



Applying AI to digital archives: trust, collaboration and shared professional ethics

Lise Jaillant ^{1,*}, Arran Rees ²

¹School of Social Sciences and Humanities, Loughborough University, Loughborough, UK

²School of Fine Art, History of Art and Cultural Studies, University of Leeds, Leeds, UK

*Correspondence: Lise Jaillant, School of Social Sciences and Humanities, Loughborough University, Loughborough, UK.
Email: l.jaillant@lboro.ac.uk

Abstract

Policy makers produce digital records on a daily basis. A selection of records is then preserved in archival repositories. However, getting access to these archival materials is extremely complicated for many reasons—including data protection, sensitivity, national security, and copyright. Artificial Intelligence (AI) can be applied to archives to make them more accessible, but it is still at an experimental stage. While skills gaps contribute to keeping archives ‘dark’, it is also essential to examine issues of mistrust and miscommunication. This article argues that although civil servants, archivists, and academics have similar professional principles articulated through professional codes of ethics, these are not often communicated to each other. This lack of communication leads to feelings of mistrust between stakeholders. Mistrust of technology also contributes to the barriers to effective implementation of AI tools. Therefore, we propose that surfacing the shared professional ethics between stakeholders can contribute to deeper collaborations between humans. In turn, these collaborations can lead to the building of trust in AI systems and tools. The research is informed by semi-structured interviews with thirty government professionals, archivists, historians, digital humanists, and computer scientists. Previous research has largely focused on preservation of digital records, rather than access to these records, and on archivists rather than records creators such as government professionals. This article is the first to examine the application of AI to digital archives as an issue that requires trust and collaboration across the entire archival circle (from record creators to archivists, and from archivists to users).

For there to be betrayal, there would have
to have been trust first.

— Suzanne Collins, *The Hunger Games*

1 Introduction

Born-digital archives¹ have been acquired and preserved for the best part of thirty years. Yet, access to these records remains a complex challenge for the institutions that hold and manage them, as well as the researchers seeking to use them. Obstacles to access are numerous and include issues with data protection, sensitivity, and copyright. In the case of government archival collections, releasing potentially sensitive and private information could threaten national security and embarrass foreign partners. However, not all collections present the same level of risk, and giving access to digital records is essential to make government accountable and enable the writing of history.

As born-digital records increasingly represent the largest part of new accessions, finding solutions to these obstacles is becoming a key priority. Unlocking vast amounts of digital data is not a task that can be done manually. Artificial Intelligence² (AI) has the potential to make born-digital archives more accessible and usable. Sensitive information can be automatically identified, making possible the release of non-sensitive data. AI can also be used when keyword search is not effective (for example in the case of web archives, which include terabytes of information). An AI-driven process of recommendation—similar to the functionality ‘customers who bought this item also bought’—could enable discovery of previously inaccessible archival materials.

Several projects have recently sought to identify key challenges and find solutions to the problem of locked born-digital archives, using AI and other advanced technologies.³ Yet, AI applied to archives remains at the experimental stage; rarely implemented beyond small data sets and collections identified for

2 Context, related work, and methodology

Substantial work has been undertaken over the past thirty years or so to establish effective practices for preserving born-digital materials in the GLAM sector (Deegan and Tanner, 2006; Delve and Anderson, 2014; Corrado and Sandy, 2017; Harvey and Weatherburn, 2018; Owens, 2018). Influential support organizations like the Digital Preservation Coalition have helped embed and sustain the work of digital preservation practitioners across a range of cultural heritage organizations. However, while preservation processes are becoming ever more integrated into the core work of GLAM institutions, processes for providing access to those same born-digital collections has been slower to develop (Jaillant, 2019). There are multiple reasons for this, from content-related legal issues such as privacy concerns and copyright, to technology issues that make it difficult to provide access to materials in appropriate formats.

