

User-centred collection metadata:

From images as information to facilitating socio-affective connections

Cassandra Kist^{a*} and Maria Economou^b

^a*Information Studies, University of Glasgow, Glasgow, UK*

^b*Information Studies & The Hunterian, University of Glasgow, Glasgow, UK*

* Cassandra.Kist@strath.ac.uk

Cassandra Kist

<https://orcid.org/0000-0001-9960-2236>

<https://www.linkedin.com/in/cassandra-kist-1a695220a/>

Twitter: @kist_c

Dr Cassandra Kist was a Research Assistant in Information Studies at the University of Glasgow while undertaking this research. She is currently a Chancellor's Fellow at the University of Strathclyde in Computer and Information Sciences. Her research combines several disciplines including Anthropology, Museum Studies, and Science and Technology Studies, to investigate the overlaps and disconnections between engagement with cultural heritage, social inclusion, and digital infrastructures.

Maria Economou

<https://orcid.org/0000-0003-1797-9262>

<https://www.linkedin.com/in/maria-economou-0895b9/>

Twitter: @MariaEcoGl

Maria Economou is Professor of Digital Cultural Heritage at the University of Glasgow, a joint post shared between Information Studies (College of Arts) and The Hunterian, the university's museum and gallery service. She is Vice-President of the UNIVERSEUM Board, the European Academic Heritage Network (2017-2024) and Co-Director of the Digital Cultural Heritage Arts Lab at the University of

Glasgow. She holds a British Academy/Wolfson Research Professorship (2022-2025) investigating emotional engagement with museum collections.

User-centred collection metadata:

From images as information to facilitating socio-affective connections

Abstract

Purpose

As museums and other memory institutions continue to invest considerably in mass-digitising collections and participating in large search portals, it is essential to understand existing and potential users, their motivations, and search needs to inform collections' documentation. In this article, we discuss insights from a collaborative project with National Museums Scotland, set up to enhance the findability of collection images and inform documentation practices by understanding the collections users and their search terms.

Methodology

The research involved interviews with National Museums Scotland staff, users and non-users of the Museums' Search our Collections portal encompassing a concept mapping and card sort activity; online surveys; and content analysis of user search queries.

Findings

The project revealed that participants are interested in searching the online collections by terms often not represented in collections metadata, including terms related to identity (their own but also others') and social context (e.g. through seasonal and social events); emotional and sensory interests (e.g. visual characteristics); and narrative themes (e.g. on under-represented histories).

Originality/value

Based on the findings, we further theorise the semantic gap in online museum collection metadata. To bridge this gap and cater to how users search, we argue for a paradigm shift in documentation practices, suggesting practitioners should not only view collection images as information but also tap into their rich potential for enabling social and affective connections.

Keywords: collection images; museum documentation practices; semantic gap; user research; cultural heritage collections; online catalogues; image tagging; search portals; object metadata; National Museums Scotland

Introduction

While museum exhibitions have, over the last few decades, shifted away from taxonomic categorizations to storytelling centred around themes and participatory opportunities, the way online collections are displayed and can be interacted with has lagged (Gil-Fuentetaja and Economou, 2019). Some practitioners have criticised the fact that collections online still represent catalogue/database card files, suggesting that museums must relinquish this format to embrace 3D, interactivity, and more immersive/emotional experiences that support users' meaning making (Newbury and Brennan, 2021). Cultural heritage institutions have struggled to cater to audiences of online collections and tend to have less confidence in the quality of collection pages compared to other digital museum offerings (One Further, 2022). This lack of confidence may be underpinned by uncertainty regarding *who* collection webpages are intended *for* (Bailey-Ross, 2021; Newbury and Brennan, 2021); a conflict between the quantity of collections online and the quality of digitisations (especially with stretched staff time) (Martin, 2021); and the ambiguity of the resulting use/value for different publics (Hopes, 2014).

A key challenge regarding the use and access of online collections is the 'semantic gap', which is frequently highlighted in studies of museum search portals. The 'semantic gap' refers to the disconnect created by expert terminologies, values, and practices of museum staff applying metadata¹ which might serve institutional needs but not necessarily those of end users outside the museum (Dahlgren, 2022; Klavans et al., 2014). This is a central challenge addressed throughout

¹ In this paper, metadata refers to the data about collection objects and associated images.

our project with National Museums Scotland (the Museums) on which this paper is based. We report here on some of the findings of our collaborative project which aimed to illuminate the types of terms online audiences use to search and use this understanding to inform documentation practices, improving access and use of collection images.

Through this research, we reveal some of the terms online audiences are interested in searching the collections by, which suggest that the online collections harbour potential for enabling social and affective forms of engagement. However, this seems to be untapped primarily because of a disconnect between these user terms and the metadata and information management practices of staff in cultural heritage institutions. The latter predominantly focus on images as information, which refers to documentation practices that strive for objectivity, accurateness, specificity, and efficiency (Dahlgren, 2022). In contrast, our findings suggest that a perception of collection images as not only information, but as enabling users' emotional and social connections is essential. Building on these findings, we argue that cultural organisations should widen their documentation practices to include sensory and affective terms, those related to social identity and cultural contexts, and broader themes that can support narrative threads. This shift in documentation practices would support the increasing efforts of cultural heritage institutions to sustain user engagement and participation with digitised collections in ways that are socially and emotionally meaningful.

Addressing the semantic gap with National Museums Scotland

During this project (2022-23), National Museums Scotland were undergoing significant changes aiming to increase both their media production and access to online collections through images and 3D digitisations. Previous research, reported in Chowdhury *et al.* (2022), revealed limited access to the online collections through the Museums' main collections portal 'Search Our

Collections (SoC): less than 2% of online collections were accessed by anyone; online collections users tended to be new and not returning; and there was a gap between the terms applied by staff to the collection metadata and those used by users to search the collections. In response, the Museums have been undertaking research, including the project we report on here, but also a staff-led project on investigating the usability of different collection search portals.

In early discussions with the Museums' Collections Data and Digitisation Team, there was a desire to improve the collections' findability and searchability by adopting a user-centred approach and applying metadata to collection images in the form of keywords and captions. Very little was known about users of the Museums' collection webpages and the types of terms they may use to search them. To complement the ongoing changes to the institution and previous research (Chowdhury *et al.*, 2022), our project aimed to make a significant first step in creating an in-depth understanding of the types of terms audiences may use to search the online collections and make recommendations on how image metadata can best support these.

Literature Review

The semantic gap and metadata practices

User-centred metadata has continuously been a challenge in museum and wider cultural heritage practices, as evidenced by the semantic gap. The semantic gap has been perceived and theorised in multiple ways: For instance, as differences between discipline-specific/'expert' terminologies applied to collection metadata and the terms used by the general public to search. The semantic gap may also be created by the values and associated practices of practitioners applying metadata which can be in opposition to the interests and needs of users. For instance, taking on a communication model perspective, Dahlgren (2022) shows how practitioners' focus on the

informational aspects of collection images shapes the terms they apply to collection metadata, striving for static, objective, accurate information without valuing their interpretive, dynamic, and subjective nature.

