From Information to Knowledge Creation in the Archive: Observing Humanities Researchers' Information Activities

Alexandra Leigh
City, University of London / The National Archives
London, UK
alexandra.leigh@city.ac.uk

Stephann Makri
City, University of London
London, UK
stephann@city.ac.uk

Alex Taylor
City, University of London
London, UK
alex.taylor@city.ac.uk

Alec Mulinder
The National Archives, UK
London, UK
alec.mulinder@nationalarchives.gov.uk

Sarra Hamdi
The National Archives, UK
London, UK
sarra.hamdi@nationalarchives.gov.uk

ABSTRACT

As primary sources, archival records are a unique information source at the very heart of humanities research. However, how humanities researchers move from information to knowledge creation by making meaning from archival records has not been the focus of previous empirical research. This is surprising, as creating new knowledge through (re)interpretation of records is a core motivation and outcome of humanities research; as representations of historical and social occurrences, archival records rely on researchers' interpretation of content, context, and structure to establish an 'archival' meaning of the record, before applying this meaning within their own work. Therefore, constructing knowledge from archival materials necessitates a dual process of knowledge creation to create novel insights from a hybrid interpretation of archival meaning and the researcher's own interests. This paper presents findings from a naturalistic empirical observation of 11 humanities researchers engaging in research at a national archive, centring on key information activities that facilitate knowledge creation from archival records: *Scanning, Relating, Capturing* and *Organising*. Through these activities, scholars integrate their research aims and objectives with archival meaning to generate new insights. Deeper understanding of the nature of knowledge creation in archives can benefit archivists, archive users and systems designers alike.

KEYWORDS

Archives; Human Information Interaction; information use; knowledge creation.

INTRODUCTION

Archives are an essential information source for humanities researchers, for whom original insights into primary material drives new interpretations of historical or social "*occurrences"* (Yeo, 2007, 2008). For these researchers, interpreting information to drive the generation of novel ideas is both a core motivation and outcome of information interaction. However, while there have been some existing user studies of Human Information Interaction within archives, these studies have focused primarily on understanding information-seeking activities (Rhee, 2015; Sundqvist, 2015) and not on information use to facilitate knowledge creation.

Archival records are a unique information source: to reliably interpret them as evidence of events or processes that took place requires not only an understanding of their content, but also their context and structure (Millar, 2018, p.9). For humanities researchers, this means undertaking a dual process of knowledge creation that requires the researcher to first interpret the record within the archival context–to invoke an 'archival' meaning of the record (Ketelaar, 2012, p.23)–before using this meaning to drive novel insights into historical or social occurrences.

No previous studies, to our knowledge, have focused on understanding key activities that support knowledge creation from information found in archives, such as close reading, capturing records through self-digitisation methods and organising and reorganising materials to support connection-making. While such activities have been studied in relation to humanities scholars' research practices in general, the unique qualities of the archive as an information environment and the centrality of interpreting primary sources to humanities research makes this a specific context worthy of in-depth examination. We report a naturalistic observation of 11 humanities researchers engaging in

research with archival records within a large national archive. We identify and elucidate two core research practices, comprising four associated information activities that facilitate both interpretation of the 'archival' meaning and creation of new knowledge from historical records. These are *Reading* (with associated information activities of *Scanning* and *Relating*) and *Collecting* (with associated activities of *Capturing* and *Organising*). Through elaboration of these activities, tensions between establishing an archival meaning and generating novel insights are examined. This aims to provide a richer understanding of the nature of knowledge creation in archives that can benefit archivists, archive users, and systems designers alike.

The rest of this paper is organised as follows: first, related literature on knowledge creation in archives and humanities scholars' information activities is synthesised. The naturalistic observation method followed is explained and justified, followed by findings, focusing on the four identified information activities of Scanning, Relating, Capturing and Organising. Next, the importance of these findings for understanding the nature of knowledge creation in archives is discussed, followed by conclusions and future work.

RELATED LITERATURE

Knowledge Creation in the Archive

Archival records serve as important sources of historical evidence, as 'witnesses' to events or processes that took place in the past (Yeo, 2007, 2008). They are also somewhat unique as information sources, as it is not only a critical understanding of their content, but also of their structure and context that is required to reliably interpret them (Millar, 2018, p.9). As such, the content of a record usually consists of a description or some other representation of an event or activity it purports to represent. Structure describes the *"physical and intellectual characteristics that define how a piece of evidence was created and maintained"* (Millar, 2018 pp.12-13), while context refers to the *"functional, organizational and personal circumstances"* surrounding the creation of the record (ibid). Collectively, these characteristics allow the record to be considered a representation of a particular event or activity and thus support the interpretation of an 'archival' meaning of the record (Ketelaar, 2012, p.23). While multiple meanings can be interpreted, each should be supported by evidence from within and beyond the record.

To use archival information effectively (and without straying far from the evidence), researchers must establish the record in relation to these elements in addition to being able to apply relevant information to their own work. Thus, archival records require a dual process of knowledge creation: establishing the 'archival' meaning–the record in relation to the activity or event it purports to evidence–while also reinterpreting this meaning in relation to their own research aims and objectives to generate new ideas and insight.

There are very few studies of information activities in archives and most of these focus on information-seeking (e.g. Daniels & Yakel, 2010; Duff, Craig, & Cherry, 2004; Duff & Johnson, 2002, 2003; Johnson & Duff, 2005). The only study on information interactions in the archive that explicitly aimed to understand 'archival meaning-making' focused on how the navigation of records transitions to interpretation of the records themselves (Duff, Monks-Leeson & Galey, 2012). Studying students' experiences of using finding aids – descriptive tools that help to highlight physical and conceptual links between items within archival collections – Duff et al. (2012) found that building both contextual knowledge and a holistic overview of the collection as a whole supported the students' interpretation of the archival records. This reflects the assertion made by Millar (2018, p.9) that it is not only the content, but also the context and structure of records that comprise a record's meaning.

