



“So How Do We Balance All of These Needs?”: How the Concept of AI Technology Impacts Digital Archival Expertise

Journal:	<i>Journal of Documentation</i>
Manuscript ID	JD-08-2022-0170.R1
Manuscript Type:	Article
Keywords:	Archives, Qualitative, digital preservation, artificial intelligence (AI), Focus Groups, Template Analysis

SCHOLARONE™
Manuscripts

Participant number	Job title & Sector	Gender Identity	Location
Phase 1 focus group: All staff from the same national institution			
FG1			
PD101	Digital Preservation Manager	Male	Ireland
PD102	Digitisation Programme Manager	Male	Ireland
PD103	Assistant Keeper, Digital Collecting	Female	Ireland
PD104	Assistant Keeper, Digital Collecting	Female	Ireland
Phase 2 focus groups: Mixed participants from different organisations			
FG2			
PD201	Researcher, Archival Project	Female	Ireland
PD202	Archivist, University Special Collections	Female	Ireland
PD203	Digital Curator, Private Business	Female	Ireland
PD204	Digital Curator, Government Library	Female	Denmark
PD205	Archives Manager, Government Archive	Male	England
FG3			
PD301	Digital Humanities Librarian, University	Female	USA
PD302	User Manager Assistant, Library Vendor	Female	China
PD303	User Services Assistant, Library Vendor	Female	China
FG4			
PD401	Project Manager, Archives/University Collaborative Project	Male	Ireland
PD402	Digital curator, Museum	Female	England
PD403	Digital curator, Moving Image Archive	Male	Ireland
PD404	Library & Digitisation Manager, Museum	Female	USA

Table I. Demographics of the focus group participants.

1
2
3
4
5
6
7
8
9 **“So How Do We Balance All of These Needs?”: How the Concept of AI Technology**
10 **Impacts Digital Archival Expertise**
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Journal of Documentation

Introduction

Explorations of the use of artificial intelligence (AI) tools have appeared in archival studies in the past few years. However, many of these articles are mostly limited to testing implementations or opinion pieces from academia. Considering this landscape, we wish to expand the discussion of AI technology in archives by empirically exploring the thoughts and opinions of archival and digital preservation practitioners.

This article attempts to fill the gap by reporting on focus groups with those working in the archives sector as practitioners about their thoughts and opinions related to adopting AI tools in archival work. Our goal is to situate the current discussion about using AI in archival practice via the perspective of working archivists.

In doing so, we hope to learn more about the challenges that may exist in a potential widespread implementation of AI technology in the archives field.

We aim to empirically explore potential social issues associated with the use of AI tools in archival work as perceived by these practitioners, rather than focus on the outcome of a specific application of the technology. We hope that this focus will add to the conversation about AI in archives at the current time.

Literature review

Defining AI

Prevailing discussions about “AI and archives” are mediated by the definition of “AI” currently being used in the discussions. The existing discussions tend to favour social and cultural definitions of AI over technical definitions that may be used in other fields. The highly cited Crawford (2021) explains that the definition of AI shifts overtime. While “AI” is frequently used in funding applications, the term “machine learning” (ML) is more frequently used in technical literature. She explains that ML can be understood as a model that can learn from data it has been given. This model can utilise ML and/or computer vision (CV). While machine learning focuses on numerical, categorical, textual and temporal (time series) data, CV utilises visual data. Crawford (2021) utilises the term ML to refer to technical approaches such as broad scale data mining, classification of data, and CV. The author uses the metaphor of an atlas to describe AI due to the technology’s far-reaching social and infrastructural implications.

Explicit definitions of AI in the context of archives have been offered by a few pieces situated in archival studies research. In their survey of archival literature, Colavizza *et al.* (2022) explain that they utilise the term AI as a proxy for ML, but also use AI “to encompass the professional, cultural and social consequences of automated systems for recordkeeping processes and for archivists” (p. 4). Also situated in archives and recordkeeping, Rolan *et al.* (2019), references text from Bellman (1978) in their definition: “we understand AI as involving digital systems that automate or assist in activities that we associate with human thinking, activities such as decision-making, problem-solving, learning [and] creating” (p. 181). In this study, we adopt Rolan *et al.*’s definition of AI.

Access and use

Viewing digital humanists as users of archival collections can also yield insight into existing thoughts and opinions about archives and AI from the digital humanities (DH) perspective. This perspective can be considered parallel to archival studies discussions about AI, as it is written from the perspective of access and use and generally does not consider the work practices of archival practitioners. Jaillant and Caputo (2022) express frustration at the inability of archival repositories to make large digital datasets available in a timely manner. Additionally, Jaillant (2022) noted that the lack of accessibility may impact end users—DH practitioners included.

Digital Humanities (DH), an umbrella term referring to humanities scholarship concerned with the use of computers as an integrated and essential part of research (Brügger, 2016), have published much more on the use of ML and CV in comparison with archival studies literature. For “genetic” reasons, most applications of AI in the field of DH are related to computational linguistics and are mostly grounded in Natural Language Processing (NLP) methods, although there is a growing interest in working with other media types, including 3D objects. Neural networks and deep learning techniques are among the most current approaches, enabling DH researchers to tackle demanding NLP and CV tasks. Examples range from more traditional use cases such as text analysis from historic and contemporary corpora (Clanuwat *et al.*, 2018; Kestemont *et al.*, 2016; Tanasescu *et al.*, 2018), image and object classification (Bermeitinger *et al.*, 2016; Wevers and Smits, 2020), to more **particular** applications like Egyptian hieroglyphs recognition, classification and translation (Barucci *et al.*, 2021) or the development of semantic analysis and comparative query of art-historic collections (Garcia and Vogiatzis, 2019; Jain *et al.*, 2021; Springstein *et al.*, 2021). Gefen *et al.* (2021) caution against the intrinsic disruptiveness of AI, which might deeply impact the way we understand, approach and produce cultural knowledge (p.196).

Digital humanists' increased use of AI has urged them to reflect on the nature of their relationship with archivists—which remained latent until recent times. Sabharwal (2017) described the concurring interests between these communities, both aiming at the life-cycle extension of the humanistic data and knowledge within the digital landscape (p. 239). Similarly, Poole and Garwood (2018) analysed the outputs from the Digging into Data 3 global challenge, with the goal of defining the roles held by librarians and archivists in highly structured DH projects. The authors stress that, although the work of archivists (including librarians) in the examined projects have low visibility, the digital curation and data lifecycle constitute a solid opportunity for collaboration between digital humanists and archival practitioners.

