Doc. J. Art Libr. Soc. North Am.* **40**(1), 21–32 (2021)
48. Offert, F., Bell, B.: Perceptual bias and technical metaphors: critical machine vision as a humanities challenge. *AI Soc* **36**, 1133–1144 (2021). <https://doi.org/10.1007/s00146-020-01058-z>
49. Petrelli, D., Clough, P.: Analysing user's queries for cross-language image retrieval from digital library collections. *Elect.*

- tron. Libr. **30**(2), 197–219 (2012). <https://doi.org/10.1108/02640471211221331>
50. Pu, H.T.: An analysis of failed queries for web image retrieval. *Int. J. Digit. Libr.* **34**(3), 275–289 (2008). <https://doi.org/10.1177/016555150708414>
 51. Roberts, H.E.: A picture is worth a thousand words: Art indexing in electronic databases. *J. Am. Soc. Inform. Sci. Technol.* **52**(11), 911–916 (2001). <https://doi.org/10.1002/asi.1145>
 52. Savolainen, R.: Approaches to socio-cultural barriers to information seeking. *Library Inform. Sci. Res.* **38**(1), 52–59 (2016). <https://doi.org/10.1016/j.lisr.2016.01.007>
 53. Seker, M., Männistö, A., Iosifidis, A., et al.: Automatic main character recognition for photographic studies. In: 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP). IEEE, pp. 1–6, (2021) <https://doi.org/10.1109/MMSP53017.2021.9733624>
 54. Sinn, D., Soares, N.: Historians' use of digital archival collections: the web, historical scholarship, and archival research. *J. Am. Soc. Inf. Sci.* **65**(9), 1794–1809 (2014). <https://doi.org/10.1002/asi.23091>
 55. Strauss, A., Corbin, J.M.: *Grounded Theory in Practice*. Sage, Thousand Oaks (1997)
 56. Stvilia, B., Jörgensen, C.: Member activities and quality of tags in a collection of historical photographs in flickr. *J. Am. Soc. Inform. Sci. Technol.* **61**(12), 2477–2489 (2010). <https://doi.org/10.1002/asi.21432>
 57. Toms, E.G., O'Brien, H.L.: Understanding the information and communication technology needs of the e-humanist. *J. Doc.* **64**(1), 102–130 (2008). <https://doi.org/10.1108/00220410810844178>
 58. Westman, S.: Image users' needs and searching behaviour. In: Goker, A., Davies, J. (eds.) *Information Retrieval: Searching in the 21st Century*, pp. 85–118. John Wiley & Sons, Chichester (2009)
 59. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., et al.: The fair guiding principles for scientific data management and stewardship. *Sci. Data* **3**(1), 1–9 (2016). <https://doi.org/10.1038/sdata.2016.18>
 60. Woll, J.: User access to digital image collections of cultural heritage materials: The thesaurus as pass-keyquasimonotonicity, regularity and duality for nonlinear systems of. *Art Doc. J. Art Libraries Soc. North Am.* **24**(2), 19–28 (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.