Though the evidence is very limited, other studies of information seeking (Duff & Johnson, 2002, 2003; Johnson & Duff, 2005) also highlight the significance of context to finding relevant information in the archive. In particular, the importance of building contextual knowledge during the course of information seeking in archives has also been pointed out (Duff & Johnson, 2002). While 'contextual knowledge' can relate to building an understanding of the circumstances in which records were created (Duff & Johnson, 2002; Duff et al., 2012), it can also refer to the expectation of historical researchers *"doing their homework"* before entering the archive (Johnson & Duff, 2005, p.127). The same expectation has also been noted of other humanities researchers, such as genealogists (Duff & Johnson, 2003; Yakel, 2004). This suggests a secondary type of 'contextual knowledge', formed by the framing the researcher themselves wishes to apply to a topic of enquiry. Indeed, Duff et al. (2012, p.84) note that students found it easier to navigate the archive if they had a predetermined framing for their research before they entered the archive. This points to the existence of dual knowledge creation processes: interpreting the 'archival' meaning and the utilisation of this interpretation within the humanities researchers own work. The broad scope of these processes means that knowledge creation should not be regarded purely as an outcome of humanities research, but as a continual and

active process of interpretation and application that occurs as a by-product of multiple information activities throughout the research process. It is these activities that are investigated in this research.

Humanities Scholars' Information Activities

While the information needs, seeking and use activities of humanities scholars have been extensively studied (e.g. Brockman et al., 2001; Stone, 1982; Toms & O'Brien, 2008; Watson-Boone, 1994), few of these studies elaborate on how these activities relate to knowledge creation. Furthermore, as explained, archival records are somewhat different to other types of primary sources and so an examination of how these studies of humanities researchers' information activities relate to archives is necessary.

Information activities are often presented as subsets of broader research practices. For example, searching, reading, and writing are all research practices that recur frequently (Brockman et al., 2001; Palmer & Neumann, 2002; Palmer, Teffeau & Pirmann, 2009). As digital technologies have had an increasing impact on humanities research, some studies have also identified personal information management as another significant research practice (Given & Willson, 2018; Kamposiori, Mahoney & Warwick, 2019; Trace & Karadkar, 2017; Antonijević & Cahoy, 2018).

Information activities related to search have been the most closely examined, with browsing and chaining identified as particularly significant to humanities scholars (Brockman et al., 2001; Toms & O'Brien, 2008; Stone, 1982). As noted above, during search humanities researchers build contextual knowledge of archival collections (Duff & Johnson, 2002). This may explain humanities researchers' preference for browsing over searching and, when searching, for recall over precision in search results (Dalton & Charnigo, 2004). In digital information environments, a preference for browsing may also reflect an awareness of digital selectivity and concern that they are not seeing everything on a topic (Coburn, 2020; Sinn & Soares, 2014).

While there has been extensive research attention on how humanities scholars find information, including how they serendipitously encounter it rather than actively seek it (Martin & Quan-Haase, 2017), less attention has been given to research practices reflecting the interpretation and use (as opposed to acquisition) of information such as reading and writing, despite their centrality to the analytic process in the humanities. Palmer & Neumann (2002) identify three types of reading particular to humanities research: scanning, rereading, and reading for writing. Though scanning is largely related to the identification of relevant material, rereading and reading for writing are recognised as at least partly interpretive processes (2002, pp.99-100). Palmer & Neumann (2002) are careful to note that activities such as note-taking and annotation often accompany reading and therefore may support knowledge creation work carried out in the archive itself.

Like many disciplines, humanities research has adapted to incorporate digital technologies, often where it supports existing practices (Bulger et al., 2011). Though some scholars have pointed to digital technologies altering reading practices, with an increase in the use of e-texts (Given & Willson, 2018; Sukovic, 2008; Talja & Maula, 2003; Toms & O'Brien, 2008), such findings are less relevant to archives as the most recent data suggests that the majority of physical archival holdings are not digitised (Poll, 2010). Furthermore, Gooding's (2016) study of webometric data from Welsh Newspapers Online also found that users' behaviours in an online environment were more indicative of changes to search than to reading.

Where digital technologies have most impacted archival research is through the ability to self-digitise large volumes of material using cameras, smartphones and tablets (Cox, 2007; Given & Willson, 2018; Rutner & Schonfeld, 2012; Trace & Karadkar, 2017). This has given rise to the increasing importance of understanding personal information management activities such as collecting and organising due to their increasing importance in humanities research (Antonijević & Cahoy, 2018; Given & Willson, 2018; Kamposiori et al., 2019; Trace & Karadkar, 2017). This importance is reflected in a significant shift in archival research since the widespread use of self-digitisation technologies, from largely interpretive activities to a greater emphasis on gathering materials for later analysis. This suggests personal information management activities, such as collecting and organising materials, should also be considered as activities that support knowledge creation from archival information.

While existing studies of humanities researchers' information activities can guide us towards those activities that may support knowledge creation in archives, they are insufficient for understanding how researchers generate novel insights from archival materials. Primary sources can vary significantly in the type and format of information they contain. Archives are no exception and it cannot be assumed that interpretation and knowledge creation occur the same way in archives as they do with other primary research materials. As archival use is central to humanities research, understanding the nature of knowledge creation in archives can enrich our understanding of humanities research more broadly.