Within archival studies, AI holds promise for managing the archival backlog. Many ideas have been proposed over the years to tame the backlog and make collections available to the public more quickly. Most notable is Greene and Meissner's (2005) more product, less process (MPLP) method of arrangement and description, in which records are not described at the item level in order to alleviate processing backlogs. Crowdsourcing with user generated tags has also been explored as a method to alleviating processing backlogs (Benoit, 2017, 2018) as has participatory archives initiatives (Eveleigh, 2014; Roeschley and Benoit, 2019) which harness the crowd and also attempt to involve users in archival work. According to Jaillant and Caputo (2022), AI offers potential to sort through backlogs more efficiently, specifically by more efficiently screening data for sensitivity. The authors state that “archival collections often close entire collections due to data protection concerns” and that “closing entire collections for an indeterminate period of time is not ethical, since archives in publicly funded organisations are meant to be open to the public” (p. 5). The authors also suggest that archivists should make it easier for (DH) researchers to utilise large datasets for their work. Writing from the archives perspective, Lee (2018) also supports the use of ML to review collections for sensitivity, as well as using ML in the appraisal and selection process. Both authors reference the ePADD email project in their work (<https://library.stanford.edu/projects/epadd>).

In their thoughts on the future of the archival field, Moss *et al.* (2018) also write of the challenges archivists face when trying to provide user access to large digital datasets. They acknowledge that large datasets have changed research methods in the humanities, using history as an example. Whereas close reading was previously the dominant research method, “distant reading” using AI technology, is becoming more common. The authors suggest that archivists may re-envision their collections as “data to be mined”, echoing Jaillant and Caputo's interest in obtaining access to large digital datasets held by archives quickly. Moss *et al.* (2018) suggest that as a result, the focus of archival work on

arrangement and description of historical collections, may need to evolve, especially if users might not find classification as useful for distant reading. However, they note that the largest cohort of archival users is family historians, not academic researchers. Proctor and Marciano (2021) suggest that these family historians would be served using CV to extract names and dates from images. The opinion that arrangement and description may not be as useful as it was previously is not shared by Randby and Marciano (2020), who state that digital curation, including description, is a necessary preparation step for application of ML algorithms.

Development of new skills in the context of Archives and AI

Developed by a team of academics, computational archival science (CAS) attempts to combine skills and knowledge from archival science, information science, and computer science to create a new interdisciplinary field. CAS is defined as:

“an interdisciplinary field concerned with the application of computational methods and resources to large-scale records/archives processing, analysis, storage, long-term preservation, and access, with the aim of improving efficiency, productivity and precision in support of appraisal, arrangement and description, preservation and access decisions, and engaging and undertaking research with archival material” (Payne, 2018, p. 2743).

CAS considers ML a computational tool, in addition to blockchain technology. Grounded in archival science, CAS focuses on the nature of the record and application of computational tools to that record, rather than focusing on data processing which is typical of computer science (Marciano *et al.*, 2018). Marciano *et al.* (2018) provide several case studies where collaboration between researchers from the different fields has enriched project findings. Proctor and Marciano (2021) provide an example of how CV can be used to process a collection using a CAS framework. CAS appears to implement what Moss *et al.* (2018) suggest: that archives may need to be re-envisioned as “data to be mined”. Computational skills may aid in this endeavour. At the least, CAS skills and knowledge can help archivists understand how their users wish to access data: as large datasets, **a perspective shared by the project Always Already Computational, resulting in the redaction of the Santa Barbara Statement on collections as data (Padilla *et al.*, 2019a). Running between 2016 and 2018, the initiative was centred on understanding and mapping “the current and potential approaches to developing cultural heritage collections that support computationally-driven research and teaching” (Padilla *et al.*, 2019).**

Another currently running initiative is InterPARES TRUST AI (<https://interparestrustai.org/>), a large project that aims to train students and professionals in these computational skills. The project goals are to explore AI technologies in the context of records and archives via case studies at locations around the world.

Social and Ethical concerns

In the past few years, archival scholars have provided initial thoughts as to how AI may impact the field. Moss *et al.* (2018) does not specifically discuss AI, but their suggestion that archives be understood as data collections to be “mined” is a nod to the technology that allows for the mining to occur. Theimer (2018) agrees with this sentiment, suggesting that archivists need to become data scientists. CAS provides one perspective, most notably that the archival science focus on the nature of the record remains central moving forward (Marciano *et al.*, 2018; Payne, 2018). Additional opinions take a wider view beyond application to archival work and ponder how AI technology may change the ways in which archives are perceived and why this may be a cause for concern.

Rolan *et al.* (2019) explored how AI technology can automate different aspects of archives and recordkeeping work. After briefly describing how ML works, the authors describe how ML might work with an Electronic Digital Records Management System (EDRMS). They note

1
2
3
4
5
6 that some proprietary commercial products have begun to offer the technology via an “AI as
7 a service” model. One such example is Preservica, which claims to integrate Microsoft Azure
8 ML into their software products ([https://preservica.com/partners/active-digital-preservation-](https://preservica.com/partners/active-digital-preservation-on-microsoft-azure)
9 [on-microsoft-azure](https://preservica.com/partners/active-digital-preservation-on-microsoft-azure)). Rolan *et al.* (2019) then discuss the need for case studies to further
10 explore how AI technology might be applied to archival work. Since 2019, some of these
11 case studies have come to fruition, such as Randby and Marciano (2020) and Proctor and
12 Marciano (2021). The authors finish by summarising some ongoing projects in Australia that
13 are trialling ML. They make one final important note: many “AI as a service” models offered
14 by large tech companies such as Microsoft, Amazon and Google typically rely on cloud
15 storage. This may conflict with recordkeeping law, which prohibits transfer of electronic
16 records outside of a specific jurisdiction. This serves as one example of the contextual
17 issues that institutions managing government archives and records may face: while “off the
18 shelf” products are affordable and may be customised via an easy to learn interface, these
19 products were initially set up for private business use and may not automatically obey local
20 recordkeeping laws. The design of AI to meet the needs of a capitalist society is one of the
21 issues that Crawford (2021) describes as a problem when AI attempts to be used in “AI for
22 good” type projects. The structure of the technology is not necessarily set up to meet the
23 needs of organisations that lack a profit driven mission. Depending on the type of archive,
24 this presents a problem that has yet to be completely addressed.

25
26 Bunn (2020) explored Explainable AI (XAI), which is described as a focus on shedding light
27 on ML models, which are often understood as a “black box” lacking transparency. Bunn
28 (2020) links XAI with “accountability, fairness, social justice, and trust” (p. 144). Similarly,
29 social justice has been of interest in archival studies, which calls into question the colonial
30 past of much of traditional archival practices and positions archives as tools to support work
31 toward an equitable society (Duff *et al.*, 2013; Punzalan and Caswell, 2016). Bunn (2020)
32 links recordkeeping and XAI via the goal of working toward explainability.

