The National Archives (UK)
Case Study

AEOLIAN
Artificial Intelligence for Cultural Instutions

Contents

The National Archives (UK)

Authors: Dr Lise Jaillant, Dr Katherine Aske, Dr Annalina Caputo

We would like to acknowledge the information provided by interviewees at The National Archives, and their colleagues who reviewed this case study prior to publication. We are very grateful for their contribution.

With this first case study on The National Archives UK (TNA), we have two main objectives. Our first objective is to raise greater awareness of current work on Artificial Intelligence (AI) applied to archives and to encourage further collaborations with other institutions on both sides of the Atlantic. TNA's projects have been developed in response to key challenges brought about by born-digital and digitised records. They range from testing existing AI-powered tools, to developing new approaches, such as using topic modelling to discover the latent or underlying topics of texts across a corpus. Here we examine a selection of TNA's AI projects and others from across the globe that are addressing similar challenges. Our second objective is to bring a critical perspective, from the viewpoint of Digital Humanities and Computer Science. Case Study 1 is written by a team composed of two digital humanists (Lise Jaillant and Katherine Aske) and one computer scientist (Annalina Caputo). Drawing on this cross-disciplinary expertise, we reviewed TNA's projects and when appropriate, drew comparison with other projects conducted by other cultural institutions.

Drawing on interviews with TNA staff as well as published materials such as reports, conference presentations and research papers, this case study is organised into three sections:

- The first section on "Rethinking the Record" examines the transition from print to digital, which has led to an explosion of born-digital and digitised records in central government departments. Established processes to deal with paper records were disrupted, leading to a need for new methods to appraise, select and screen digital records before transfer to TNA. In this section, we will focus particularly on AI for digital selection, for sensitivity review, and for discovery of relevant records.
- The second section is on "Openness, Access and Use". We will start with the need to strike a balance between risk and access, before turning to the development of tools to make collections more accessible, and finally the need to prevent harmful use of collections through risk management.
- The third section deals with "Risk, Uncertainty and Trust". We will look at blockchain to establish the authenticity of records; at the need to balance risk and access; and at explainable AI to build trust.
- The conclusion focuses on ethical uses of AI in the context of large archival collections, at TNA and elsewhere.

Introduction

AI applied to archival records is not a new thing, and innovation initially came from outside the LAM sector (Libraries, Archives and Museums). At the turn of the twenty-first century, lawyers had to deal with an explosion of documents in digital form. In 2003, it was estimated that 93% of documents were created electronically of which over 70% were never converted into hard copy.[1] In legal proceedings, it had always been essential to find relevant documents in a mass of records, and the change of scale brought by the digital revolution led to the development of new software, "eDiscovery", which was developed to allow legal professionals to identify evidence proving or disproving a case. These AI-powered tools rely on predictive coding, which consists of two learning methods: supervised learning and active (unsupervised) learning. With supervised learning, a lawyer chooses a subset of documents, and this selection enables the analytics system to rank the remaining documents in the collection based on their similarity with the initial subset. In the case of unsupervised learning, the machine selects a subset of all case documents using sophisticated algorithms. The lawyer reviews this subset to determine its relevancy and submits it to the system. The machine then analyses the selected documents to identify and code key trends or patterns, before turning to the rest of the collection.

In the early 2000s, these new tools attracted the attention of the National Archives and Records Administration (NARA) in the US, who were, at a time, experiencing a huge increase in born-digital records such as emails, Word documents, PDFs and digital audio-video files. The boom in born-digital files was leading to significant challenges in terms of appraisal, selection and sensitivity review. Following a trial of eDiscovery software, NARA concluded that these tools could be used to identify valuable documents thus aiding with appraisal and selection, as well as assisting with sensitivity review for confidential and problematic documents.[2] On the other side of the Atlantic, The National Archives UK (TNA) were also conducting trials with eDiscovery software, resulting in a 2016 report.[3] The past few years have seen a sustained increase in the number of AI tools and technologies being developed, and as these new tools develop, the use of AI in cultural heritage institutions around the globe