Research suggests that user motivations for searching online collections extend beyond the informational focus of museum practitioners, and hint at an interest in the social and affective connections that collections harbour. For instance, Bailey-Ross (2021, p. 2), drawing on a wide range of previous studies, e.g., from the Metropolitan Museum of Art (Villaespesa, 2019), Europeana (Clough et al., 2017), and National Museums Liverpool (Walsh, et al. 2020), summarised diverse user motivations for visiting online collections as: casual use (browsing, which may also include inspiration/creativity); personal interest (searching for specific information); and scholarly and professional research (highly motivated users looking for specific information). The category of ‘casual use’ in particular, which encompasses creativity and inspiration may be closely tied to emotional and aesthetic motivations for searching images of art (Villaespesa and Stack, 2015). Given the potential significance of online collections for the emotional interest/needs of users, Bailey-Ross (2021) further suggests there is an increasing need to also understand the role of digital collections for health and wellbeing post-COVID-19. Park (2021; 2023) also revealed that the public are interested in understanding the rich connections of collections with other media and social context – information that is typically not supported or included in the museum’s metadata.

As others have pointed out, online engagement with collections can also be closely tied to the formation of identity (Dallas, 2018) and connections between groups/individuals, social contexts, and place (Kuntsman, 2017; Budge and Burness, 2018). As encapsulated by Meecham (2013, p. 51) discussing digital replicas of art, collections ‘contribute to our sense of self, and

their very familiarity across forms gives us a sense of place and authenticity'. Yet, despite this understanding of searching and engaging with online collections as subjective and closely connected to identity, emotions, and social contexts, institutions continue to apply collection metadata that prioritises images as information, catering towards institutional needs as opposed to the interests/needs of users (Martin, 2021). This is partly related to how legacies of technological systems for collections management (originally built for internal management and audit purposes in the 1960s-70s) continue to shape documentation practices today (Chapman, 2015). Collections management systems, as closed/static models, shape how and what information can be documented and can hinder the representation of the social dimensions of collections (Park, 2021; 2023).

Underpinning image description in several metadata systems today is the visual analysis research pioneered in the 1970s, epitomised by Panofsky's framework. We used and expanded this framework to pinpoint the challenges documentation professionals encounter in describing images and identify categories of terms they mainly use, as well as the ones they leave untapped, thus, contributing to the semantic gap. In his framework for analysing Renaissance artworks, Panofsky (1972) suggested three levels: 1) generic description - 'pre-iconographic description'; 2) identification of specifics – 'iconographic identification', requiring knowledge about art; and finally, 3) 'iconology', an interpretive level that is associated with abstract concepts and interpretation. Shatford (1986; 2002) built upon this model for describing all types of images, suggesting the incorporation of the 4Ws: who, what, where, and when across Panofsky's three categories. However, cultural heritage institutions' informational focus may cause staff to overlook or indeed, avoid the categories of terms related to abstract concepts (iconology) and

generic (pre-iconographic) descriptions, favouring technical and specialised terms for collection images (iconography).

One of the reasons for this oversight is that terms related to the iconology stage can be challenging to apply. Shatford (1986, p. 45) suggests that ‘[t]his level [iconology] of meaning cannot be indexed with any degree of consistency’, and indeed, Panofsky (1972) acknowledges the difficulties of applying this level of interpretation to works of art. For this reason, Shatford (1986) suggests iconographic (technical and specialist) terms may, therefore, be prioritised. This is a common documentation practice still followed today despite evidence that when searching the collections, users often use terms that differ from those of cultural heritage professionals (Martin, 2021; Dahlgren, 2022).

Klavans, LaPlante, and Golbeck (2014) expanded a version of the Panofsky-Shatford framework² to analyse crowdsourced tags of collection images, adding ‘visual elements’ encompassing colours or shapes, and an additional ‘other’ category for terms such as type of image/shot. We adapted this framework (Table 2) to analyse the types of terms users used to search the National Museums Scotland’s Search our Collections (SoC) webpage (from a 2020-21 dataset) and to tag collection images in response to a series of survey questions described in Methods. In turn, we further the sector’s understanding of the types of terms users are interested in searching the collections by, and particularly, clarifying if/how they fall into the lesser used image description categories: ‘pre-iconographic’ (general terms) or ‘iconographic’ (interpretive terms), raising implications for future documentation practices, which we explore in the Discussion.

² Reworked by Armitage and Enser (1997).

Despite recent work on increasing accessibility of online collections (e.g., generous interfaces (Hall and Walsh, 2021) or potentials of AI to automate tagging (Villaespesa and Crider, 2021)), the gap between staff metadata practices and how users search persists. In order to take advantage of these recent research and technological developments, there is a paramount need to first understand the types of terms users search by. Our research serves as a significant step in this direction and an impetus for broadening cultural heritage documentation practices from an information focus to also considering collection images' potential for enabling social and affective connections.

Methods

For this project we employed a mix of qualitative and quantitative methods, including semi-structured interviews with National Museums Scotland's staff, deductive content analysis of user search queries from the Museums' Search our Collections webpages, public surveys, and interviews with 10 participants involving a card sort and concept mapping activity. The goals of using each method differed slightly (as outlined in Table 1) but all supported and were triangulated to develop a thick understanding of who uses the online collections, the types of terms and themes users/potential users use to search these, and the implications of these findings for applying metadata in the form of keywords and captions to collection images.

Table 1. Relationship between methods and research goals

	Goal 1 Types of terms	Goal 2 Who is searching the collections	Goal 3 Themes/categories
Method 1 Content analysis of Search our Collections queries	Understand the types of terms searched (general/specific/abstract – who, what, where, when, or visual).		Understand which collection categories were being searched.
Method 2 Surveys of users/non-users	Understand types of terms users may search the collections by (general/specific/abstract – who, what, where, when, or visual).	Understand who is searching the collections.	Understand what themes participants are interested in searching the collections by.
Method 3a Interviews: Concept map activity	Identify types of terms related to underrepresented histories that users would search by.		To explore what ‘underrepresented histories’ means to participants (a collections theme popular in the survey results).
Method 3b Interviews: Card sort activity	Identify terms associated with broader collection categories.		Understand how participants might group together collections into broader categories /themes that go beyond current collection categories/departments.

The research methods are broken down in further detail below and also in the project report (Kist *et al.*, 2023). This paper draws from the findings of the user research, while a forthcoming paper will discuss museum staff perspectives; therefore, staff interviews are not covered in detail here. The project received ethical approval by the University of Glasgow Arts Committee.³

Content analysis of users' queries on the Search our Collections portal

To understand how users search the collections, we investigated the types of terms used to search the Search our Collections (SoC) webpage, and the types of collections being searched. To do this, we drew on an anonymised dataset generated from a previous project with the Museums (Chowdhury *et al.*, 2022). This dataset was created by the Museums' staff by extracting the search queries from the SoC portal and documenting their frequency from 2019-2020 and 2020-2021. Due to time limitations, we analysed a sample from the large dataset of user queries from 2020-2021.

Overview of the SoC users' query dataset

- Total search queries including repetitions for 2020-2021: 570,990
- Total unique search queries in the dataset for 2020-2021: 70,874
- Total queries in our sample including repetitions: 168,970
- Total unique queries in our sample: 1,779

The sample encompassed search queries with:

³ Ethics Application 100220077

- Search frequencies from 90 - 1909 (search queries with the highest repetitions)
- Search frequencies from 15 - 30 (search queries with lower repetitions)

We applied a deductive content analysis to this dataset based on the framework reworked by Klavans, LaPlante, and Golbeck (2014) for defining level of detail and describing images. We adapted and extended this framework to suit the specific context of the Museums, adding our own examples, referred to from here as the image description framework (Table 2). We analysed the sample for the most popular or most frequently searched terms, and subsequently, less frequently searched or popular search terms. Like Klavans, LaPlante, and Golbeck (2014), we applied only one perceived main code or category to each user query.