33 Colavizza *et al.* (2022) explored thoughts and opinions of AI and archives via the approach
34 of an environmental scan of literature. The authors used 53 articles published on the topic of
35 archives and AI between 2015 and 2020 as their corpus. They found four broad themes
36 dominating the literature which they discuss using the framework of the Records Continuum
37 model: theoretical and professional considerations, automating recordkeeping processes,
38 organising and accessing archives, and novel forms of digital archives. Overall, the authors
39 find that there is a trend in AI being used to probe the traditional definition and concepts of
40 what an archive is, which would put traditional archival principles such as provenance and
41 original order “under pressure.” Other scholars have also highlighted the colonialist
42 overtones of provenance and original order, explaining that they can conflict with diversity
43 and inclusion goals (Punzalan and Caswell, 2016; Steinmeier, 2020).

44 Chabin (2020) uses case study to highlight a point also found by Bunn (2020) and Colavizza
45 (2022)—that other archival and recordkeeping principles may be of use to the AI field,
46 specifically authenticity. Chabin (2019) uses the case of the data processing associated with
47 the great debate in France to explain how diplomatic analysis can be used to enrich a
48 dataset that is used in ML or CV. When compared with discussions over original order and
49 provenance, it is clear that Colavizza *et al.*'s (2022) analysis that AI technology is being used
50 to trigger conversations about the application of traditional archival principles is accurate.
51 The lingering question from this review is how the archival field may progress this debate
52 about the archives field in the context of novel technology such as AI.
53
54
55

56 Method

57 A series of focus groups held via Zoom were used to gather data about how working archival
58 and digital preservation practitioners think and feel about using AI technology in archival
59 practice. **All focus groups received ethics clearance from University College Dublin.** In
60

1
2
3
4
5
6 phase one of data collection, focus group one was held to explore reactions to the use of AI
7 computer vision for assisting with item level metadata creation for historical photograph
8 collections as a way to understand the larger AI issues related to digital preservation work.
9 This focus group was largely exploratory in nature and was used to develop a baseline of
10 understanding of relevant issues that would be used to develop questions for the next round
11 of data collection.

12 In phase two of data collection, focus groups two to four of professionals working at different
13 institutions in the archival and digital preservation field were asked for their thoughts and
14 opinions about using AI technology to complete their work. Participants were recruited via
15 word of mouth, and posts made to professional listservs and Twitter.
16

17 *Participants*

18
19
20 Please see table I below for a description of participant demographics.
21
22

23
24 **The volunteer basis of this study meant that the sample was not representative or**
25 **generalisable. Focus groups are not designed to provide statistical validity since they**
26 **mostly privilege thickness and richness of information (Morgan, 1998). Instead of**
27 **trying to achieve saturation (*sensu* Glaser and Strauss, 2017) as a criterion of**
28 **representativity, we sought a good trade-off between the quality of the insights and**
29 **the number of focus groups to run – as suggested by Carlsen and Glenton (2011).**
30

31 *Procedure*

32 In phase one, four staff members of a national library responded to prompts that consisted of
33 Microsoft Azure produced tags and titles for four photographs from one of the library's
34 digitised photograph collections. These tags were created as part of a class project for the
35 Digital Curation module at UCD. The purpose of the focus group was not to "test" the use of
36 Microsoft Azure Cognitive services for CV on the collection, but to gather participant
37 response to some of the tags and descriptions as a way to prompt further discussion about
38 the use of AI technology in archival and digital preservation work. All staff members were
39 provided an information sheet about the study and consented to participation and audio
40 recording of the focus group.
41

42 In phase two, three different focus groups were held with archival and digital preservation
43 professionals to gather their thoughts and opinions on using AI technology in their work.
44 Some of these participants had familiarity with using AI in an archival setting, some did not.
45 **Once participants responded to the call for participation, they were provided with an**
46 **information sheet outlining their participation in the study, and provided verbal**
47 **consent to participate and be audio-recorded before the focus groups began.** As an
48 ice breaker, participants were first asked to provide their "idealised or dystopian"
49 implementation of AI in their field of work and were then provided a working definition of AI
50 technology that would be used in the focus group. Participants were then asked if they
51 wanted to share any positive or negative experiences that they had had with AI technology.
52 Next, participants were presented with three prompts about the use of AI in archival and
53 digital preservation work.

54 Prompt one detailed a scenario in which the participant was asked to provide users access
55 to a partially digitised photograph collection, which has not been described in any detail. AI
56 technology is suggested as a tool to speed up description of the collection. Participants were
57 asked to respond with their thoughts about the idea and what they might do in response to
58 the suggested use of the technology.
59
60

Prompt two adds further detail to prompt one, in which the participant has noted that several of their colleagues do not fully trust an implementation of the AI technology due to concerns about job losses, lack of training available to learn the new system, and concerns over reliability of the system and whether the technology may become obsolete quickly. The participant was asked if they relate with any of these concerns and if they had other concerns and why.

In the final prompt (three), participants were told that in a continuation of the scenario, their work using AI technology to describe the collection was featured in an online newspaper article in their community. The author of the newspaper article mentions general concerns about privacy, transparency, and ethical use of data about AI in the article not specific to their team's use of the technology. The participants were asked to respond to the article. To conclude, participants were asked to summarise their position on the use of AI to complete their work and if there was anything that was missed, that they would like to add to the discussion.

Analysis

Template analysis, a style of thematic analysis, was used to analyse the data (Braun and Clarke, 2021; King, 2012; King and Brooks, 2017). We used a deductive process to identify themes: a scan of literature about AI and archival studies was completed a priori, and tentative themes from the literature were identified.

The focus group one transcript was read and re-read by the PI to gather information to develop the prompts in the second phase of focus group data collection. After focus groups two thru four were held, all focus groups were read and re-read to develop tentative themes for an initial template. Both researchers worked together to refine the final template, which was complete after four rounds of revision. All focus groups were then coded using nVivo 12 to identify evidence for patterns. After this process, a final tentative template with 4 themes was developed and accounts were written describing the details of each theme.

Findings

Fitting AI into the day to day practice

Most of the concern about "fitting AI in" to existing work centred around additional duties that use of AI technology may require, such as checking outputs before making them available to users. The fact that focus groups two thru four were asked to respond to hypothetical prompts could account for the lack of specific information about the practicalities of using AI technology in digital archives and preservation settings.

Similarly, the participants from focus groups two thru four did not discuss specific benefits that could result from the implementation of AI in the day to day practice, but still considered the potential for AI technology to be valuable. However, they described the technology as having the potential to create more tasks to be performed by humans. Commenting on the use of AI technology to generate descriptive metadata (prompt one, option one), PD204 proposed an alternative scenario, tackling the need for manual validation. While suggesting that AI could support the generation of administrative metadata to enable another option presented in prompt one, the participant clearly defined when additional human performed tasks might be necessary and why:

I would also like to throw in a bit of crowdsourcing in (option) number one because the AI can describe what it sees but it can't tell us anything about the context or the event leading up to what's happening on the picture so we need some again some human intelligence involved.