[1] Isaacson, S. (2013) *Computer Technology Review*, March.
[2] Baron, J.R. (2005) 'Toward a Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery', *Sedona Conference Journal* 6, pp. 237–239. Available at: https://thesedonaconference.org/sites/default/files/publications/237-246%20Baron_237-246%20Baron.gxd__0.pdf (Accessed 7 September 2021).
[3] The National Archives UK [hereafter, TNA]. (2016) *The application of technology-assisted review to born-digital records transfer, Inquiries and beyond* [online], available at: https://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf (Accessed: 8 September 2021).

is demanding a more collaborative approach, where technologies, results and best practices can be shared across sectors.

AEOLIAN is a UK/US network on AI applied to cultural institutions.[4] As part of this project, funded by the Arts and Humanities Research Council (AHRC) on the UK side and the National Endowment for the Humanities (NEH) on the US side, we are organising six workshops to bring together Digital Humanists, Computer Scientists, archivists, librarians and other stakeholders. We are also writing five case studies on UK and US cultural institutions that have done pioneering work on applied AI. As Gregory Rolan et al. point out in their 2019 article on Artificial Intelligence in the archive, there is currently a "lack of compelling case studies": "the literature is rather thin on the ground, and there are few clear success stories being trumpeted".[5] It is precisely this gap that we are trying to fill with this open-access report for a diverse audience of academics and practitioners in libraries, archives, museums and government departments.

In the past decade, TNA has led several AI-driven projects, which has resulted in a substantial portfolio of work that forms the basis for this first AEOLIAN case study. As a non-ministerial government department, TNA is the official archive for the UK government and for England and Wales. There are separate national archives for Scotland (the National Records of Scotland) and Northern Ireland (the Public Record Office of Northern Ireland). TNA's collections include records of central government from the Middle Ages onwards, documents such as wills, naturalisation certificates and criminal records, and many others. Since 2003, TNA has also actively curated the UK Government Web Archive, which captures, preserves, and makes accessible UK central government information published on the web. The web archive collects born-digital records such as websites but also videos, images and tweets.

[4] www.aeolian-network.net

[5] Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., Stuart, K. (2019) 'More human than human? Artificial intelligence in the archive', *Archives and Manuscripts*, 47, pp. 179–203. Available at: https://doi.org/10.1080/01576895.2018.1502088 (Accessed: 8 September 2021).

(I) Rethinking the Record

The transition from print to digital has led TNA, like other cultural institutions, to rethink the record. As part of this priority research theme, a core challenge is to focus on "digital recordkeeping at scale".[6] To deal with the boom in digital records, old approaches – such as manually reviewing collections to identify sensitive documents – can no longer be applied. Digital recordkeeping at scale has led TNA to rethink their practices and explore computational methods and other advanced techniques. This requires close collaboration with central government departments.

Following an amendment of the Public Records Act, the UK government is now required to transfer records of historical value to TNA after 20 years for permanent preservation. Before that, records stay within government departments, first as living records (until Year 7) and then as archival records kept in internal archives (from Years 7 to 20). Good record management is essential both before and after transfer to TNA. As Sir Alex Allan explained in his 2015 review of government digital records:

> Records are needed to support policy development; to help assess the impact of policies; to provide accountability for decisions; to share knowledge across government; to enable departments to provide accurate and comprehensive evidence to inquiries or in legal actions; to answer Freedom of Information requests; and eventually to provide the historical background to government.[7]

The key challenge is that digital records are seldom well-organised. The 2017 "Better Information for Better Government" (BI4BG) report – authored by the Cabinet Office, in partnership with TNA – declares: "much of what has accumulated over the past fifteen to twenty years is poorly organised, scattered across different systems and almost impossible to search effectively".[8] It attributes this digital disorganisation to the lack of incentive for civil servants to sort out their mass of digital data – a time-consuming task that has no or few rewards. As the Allan report had done before, the BI4BG report recommends enlisting the help of senior decision makers to improve the management of digital records.