The legal concerns mostly relate to the General Data Protection Regulations (GDPR) in the European Union, the Data Protection Act 2018 in the UK and Copyright legislation. The GDPR and Data Protection legislation place a responsibility to build in ‘privacy by design’ on organizations and institutions handling personal data, but also provide exemptions to aspects of the regulations when personal data is being archived in the public interest (Lomas, 2019). Whilst The National Archive’s *Guide to Archiving Personal Data* (2018) makes clear that the updated data protection does not prevent archiving, it does influence how archival institutions need to articulate the purposes of archiving personal data and how they provide access to collections containing it (The National Archives, 2018). Records deemed to contain personal data are carefully managed and normally kept closed from public access during the lifetime of the individual the data is relating to (The National Archives suggests that a lifespan of 100 years should be presumed, unless proven otherwise).⁴ However, in some circumstances access to these records can be granted, and decisions need to be taken with respect to the safeguarding measures defined in the Data Protection legislation, and fully documented by the archive (The National Archives, 2018).

Additional concerns relate to Copyright legislation which has an impact on how archival material can be processed and made accessible. The application of copyright law on born-digital materials in GLAM institutions is an ongoing and complex issue. Many born-digital collections contain content from multiple copyright holders, described at varying levels of detail and often without proof of consent. For example, an archived website may have personal information, images, creative writing, or videos embedded within it. It is not possible to assume that all the content belongs to the

person or organisation that created it, and it is often not possible to trace who the copyright holder is. Many GLAM institutions with web archives face difficult decisions about the amount of their collections they can make available whilst managing the potential risks of making it available (Hockx-Yu, 2014; Vlassenroot *et al.*, 2019). Copyright legislation in the UK and in the EU has acknowledged the digitization of physical material as a form of preservation in GLAM institutions and has developed exemptions to support this. However, the management of born-digital material is largely ignored in the legal context. Creating preservation and access copies of born-digital content remains a grey area (Koščík and Myška, 2019).

With so many digital collections closed to users due to these issues with data protection and copyright, AI has been presented as a possible solution to the problem of locked archives (Jaillant and Caputo, 2022; Jaillant 2022a, 2022b). For the purposes of this article, we refer to AI in a broad sense. Instead of focusing on particular subsets of AI such as machine learning, computer vision, or natural language processing, we consider AI as a broad suite of computational approaches used to make decisions and complete tasks. Although often discussed through the lens of new and emerging technologies, AI has a longer history, with waves of excitement and periods of disinterest (Jackson, 2019; Agar, 2020). The current period is characterized by sustained development, and AI technologies are increasingly featuring in all aspects of our everyday lives. The role of AI technologies in making born-digital archives more accessible and usable is increasingly acknowledged as an important direction for record creators and managers (including civil servants responsible for knowledge, information, and records management), GLAM professionals, and researchers interested in using the archival material.

Examples of AI tools used to increase the accessibility and usability of digital archives include: the automatic sensitivity classification of government records before being released through Freedom of Information requests or transferred to national archives (The National Archives UK, 2016; McDonald *et al.*, 2020a; Baron *et al.*, 2020); experimentations with computer vision to enhance the discovery of visual archival collections (Angelova *et al.*, 2020); and algorithmic search tools (Nix and Decker, 2021). Additionally, scholarship and research networks, such as AURA, AEOLIAN and AI4LAM bring together multiple stakeholders interested in the application of AI to archival collections.

As awareness of AI technologies increase, so does public scrutiny, and ethical questions, including ones on transparency and bias, begin entering public discourses (Fast and Horvitz, 2017; Cave *et al.*, 2019; Kelley *et al.*, 2021). Calls to develop more widely

that all professionals working in the civil service, GLAM sector and academia consult their codes of practice or ethics before they make new decisions or undertake a new task, but that these codes broadly influence the principles and areas of concern for the sectors.