METHOD

To identify the information activities carried out in archives that are key to knowledge creation, along with how and why they are carried out and how this supports knowledge creation, in-person observations inspired by a Contextual Inquiry approach (Holtzblatt & Beyer, 2017) were conducted with 11 humanities researchers at the main public site of a national archives. A Contextual Inquiry-inspired approach to observation (Holtzblatt & Beyer, 2017, pp.72-73) was chosen because knowledge creation is inherently difficult to observe and would therefore require probing, through question asking and answering, to understand. As such, we felt an approach that allowed for greater researcher intervention, as a curious 'research apprentice' rather than passive observer could provide greater insight into information activities that incorporate or give rise to knowledge creation. It was also likely to provide greater insight into motivations behind the participants' activities (Boren & Ramey, 2000). Though researcher questioning poses a potential disruption to observation that is 'as natural as possible', this risk was mitigated by ensuring interventions were limited to where we believed this would provide greater insight and would not influence the participant's actions. Both these mitigations had resulted in useful data in a previous observation study conducted by the researchers (Makri, Blandford & Cox, 2011).

In this section, the participant recruitment approach is discussed, including key ethical considerations, and data collection and analysis approaches explained and justified. The study received ethical approval from our departmental Research Ethics Committee.

Participant Recruitment

Participants were approached on the basis that they were currently conducting research within the archive, which contained historical records covering a period of 1000 years. Participants were recruited either in advance (through the archive's research newsletter and social media channels) or in-person, at the archive itself, pre-pandemic. All except one were recruited ad-hoc on the day; this was the most successful recruitment approach. They were asked if they would like to be observed and, if so, to confirm that they would be conducting research using the archive that day and to explain the topic of their research, to ensure a breadth of humanities research. The topic of the advance recruit was also documented. Participants were professional historians, academics, research assistants, PhD students, and genealogists. They were a mix of newer and seasoned researchers. Their research topics ranged from Anglo-German relationships under Henry IV, to land and housing policy in Malaysia; genealogy and family history, to the nuclearization of British society under Margaret Thatcher. Most were collecting materials for further detailed analysis, though several could be better characterised as 'fact-finding', e.g. whether an individual was awarded a posthumous medal in WWII (P08).

The study was naturalistic in the sense that none of the topics were prescribed: they were what the humanities researchers had already planned to research within the archive. Participants were also asked whether they would be working with physical or digital archives (or both) that day, to ensure findings were not restricted to a particular materiality of the record. During the study, 6 of 11 worked exclusively with physical records and 5 exclusively with digital records, though this solely comprised digitised rather than born digital material. Data collection took place over 6 weeks, from October-December 2019. Rather than determining sample size in advance, the principle of 'information power' (Malterud, Siersma & Guassora, 2016) – where sampling continues until a (subjectively) rich insight is gained to address the research aims – was adhered to. Although participants were recruited from a single archive, which mostly contained textual documents, the nature of the information activities identified does not appear to be archive-specific. However, we only make limited claims of generalisability to archives with predominantly non-textual (i.e. image, video, or audio) materials.

Data Collection

Before the observation, participants were told that the total session, including obtaining informed consent and any follow-up questions, would last around an hour (mean = 52 min., 7 secs.; s.d. = 8 min., 2 secs.) and that they would be notified as the end of the session approached. As well as describe their task, participants were asked to provide background context to their research. The archive advised users to order material prior to visiting, so all participants had a predefined task in mind. During the observation, participants were asked to carry out their chosen research task. These included but were not limited to: consulting online databases only accessible within the physical archive; creating conceptual links between existing research materials and new information found; and making personal copies of documents. It was stressed to participants that the study was interested in understanding the routine research activities of humanities researchers in archives and that they should carry out their research as they normally would. After the observation, follow-up questions were asked to expand on or clarify participant actions and test researcher

assumptions. Participants were also asked to explain how the research activities observed fit into their wider research process. No fixed questions were asked, allowing the researcher to follow-up on intriguing or seemingly important comments or actions made by the participant.

Data was audio recorded, de-identified prior to transcription by assigning participant numbers, and transcribed in full. Any identifying details from transcripts were also removed, such as references to personal names where participants were carrying out genealogical research. Video data was not collected, due to data collection happening in public areas of the archive and the need to protect the privacy of non-participants. Brief notes were taken to contextualise participants actions, e.g. *"P07 has drawn boxes around the individual images to highlight which sections of the volume will be of interest to them."* These were not included for coding during data analysis.

Data Analysis

Transcription was supported by Otter.ai, a GDPR-compliant, automated transcription tool. The transcripts were not stored on Otter's server, but on the researcher's encrypted and password-protected computer. We did not grant Otter permission to use the transcripts for machine learning purposes. The analysis itself was supported by NVivo. Analytic coding was partially inspired by Thematic Analysis (TA) (Braun & Clarke, 2013). Early stages of coding followed TA closely: an initial total coding of activities was carried out primarily inductively, through which codes relating to information activities were identified, with codes compared for similarity and some being merged or further split accordingly. For example, *"there is a particular case in here, which I think is directly comparable"* (P08) was coded as '*Relating to other records'*; this was later grouped with other codes referring to a relational style of reading, such as '*Relating to existing information'* and *'Reading things through'*. During analysis, comparison was made with existing classifications of information activities (e.g. the classification by Palmer et al., 2009) and a hierarchical approach to classifying the information activities was adopted at this stage, with the activity of Relating grouped with Scanning under Reading. Excerpts related to each activity were extracted from NVivo and re-ordered to construct a narrative explaining what the practices involved, how and why they were carried out and how they facilitated knowledge creation as a by-product. Inter-rater reliability was not sought as this is less appropriate for assessing the validity of highly interpretive and inductive research than for deductive research (Braun & Clarke, 2013, p.279; Yardley, 2008, p.237).