Table 2. Image description framework based on Klavans, LaPlante, and Golbeck (2014); Armitage and Enser (1997); Shatford (1986); Panofsky (1972) with our changes/examples in blue and justification in italics

	(G) Generic (Pre-iconography)	(S) Specific (Iconography)	(A) Abstract (Iconology)
1. Who	G1: Kind of person or thing (<i>maid, ring, sword</i>)	S1: Individually named person, group (<i>culture</i>), thing (<i>Lewis Chess pieces</i>), specific materials (<i>alabaster</i>), species names (<i>Lynx pardinus</i>), complicated/technical terms (<i>daguerreotype</i>) <i>Named objects and individual people were uncommon in the SoC queries, but it was useful to use this 'specifics/detailed' category to place species' names, technical or complicated terms, culturally specific terms for objects, and very specific materials⁴</i>	A1: Mythical or fictitious beings (<i>unicorn, mermaid</i>), religious/spiritual and magical figures (<i>god, goddess</i>), and objects (<i>amulet, votive</i>) <i>Spiritual objects and beings were grouped together, as they could be connected to more abstract ideas</i>
2. What	G2: Kind of event, action, condition (<i>plague, knitting</i>)	S2: Individually named event or action (<i>WW1</i>)	A2: Emotion or abstraction, <i>values, themes/topics (grief, fortune)</i>
3. Where	G3: Kind of place: geographical/architectural (<i>space, croft</i>)	S3: Individually named geographic locations – city, country, or place (<i>Rome, Hawaii</i>), exhibition title (<i>Fashion and Style</i>)	A3: Place symbolised (<i>Mount Olympus</i>)
4. When	G4: Cyclical time & <i>broad time periods</i> : season, <i>holidays</i> , time of day (<i>Renaissance, winter</i> ,	S4: Linear date, time period or <i>numerical dates (1990)</i>	A4: Emotion or abstraction symbolised by time (<i>spring for</i>

⁴ A limit to this coding is our own knowledge and bias regarding technical language but we tried to follow the Museums' staff learning accessibility requirements as understandable for 7-to-8-year-olds.


	Halloween) <i>We grouped words for holidays, seasons, and general time periods (e.g., medieval) together here - as they were perceived by staff to be general and ambiguous. Therefore, we placed only specific numeric dates/date ranges in S4</i>		youth)
Visual elements	V: Shapes, size, age, design, style, and colours (colourful, miniature)		
Other	O: Items that do not fit into the above categories (type of image/shot)		

User/non-user surveys

In order to also understand who is using the collections, the type of terms online audiences may use to search the collections by, and gauge interest in collection themes, we distributed two surveys for a period of around 7 weeks (3 April – 26 May 2023): one through social media/our project team’s networks and one as a pop-up on the Museums’ SoC portal. Both surveys were similar (so only the one distributed over social media is included in Supplementary Material 1), with the exception that the latter allowed people who had never used the collections portal to participate, aiming to explore their interests in cultural heritage/collection images. Apart from understanding who is using the online collections, the key focus of both surveys was the themes and types of terms users are interested in searching the collections by, investigated through an image keywording activity. This asked users to suggest searchable terms for four collection images, that could encapsulate what each image was of and about (Supplementary Material 1, p. 12-21; for an example of one, see Figure 1), which were subsequently analysed using the image description framework (Table 2).

Section 4: Labelling activity

8. Help us make the following images searchable. Please attribute the keywords you may expect to search the following images by.



Choose some key words that describe what the image is of and about:
(open-ended)

Figure 1. One of four image keywording activities included in the online survey

Online surveys can pose limitations (Nayak and Narayan, 2019)—for us, a particular challenge was the equality and diversity of participants rather than the size of the sample. 132 surveys were filled out, with 81 having used the Museums’ collections and 51 not having used them. Interestingly, there was no significant difference in the types of terms mentioned by users and non-users of the SoC, so the findings reported below refer to the total of 132. Surveys exploring questions qualitatively or through mixed methods tend to have a sample size ranging from 20–49 (low range) or 60–99 (mid-range), to over 100 participants (upper range), where our study falls (Braun et al., 2021, p. 649). The survey sample enabled us to qualitatively explore the types of terms participants might search the collections by, complementing the analysis of the large sample of SoC users’ queries.

However, the diversity of our survey participants was limited as the majority identified as female and from Scotland. Moreover, the most common motivations for visiting the collections were for work (cultural heritage profession), school, and lastly, personal interests (e.g., for

writing historical fiction). This could be a symptom of the survey distribution through our own networks and the Museums' social media followers, and/or in-part, the profile of users using the SoC portal. This bias was taken into account in our analysis, and it would be useful for future research in the same direction to include broader participation of a more general public.

However, this does not undermine the usefulness of the findings, as these survey participants, including archaeologists, archivists, museum professionals and students/researchers in related disciplines are a core user group of the online collections.

Moreover, in opposition to what we would expect of such groups in terms of 'expert' terminologies, it was interesting to observe that general terms (pre-iconography) were prevalent across survey respondents with different motivations for visiting the collections. This suggests that the collections terms currently drawn on by the SoC API are failing both users/potential users with expertise in cultural heritage subjects, and a more general curious public (those who identified as 'curious individuals' or 'cultural heritage/art enthusiasts' in their responses).

Due to the composition and size of the survey sample, the findings are not representative of a general population of online museum collection users. Rather, the survey findings provided an important first step in understanding the current and potential users of the Museums' online collections and how collection image metadata can best meet their needs.

User/non-user interviews with concept map and card sort activity

Following the surveys, we undertook 10 interviews with mainly survey participants (curious individuals/cultural heritage enthusiasts but also 'professionals' as these are a core user group of the online collections) and a few participants from our networks, encompassing:

- 3 museum professionals

- 1 librarian/archivist who is also a fiction author
- 1 PhD student (in Information Studies)
- 1 post-doctoral researcher (with a background in Information Studies)
- 4 cultural ‘heritage/art enthusiasts’ and ‘curious individuals’

The interviews, as one part of three methods, were intended to contribute to a ‘deeper dive’, complementing the user surveys and content analysis of search queries. Ten interviews were a plausible number to explore topics in-depth within the limited time frame of the project and despite the small size, enabled rich insights. Within this small sample, we were able to represent different levels of abilities/experiences including students, professionals, and members of the public.

In answer to the multiple-choice survey question about the theme(s) that participants would be most interested in when exploring images of the Museums’ collections (Supplementary Material 1, p. 7), the most popular was ‘under-represented histories’. Based on this finding, we carried out interviews encompassing two parts: a concept map and card sort activity, during which participants used Miro⁵ or pen and paper and verbally walked the interviewer through their thought process.

During the concept map activity, taking inspiration from cognitive mapping (Priestner, 2018), interviewees were invited to explore what ‘under-represented histories’ meant to them, and what themes and keywords they might connect these to (Figure 2, 3, 4). This is a method commonly used to understand users’ thoughts, experiences, and preferences (Priestner, 2018) and enable access to their world (Gibbons, 2019). Participants were encouraged to switch either

⁵ Miro is an online visual collaborative board (<https://miro.com>).

sticky note or line colour every two minutes to enable easier observations of what concepts/terms came to mind first.

Then, the interviews used a card sort activity to further understand how and why users might group together collections into broader categories (Rainbow *et al.*, 2012) (Figure 5, 6, 7). Participants were invited to sort 16 collection images (included with their description in Supplementary Material 2) from the eight most popular collection categories identified in the SoC queries analysis.