---

[6] TNA. *Priority Research Themes.* Available at: https://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/our-research-priorities/priority-research-themes/ (Accessed: 30 August 2021).
[7] Allan, A. (2015) *Review of Government Digital Records.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/486418/Report_-_Digital_Records_Review.pdf (Accessed: 29 August 2021).
[8] Cabinet Office (UK). (2017) *Better Information for Better Government.* Available at: https://www.gov.uk/government/publications/better-information-for-better-government (Accessed: 29 August 2021).

Since born-digital records are often scattered across different systems (for example, email accounts, records management systems and shared drives), duplicates or near-duplicates are frequent. The volume of data makes the task of searching for specific information extremely difficult. As Andrew Prescott and Jane Winters have shown in a 2019 article, keyword search is not effective with very large datasets.[9] When a search query produces hundreds of thousands of results, potentially ranked only by date, it is difficult to know where to start. While this is an issue that can be approached by the semantic web (a technological effort to make web content more meaningful and readable to machines), it is limited by its inability to identify how knowledge content can change depending on context and use.[10] Moreover, the semantic web presents drawbacks because it can be directly applicable to metadata – provided that an appropriate knowledge base is provided (such as Dbpedia[11]) and that the metadata can be linked to it. However, its application to the textual content of emails, for example, is not so naive, since it requires techniques of Natural Language Processing (NLP) for the extraction of concept and entities from the text, disambiguation, and then for the linking of such entities to the relevant knowledge base. That said, if methods to search and retrieve information are not effective across the broad range of formats of born-digital records, there will be implications not only for usability and access, but also for archives responding to Freedom of Information (FOI) requests and inquiries, which could present further issues as time goes on.

The BI4BG report mourns the golden age of paper records that were neatly filed according to established processes: "Files and filing were at the centre of how work got done: they were intrinsic to the flow of work, not an overhead on it".[12] But is that really the case that the lifecycle of paper records, from creation to preservation, was much more robust? In a recent interview, Anthea Seles, the Secretary General of the International Council on Archives, said:

> There's this notion that exists out there, it's like, Norman Rockwell's lovely paintings of the United States at a particular golden era and people have this notion about paper … [yet] we didn't get it right with paper … it's less discoverable to some degree.[13]

[9] Winters, J., and Prescott, A. (2019) 'Negotiating the born-digital: a problem of search', *Archives and Manuscripts*, 47, pp. 391–403. Available at: https://doi.org/10.1080/01576895.2019.1640753 (Accessed: 27 August 2021).
[10] Fesharaki, Mehdi N., et al. (2020) 'A Conceptual Model for Socio-Pragmatic Web Based on Activity Theory', *Cogent Education*, 7(1), p. 6. Available at: https://doi.org/10.1080/2331186X.2020.1797979 (Accessed 6 October 2021).
[11] See https://www.dbpedia.org.
[12] Cabinet Office (UK). (2017) *Better Information for Better Government.*
[13] Seles, A. (2021) Interview for the AURA project (Archives in the UK/ Republic of Ireland and AI), 28 May.

Seles recommends adjusting expectations to make clear that no system of appraisal and selection is ever going to be perfect. Accepting a certain level of risk and imperfection is essential in order to move forward in the digital age.

1) AI for Digital Selection

How can government departments receive the help they need to select digital files of long-lasting value? TNA recently conducted a project called "AI for Digital Selection" to evaluate existing AI tools that could be used for appraisal and selection of digital records (including emails and datasets) held across government sectors.[14] After choosing a few relevant tools, TNA tested them on a set of their own corporate records – rather than records from the Cabinet Office or other central government departments. TNA's corporate files had already been sensitivity reviewed and had also been assigned to retention schedules which indicated how long they should be kept, in some cases this being permanently. The tasks assigned to the AI tools were to review the content of test documents and to predict whether they should be preserved or not.