1
2
3
4
5
6 In relation to the same prompt, PD402 agreed with PD204 on the use of AI to generate
7 descriptive metadata, (in this case to enhance accessibility), while also expressed a concern
8 about their job security:

9 I guess if you can automate, you know, describing, you know, basic description for
10 each (photograph) that I mean, I guess that would help with, you know, making the
11 collection somewhat accessible for people to actually view, you know, what's, what's
12 in the collection, em it might put me out of a job, but [laughs], I dunno, I think I'd go
13 for (option) number one.
14

15 PD404 added to the reflection, cautioning against the lack of supervision of machine-
16 generated outputs:

17 I think that all the results that come out of these (AI) systems need to be reviewed,
18 and in theory they're going to do things that, at scale um, that they can do; and so
19 there's all the things that they can't do, that we now have the time to do, because
20 those things are doing that. I think there is always that risk, depending on how we
21 frame the use of these tools up the chain, um, as to what they're really capable of
22 doing.
23
24

25 From these participant quotes it is clear that there is optimism that AI technology can be of
26 use, but there is also concern about how to fit it into existing workloads. This concern was
27 largely framed as a need to "check" the "work" or output of the AI technology.

28 In light of these expectations, many participants expected the use of AI to be central to
29 archival work in the future. As PD203 highlights, "it's gonna be such a huge part of what
30 every librarian has, or information professional and general archivist librarian, whatever role
31 you're involved in". Participants expressed major concerns about who would be responsible
32 for upkeep associated with the AI systems and what the upkeep would require. At times,
33 these anxieties were expressed via binary choices: outsourcing of the AI system
34 development/maintenance versus the use/upkeep of the same by the in-house staff. Both
35 the reliability of the software and the effective role(s) of the person(s) expected to secure the
36 functionality of the system were questioned by PD201: "I'd be much more concerned with
37 how reliable the software is and also, if this is a tool, what practices are in place eh, that, to
38 keep it going [...]". PD404 agreed:

39 And so that is really where my concerns come in around whenever this comes up,
40 they're like, we're going to just bring in somebody and drop this on your lap. I, I who
41 is taking care of it, who is maintaining it, who is growing it and who is dealing with its
42 outputs? So that's, that's always me, and again that reliability of the software, what is
43 it really doing? Is it doing what I think it's doing? And what happens if something
44 goes wrong?
45
46

47 *Responsible use of technology as expertise*

48 Within the focus groups, "responsible use of technology" was discussed as a human-
49 machine partnership - often found in literature as human-robot collaboration. However, who
50 should be responsible for the AI system and with which modalities, was not clear.
51 Interestingly, the awareness about the pitfalls commonly associated with AI emerged across
52 the focus groups from both of the phases; the role of human agency emerged while
53 participants discussed the tasks that could be "safely" performed by the machine.
54 As recognised by PD204, AI is "not just simply a tool, it creates issues around sensitivity,
55 privacy; those two things are different". PD205, commenting on the application of AI to
56 manage legal records including recent crimes (having the relatives of the victims eventually
57 exposed to the AI system outputs), tackled the need for a critical design to prevent ethical
58 issues:
59
60

I think AI can help us, but it's, it's how you think through each of those steps and, and structure something around it, so that it's, it's helping rather hindering kind of human interaction.

Participants were sceptical about the adoption of AI to address sensitive matters, (such as records that should not be made public) or decision-making process: PD201 argued that “automated systems aren’t really sophisticated enough to do a human’s job.” She kept elaborating on the theme of a human-machine partnership:

So, I think, in general, the consensus here is that AI technologies are useful for making jobs more efficient and for helping the humans, the experts in the field that to (parse?) large bodies of data. But a lot of these ethical questions, em, and special interest questions need to be dealt with by humans.

PD404 asserted “I’m not trusting AI to make curatorial decisions [...]. I’d want to have a better sense of the collection, to see, you know, how much information is there, really, to get from, from these pieces to, to actually transcribe and acquire and whether that would be the most useful thing”. PD204 proposed a different point of view, commenting on quality delivery in discussion of the lack of transparency (or explainability) of the AI systems on use, which required extra-caution in proposing unreviewed AI outputs to their customers (researchers):

I’m running a project where we’re looking at the use of two AI, AI systems in literature reviews and we are incredibly sceptical regarding the quality we are getting back. We can’t understand, using our librarian ninja skills, that we can’t understand how, ehm, the string that’s, ehh, how relevance is considered, ehm, in these two systems we’re using and the quality of the sources, the systems are searching in as well is also quite, ehm, it’s almost a trade secret. We can’t get a full list of where these, eh, systems are searching, ehm, so we are very cautious before we start, eh, introducing these two systems to our researchers. We are testing, testing, testing to make doubly sure we as a library can still deliver quality products.

Participants experienced a lack of trust toward using AI technology, characterised by concern over the ability of the AI technology to redact personal/sensitive information, or complete a task to the same level of an archival and/or digital preservation professional. As a result, participants were more comfortable with the potential concept of a human-machine partnership, in which the human had the ability to check the AI technology’s output for potential issues before releasing the collections to users.

Managing expectations

Another subject that was frequently discussed in all focus groups was the belief that the use of AI technology (including ML) would require practitioners to “manage expectations” of “higher ups” (their line managers, and those managing their organisations). In focus groups two thru four, prompt three, which asked the participants to imagine that a newspaper article was being written about their organisation’s use of AI, was the prompt most likely to result in discussion of managing expectations.

Participants were most concerned with managing the expectations of those that work above them because these “higher ups” controlled resources and funding of specific departments in the organisation. The use of AI technology was frequently compared to the early days of the digitisation “era” in the mid-1990s, when heritage institutions experienced a radical transformation in the way they delivered access to collections. It may be significant that AI technology is being considered by GLAMs at a period of time nearly 30 years after the transition to digitisation and mass availability of digital resources. Those beginning their careers in the mid-1990s that experienced the difficulties that came with this disruptive

1
2
3
4
5
6 technology first hand may still be on the job and more aware of how important managing
7 expectations of those who allocate resources can be. According to PD205:

8 Yeah it's linked to, to training, but also awareness, especially at the top of the
9 organisation, where people could often see the, the digital solution to be the answer
10 to all the problems, em, and, and that creates quite a different shift in the way an
11 organisation can put it through resources.
12

13 PD202 also compared the potential adoption of AI in the sector to previous adoption of
14 digitisation:

15 Just thinking of management em, and the higher ups of the organisation, when
16 digitisation started being used in archives, one of the issues was that the term
17 digitisation was being bandied about a lot. So, it was almost seen as replacing the
18 physical archives and replacing the work of cataloguing and arranging and, you
19 know, as as mentioned earlier, it isn't just about the specific document you're looking
20 at, you need the context um, it's a huge, hugely important part of our job is to provide
21 that context.
22