Concepts of trust, transparency, and professional ethics emerged as a recurring theme throughout ‘Unlocking our Digital Past’. The project was part of an ongoing collaborative relationship between Loughborough University and the Cabinet Office of the UK Government that focuses on the application of AI to born-digital archives. The project facilitated a series of workshops aimed at increasing interdisciplinary dialogue between civil servants, GLAM sector professionals, and academics. After obtaining approval from the Ethics Committee at our institution, our team of Digital Humanists conducted thirty interviews with key thinkers and practitioners who have a stake in the accessibility and usability of born-digital archives now and in the future. Our background is in cultural history, literary studies and museum studies, rather than computer science. This has informed the nature of our questions: instead of focusing on the technical aspects of AI, the questions we asked centred on the key obstacles to making born-digital and digitised collections more accessible, and on the possible solutions to these issues. We were interested in the day-to-day practice of professionals—from record creators who need to prepare archival records ahead of transfer to The National Archives and other repositories, to archivists who need to preserve and make accessible these records. We also interviewed historians who need access to archives (including born-digital archives) for their research. A common thread of these interviews was the lack of trust in other stakeholders and in technology. In turn, this mistrust has an impact on the accessibility and usability of digital archives.

3 Mistrust of other stakeholders and of technology

Collaborative discussions between professionals in government, GLAMs and academia are the exception rather than the norm. This, in turn, has an impact on the way professionals see other professionals, often leading to misunderstandings. James Baker, who studied history and worked as a curator at the British Library before moving to academia, gave an example of this lack of dialogue between archivists and researchers. While archivists see appraisal and selection of records as a central aspect of their role, researchers often expect complete archives. In other words, academics view archivists as record keepers, rather than professionals who must make a selection and throw

away records that lack lasting value. This task of appraisal is particularly complicated due to the scale of born-digital archives. Baker said:

I always feel very sorry for my archivist friends every time historians turn round and say, well, you must collect everything, and why aren't you trying, you know, we need all this stuff and it's going to be a digital dark age. And they're like, it's not going to be a digital dark age, we always do selection, you know, it's just that selection issues are kind of way more complicated now in the age of the born digital.¹³

There was a consensus among our interviewees who engage closely with born-digital records: keeping and describing everything is not a viable option. Clifford Lynch, Executive Director of the Coalition for Networked Information, pointed out that in the very early days of the internet, there were attempts to apply library-style cataloguing to internet resources. But these ‘were rapidly abandoned because the scale was just unmanageable’.¹⁴ The Google-style, computational approach to records became the norm to search records at scale.

The problem of scale is particularly acute for government digital records. Jason Baron, who served for 13 years as Director of Litigation for the US National Archives and Records Administration (NARA) before moving to the private sector and then academia, gives the example of US presidents' email records:

The National Archives in the US today has 500 million emails from the Regan administration to the Obama administration. The Trump emails haven't been counted yet or haven't been fully processed. But it has 500 million distinct emails, and then more than a billion pages because those emails have attachments.¹⁵

In this role at NARA, Baron pushed for an approach that would diminish the number of government emails transferred to archival repositories. This CAPSTONE approach adopted in 2013 led to an email preservation policy based on the seniority of the record creator.¹⁶ Individuals at the top or near the top of an agency have all their emails permanently preserved, while others see their emails destroyed after a certain period (generally seven years).

CAPSTONE addressed the issue of email exponential increase, but emails are of course not the only born-digital records produced by US government. Leslie Johnston (Director of Digital Preservation at NARA) notes that the Trump administration produced about 500 terabytes of archival records—compared to

200 terabytes for the two terms of the Obama administration.¹⁷ Libraries and archival institutions lack the staff necessary to handle enormous amounts of data. Baron gives the example of the Clinton Presidential Library, which has six archivists and ‘a 10 million document queue estimate for Freedom of Information Act requests’.¹⁸ This leads to a lack of accountability in the short term, and risk impacting the cultural memory in the long term.