A limitation of both methods was the use of Miro which was challenging for some participants to use. Although in our case, both methods provided insight into how participants categorised collections and explored different themes, primarily ‘under-represented histories’, this may not necessarily reflect how they would actually search (Rainbow *et al.*, 2012).

Findings

Identity and social context

The analysis of both the sample of Search Our Collections (SoC) queries and survey responses to the keywording activity questions showed that collection images are not only seen as information by users. Rather, they also represent possibilities for social connectivity, in terms of how they can be associated with current social events, holidays, and even personal experiences. The terms used on the SoC portal and applied by survey participants to collection images suggest an interest in how the images relate to social context and identity. The terms observed fall into the pre-iconography/general category (G1 ‘who’ & G4 ‘when’) and the iconology/abstract category related to spirituality (A1) of the image description framework (Table 2).

For instance, in an image keywording activity which included a mariachi figure as part of the survey (Figure 1), participants applied keywords such as ‘spring’, ‘Halloween’ or ‘Day of the Dead’. Similarly, in our analysis of SoC queries, while infrequent, users used search terms related to seasonal events such as ‘snow’ and ‘Christmas’, contemporary events such as ‘F1’ (Formula 1), and events that may be relevant to individual lives, including ‘wedding’ and ‘funeral’. These findings are aligned with the previous research project at the Museums (Chowdhury *et al.*, 2022, p. 33) which showed that collection searches increase over certain seasons – with searches related to seasonal holidays. Moreover, although further comparisons with other years would be revealing, in our SoC queries sample there were several searches of spiritual and death-related objects, including ‘death’ and ‘plague’, which could be related to the sample’s occurrence over the COVID-19 pandemic. Spiritual objects included votives, charms, caskets, coffins, urns, and others.

Beyond this entanglement with social context, in the participant interviews, respondents mapped out under-represented histories and sorted images of the collections in ways that were shaped by their personal interests and how they related collections to their own identities and those of other people. Firstly, when asked to mind map what under-represented histories means to them, most interview participants first noted down aspects of identity associated with under-representation (e.g., socio-economic class, gender, sexuality, and race). Secondly, they also connected these aspects of identity to histories and experiences that were of interest to them personally. For instance, interviewee Anne⁶ - a retired medical professional was interested in both medicine and rewilding which impacted the under-represented histories they noted down in their concept map (Figure 2):

⁶ Pseudonyms have been used for interviewee participants.

I was a doctor before I retired, so I am interested in women in medicine. And actually, Edinburgh is quite famous; there were riots because five women applied to Edinburgh Medical School and came out in top on the exams, and the men rioted because they didn't think women should come. It was Sophia Jacks Blake – and that's actually quite a nice story that could be represented there or somewhere about what women have done.

On rewilding and socio-economic class, they noted:

If it's only ever done by people who have thousands of acres, what relevance does it have to some kid, living in some grotty council estate in Glasgow? If they can see, they can plant some tree here or some bushes here, it can then relate to them, and they value it as well.

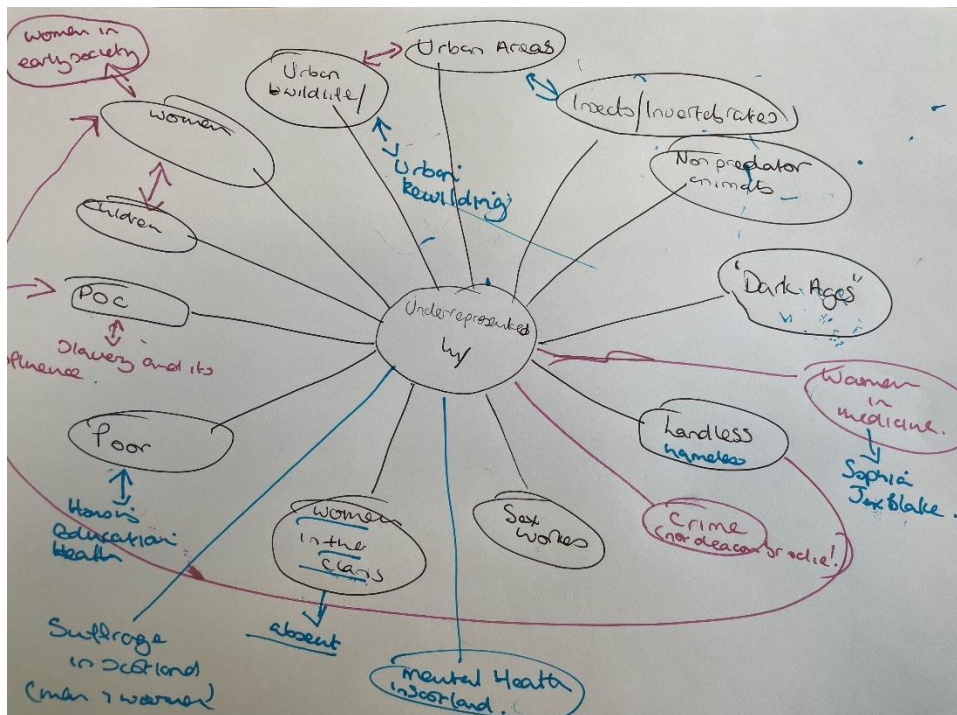


Figure 2. Anne's concept map

These interests and facets of her own identity and those of others came to light and shaped the under-represented histories that came to this participant's mind. Similarly, interviewee Samantha who works with jewellery was interested in small things that may go unnoticed, referencing the many layered histories that even the smallest of objects can hold. In particular, they reflected on

the stories such objects could tell about women (Figure 3):

I was thinking about under-represented things, so I was thinking a lot about women's work and the small objects that are maybe kind of lost or how some objects can tell us about how women in the past would have functioned in the home as well – like how homes are created. ... I think a lot of things I was thinking of probably come from my background because I deal with jewellery and metal work, beading, jewellery that's made from whatever is on hand.

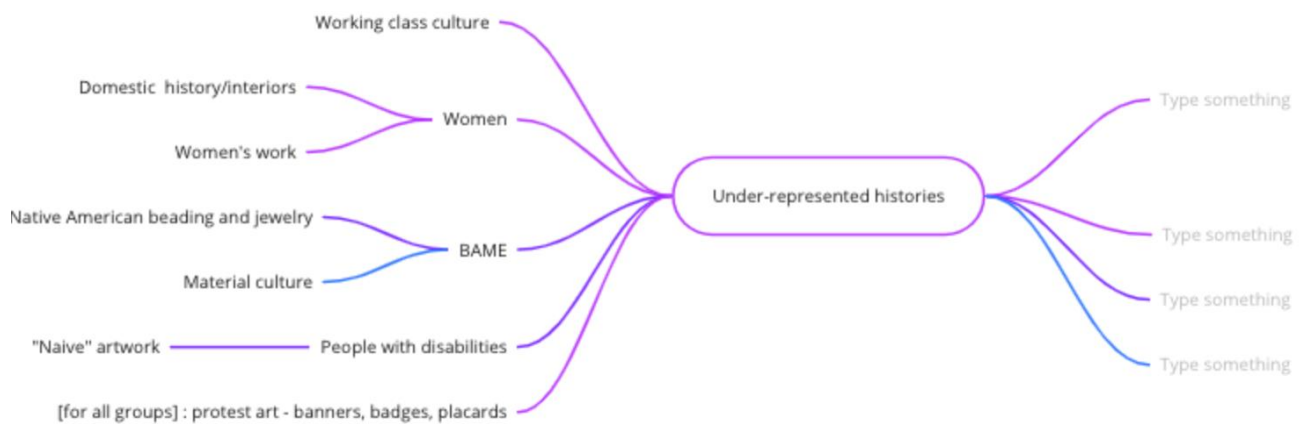


Figure 3. Samantha's concept map

Moreover, an interest in objects and their connection to identity was also suggested in the SoC queries through searches involving a gender or role. For example, users often searched for terms like women, woman, slave, maiden, military, and modified objects with identity factors such as belonging to a baby, actor, chief, or a man. Such instances expose the interconnections participants made between collection objects, personal interests, and facets of identity, including those connected to under-representation but also the identities of participants through their personal interests and experiences.