23
24 The concern that AI could divert resources away from other day to day work practices is
25 understandable, considering that putting effort in at the start will be necessary. However,
26 PD202's concern that AI technology may leech resources "context" building tasks is worth
27 further investigation. PD202 explained context as expertise in the following statement as
28 "making all of those kinds of connections":

29 If we're going to use AI at all in the archive, it should be assisting and never
30 overriding the expertise of the, of, of the professionals, and as PD203 alluded to
31 there, talking about librarians would actually be the archivist involved in something
32 like this, you know our whole training is all about appraising collections and deciding
33 what's kept and what's not kept and why, em, and arranging collections and making
34 all of those kind of connections, so AI should be able to assist us with that, but
35 absolutely never override that.
36

37
38 These comments also express the concern that management may fixate on a new
39 technology as a way to improve efficiency, but that this efficiency would come at the cost of
40 other essential duties that require digital archival expertise, similar to "when digital
41 happened". PD404 also echoed this sentiment:

42 I am concerned, again, cultural heritage continues to get more and more and more
43 breadth in all the things that we're able to do, and we saw it when digital happened to
44 begin with, you know, everyone who does the acquisitions work and the collection
45 development work and the descriptive work and our physical preservation team, and
46 now we had to support these digitisation folks and the digital preservation folks. None
47 of it is well funded and now we're looking at AI. So how do we balance all of these
48 needs?
49

50
51 Similarly, PD404's description of acquisitions, collection development and descriptive work is
52 also central to what the archivists do and how they demonstrate their expertise.
53

54 *Bias*

55 Across all focus groups, there was discussion about concerns over bias. Others have
56 discussed how algorithms can become embedded with the bias of the programmers who
57 train them and how algorithms can exacerbate bias against marginalised communities
58 (Crawford, 2021). In the context of archives and digital preservation practitioners, concerns
59 about bias were linked with the ways in which AI technology could allow collections to be
60 used in ways that would further marginalise underrepresented communities, misrepresent

collections, and conflict with institutional diversity aims and objectives and the archival and digital preservation field in general.

Throughout the focus groups, participants were asked to reference what they knew of AI technology. In broaching the issues of AI and bias, participants in two different focus groups referenced the story of how Timnit Gerbu, a prominent AI Ethics researcher, was forced out of Google as a culmination of her criticism of the company's policies and practices associated with AI technology in relation to diversity (Hao, 2020). While participants did not mention Gerbu by name, the "Google case" was discussed: "a woman who was researching the ethical implications and biases of Google's algorithms was fired by Google and she wrote some- and she wrote some op eds and stuff. Anyway, what they did to her was not good (PD201). In addition, PD101:

So, it's worth looking into the Google researcher that was recently fired, asked to leave over the paper that she wrote. And I think that was a huge petition to kind of support her and things like that. And it was around questions of diversity and how these models can uphold existing structures of inequality and all that and also climate issues. And I think it was rejected by Google and the kind of a dodgy way they didn't want to publish it. And then she had to leave the company. And she's a very well respected researcher, who also happened to be a person of colour as well. So it was, it was really like, really troubling.

There was also a concern that users might extend their existing understanding of bias and AI to concern about use of AI in the archive for any purpose:

Yeah, you know race relations being such a forefront in academia, right now, like AI is just very untouchable for a lot of American academia at the moment. Especially like in the college where I'm working at, if I phrased it as AI it would not be good. If I phrased it as stylometry they might be interested. So, you know machine learning - maybe, but AI would definitely be a no go. (PD301)

PD301 was concerned about use of AI on the existing collections because users (academics and students in her case) would be concerned about AI and racial bias. This prior knowledge of AI technology and bias was applied to practitioners' own context in working with digital collections. For example, several participants acknowledged the colonialist past of some aspects of archival theory and practice (Punzalan and Caswell, 2016). There was an acute awareness that what they knew about AI and bias would further exacerbate marginalisation. According to PD404:

I would just say the bias question is a huge one, depending on what the data set is and what you're looking at and what it's being trained on, which continues to be my, my deepest fear about AI, even in sort of innocuous areas, like is this a bird or is this a plant, which happens more in my work. When we're talking about what is really going on in this image and if, since we're working in what is still a predominantly white field, and is still a field that is, is full of legacy data that we haven't been able to clean up before training the AI on records that were created seventy years ago that we haven't touched since I think it's really, really disturbing, especially when we're talking about descriptive, descriptive information if we're going beyond transcription translation, that kind of work and we're actually saying what do you see in this image AI and what do you think this is, that always makes me very squeamish, and I always want to do that, that thorough review before it goes up, um, and so it's it's definitely one of my big concerns.

For PD404, there was a concern that digitised collections in her care might need to be "cleaned up": a subtle reference to the concept that what was acceptable in the past is no longer acceptable now and if collections were not "cleaned up" harmful stereotypes could be

perpetuated. This also demonstrates some of the additional tasks a practitioner might need to take on when using AI technology. PD305 had similar concerns, which were expressed as “very strange views of the world”:

... There's instances of this happening, where the records have come through the system, and the, there is data in, in which there's personal nature, and then they have to be taken back out again from the public domain and that's with humans being involved in the process. So almost, it's a question of can the AI recognize the type of collection which is likely to have the things in it which might need a close look from a human, because clearly the examples you put up there are things which might not be obviously interpretable by a machine, so you could have like saying physical representations of people, or the language that was used to create those records, or which is in those records, say, colonial papers from the 1930s, are going to have some very strange views of the world. Em, which we wouldn't share today, some of us wouldn't anyway.

Viewed in the context of digital archival expertise, comments from PD404 and PD305 also suggest concern that an AI system may not have enough relevant expertise to understand stereotypes that were common in the past and that these stereotypes are no longer acceptable to reference in description of a collection today. The relevant expertise in this example is understanding of harmful stereotypes that could perpetuate marginalisation, as well as expertise about current values held in modern archival and digital preservation studies, such as a commitment to social justice.

PD401 was also concerned with the way in which AI technology could preference the voices of elites in a collection. This is exemplified in discussion of a hypothetical situation in which AI technology was used to search a large, digitised collection to meet a user request for information:

So, 85% of the population are missing from the archive. So, these are all balances that we have to come to, because somebody asks a question of the archive, what is the answer that they're getting back, are they hearing the voices of the people in charge, or are they hearing the voices of the people? And how to use and interpret that answer into something that you think is either fair and accurate? So yeah these are things that we do have to, we have thought a lot about, and we are careful as to what we will include on the bias of what we're including, and have lots of disclaimers, and this is what you're looking at kind of commentary throughout it. (PD401)

PD101 held a similar concern, and contextualised the concern in reference to his institution's diversity policy:

I've been hearing more about lately is the things like all of the bias that's involved when it comes to image recognition it is severely biased towards white men in particular, people of colour and women tend to be significantly underrepresented. And that's something that's quite interesting, I suppose, in a general sense in our fields, but I suppose specifically in [institution] where we are quite serious about our diversity policies and stuff like that. Which...that would be pretty much it, I guess, my understanding would be that, you know, the machine learning and the computer vision is really only as good as whatever like model you have like the actual like, what what data are you actually basing all this on and stuff like that, but what corpus...is it are you working with?