FINDINGS

The findings highlighted that humanities researchers undertake a dual process of knowledge creation to create novel but evidence-based insights from the archive. This involves making an integrated interpretation of archival meaning and the researcher's own interests. As such, the archival meaning and researcher's interpretation become entangled through the researchers' information activities, resulting in the making of a hybridised meaning from information found in the archive.

Reading	*Scanning*	Scanning both involved looking for cues relating to a project or area of interest (Palmer & Neumann, 2002, p.99) and constructing an initial, tentative meaning of the record by interpreting these in the context of the researcher's own project.
	Relating	A more in-depth style of reading than Scanning, Relating included an interpretive element that involved constructing the meaning of the record in relation to the wider archival context and the user's evolving knowledge base. It included both in-depth reading of a single document and working through a collection of records stored in a file to gain a holistic understanding.
Collecting	*Capturing*	Capturing materials, primarily through photography on a phone, tablet or digital camera, to work on them elsewhere. Capturing was undertaken as other means of access were not possible; borrowing from archives is generally not permitted and this material had not yet been digitised.
	Organising	(Re)arranging materials in a purposeful order to support interpretation or creation of meaning. Participants drew on the hierarchal structure of the archive (series→file→item) in their organisation of personal copies of archival material. Some indicated they would later break with this structure when moving to a later stage of analysis in their research.

Table 1. Description of Archival Information Activities to Support Knowledge Creation

Reading

Reading is a key way in which archival users assign meanings to records (Ketelaar, 2012, p.25) and our findings reflect this. The practice of Reading to support knowledge creation is examined through discussion of two distinct information activities: Scanning and Relating. Scanning a record involved finding information and determining its significance in relation to the user's own research interests. This builds on Palmer & Neumann's (2002) definition of identifying cues *"that intersect with their line of inquiry"* (p.99), going beyond this to incorporate constructing an initial, tentative meaning of the record through an interpretation of these cues in the context of the researcher's own project. Relating is used to describe a type of reading that was particular to the archival interactions observed here. It is based on descriptions of *"reading through"* (P04), which were used by several participants to refer to a more interpretive style of Reading that relied on constructing the meaning of the record in relation to both the wider archival context and the user's evolving knowledge base. This went beyond merely relating the information found to that already known, to making an integrated interpretation of archival meaning and how this meaning relates to the researcher's aims and objectives.

Scanning

Due to pressures involved in visiting the archive, such as time, distance, and cost, several users spent the majority of their visit scanning material, aiming to maximise efficiency by covering as much material as possible and reducing their potential number of visits to the archive. Scanning served not only as a relevance check (to determine the potential importance of the document to the researcher), but also began to align the record to participants' own research interests – a conceptual 'shaping' that went beyond merely reading what was written, to engaging with what was written and what was already known.

All participants arrived at the archive with some level of existing knowledge on their topic, though this varied widely in formality from detailed research guides that determined "exactly what to be looking out for" (P05) to "a few different subject areas" (P06) to be explored. Through the activity of Scanning, participants simultaneously brought this information to bear on the record while bringing the record within their own specific framing of a research topic. For example, P04 was at the archive to scope information on the topic of WWII spies. Attempting to confirm her *"growing hypothesis that a lot of these people who were bilingual acted as spies, as well"* she scanned the record for any possible clues that might confirm this. She picks out *"two clues",* the first relating to the content*: "the person here is talking about what might happen, if she is questioned by the Russians, and... what she might say to them"*. The second clue is the context in which the record was produced*, "I understand that the captain who is writing the letter is trying to report almost verbatim the conversation, so I would assume that this is what she herself said, her 'very special work' which would be an understatement of what it really was".* The details that P04 picks out fit the original framing or hypothesis that she had developed through reading before she entered the archive, though the precise details come from the record itself. Chosen from the wider context of the document precisely because they suit her framing, she shapes the record according to her own particular interests. This was described as a process of *"hunting, searching for clues"* (P02). This comparative practice goes beyond a simple relevance check to actively create a particular meaning of the record that aligns the record to the users' own research interests by defining the record in relation to those interests. It follows that this process is also selective, in that the user does not pick up on everything from the record: *"It's basically you know, maybe someone might read it and not really see it and someone else might kind of notice it"* (P11). Through this initial interaction, P04 defines the way the record is read and how it relates to her interests.

Relating

A second style of Reading within the archive was identified through several participants' references *to "working my way through"* (P08) or *"reading through the file"* (P04). Such phrasing could refer either to the level of the record, where participants stopped to read through the record to interpret an intriguing piece of information in more detail, or to the level of the file, skimming through the file and gaining an impression of the whole. We use the term 'Relating' to refer to this activity, as participants related information found to the context of the record, to information found elsewhere (i.e. outside of the archive), or to other records within the archival file.

When slowing down to read a record in depth, participants frequently compared it with other records they had recently read within the same file, whether mentally recalling these, or physically skipping back through the file. P04 often referred to information read in previous documents, reflecting a growing understanding of the context the records were produced in as she progressed through the file. Information found in the record was also compared to that which the participant had found before entering the archive, as in the case of P03. After finding a newspaper announcement

for *"twenty-three substantial brick dwelling houses",* he invoked his wider knowledge of the period, that, *"at this particular time, that area was going down so they're obviously pointing this out because anybody reading the advertisement might assume because they were in that area, they weren't any good."* Such an assertion is rooted in the details provided by the record itself. Yet it gains a deeper significance for P03 when it is brought into contact with information previously found or known. The meaning of the record is thus created by P03's recontextualisation of the information within his own framing. This goes beyond situating or scaffolding the information found within prior knowledge, which involves comparison or incremental knowledge-building; the information found *only* gains important meaning when understood in the context of prior knowledge. Through the practice of reading, the record can be framed by the context provided by other records within the archive. Alternatively, context may also be provided by the researcher's own evolving and active knowledge of the topic. The interpretation of archival meaning is thus never solely 'archival' but a hybrid, shaped by the researcher themselves, and provides a basis for knowledge creation.