Sensory exploration and affective connections

As reflected in both the surveys and analysis of user search queries, the majority of terms observed could be described as falling into the generic/pre-iconography categories of the image description framework (Table 2). This includes terms that have significance for social context and identity as discussed above, but also terms that have significance for sensory and affective connections, including terms that not only fall into the ‘who’ category (G1) but also the visual category (V) (such as colour or pattern).

For instance, participants used terms related to colours (e.g., colourful, multi-coloured, pink, black), patterns (e.g., paisley, tartan), or something generically represented - a type of object or thing (e.g., ring, clothing, vase). As opposed to searches for very specific objects and using specialised language (e.g., Lewis Chess piece – ‘S1’ from the iconography column of the image description framework), these terms could have significant relevance for affective or sensory-rich connections of users by bringing together very different objects of a certain colour, pattern, or similar thing(s) represented, opening new connective potentials.

Further, in the open question for ‘additional suggestions’, included in the surveys, users expressed a desire for: more images (*‘Add more images to the online collection. This is singularly the most important thing’*); more descriptive details about collection objects; linking different media and collections together; and enabling image sharing/re-use. Such suggestions would further enhance the potential to connect and link collections in new ways. This interest in connecting collections and particularly enhancing access through images is also reflected in participants’ desire to explore the collections through themes, primarily that of under-represented histories, enabling new narrative threads, discussed below.

Narratives and under-represented histories

Our findings showed a common interest among users in the potential to interconnect online collections through narrative threads that relate to terms that fall into the general/pre-iconography category of the image description framework (Table 2), such as ‘who’ (G1), but also the more abstract /iconology category such as terms encompassing emotions, values, and themes (A2). In our surveys, participants expressed a desire to explore the collection images primarily by under-represented histories, followed by the figures or things represented, or politics and society. Furthermore, in an additional suggestions section of the survey, they requested more specific information such as provenance, but also more narratives and alternative themes to search the collections by, for example through ‘colonisation and imperialism’, or through ‘a good story’. As mentioned in the Identity and Social Context section above, in exploring what under-represented histories means to interview participants, while facets of identity often came to their mind first, they also noted and reflected on the experiences of communities, individuals, and groups who may be under-represented in cultural heritage and social contexts (e.g., Figure 4). For example, these included:

- Wildfire and flooding
- Work, trades, domestic work
- Mental health
- Poverty, homelessness
- Racism

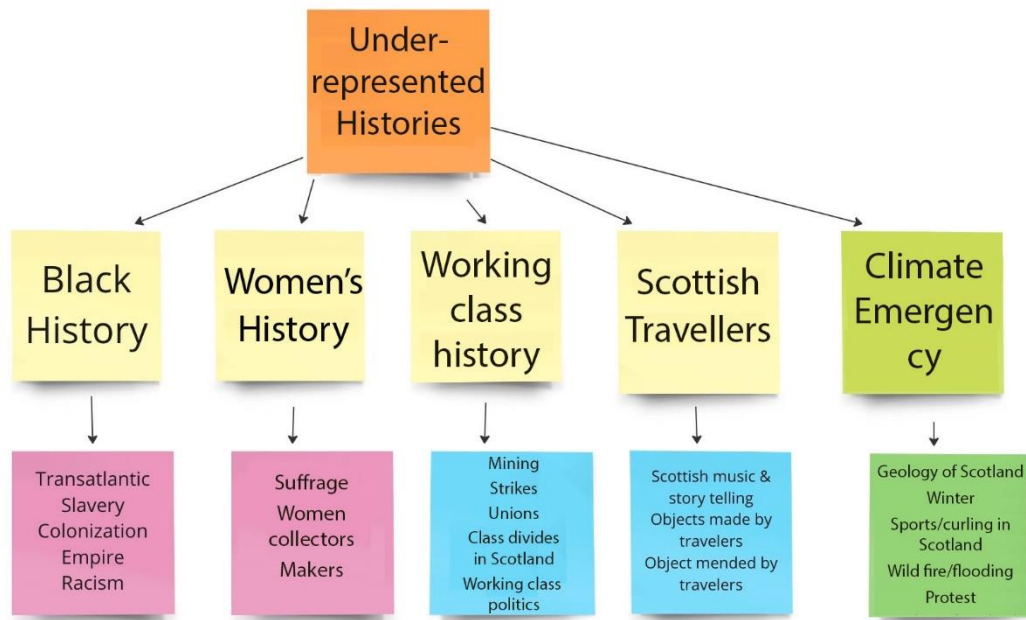


Figure 4. Amelia's concept map

An interest in following narrative threads through the SoC was also suggested through the card sort activity involving collection images. During this activity, several participants grouped together images based on traditional museum collection areas (e.g., decorative art). However, several participants also reflected on and grouped together collections based on narratives that could be told by bringing together different collection images. For instance, interviewee Jordan grouped together the mining helmet and the asteroid, which they suggested could tell multiple narratives – for example, about geology or about work (Figure 5), while Amelia grouped these together to tell a story about extraction (Figure 6). On the other hand, interviewee Kimber suggested that the scent bottle and ambrotype photo could tell stories about women/domestic life (Figure 7), while Amelia similarly suggested the piano, vacuum, photo, and cup could represent women's histories (Figure 6).