Discussion

One of the common threads that run throughout the different themes identified in the focus group data is the nature of digital archival expertise and how AI challenges and/or supports

1
2
3
4
5
6 the ways in which participants conceive of digital archival expertise.

7 For example, in discussing “fitting AI in the day to day practice”, there was concern that
8 using AI technology might result in additional work tasks for practitioners because the AI
9 system would lack certain expertise and the outputs would need to be “checked” by a
10 human. The unclear division of responsibilities towards the upkeep of an AI system to
11 ensure its reliability through time and the lack of specifics to the expertise of the person(s) in
12 charge of the task emerged among the major anxieties.

13 Following, in discussion of responsible use of AI technology in the context of archives,
14 participants discussed the concept of “checking” outputs; that led to suggestions of a human-
15 machine collaboration in which archives and digital preservation practitioners would provide
16 the expertise needed to make sure sensitive/personal records were not accidentally “missed”
17 and/or released to users when they should not be.

18 When participants discussed the need to manage the expectations of organisation
19 management and “higher ups,” this “managing of expectations” was discussed as an “art”-
20 the ability to explain the limits of what the AI technology was capable of and how use of the
21 AI technology would require a human to “check” the work, while simultaneously signposting
22 the importance of traditional archival activities such as arrangement and description of
23 collections. This need to “not forget” about traditional archival activities was framed as a
24 lesson learned from the adoption of digitisation in archives in the mid-1990’s which radically
25 changed the way users engage with and their expectations for access and use of archival
26 collections.

27 Finally, in discussing bias, this expertise was framed as knowledge of harmful stereotypes
28 and context surrounding information in digital collections which might act as datasets for
29 application of AI, the ability to steer how AI technology should be applied to digital archival
30 collections, as well as digitisation priorities in the context of the archives social justice
31 movement. Punzalan and Caswell (2016) include “inclusion of underrepresented and
32 marginalised sectors of society” as one of their five areas of social justice in archival studies.
33 Participants expressed concern about this specific issue-how AI technology may
34 inappropriately be applied by users in the context of marginalised and underrepresented
35 groups represented in the collections for which they care.

36 Much of the exiting archives literature about AI technology is focused on how AI may impact
37 the archives field: this includes how AI may change traditional archival work such as
38 arrangement, description and appraisal via automation (Lee, 2018); how traditional archival
39 theory such as original order and provenance may change (Colavizza *et al.*, 2022); and how
40 AI technology will require archivists to learn new computational skills (Marciano *et al.*, 2018;
41 Payne, 2018). Moss *et al.* (2018) and Theimer (2018) specifically address the concept of
42 archival expertise and how it may change in the context of AI, suggesting that expertise
43 would grow to closely resemble data science.

44 Susskind and Susskind (2015) in *The future of the professions: How technology will*
45 *transform the work of human experts* predict the decline of professions that require expertise
46 like law and medicine because of the rise of advanced technology, including AI. Theimer
47 (2018) applies this perspective to the archival profession by predicting that, facing the
48 growing use of AI technology, archivists will attempt to challenge the rise of automation by
49 highlighting the work that a machine can’t do which requires “creative thought” (p. 11) and
50 making the argument that “a machine can’t do all the parts of my job” (p. 11). Archival tasks
51 that are difficult to automate, that require “a human touch” (p. 12) will be highlighted.

52 The thought that the “machine can’t do all the parts of my job” was present throughout the
53 focus group participants’ discussion. For example, in all themes, digital archival expertise
54 was framed as expertise needed to “check” outputs of the AI, the ability to provide context
55 for collections, and to steer the use of algorithms to tasks that would not conflict with social
56 justice values held by the archival profession, including not compounding marginalisation of
57 under-represented groups.
58
59
60

1
2
3
4
5
6 **Theimer (2018) counters that while archivists may try to make the “a machine can’t do**
7 **all the parts of my job” argument, Susskind and Susskind (2015) argue that**
8 **eventually, all tasks will become routinised in one way or another, such as tasks**
9 **being completed in an entirely** different manner in which the end result is similar enough
10 to the “non-routinised task” (the parts the machine can’t do). She also predicts that what will
11 matter to users is “delivering an acceptable level of service as freely and broadly as
12 possible” (p. 12) which contradicts the argument that users will always value the “human
13 touch” work that archivists do (and machines can’t currently do)—the digital archival
14 expertise.

15 In the focus groups, there was little concern that AI technology would completely replace the
16 work of a professional archivist—this was underpinned by a focus on digital archival
17 expertise: the ability to manage expectations, the contextual expertise an archivist can
18 provide to collections, and the dedication to upholding social justice in archives principles,
19 namely, to not continue to marginalise underrepresented groups and perpetuate bias.

20 However, in light of the Theimer piece, this discussion of “digital archival expertise” can be
21 explained as an offensive mounted against the rising use of intelligent systems such as AI
22 technology.

23 As such, when one asks, what are the thoughts and opinions of archival practitioners about
24 AI, as a way to understand how AI adoption may impact archival work? The focus group
25 data suggests that AI, as a disruptive technology, impacts the ways in which archivists
26 characterise and communicate digital archival expertise. AI may cause a reaction to revise
27 the ways in which digital archival expertise is highlighted and presented to “the higher ups,”
28 users and the public. In this sense, AI technology in archives acts as a *counter to balance*
29 *digital archival expertise against*. This is the potential “impact” of AI on digital archives
30 practice—it will cause a change in the characteristics of digital archival expertise and the
31 elements of the expertise that are advocated and communicated to interested parties. Our
32 findings suggest that the way that digital archival expertise is characterised will slowly evolve
33 to represent “what the machines can’t do.” For our participants, digital archival expertise was
34 described as arrangement, description and appraisal, managing expectations, understanding
35 and contextualising user needs and a dedication to social justice initiatives. We predict this
36 will change, and the speed with which it will change will depend on how quickly those tasks
37 can become routinised.

38 The focus of this article was to explore thoughts and opinions of AI technology in archival
39 practice, as a way to learn more about how AI technology may impact archival work. Results
40 suggest that the impact may rest in *how* digital archival expertise is characterised,
41 highlighted, and communicated to interested parties. The specific characteristics of the
42 digital archival expertise and how it may evolve in the future is beyond the scope of the
43 current work and could be explored in future research projects. The *impact* of AI on archival
44 and digital preservation practice in this study may be summarised as a force that triggers an
45 evolution in how digital archival expertise is characterised, discussed and highlighted.