Participants also engaged in Relating as they sought to interpret the potential significance of the file structure, providing an understanding of the file greater than the sum of its parts (the individual records). P08 had initially approached the files he was working with as a fact-finding task: to determine whether the person he was researching had received a posthumous award. While the record did not provide evidence of this, reading the file as a whole suggested an unanticipated relationship between the records within the file. As he put it, the file provided him with *"evidence I hadn't expected to find, and hadn't even thought about",* providing him with *"a clearer idea of the decision-making process"* for the award. Through the act of interpreting connections between the records in the file, P08 gained a very different understanding than that suggested by his own expectations of what information the file would hold. This holistic reading can be seen as a reading of the archival structure particular to P08: combining his own particular interest in the file with the significance of the archival structure.

Collecting

The research practice of Collecting comprises the information activities of Organising and Capturing (Palmer et al., 2009). While Palmer et al. (2009) refer to 'Gathering' rather than Capturing, a narrower term has been used here to reflect that participants primarily sought to create surrogates or copies of archival material, often through digital photography but also occasionally through note-taking (where no interpretation was involved and text was copied verbatim).

The archival meaning of the record, as inferred through its content, context, and structure, partly relies on arrangement to preserve the meaning of the record. Archival classification is largely based on provenance, meaning that records from different creators should not be mixed and the original order of records preserved wherever possible (Meehan, 2014, pp.65-68). When Collecting, participants chose to what extent they preserved this archival arrangement. Sometimes they chose to closely replicate it in their own organisation of materials so that they could read the archive in the same manner later. Other times, they opted to rearrange materials in a way that better facilitated interpretation, to support the creation of new insights.

Capturing

Amassing personal copies of archival records is not new (Cox, 2007; Orbach, 1991). However, capturing data to work on outside the archival reading room is becoming more prevalent over other forms of interaction in the archive and occurring at a much larger scale, due to facilitation through digital technology (Bulger et al., 2011; Rutner & Schonfeld, 2012; Trace & Karadkar, 2017). This was reflected in the findings, with almost all participants capturing information to work on outside of the physical archive.

Participants' careful preservation of the record and its order and grouping within the wider file demonstrates awareness of both context and structure that is provided by archival practices, and how they support creation of a particular interpretation of the record. In retaining this, participants sought to maintain connections between the record and its form, as well as between records, to allow them to fully and faithfully interpret the record outside of the physical archive. Though Capturing approaches varied widely, most participants sought to capture at some level *"the context it's found within"* (P11) and several, such as P09, suggested *"normally if it's not too big, [...] I just take the whole thing",* thus preserving the structure of the individual record. Occasionally, participants collected more than one document to preserve the relational nature of the relevant information: its meaning in its wider context. P07 used personal letters as an example to demonstrate how meaning is often spread over more than one document. The variety in participants' Capturing suggests that each held an individual conception of what was necessary or relevant to perform the same meaning of the record derived within the archive at a later date and in a different location. Such decisions and personal views subtly shape the meaning rendered by the record, presenting a hybrid between the

framing that the record's structure and the archival context provide for the record's content, and that provided by the user themselves.

A few participants demonstrated awareness of these subtle shifts in meaning and attempted to mitigate potential risks by capturing further details beyond the digital copy. For example, P05 used note-taking to supplement photography when capturing archival materials, to allow her to reconstruct a more thorough interpretation of context. She wrote *"a brief description of what the folder as a whole contained",* stored in a spreadsheet. Motivated by the belief it would be *"good to get it read as much as possible, in one go",* her written summary provided an immediate impression of the file as a whole, based on her subjective experience of examining it while in the archive. Although this impression remains subject to her selectivity in reading, the textual summary enabled her to preserve a point of reference for the whole collection, upon which future interpretations could be based.

Organising

Some participants expressed frustration with how the archive was arranged: P05 noted that it *"messes up the ideas in my head",* while P11 said she *"might've organised it differently because of my interests"*. This reflects an incongruence between the personalised meaning researchers wished to create, and the provenance-based classification system employed by the archive to preserve the context and structure of the records. Nonetheless, it was not often practical to rearrange materials when working in the archive. For those accessing physical documents, only one bound volume or three files could be brought to the reading room table at a time. This limited participants' ability to reorganise material and restricted direct comparison of materials housed in different files while the user remained within the physical archive building.

Despite having difficulties in navigating the archival arrangement at times, the same participants also saw value in preserving this arrangement in the organisation of their own research materials. At the most basic level, record numbers were used to maintain the archival arrangement of materials. P06 ensured she viewed and photographed materials in order of record number so they could be retrieved in the same order, while P05 used record number to document files in a spreadsheet, likewise preserving the arrangement of records as they appeared in the archive. Though P11 was not observed setting up her organisational approach during the session, she too claimed that, *"my initial, let's say categorisation is very much based on the categorisation I find them in, here, and actually, it's quite sensible because a lot of times it's easier to read things, how they've been categorised"*. The example P11 discusses here involves her attempts to uncover information on an architect involved in several colonial building projects. For her, the fact that *"I found [his] resume, found here within these documents"*– i.e. government documents – was significant as *"it tells you a larger story"* about the use of the records. Conversely, it might be imagined that if this record were found among his family archive or personal affects, it would not hold the same meaning. As such, we see how it becomes important to P11 to preserve not only the record's content, but also its context to support the generation of a specific insight that the provenance-based archival classification preserved.