In regard to preservation, many libraries apply a faceted classification system to organise their materials into categories based on multiple characteristics, such as subject, form, place etc. However, when archives are dealing with a diverse range of materials, these types of classification systems can present limitations when it comes to preservation selection.[15] Through the "AI for Digital Selection" project, TNA learned what metadata should be captured about the AI tools and processes, to help end users understand and use government records selected via these methods. Moreover, TNA is now in a better position to assist government departments in automating the selection of born-digital documents ahead of transfer for permanent preservation and presentation. However, as Santhilata Venkata (Digital Archiving Researcher at TNA) points out, while the project concluded that AI tools can assist record managers, the machine cannot replace human input.[16]

2) AI for Sensitivity Review

Identifying sensitive materials in large digital collections requires technology-assisted review with human oversight. In its 2016 report on eDiscovery tools, TNA discussed the issue of born-digital records often containing sensitive information, such as contact details of individuals or

[14] Venkata, S., Young, P., Bell, M. and Green, A. (2021) 'Alexa, is this a Historical Record?', accepted for publication in the special edition *Computational Archival Science (CAS) of Journal on Computing and Cultural Heritage*.
[15] See Hoffman, G. L. (2019) *Organizing Library Collections: Theory and Practice*, London, Rowman & Littlefield. See also, Mas, S., Maurel,D., and Alberts, I. (2011) 'Applying Faceted Classification to the Personal Organization of Electronic Records: Insights into the User Experience', *Archivaria*, 72, pp. 29–59.
[16] Venkata, S. (2021) Interview for the AURA project (Archives in the UK/ Republic of Ireland and AI), 21 May.

financial details.[17] In the case of FOI requests, this kind of information falls under the category of "exemptions" and cannot be disclosed. Around three quarters of exemptions to release relate to personal information, so this is clearly a priority area for government departments. After transfer to TNA, it is also essential that no personally identifiable information is released to the public.

AI-powered tools can sort documents according to their sensitivity level: when no sensitive information is identified, documents can be released – although human input is often necessary to prevent any false negatives (in the case of a personal name spelled in various ways, for example). As the report points out, "technology-assisted review is never going to be 100% accurate – departments will need to define and accept their risk appetite".[18] When sensitive information is identified, documents can be closed for a specific period. Another approach is to redact sensitive/personal information using digital forensics tools.[19] The open-source tool BitCurator[20] offers a bulk extractor functionality that lexically analyses text looking for sensitive features, such as email addresses, phone numbers, and other personally identifiable information.

So how does automatic sensitivity review work in practice? As Graham McDonald et al. note, keywords are not enough to identify sensitive information.[21] However, the relationships between terms and entities in the discourse, in addition to single keywords, can help disclose sensitivities. In other words, the context is as important as the text itself when evaluating the sensitivity of a document. To capture contextual information, and at the same time overcome the ambiguity of language, word embedding features can replace or be juxtaposed with simple keywords. In NLP, *word embedding* is a representation type that links a word with other words with similar meanings. For example, "terrorism" and "radicalism" should be closer than "terrorism" and "agriculture". In their study, categorising a collection of c. 3,800 government documents as either sensitive or not-sensitive, McDonald et al. showed that the inclusion of word embeddings significantly increased the accuracy of the classifier.

[17] TNA. (2016) *The Application of Technology-Assisted Review to Born-Digital Records Transfer, Inquiries and Beyond.*
[18] Id.
[19] Woods, K., and Lee, C.A. (2015) 'Redacting Private and Sensitive Information in Born-Digital Collections' in: *Archiving 2015 Final Program and Proceedings*, *May 2015*, Los Angeles: Society for Imaging Science and Technology, pp. 2–7.
[20] https://bitcurator.net/
[21] See Mcdonald, G., Macdonald, C., and Ounis, I. (2020) 'How the Accuracy and Confidence of Sensitivity Classification Affects Digital Sensitivity Review' ACM Transactions on Information Systems, 39, 4, pp. 1-34. Available at: https://doi.org/10.1145/3417334 (Accessed: 7 September 2021). McDonald, G., Macdonald, C., Ounis, I. (2020) 'Active Learning Stopping Strategies for Technology-Assisted Sensitivity Review', in: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York (USA): Association for Computing Machinery, pp. 2053–2056.