46 However, it is not acting alone: several of the focus group participants referenced the time
47 period when digitisation “happened” in the mid 1990s in discussing potential AI adoption. As
48 such, AI technology, *combined with* understanding of how to adopt a disruptive technology
49 “better” by applying lessons learned from the digitisation era of the mid-1990s may be
50 prompting the desire to re-evaluate and change the ways in which digital archival expertise
51 is discussed.

52 AI technology, along with the desire to apply lessons learned from the implementation of
53 mass digitisation, are acting as a trigger for archival practitioners to re-evaluate the way they
54 discuss their contributions and their practice, which may have continued knock on effects for
55 activities such as advocacy and outreach. Of course, AI technology will impact workflows
56 and archival activities, but that was not found to be of most concern to our participants when
57 we spoke to them.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Lastly, the results can be situated in the context of a social construction of technology perspective to produce greater insight. According to Baym (2015), in brief, the social construction of technology (SCOT) perspective “focuses on how technologies arise from social processes” (p. 44). Applied to our example, a SCOT perspective would have investigated the social contexts of archivists to understand their use and adoption of AI. We would not say that we explicitly used a SCOT theoretical framework to organise this study—we have focused on opinions and perceptions which are only part of a greater social context. That being said, we can use SCOT as one lens through which to view our results.

In contrast to SCOT, technological determinism aligns with the belief that technology changes us. We would not go as far as to qualify participant concerns in this study as techno deterministic—AI technology is not *causing* a change in expertise. In contrast, AI technology combined with hindsight about mass digitisation as a disruptive technology are *affording* archival practitioners the opportunity to reinterpret their concepts of digital archival expertise. This perspective is more aligned with a social construction of technology (SCOT) perspective (Baym, 2015). Future research could explore this overlap using an explicit SCOT theoretical framework. In addition, we would argue that any AI adoption in the archives sector will need to address the issue of evolving concepts of digital archival expertise to move forward with large scale adoption of any new technology.

Conclusion

We began this project posing the following questions: what are the thoughts and opinions of archival and digital preservation practitioners concerning the use of AI technology in their work? In what ways can these thoughts and opinions help us understand the impact of AI technology on archives and digital preservation work? We conducted four focus groups across two phases of data collection. As a result, findings were not generalisable. Using template analysis, we were able to identify four themes from the data, which suggests that AI technology *combined with* hindsight about digitisation as a disruptive technology may prompt archival and digital preservation practitioners to change the way they characterise and communicate digital archival expertise. The specific content of this expertise is beyond the scope of this article but could be addressed in future work. **Investigating further the adoption of AI in specific archival contexts would benefit the understanding we have of the evolution of digital archival expertise, going beyond sentiment and perception mapping.**

Bibliography

- Barucci, A., Cucci, C., Franci, M., Loschiavo, M. and Argenti, F. (2021), “A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification”, *IEEE Access*, Vol. 9, pp. 123438–123447, available at: <https://doi.org/10.1109/ACCESS.2021.3110082>
- Baym, N.K. (2015), *Personal Connections in the Digital Age*, John Wiley & Sons.
- Bellman, R. (1978), *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company.
- Benoit, E., III. (2017), “#MPLP Part 1: Comparing Domain Expert and Novice Social Tags in a Minimally Processed Digital Archives”, *The American Archivist*, Vol. 80 No. 2, pp. 407–438, available at: <https://doi.org/10.17723/0360-9081-80.2.407>
- Benoit, E., III. (2018), “#MPLP Part 2: Replacing Item-Level Metadata with User-Generated Social Tags”, *The American Archivist*, Vol. 81 No. 1, pp. 38–64, available at: <https://doi.org/10.17723/0360-9081-81.1.38>
- Bermeitinger, B., Freitas, A., Donig, S. and Handschuh, S. (2016), “Object Classification in Images of Neoclassical Furniture Using Deep Learning”, in Bozic, B., Mendel-

- Gleason, G., Debruyne, C. and O'Sullivan, D. (Eds.), *Computational History and Data-Driven Humanities*, Springer Publishing, Cham, pp. 109–112, available at: https://doi.org/10.1007/978-3-319-46224-0_10
- Braun, V. and Clarke, V. (2021), *Thematic Analysis: A Practical Guide*, SAGE.
- Brügger, N. (2016), "Digital Humanities", edited by Pooley, J.D. and Rothenbuhler, E.W. *The International Encyclopedia of Communication Theory and Philosophy*, John Wiley & Sons.
- Bunn, J. (2020), "Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI)", *Records Management Journal*, Vol. 30 No. 2, pp. 143–153, available at: <https://doi.org/10.1108/RMJ-08-2019-0038>
- Carlsen, B. and Glenton, C. (2011), "What about N? A methodological study of sample-size reporting in focus group studies", *BMC Medical Research Methodology*, Vol. 11 No. 26, pp. 1–10.
- Chabin, M. A. (2020), "The potential for collaboration between AI and archival science in processing data from the French great national debate", *Records Management Journal*, Vol. 30 No. 2, pp. 241–252, available at: <https://doi.org/10.1108/RMJ-08-2019-0042>
- Clanuwat, T., Lamb, A. and Kitamoto, A. (2019), "KuroNet: Pre-Modern Japanese Kuzushiji Character Recognition with Deep Learning", *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 607–614, available at: <https://doi.org/10.1109/ICDAR.2019.00103>
- Colavizza, G., Blanke, T., Jeurgens, C. and Noordegraaf, J. (2022), "Archives and AI: An Overview of Current Debates and Future Perspectives", *Journal on Computing and Cultural Heritage*, Vol. 15 No. 1, pp. 1–15, available at: <https://doi.org/10.1145/3479010>
- Crawford, K. (2021), *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven.
- Duff, W. M., Flinn, A., Suurtamm, K. E. and Wallace, D. A. (2013), "Social justice impact of archives: a preliminary investigation", *Archival Science*, Springer, Vol. 13 No. 4, pp. 317–348.
- Eveleigh, A. (2014), "Crowding Out the Archivist? Locating Crowdsourcing within the Broader Landscape of Participatory Archives", pp. 211–230.
- Garcia, N. and Vogiatzis, G. (2019), "How to Read Paintings: Semantic Art Understanding with Multi-modal Retrieval", in Leal-Taixé, L. and Roth, S. (Eds.), *Computer Vision – ECCV 2018 Workshops*, Springer, Cham, pp. 676–691.
- Gefen, A., Saint-Raymond, L. and Venturini, T. (2021), "AI for Digital Humanities and Computational Social Sciences", in Braunschweig, B. and Ghallab, M. (Eds.), *Reflections on Artificial Intelligence for Humanity*, Springer, Cham, pp. 191–202.
- Glaser, B.G. and Strauss, A.L. (2017), *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Routledge, New York, available at: <https://doi.org/10.4324/9780203793206>.
- Greene, M. and Meissner, D. (2005), "More Product, Less Process: Revamping Traditional Archival Processing", *The American Archivist*, Vol. 68 No. 2, pp. 208–263.
- Hao, K. (2020), "We read the paper that forced Timnit Gebru out of Google. Here's what it says.", *MIT Technology Review*, 4 December, available at: <https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/> (accessed 2 August 2022).
- Jaillant, L. (2022), "How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions", *Archival Science*, available at: <https://doi.org/>
- Jaillant, L. and Caputo, A. (2022), "Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data", *AI & Society*, pp. 1–13, available at: <https://doi.org/10.1007/s00146-021-01367-x>