Some participants also suggested that their organisational methods would later change to reflect different research stages. Interestingly, this process often took place after participants had established an interpretation of the record as suggested by the archival arrangement: P11 described how she would rearrange the records *"once I have all the documents, and I know what's in each one",* having initially stored them in a manner very similar to the archival arrangement. Likewise, P07 envisioned sorting out material thematically later, as *"the sampling frame for that would be different I think."* Once the record's meaning within the archival context was established, P11 described how she sought to reorganise any notes and images collected and *"would copy paste into a separate folder that's about…a specific part that I'm looking at"*. Other participants hinted at a similar process of re-organising information to fit the evolving research task: P03 initially stored his notes according to *"different physical locations"* where he had accessed records, before subsequently copying them into *"whatever the topic is that I'm writing at that time"*. This relocation of the record and other information further highlights the importance of the context in which materials are stored. While initially preserving the archival arrangement – and thus the contextual interpretation afforded by this through their Organising, subsequent rearrangement allows the researcher to break with the archival structure to support them in using this information to shape their research. This suggests that some participants may perceive the initial interpretation of the record as separate to applying archival information within their own research. Using the document as part of their own broader argument is seen here as a later stage of research and possibly supports a distinction between interpretation of the record and later knowledge creation.

DISCUSSION

The findings provide a rich insight into how knowledge creation occurs within (and beyond) the physical archive, through the enactment of four information activities – Scanning, Relating, Capturing and Organising. These activities are not new. Indeed, they have been identified and discussed in relation to several disciplines, including the humanities (e.g. Palmer et al., 2009; Meho & Tibbo, 2003). What is new, however, is an understanding of how and why they are undertaken in the archive: a unique primary information source at the heart of humanities research. While existing empirical research has focused on understanding information acquisition in archives, this naturalistic observation study focused not on understanding how information is found, but how it is leveraged to create original insights. In that respect, this research regards information acquisition as a means to an end; as a facilitator of knowledge creation.

The findings affirm that creating new knowledge through the (re)interpretation of records is a core motivation and outcome of humanities research; in line with Millar (2018), humanities researchers were found to use not only the content of archival records, but also their context and structure to establish an 'archival' meaning of the record, then apply this meaning within their own work. Therefore, constructing knowledge from archival materials was found to necessitate a dual process of knowledge creation to create novel insights from a hybrid interpretation of archival meaning and the researcher's own interests. This involved striking a simultaneous balance between interpreting an 'archival' meaning of the record through gaining an understanding of its content, context and structure and considering the relationship between information found within and outside the archive to construct new knowledge.

Most of the information activities discussed here revolve around an 'ex situ' mode of archival research (Trace & Karadkar, 2017, p.500): researchers primarily sought to gather large amounts of archival material to interpret it offsite. While Trace & Karadkar (2017) define this in opposition to an 'in situ' mode, the information activities we examined suggest a blurring of the archive's boundaries and reflect a more continuous interpretive process, occurring both on and off site. The effect of this blurring is to illustrate a tighter integration between the two knowledge creation processes discussed in this paper: interpretation of an 'archival meaning' (Ketelaar, 2012, p.23) and generation of new insights through knowledge creation.

The information activities identified in this study were undertaken uniquely in an archival context; Scanning an archival document does not merely involve extracting potentially relevant information to the research task at hand (Palmer & Neumann, 2002, p.99), but re-interpreting the text through the lens of the research topic and other information acquired from within and outside the archive. Likewise Relating does not simply involve making mental connections to prior knowledge or previous information acquired (Kuhlthau, 1991), but actively shaping the information found based on the content, context and structure of the archive. While Relating might usually be considered an activity that operates on informational content, it was found to be strongly influenced by the record's structure and context: as with P08's experience of the file as a whole lending a different interpretation than that provided by the individual records.

Capturing does not only involve taking snapshots of potentially useful information as a workaround to physical access constraints and to minimise visits to the physical archive. It also involves deliberately preserving (and forsaking) some of the archival context and structure to support the creation of new knowledge outside of the physical archive. However, while participants regarded themselves as interpreting the record once they had left the archive, decisions made about what information to capture was found to shape their subsequent ability to reconstruct the content, context and structure of the record. This can be seen in the significant variety in the ways that participants tried to preserve these characteristics of archival information. At the most basic level, participants ensured that items were captured in order, allowing associated items from the same file to be browsed together. Others, such as P05, had more complex systems of Capturing and Organising that allowed them to reconstruct the archival structure at the level of both series and file. This suggests each user has a different conception of what is necessary to recreate the record and allow for interpretation outside the physical archive, making this always a partial and unique enactment of potential meanings from the archival record. This should not be seen as indicative that working 'ex situ' will result in an inaccurate creation of archival meaning, but as highlighting how such interpretations are always generated at the intersection of the connections the archive seeks to preserve and the original insight the user is trying to create.

Finally, Organising archival materials does not solely involve creating and re-arranging personal classification systems, but deliberately emulating (and eventually breaking away from) the original archival system to generate new insights from beyond the archive's physical borders. While emulating the archive's (somewhat arbitrary) structure might at first seem counter-intuitive, considering this action in the context of preserving structure to support knowledge creation helps us understand how it can be useful. Similarly, breaking away from a structure that has helped

spur insight may seem strange on the face of it. However, when considered as a transition to a new form of knowledge creation support best suited to later research stages, the rationale becomes clearer.