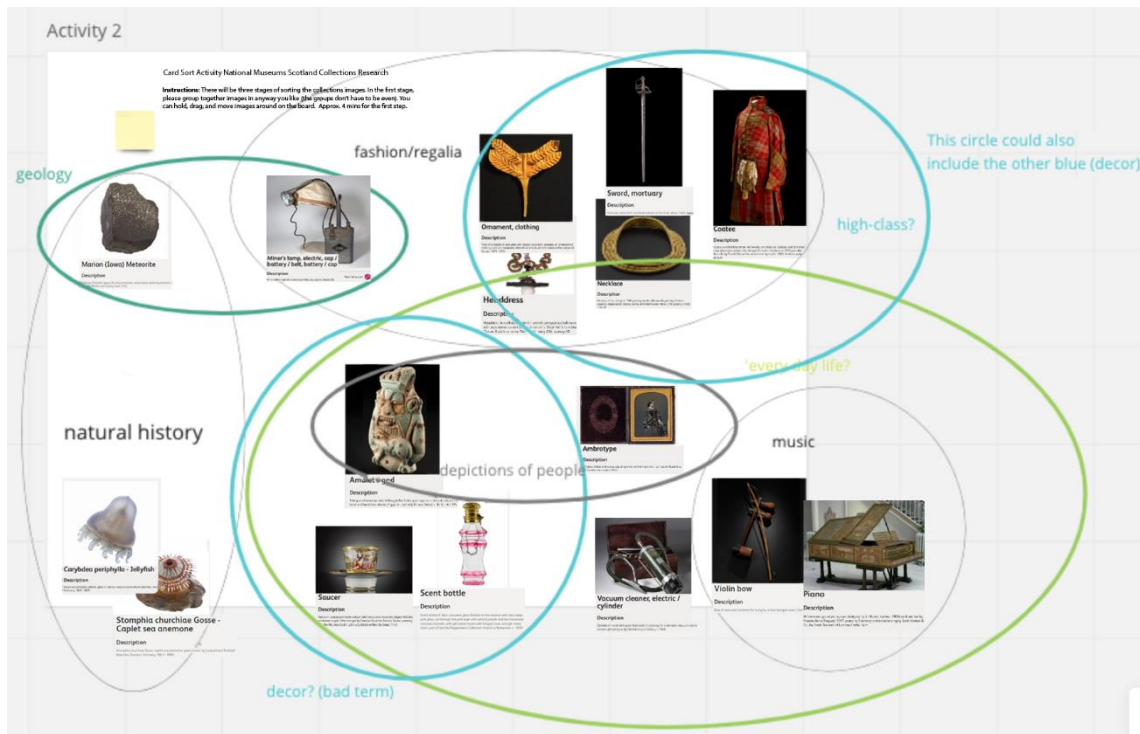


Figure 5. Jordan’s card sort (collection images copyright © National Museums Scotland)

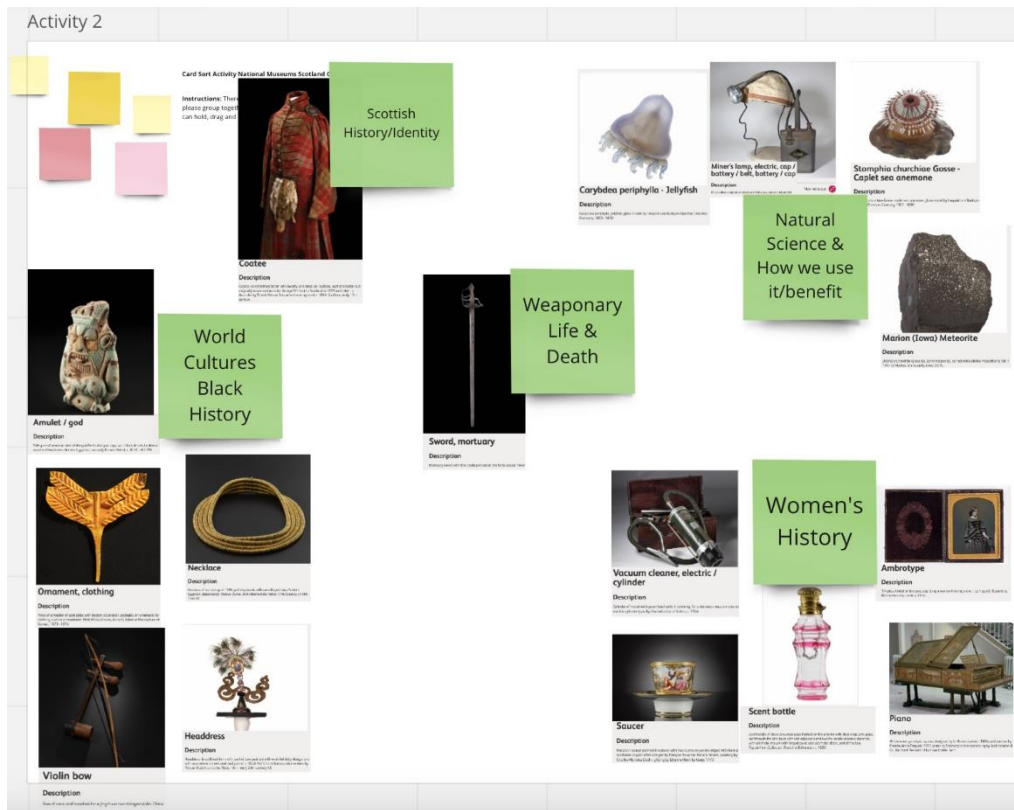


Figure 6. Amelia’s card sort (collection images copyright © National Museums Scotland)

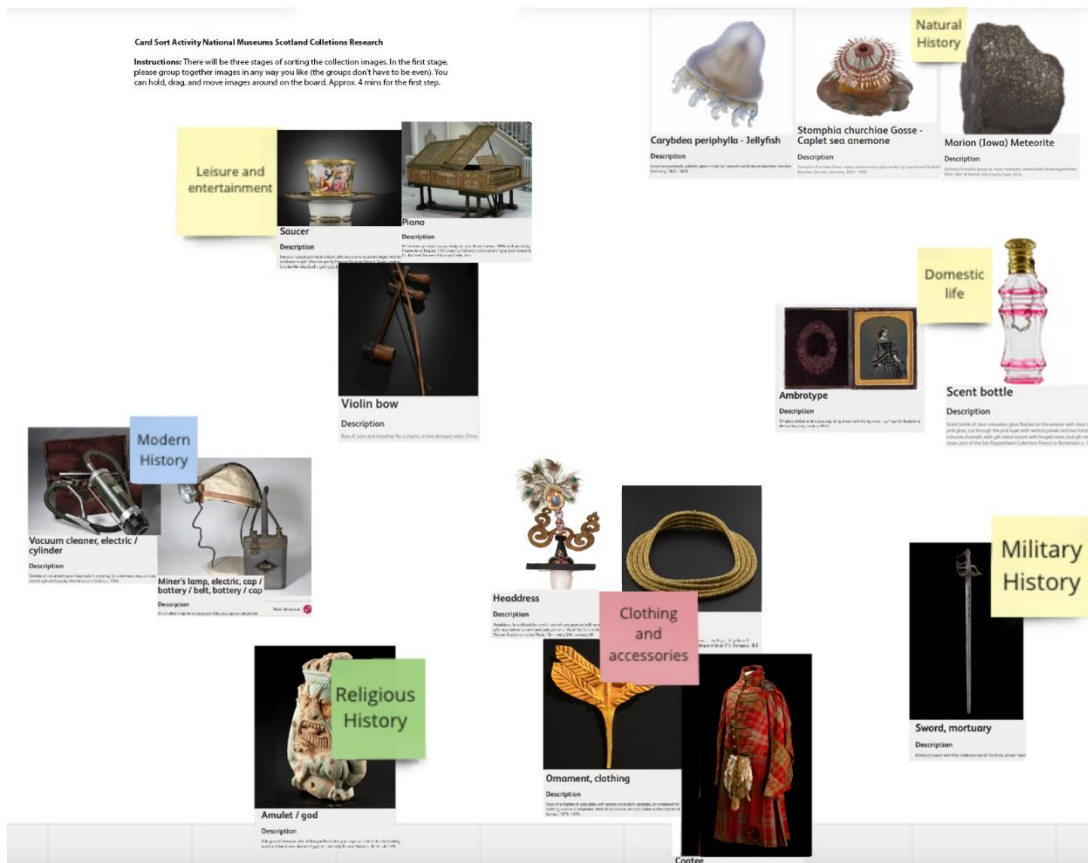


Figure 7. Kimber’s card sort (collection images copyright © National Museums Scotland)

Participants’ interest in experiences of under-represented individuals and groups, and groupings of collection images based on stories they could tell, suggest a desire to follow narrative threads through the SoC webpages. While narratives and rich interpretive material exist on another section of the Museums’ website—the Museums’ Stories webpages, these are not always interconnected with the SoC object webpages. Moreover, interviewees’ interest in the experiences of under-represented individuals and groups and the narratives that can be created by bringing together diverse collections are indicative of the connective potential of digitised collections to create stories that are rich, human, and relevant to contemporary social contexts.