Jason Baron and others have argued that automation is a necessity rather than a choice. ‘How do you search a billion objects?’, asks Baron. ‘Well, you don’t do it manually, as an archivist, and you don’t do it with keyword searches, although you can try, but those are terribly inefficient’.¹⁹ Indeed, an approach based on keyword search does not work well with collections characterized by their huge size and lack of metadata. For instance, a search for ‘Brexit’ on the UK web archive gives 63,968,885 results.²⁰ As Leontien Talboom (Doctoral Researcher at The National Archives UK and University College London) told us, analysing each record manually is not an option.²¹ Instead, a computational approach is the only way to make sense of this mass of results. ‘How can they have any success of opening a substantial amount of history for the 21st century and before, if they don’t use machine learning techniques to perform searches?’ Baron inquires.²² It should be noted that AI-assisted searches are not perfect solutions, since they might also miss relevant results.²³

The application of AI and machine learning to archival records is not a new thing. In the early 2000s, Jason Baron was part of the government legal team involved in a giant tobacco suit against Philip Morris and other tobacco companies. Key information was scattered in the Clinton email archive, which included 30 million emails. Baron decided to seek out information scientists and computer scientists to figure out a better way for lawyers to search, using other techniques than keyword searching.²⁴ At around the same time, commercial firms started marketing eDiscovery software that drew on machine learning to search and find information in vast amounts of data. The archive sector on both sides of the Atlantic then investigated the potential of eDiscovery in solving the challenges of born-digital records. In 2016, a report from The National Archives (TNA) in the UK concluded that ‘technology-assisted review using eDiscovery software can support government departments during appraisal, selection and sensitivity review as part of a born-digital records transfer’ to TNA (*The National Archives UK, 2016*, p. 5).

Researchers such as Graham McDonald (Lecturer in Information Retrieval at the University of Glasgow) have shown that AI and machine learning can indeed unlock government archives, using two main approaches: ‘protect then search’ or ‘search then

protect’.²⁵ McDonald’s work has used mostly the first approach. Tools that his team developed can help humans identify the sensitivities in a particular collection, and then make informed decisions before transferring records to archival institutions. For example, record creators can choose to protect sensitive information by redacting it, ahead of transfer to archives. This process lowers the level of risk in providing access to that collection.

The second approach consists in making a collection entirely accessible, with a search framework that hides sensitive information when it comes across it. The machine is trained using a subset of data identified as sensitive or confidential and is then expected to identify sensitivity in larger datasets. The problem is that there is no universal definition of sensitivity. It largely depends on the context, and the machine is not always able to correctly identify contextual information that makes a record sensitive.²⁶

While Graham McDonald recognizes that sensitivity review is still an imperfect process, the risk of giving access to archival records is worth taking, according to Jason Baron. AI ‘will get a tremendous amount of return’ even though it comes with a risk of bias.²⁷ The most important thing is to allow researchers, journalists, historians, and other users to make discoveries that will serve the public good. The stakes are lower than in the criminal justice system, where AI has in the past led to disastrous results [e.g. algorithms used in the USA to evaluate the risk of prisoners to re-offend were shown to be racially biased (*Wadsworth et al., 2018*)]. Despite the risk that comes with AI, the police and justice system on both sides of the Atlantic see advanced technologies as a necessary tool in the fight against crime. In 2014, the UK Home Office released a report on eDiscovery in digital forensic investigations (*Lawton et al., 2014*). Four years later, it announced the development of new technology to automatically detect terrorist content on online platforms.²⁸

If AI is routinely deployed by police officers and judges to analyse large datasets, why is it rarely applied to government archives? It is still exceptional to use AI for sensitivity review, or to find results in vast digital collections. In order to train AI systems, access to data is needed. And this notion of access (even limited access to training datasets) is problematic for many record creators and archivists. When users are not trusted to make the right decisions, it seems logical to close entire collections—or to restrict access only to groups who are seen as trustworthy. For example, web archives generally require users to travel to the library due to copyright reasons. Content is available onsite via browsers, rather than offsite via downloading data. Jane Winters, Professor of Digital Humanities at the University of London, notes that Denmark has a

researcher exception to this restrictive access policy. Researchers attached to a Danish university, as well as overseas collaborators, can get remote access to web archive data anywhere in the world. ‘You have to sign waivers and so on that you will treat the data appropriately’, Winters said. This can be seen as a first step ‘to start to think about opening things up’.²⁹