What is abundantly clear is whether Scanning and Relating within the physical archive or Capturing and Organising materials to consult offsite, participants wanted to work with the totality of the record: not only its content but also its context and structure. It was embracing the totality of the record that helped support the generation of original knowledge: knowledge not solely derived based on content. It was not, however, only the sum of content, context and structure that resulted in the creation of new insight; the research aims and objectives, prior knowledge held by the researcher and previous information found within and outside the archive played a combined role. This highlights the entangled nature of 'archival meaning,' researcher interests, prior knowledge and knowledge creation. Nonetheless, some participants believed knowledge creation took place in a later stage of research, such as analysis or writing, rather than being seeded throughout the information acquisition activities they were observed undertaking. This suggests that humanities researchers may seek to first establish an 'archival' interpretation of a document before applying it to their own research.

These findings highlight the complex nature of knowledge creation in archives, explaining how it can be facilitated through seemingly mundane information activities such as Scanning, Relating, Capturing and Organising. They expand our existing understanding of archival information activities, by moving beyond seeking and towards knowledge creation. Though we can trace the origins of knowledge creation to these activities, it is possible most knowledge creation does occur in later stages of research. As such, a follow-up study is proposed that will examine humanities researchers' information activities once they have left the physical archive. It is hoped that this study will complement the current findings by offering a holistic view of archive-based information activities related to knowledge creation; it is also anticipated that further information activities may be identified through this second study. A rich understanding of information activities and their interconnection can help to inform the design of the next-generation of digital archives by encouraging a shift in thinking and approach, away from merely preserving and providing access to information and towards enhancing opportunities for knowledge creation both within the archive and beyond.

CONCLUSION

This paper presented an observational study of humanities researchers conducting research in a national archive, focusing on elucidating the information activities key to knowledge creation, how and why they are carried out and how this supports knowledge creation. These activities comprised Scanning, Relating, Capturing and Organising and were identified as part of the wider research practices of Reading and Collecting. These activities were highly influential in supporting knowledge creation, both in establishing the meaning of an archival record in relation to its content, structure, and context as well as in relation to researchers' prior knowledge gained from information previously found in and outside the archive. These two areas of knowledge creation were found to be strongly intertwined, with the researcher's framing of the research aims and objectives and the seeking of original insight tightly integrated within the process of understanding the evidential value of the archival records. An enriched understanding of the nature of knowledge creation in archives can benefit archivists, archive users and archive designers alike; reflecting on these findings can support archivists in labelling and structuring archives to better support knowledge creation, users in enhancing their knowledge creation practices and archive designers in more explicitly supporting knowledge creation in archives.

REFERENCES

Antonijević, S. and Cahoy, E. S. (2018). Researcher as Bricoleur: Contextualizing humanists' digital workflows. *Digital Humanities Quarterly* 12 (3). Retrieved from: http://digitalhumanities.org:8081/dhq/vol/12/3/000399/000399.html.

Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication* 43(3). pp.261–278. https://doi.org/10.1109/47.867942.

Braun, V. and Clarke, V. (2013). *Successful Qualitative Research: A practical guide for beginners*. Los Angeles, London, New Delhi, Singapore: SAGE Publications.

Brockman, W. S., Neumann, L., Palmer, C. L. and Tidline, T. J. (2001). *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington DC: Digital Library Federation.

Bulger, M., Meyer, E. T., de la Flor, G., Terras, M., Wyatt, s., Jirotka, M., Eccles, K., and Madsen, C. (2011). *Reinventing Research? Information Practices in the Humanities*. London: The Research Information Network.

Coburn, J. (2020). Defending the Digital: Awareness of digital selectivity in historical research practice. *Journal of Librarianship and Information Science*. pp.1-13. https://doi.org/10.1177/0961000620918647

Cox, R. (2007). Machines in the Archives: Technology and the coming transformation of archival reference. *First Monday* 12 (11). https://doi.org/10.5210/fm.v12i11.2029.

Daniels, M. G. & Yakel, E. (2010). Seek and You May Find: Successful Search in Online Finding Aid Systems. *The American Archivist* 73 (2). pp.535-568. https://doi.org/10.17723/aarc.73.2.p578900680650357.

Dalton, M. S. and Charnigo, L. (2004). Historians and their Information Sources. *College & Research Libraries* 65 (5). pp.400-425. https://doi.org/10.5860/crl.65.5.400.

Duff, W. M., Craig, B. & Cherry, J. (2004). Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age. *The Public Historian* 26 (2). pp.7-22. https://doi.org/10.1525/tph.2004.26.2.7.

Duff, W. M. and Johnson, C. A. (2002). Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives. *The Library Quarterly* 72 (4). pp.472-496. https://doi.org/10.1086/lq.72.4.40039793.

Duff, W.M. and Johnson, C. A. (2003). Where Is the List with All the Names? Information-Seeking Behavior of Genealogists. *The American Archivist* 66 (1). pp. 79-95. https://doi.org/10.17723/aarc.66.1.l375uj047224737n.

Duff, W. M., Monks-Leeson, E. and Galey, A. (2012). Contexts Built and Found: A Pilot Study on the Process of Archival Meaning-Making. *Archival Science* 12 (1). pp.69-92. https://doi.org/10.1007/s10502-011-9145-2.

Given, L. M. and Willson, R. (2018). Information Technology and the Humanities Scholar: Documenting Digital Research Practices. *Journal of the Association for Information Science and Technology* 69 (6). pp. 807-819. https://doi.org/10.1002/asi.24008.

Gooding, P. (2016). Exploring the information behaviour of users of Welsh Newspapers Online through web log analysis. *Journal of Documentation* 72 (2). pp.232-246. https://doi.org/10.1108/JD-10-2014-0149

Holtzblatt, K. and Beyer, H. (2017). *Contextual Design: Design for Life* (2nd ed.). Cambridge, MA: Morgan Kaufman.