- Jain, N., Bartz, C., Bredow, T., Metzenthin, E., Otholt, J. and Krestel, R. (2021), "Semantic Analysis of Cultural Heritage Data: Aligning Paintings and Descriptions in Art-Historic Collections", in Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., et al. (Eds.), *Pattern Recognition. ICPR International Workshops and Challenges*, Vol. 12663, Springer Publishing, Cham, pp. 517–530.
- Kestemont, M., de Pauw, G., van Nie, R. and Daelemans, W. (2016), "Lemmatization for variation-rich languages using deep learning", *Digital Scholarship in the Humanities*, pp. 797–815.
- King, N. (2012), "Doing Template Analysis", *Qualitative Organizational Research: Core Methods and Current Challenges*, SAGE, London, pp. 426–450.
- King, N. and Brooks, J.M. (2017), *Template Analysis for Business and Management Students*, *Template Analysis for Business and Management Students*, SAGE, London.
- Lee, C.A. (2018), "Computer-Assisted Appraisal and Selection of Archival Materials", *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2721–2724.
- Manovich, L. (2017), "Cultural Data. Possibilities and limitations of the digital data universe", *Museum and Archive on the Move. Changing Cultural Institutions in the Digital Era*, De Gruyter, Berlin, pp. 259–276.
- Marciano, R., Lemieux, V., Hedges, M., Esteva, M., Underwood, W., Kurtz, M. and Conrad, M. (2018), "Archival Records and Training in the Age of Big Data", in Percell, J., C. Sarin, L., T. Jaeger, P. and Carlo Bertot, J. (Eds.), *Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education*, Vol. 44B, Emerald Publishing, pp. 179–199.
- Morgan, D. (1998), "What Do You Get from Focus Groups?", *The Focus Group Guidebook*, SAGE Publications, Inc., Thousand Oaks, pp. 55–64.
- Moss, M., Thomas, D. and Gollins, T. (2018), "The Reconfiguration of the Archive as Data to Be Mined", *Archivaria*, Association of Canadian Archivists, Vol. 86 No. 86, pp. 118–151.
- Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E. and Varner, S. (2019), "Final Report --- Always Already Computational: Collections as Data", Zenodo, available at: <https://doi.org/10.5281/zenodo.3152935>.
- Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E. and Varner, S. (2019a), "Santa Barbara Statement on Collections as Data --- Always Already Computational: Collections as Data", Zenodo, available at: <https://doi.org/10.5281/zenodo.3066209>.
- Payne, N. (2018), "Stirring The Cauldron: Redefining Computational Archival Science (CAS) For The Big Data Domain", *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, Seattle, USA, pp. 2743–2752.
- Poole, A.H. and Garwood, D.A. (2018), "'Natural allies': Librarians, archivists, and big data in international digital humanities project work", *Journal of Documentation*, Emerald Publishing, Vol. 74 No. 4, pp. 804–826, available at: <https://www.doi.org/10.1108/JD-10-2017-0137>
- Proctor, J. and Marciano, R. (2021), "An AI-Assisted Framework for Rapid Conversion of Descriptive Photo Metadata into Linked Data", *2021 IEEE International Conference on Big Data (Big Data)*, IEEE, Orlando, USA, pp. 2255–2261.
- Punzalan, R.L. and Caswell, M. (2016), "Critical Directions for Archival Approaches to Social Justice", *The Library Quarterly*, The University of Chicago Press, Vol. 86 No. 1, pp. 25–42, available at: <https://www.doi.org/10.1086/684145>
- Randby, T. and Marciano, R. (2020), "Digital Curation and Machine Learning Experimentation in Archives", *2020 IEEE International Conference on Big Data (Big Data)*, IEEE, Atlanta, USA, pp. 1904–1913.
- Roeschley, A. and Benoit, E.I. (2019), "Chapter 14. Degrees of mediation: a review of the intersectionality between community and participatory archives", *Participatory*

- 1
2
3
4
5
6 *Archives*, 1st ed., Facet Publishing.
- 7 Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsouпова, T. and Stuart, K.J. (2019),
8 “More human than human? Artificial intelligence in the archive”, *Archives and*
9 *Manuscripts*, Vol. 47 No. 2, pp. 179–203, available at:
10 <http://www.doi.org/10.1080/01576895.2018.1502088>
- 11 Sabharwal, A. (2017), “Digital humanities and the emerging framework for digital curation”,
12 *College & Undergraduate Libraries*, Routledge, Vol. 24 No. 2–4, pp. 238–256,
13 available at: <https://www.doi.org/10.1080/10691316.2017.1336953>
- 14 Springstein, M., Schneider, S., Rahnama, J., Hüllermeier, E., Kohle, H. and Ewerth, R.
15 (2021), “iART: A Search Engine for Art-Historical Images to Support Research in the
16 Humanities”, *Proceedings of the 29th ACM International Conference on Multimedia*,
17 Association for Computing Machinery, New York, USA, pp. 2801–2803.
- 18 Steinmeier, D. (2020), “Diversity, Inclusion, and Digital Preservation”, *Patterns*, Vol. 1 No. 9,
19 pp. 1–4, available at: <https://www.doi.org/10.1016/j.patter.2020.100152>
- 20 Susskind, R.E. and Susskind, D. (2015), *The Future of the Professions: How Technology*
21 *Will Transform the Work of Human Experts*, Oxford University Press.
- 22 Tanasescu, C., Kesarwani, V. and Inkpen, D. (2018), “Metaphor Detection by Deep Learning
23 and the Place of Poetic Metaphor in Digital Humanities”, *The Thirty-First International*
24 *Florida Artificial Intelligence Research Society Conference (FLAIRS-31)*, Association for
25 the Advancement of Artificial Intelligence, pp. 122–127.
- 26 Theimer, K. (2018), “It’s the end of the archival profession as we know it, and I feel fine”,
27 *Archival Futures*, Facet Publishing, London, pp. 1–17.
- 28 Wevers, M. and Smits, T. (2020), “The visual digital turn: Using neural networks to study
29 historical images”, *Digital Scholarship in the Humanities*, Vol. 35 No. 1, pp. 194–207,
30 available at: <https://www.doi.org/10.1093/lc/fqy085>
- 31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60