The risk of releasing potentially sensitive data should be balanced against the risk of keeping records dark and inaccessible. Indeed, access is at the centre of ethical codes in the GLAM sector. The Museums Association puts ‘public engagement and public benefit’ as the first principle in its Code of Ethics. More specifically, the code states that museums and other cultural heritage organisations should ‘provide public access to, and meaningful engagement with, museums, collections, and information about collections without discrimination’. Likewise, the UK’s library and information association CILIP declares that ‘preservation and continuity of access to knowledge’ is central to libraries’ mission. Similarly, the International Council on Archives asserts that ‘archivists should promote the widest possible access to archival material and provide an impartial service to all users’. Of course, access is not the only aspect that GLAM professionals need to consider. Codes of ethics include references to privacy and copyright as limits to access. For the ICA, ‘archivists should respect both access and privacy’. Yet, the balance is often skewed towards privacy and data protection.

While archivists must respect legal frameworks and protect the private data of individuals, they could also take limited risks and make some collections available without necessarily checking all materials. The need for trust and shared codes of practice and conduct was highlighted by our interviewees. Jane Winters (University of London) gave the example of personal archives, rather than web archives and other large-scale collections. ‘Relationships of trusts between depositing authors, archivists, and researchers’ are central. However, these relationships are also ‘time-consuming’ and ‘excluding because if you happen to know the librarians it’s much easier to have those conversations than if you’re coming to it cold’. In other words, access is often decided on a one-to-one basis based on preliminary discussions with archivists. ‘Having those negotiations and building relationships of trust around the use of this material’ can unlock previously closed collections.³⁰ Winters’s comments echo the experience of literary scholars who have had access to previously closed collections. Lise Jaillant has thus written about her experience of accessing the email archive of the British writer Ian McEwan at the Harry Ransom Center (HRC) in Texas, which is officially inaccessible to researchers. Access was granted through personal contacts, and the fact that the HRC archivist was participating in the same research project (Jaillant 2019).

Providing ad hoc access to known users is not a longstanding solution since it excludes entire groups—for example, independent scholars or family historians who lack a university affiliation and professional networks. As we have seen, access without discrimination is central to GLAM codes of ethics. Adam Nix, Lecturer in Responsible Business at the University of Birmingham, favours a systematic policy of access for those who respect certain codes of behaviour:

I would like to see more focus on sharing that duty of care between the archives and the users. I think users need to take a far more active and conscious role in maintaining ... the integrity of the archival discovery process. And I think that can be done by having well developed and well-respected codes of conduct that the user very consciously agrees with beforehand, which means that even if they see sensitive information, that sensitive information goes no further than that individual researcher.³¹

This proposed process of opening up archives to users who behave in an ethical manner is based on trust. But it is also based on the possibility of sanctions (including legal sanctions) for those who transgress codes of conduct. A similar system of trust and possible sanction is described in Frances Harris and Fergus Lyon’s research on collaborations across professional cultures. Their central findings are that first, trust is a vital ingredient when collaborating across disciplines and sectors. Second, maintaining collaboration across professional cultures presents particular challenges for building trust. Third, trust in interdisciplinary and transdisciplinary teams is based on norms, information, sanctions, and controls (Harris and Lyon, 2013).

John Sheridan, Digital Director at The National Archives UK, is in favour of quantifying risk and therefore trust. ‘Looking at our processes around digitization, all tend to lean much more heavily on managing the risks through expert knowledge, than on systems’, Sheridan said.³² He advocates for a computational approach using statistical models to assess risk. To this purpose, TNA has explored a Bayesian network, i.e. a graphical model that represents a set of variables and conditions, often used for probability analysis (Barons *et al.*, 2021). For Sheridan, archival institutions should strike the right balance between risk and access to potentially sensitive materials. It is not a case of ‘transparency above everything’, as:

archives are not Wikileaks, and we’re not in the WikiLeaks business. ... It’s not responsible to data subjects; it’s not responsible to other people’s intellectual property rights; it’s not lawful. So, we then

An example of this can be seen through the Digital Sensitivity Review project being undertaken between the Foreign, Commonwealth and Development Office (FCDO), and the consultancy company SVGC in partnership with the University of Glasgow.⁴⁴ Andrew Dixon, Managing Director of SVGC, explained that in using AI tools to make decisions about public records to be archived, the public has a right to know how those selections are being made. This is of course derived from the civil service's professional code requirement for accountability and transparency. Dixon explained that this influenced the tools they use in the project: 'we're not using neural networks which are unexplainable... we've avoided technologies that create random outcomes'—instead opting to use the kind that are 'predictable and repeatable technologies... so you can determine the outcome'.⁴⁵ Here we see how the principles and professional codes of ethics that govern the civil service, GLAM sector and academic researchers are influencing the technology being used.