Johnson, C. A. and Duff, W. M. (2005). Chatting up the Archivist: Social Capital and the Archival Researcher. *The American Archivist* 68 (1). pp.113-129. https://doi.org/10.17723/aarc.68.1.h1l2r87kl1846417.

Kamposiori, C., Mahoney, S. and Warwick, C. (2019). The Impact of Digitization and Digital Resource Design on the Scholarly Workflow in Art History. *International Journal for Digital Art History* 4. pp.3.11-3.27. https://doi.org/10.11588/dah.2019.4.52795

Ketelaar, E. (2012). Cultivating Archives: Meanings and Identities. *Archival Science* 12 (1). pp.19-33. https://doi.org/10.1007/s10502-011-9142-5.

Kuhlthau, C. C. (1991). Inside the Search Process: Information seeking from the user's perspective. *Journal of the American Society for Information Science* 42(5), pp.361-371.

Malterud, K., Siersma, V. D. & Guassora, A.D. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research* 26 (13). pp. 1753-1760. https://doi.org/10.1177/1049732315617444.

Makri, S., Blandford, A. and Cox, A. L. (2011). This is What I'm Doing and Why: Methodological reflections on a naturalistic think-aloud study of interactive information behaviour. *Information Processing and Management* 47. pp.336-348. https://doi.org/10.1016/j.ipm.2010.08.001.

Martin, K. and Quan-Haase, A. (2017). "A process of controlled serendipity": An exploratory study of historians' and digital historians' experiences of serendipity in digital environments. *Proceedings of the Association for Information Science and Technology* 54 (1). pp.289-297. https://doi.org/10.1002/pra2.2017.14505401032.

Meehan, J. (2014). 'Arrangement and description: between theory and practice.' in Brown, C. (Ed.) *Archives and Recordkeeping: Theory into Practice.* London: Facet Publishing. pp.63-99.

Meho, L. I., & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology* 54 (6). pp.570-587. https://doi.org/10.1002/asi.10244.

Millar, L. A. (2018). *Archives: Principles and Practices* (2nd ed.). London: Facet Publishing.

Orbach, B. (1991). The View from the Researcher's Desk: Historians' Perceptions of Research and Repositories. *The American Archivist* 54 (1). pp.28-43. https://www.jstor.org/stable/40294400.

Palmer, C. L. and Neumann, L. J. (2002). The Information Work of Interdisciplinary Humanities Scholars: Exploration and Translation. *Library Quarterly* 72 (1). pp.85-117. https://doi.org/10.1086/603337.

Palmer, C. L., Teffeau, L. C., and Pirmann, C. M. (2009). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Dublin, OH: OCLC Research.

Poll, R. (2010). NUMERIC: statistics for the digitisation of European cultural heritage. *Program: Electronic Library and Information Systems* 44 (2). pp.122-131. https://doi.org/10.1108/00330331011039481.

Rhee, H. L. (2015). Reflections on Archival User Studies. *Reference and User Services Quarterly* 54 (4). pp.29-42. https://doi.org/10.5860/rusq.54n4.29.

Rutner, J. and Schonfeld, R. C. (2012). *Supporting the Changing Research Practices of Historians*. New York, NY: Ithaka S+R.

Sinn, D. and Soares, N. (2014). Historians' Use of Digital Archival Collections: The Web, Historical Scholarship, and Archival Research. *Journal of the Association for Information Science and Technology* 65 (9). pp.1794-1809. https://doi.org/10.1002/asi.23091.

Sukovic, S. (2008). Convergent Flows: Humanities Scholars and their Interactions with Electronic Texts. *The Library Quarterly* 78 (3). pp.263-284. https://doi.org/10.1086/588444.

Stone, S. (1982). Humanities Scholars: Information Needs and Uses. *Journal of Documentation* 38 (4). pp.292-313. https://doi.org/10.1108/eb026734.

Sunqvist, A. (2015). Conceptualisations of the use of records. *Tidsskriftet Arkiv* 6. pp.1-15. https://doi.org/10.7577/ta.1358.

Talja, S. and Maula, H. (2003). Reasons for the use and non-use of electronic journals and databases: A domain analytic study in four scholarly disciplines. *Journal of Documentation* 59 (6). pp.673-691. https://doi.org/10.1108/00220410310506312.

Toms, E. G. and O'Brien, H. L. (2008). Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation* 64 (1). pp.102-130. https://doi.org/10.1108/00220410810844178.

Trace, C. B. and Karadkar, U. P. (2017). Information Management in the Humanities: Scholarly Processes, Tools, and the Construction of Personal Collections. *Journal of the Association for Information Science and Technology* 68 (2). pp.491-507. https://doi.org/10.1002/asi.23678.

Watson-Boone, R. (1994). The Information Needs and Habits of Humanities Scholars. *RQ* 34 (2). https://www.jstor.org/stable/20862645.

Yakel, E. (2004). Seeking information, seeking connections, seeking meaning: genealogists and family historians. *Information Research* 10 (1). Retrieved from: http://informationr.net/ir/10-1/paper205.html.

Yardley, L. (2008). Demonstrating validity in qualitative psychology, in J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (2nd ed. pp.135-251). London: Sage.

Yeo, G. (2007). Concepts of Record (1): Evidence, Information, and Persistent Representations. *The American Archivist* 70 (2). pp.315-343. https://www.jstor.org/stable/40294573.

Yeo, G. (2008). Concepts of Record (2): Prototypes and Boundary Objects. *The American Archivist* 71 (1). pp.118-143. https://www.jstor.org/stable/40294496.