Discussion

Our findings showed that current and potential users may wish to search the collections using terms that are relevant to their own identities, but also of other people (e.g., gender, socio-economic status), and related to social context through social and seasonal events (e.g., wedding, snow, Halloween). Further, it was clear that they desire to link and bring together collections in ways that could support sensory and emotive experiences, for example by using general terms for objects (e.g., ring, vase) and visual terms (e.g., colour, pattern) and by following interconnections across collection media. This also included linking the collections through narrative threads which may represent for example, the experiences of under-represented individuals or groups. These findings suggest that in order to support users' search interests, the terms used to document collection images must have the capacity to be connected to social context and identity, support sensory/affective connections and narrative threads.

This is in opposition to how collection images and associated metadata have typically been envisioned and practiced in cultural heritage institutions until now. Currently, the prevalent documentation practices still focus on images as information and a pursuit for objectivity and accurateness, perceiving metadata as static and needing to be specific and exhaustive (Dahlgren, 2022). That is, there is a drive for metadata that is 'true' and disconnected from the influences of the personal identities, interests, and social context of both users and staff cataloguing the collection, resulting in metadata that is technical and specialised, leaning towards the iconography category of terms in the image description framework (Table 2).

In direct contrast, our findings, as situated in current literature on user search/engagement with digital collections, highlight a need for documentation practices to focus on the potential of collection images to enable social and affective connections.

Implications for collections image documentation practices

Drawing from our results, we suggest an application of metadata to collection images which embraces subjectivity and includes terms represented in the generic/pre-iconography and abstract/iconology categories of the image description framework (Table 2). Firstly, generic/pre-iconography terms can enable the bringing together of diverse collection objects through the SoC portal (in the case of the Museums) and other similar pages (in the case of other cultural institutions), including by features such as general things and people represented in images, visual aesthetic qualities, and associated social events.

Further, we also advocate for the application of more abstract terms and thematic categories of the iconology category that speak to narrative threads and the collections' potential to foster inter-connections with facets of identity and social context. For example, this could encompass terms relating to experiences of under-representation such as migration or mental health. The use of these two categories of terms—pre-iconography and iconology—have been recognised as challenging to apply and therefore, traditionally less of a priority compared to specific and technical subject terms (iconography) (Shatford, 1986). However, our findings and current literature show that these are proving essential to support the collections' social and affective connective potential. In this way, we align with Dahlgren's (2022) research -participants, the humanities scholars interested in the historicity and social context of cultural heritage metadata, who support that the interpretative nature and subjectivity of metadata should be embraced rather than avoided. To implement this perspective and our recommendations in the Museums' documentation practices, we created a flowchart for guiding the application of keywords and captions to images for the Museums (Kist *et al.*, 2023, section 6), based on the pre-iconography and iconology categories of the image description framework. We envision

expanding this further in an upcoming paper which will refer also to the infrastructural potentials and challenges that applying this involves.

Applying collection image metadata that considers images' potential to enable social and affective connections, broadens documentation practices from just information towards honouring the subjective meaning-making processes of users/visitors and sustaining cultural heritage objects' connectivity. This could support users' ability to search and engage with cultural heritage collections in a wide range of ways that are sensitive to their emotional and social needs. As described by Giaccardi & Plate (2016), digitised collections can be engaged with in personal and affective ways through user sharing and interpretation, causing online collections to become 'sticky' with new stories and perspectives. However, for users to find online collections in the first place and engage with them in these ways, image metadata as our researched showed, must encapsulate terms that support this social and affective connective potential, falling into the pre-iconography and iconology categories. In turn, we argue that broadening heritage documentation practices in this way, beyond just the informational role of collection images, can sustain this connectivity and 'stickiness' of online collections and act as a vital step towards closing the semantic gap.

While we focus on National Museums Scotland here, based on the alignment of our findings within wide-ranging current literature on engagement with digital cultural heritage, it is evident that a socio-affective approach to documentation could benefit other heritage institutions. This approach has particular significance for heritage institutions with a similar audience base (combination of experts and curious publics) and diversity and size of collections to that of the Museums. In the case of institutions who are very specialised in one collection area (e.g., medical collections), and/or oriented towards serving experts (e.g. university museums), it would

be interesting for future research to explore whether expert users/visitors may use more specific/expert technical terms encompassed by the ‘Iconography’ category that was not apparent in our findings. However, from our experience and the communications from museums and international and national/regional professional museum organisations over the last years, it is evident that even specialist cultural heritage organisations are in most cases also trying to reach wider, less specialised audiences beyond their core specialised users. For the former type of user, this approach would be particularly relevant. But even in the case of more specialist users, although requiring further research to test and confirm in different cultural heritage contexts, we argue that being able to find and explore collections using pre-iconographic and iconology terms is likely to be beneficial and open up new pathways of exploration. This is because the broadening of documentation practices that we are proposing here is complimenting and extending, rather than replacing the use of specialist iconographic terms. We acknowledge that using this approach has implications for the prioritisation of resources, even more so in the case of small organisations. This is closely connected with their overall strategic aims of the organisation and the way it wants to engage with its audiences.

Summing up, in this article, we contribute to an understanding of the semantic gap in online museum collection metadata, and ways to bridge it by investigating existing and potential users’ search terms with the National Museums Scotland’s Search our Collections portal. Through multifaceted methods, including content analysis of a sample of user search queries, interviews with existing and potential users encompassing a concept map and card sort activity, and online surveys, our research revealed the terms participants use and are interested in searching the collections by. Building on the findings of the study, we suggest that the application of image metadata should embrace the collections’ interpretive and social/cultural

context and include general and abstract (pre-iconography and iconology) terms to support user search interests, embracing the dynamic potential for social and emotional connections that online collections harbour.

To achieve this, however, represents not only a technical challenge involving the expansion and updating of expert terminologies but also a social one requiring changing long-established documentation practices and legacies of value in cultural institutions.

Acknowledgements

We would like to thank the following for sharing their time, expertise, and advice: all the National Museums Scotland staff, especially the Collections Data and Digitisation Team, as well as external workshop, survey, and interview participants who generously gave their time and offered us valuable insights; Our colleagues at The Hunterian and the University of Glasgow for supporting us throughout the project, providing us with feedback and advice, pointing us to relevant literature, and sharing their personal experiences and advice on methods.

This project received funding from the University of Glasgow's Global Knowledge Exchange Fund (GKEF) with contribution from National Museums Scotland.

Disclosure

The authors report there are no competing interests to declare.

Data availability statement

The authors confirm that some of the anonymised data supporting the findings of this study are available within the article. Unprocessed data is not shared publicly as the participants did not give written consent for this. Aggregated and anonymised data from the surveys, user queries, and participant interviews are available at the University of Glasgow Enlighten research data depository:

<http://dx.doi.org/10.5525/gla.researchdata.1490>.