McDonald, who also works on this project, pointed out that intuitive explanations of what is happening when you apply an AI tool is crucial for building trust—especially when those using the tool are not the computer scientists who have written the algorithms, tested, and evaluated them.⁴⁶ Successful AI projects like these are a product of collaborations between the civil servants who manage the sensitivity review process, with archival partners, academic researchers, and computer scientists who provide expertise on the technology. Stakeholders effectively communicated professional principles and codes, and through that, AI tools that are congruent with the notions of accountability and transparency were implemented.

3.2 Control

The second tension that emerged through our interviews focused on the lack of control and trust in AI tools. Eleven of our interviewees brought up the idea of control in discussing the application of AI tools to digital archives. Although we do not want to present colleagues in the civil service, GLAM sector and academia as control obsessed, it should be acknowledged that in these professions, control over what is created, released, archived, described, included, and excluded is important. For civil servants and GLAM professionals, there is a professional responsibility toward maintaining privacy and managing sensitive information. James Lappin (Cabinet Office) explained how once a decision is made to release a document, there is no going back: 'If you use AI to determine access permissions, to open up access... it's an irrevocable decision, because once you've given me that access, there isn't much point in then taking it away from me, and the horse has bolted, if you like'.⁴⁷ Similarly, Leontien Talboom (The

National Archives UK/University College London) discussed how archivists tend to default to wanting material to only be available in reading rooms where there can be more control over the type of access to it.⁴⁸ Likewise, Andrew Riley, Senior Archivist at the Churchill Archival Centre, explained how archivists value control and that the loss of that through the use of AI with respect to sensitive data was one of his concerns.⁴⁹

These responsibilities towards protecting privacy and sensitive information connects to the code of practice and professional ethics that govern the sectors. The Civil Service's principle of integrity is particularly relevant here as it references the obligation to undertaking duties responsibly, professionally, and also taking into account the ethical standards that govern other professionals.⁵⁰ The ICA Code of Ethics is more explicit in the archivist's responsibilities towards privacy, specifically noting that archivist should 'take care that corporate and personal privacy as well as national security are protected without destroying information, especially in the case of electronic records where updating and erasure are common practice' ([International Council on Archives, 1996](#)).

The fear over a loss of control was articulated differently by the Humanities academics interviewed during the project. In most cases, the lack of control came down to not knowing what was being missed. Historians highlighted the importance of serendipitous findings when accessing a box of physical archival items. Lindsay Aqai, Research Fellow at the University of Westminster, explained how she felt an 'immediate discomfort' around AI-assisted searches because she 'wouldn't want there to be something in the process that meant 10% was being excluded because it was deemed irrelevant or didn't match the search criteria in a way that I thought it would'.⁵¹ This sentiment was echoed by Emily Robinson, Senior Lecturer in Politics at the University of Sussex, who noted that using AI-driven search tools may prove a barrier to finding the things you did not know you were trying to find in the archive.⁵² Helen McCarthy, Professor in Modern and Contemporary British History at the University of Cambridge, described her feelings towards the application of AI to digital archival collections as ambivalent: simultaneously open to the potentials, but nervous about the loss of control for civil servants and archivist, and the potential losses in material that may come of this.⁵³

One way in which computer scientists are attempting to address this concern over a loss of control is by flipping the narrative on the use of AI. The term 'human-in-the-loop' is often used to refer to the way in which AI tools are trained with human interaction. Graham McDonald explained how 'if you're putting across a