References

- Armitage, L.H. and Enser, P.G.B. (1997) 'Analysis of user need in image archives', *Journal of information science*, 23(4), pp. 287–299. Available at: <https://doi.org/10.1177/016555159702300403>.
- Bailey-Ross, C. (2021) *Online User Research Literature Review: UK Gallery, Library, Archive and Museum (GLAM) Digital Collection*. Zenodo. Available at: <https://doi.org/10.5281/zenodo.5779826>.
- Braun, V. et al. (2021) 'The online survey as a qualitative research tool', *International journal of social research methodology*, 24(6), pp. 641–654. Available at: <https://doi.org/10.1080/13645579.2020.1805550>.
- Budge, K. and Burness, A. (2018) 'Museum objects and Instagram: agency and communication in digital engagement', *Continuum*, 32(2), pp. 137–150. Available at: <https://doi.org/10.1080/10304312.2017.1337079>.
- Chapman, M. (2015) 'Managing Collections or Managing Content?: The Evolution of Museum Collections Management Systems', in S. MacDonald et al. (eds) *The International Handbooks of Museum Studies*. New York, NY, USA: Wiley-Blackwell, pp. 267–292. Available at: <https://doi.org/10.1002/9781118829059.wbihms212>.
- Chowdhury, G. et al. (2022) *Digital footprints and search pathways: working with National Collections in Scotland during Covid-19 lockdown to design future*. Zenodo. Available at: <https://doi.org/10.5281/zenodo.6602365>.
- Clough, P. et al. (2017) 'Europeana: What Users Search for and Why', in *Research and Advanced Technology for Digital Libraries*. Cham: Springer International Publishing, pp. 207–219. Available at: https://doi.org/10.1007/978-3-319-67008-9_17.
- Dahlgren, A.N. (2022) 'Image metadata. From information management to interpretative practice', *Museum Management and Curatorship*, pp. 1–21. Available at: <https://doi.org/10.1080/09647775.2022.2073562>.
- Dallas, C. (2018) 'Heritage encounters on social network sites, and the affiliative power of objects', in S. Antoniadou et al. (eds) *Culture and Perspective at Times of Crisis: State Structures, Private Initiative and the Public Character of Heritage*. Oxbow Books, pp. 116–131. Available at: <https://doi.org/10.2307/j.ctvh1dph.12>.

- Giaccardi, E. and Plate, L. (2016) 'How memory comes to matter: From social media to the internet of things', in M. László, P. Liedeke, and S. Anneke (eds) *Materializing memory in art and popular culture*. New York and Abingdon: Taylor & Francis, pp. 65–88.
- Gibbons, S., (2019). *Cognitive Mapping in User Research*. NN/g Nielsen Norman Group. 11 August 2019 (blog). <https://www.nngroup.com/articles/cognitive-mapping-user-research/>
- Gil-Fuentetaja, I. and Economou, M. (2019) 'Communicating Museum Collections Information Online: Analysis of the Philosophy of Communication Extending the Constructivist Approach', *Journal on computing and cultural heritage*, 12(1), pp. 1–16. Available at: <https://doi.org/10.1145/3283253>.
- Hall, M.M. and Walsh, D. (2021) 'Exploring digital cultural heritage through browsing', in *Information and Knowledge Organisation in Digital Humanities*. Routledge.
- Hopes, D. (2014) 'Digital CoPs and Robbers: communities of practice and the use of digital artefacts', *Museum Management and Curatorship*, 29(5), pp. 498–518. Available at: <https://doi.org/10.1080/09647775.2014.959703>.
- Kist, C., Economou, M., St Clair Inglis, C., Babes, P., Kneale, A. and Norfolk, H. (2023) *Surfacing the National Collections: adapting image cataloguing standards to transform access to National Museums Scotland's online collections*. University of Glasgow and National Museums Scotland. Available at: <https://doi.org/10.36399/gla.pubs.302242>.
- Klavans, J.L., LaPlante, R. and Golbeck, J. (2014) 'Subject matter categorization of tags applied to digital images from art museums', *Journal of the Association for Information Science and Technology*, 65(1), pp. 3–12. Available at: <https://doi.org/10.1002/asi.22950>.
- Kuntsman, A. (ed.) (2017) *Selfie Citizenship*. Cham: Springer International Publishing.
- Martin, J.M. (2021) 'Records, Responsibility, and Power: An Overview of Cataloging Ethics', *Cataloging & Classification Quarterly*, 59(2–3), pp. 281–304. Available at: <https://doi.org/10.1080/01639374.2020.1871458>.
- Meecham, P. (2013) 'Social Work: Museums, Technology, and Material Culture', in K. Drotner and K. Schröder (eds) *Museum Communication and Social Media*. Routledge, pp. 33–53.
- Nayak, M. and K. A., Narayan. (2019) 'Strengths and Weakness of Online Surveys', 24, pp. 31–38. Available at: <https://doi.org/10.9790/0837-2405053138>.

- Newbury, D. and Brennan, D. (2021) *Digital Collections as Product, Museum Computer Network*. Available at: <https://mcn.edu/resources/digital-collections-as-product/> (Accessed: 13 October 2022).
- One Further (2022) *The Cultural Content Report 2022, One Further*. Available at: <https://onefurther.com/blog/cultural-content-survey-2022> (Accessed: 21 October 2022).
- Panofsky, E. (1972) *Studies in iconology: humanistic themes in the art of the Renaissance*. Boulder, Colo. ; Oxford: Westview Press (Icon editions).
- Park, J. (2021) ‘An Actor-Network Perspective on Collections Documentation and Data Practices at Museums’, *Museum and Society*, 19(2), pp. 237–251. Available at: <https://doi.org/10.29311/mas.v19i2.3455>.
- Park, J. (2023) “‘We want to know more than that’: lessons learnt from the public workshop on collections data at the V&A’, *Museum Management and Curatorship*, 38(2), pp. 141–156. Available at: <https://doi.org/10.1080/09647775.2022.2111331>.
- Priestner, A. (2018) *UX In libraries: Cognitive Maps, CILIP: the library and information association*. Available at: <https://www.cilip.org.uk/page/CognitiveMapsMember> (Accessed: 17 May 2023).
- Rainbow, R., Morgan, M. and Morisson, A. (2012) *Providing Accessible Online Collections*. Available at: https://www.museumsandtheweb.com/mw2012/papers/providing_accessible_online_collections.html (Accessed: 17 April 2023).
- Shatford, S. (1986) ‘Analyzing the Subject of a Picture: A Theoretical Approach’, *Cataloging & Classification Quarterly*, 6(3), pp. 39–62. Available at: https://doi.org/10.1300/J104v06n03_04.
- Shatford, S. (2002) ‘Subject Access to Art Images’, in M. Baca (ed.) *Introduction to Art Image Access: Issues, Tools, Standards, Strategies*, pp. 1–20. Available at: <https://www.getty.edu/publications/resources/virtuallibrary/0892366664.pdf>.
- Villaespesa, E. (2019) ‘Museum Collections and Online Users: Development of a Segmentation Model for the Metropolitan Museum of Art’, *Visitor studies*, 22(2), pp. 233–252. Available at: <https://doi.org/10.1080/10645578.2019.1668679>.
- Villaespesa, E. and Crider, S. (2021) ‘A critical comparison analysis between human and machine-generated tags for the Metropolitan Museum of Art’s collection’, *Journal of*

Documentation, 77(4), pp. 946–964. Available at: <https://doi.org/10.1108/JD-04-2020-0060>.

Villaespesa, E. and Stack, J. (2015) ‘Finding the motivation behind a click: Definition and implementation of a website audience segmentation | MW2015: Museums and the Web 2015’, in *MW2015. Museums and the Web 2015*, Chicago, USA: Museums and the Web. Available at: <https://mw2015.museumsandtheweb.com/paper/finding-the-motivation-behind-a-click-definition-and-implementation-of-a-website-audience-segmentation/index.html> (Accessed: 31 January 2024).

Walsh, D. *et al.* (2020) ‘Characterising online museum users: a study of the National Museums Liverpool museum website’, *International Journal on Digital Libraries*, 21(1), pp. 75–87. Available at: <https://doi.org/10.1007/s00799-018-0248-8>.