3) AI for Discovery

AI can be used by creators of data and archivists for selection and sensitivity review, but also by researchers to discover relevant information. To complement or replace keyword searches, topic modelling can group words into clusters based on similarity. Drawing on unsupervised and supervised machine learning techniques, this text mining method can be used to highlight underlying topics across a dataset – for example, on catalogue metadata describing a large collection.

While metadata is often seen as a way to enhance the human findability of archival material, detailed item descriptions also offer vast corpuses of machine-readable data to analyse. Christopher Day (Head of Modern Domestic Records at TNA) has undertaken research on the catalogue data of the General Board of Health records, a collection comprising c. 89,000 items of correspondence, individually described. In the mid-nineteenth century, the rapid development of capitalism led to overcrowded, poorly drained cities, creating an environment ripe for diseases. The 1848-1849 cholera epidemic in England and Wales claimed around 52,000 lives. In response, the government passed the Public Health Act of 1848, creating a General Board of Health to oversee sanitary measures throughout the country.[22] Drawing on a test corpus of the 1,967 descriptions dated 1848, Day used an algorithm called *Latent Dirichlet Allocation*, in which the machine applies probabilistic statistics to discover topics across a corpus and sorts them into a number of groups defined by the user. Topics were then visualised using the Python library pyLDAvis. The results revealed topics such as sanitary inspections, which were central to the activities of the General Board of Health during its first year.[23]

Cholera may no longer be a major risk in Britain, but the COVID-19 pandemic has reminded us of the centrality of government in designing and implementing public health measures. AI-powered approaches such as topic modelling will be invaluable to analyse the other large-scale collections that TNA continues to collect. During the pandemic, TNA set out to capture a detailed record of the government's response to COVID-19 on the web, using high intensity and in-depth web archiving. Other web archiving initiatives (such as the Internet Archive or UK Web Archive) risked missing this content, which could have been lost to posterity in a rapidly changing context. TNA's COVID-19 collection contains over 50 TB of born-digital material, which could be used as evidence for a future Public Inquiry into the pandemic. As John Sheridan (Digital Director at TNA) puts it, "How does a Public Inquiry begin

[22] Day, C. (2021) *Cholera! Public health in mid-19th century Britain.* Available at: https://media.nationalarchives.gov.uk/index.php/cholera-public-health-in-mid-19th-century-britain/ (Accessed: 7 September 2021).
[23] Day, C. (2020). *Computing Cholera? 'Distant Reading' General Board of Health catalogue data.* Available at: https://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/our-research-projects/2020-annual-digital-lecture-staff-research-poster-exhibition/#computing-cholera (Accessed: 7 September 2021).

to grapple with a collection of this size and scale? What role do AI tools have as we provide the mediation layer between the evidence on one side and the big questions that the Inquiry will be exploring on the other?"[24]

Other large-scale collections are regularly transferred to TNA. In June 2020, the Lord Chancellor and Secretary of State for Justice announced that TNA will be the institutional home for Court Judgments and Tribunal Decisions for England and Wales from April 2022. TNA will inherit a large existing digital collection of judgments and decisions, which will then expand rapidly. For Sheridan, it is essential to think of the contribution that AI can make to improve TNA's intellectual control over this material.[25] Indeed, letting the public access this material (previously not available for re-use) is not without risk. TNA needs to enable that access whilst protecting against potential harms to the justice system. For example, an unscrupulous user could design an algorithm to game the justice system,[26] which would impact on public trust in legal decision-making processes.

[24] Sheridan, J. (2021) Correspondence with L. Jaillant.
[25] Id.
[26] Sheridan, J. (2021) Interview for the "Unlocking our Digital Past" project, Loughborough University, 29 June.

(II) Openness, Access, and Use

TNA are committed to making their collections as accessible as possible to their users, from scholars conducting research, to the public searching for their family histories. They are finding innovative ways to present their collections in consideration of the rapidly increasing volume of digital records, and how the expectations of users are changing.[27] The usability of archival records is therefore central to their digital strategy: "Archives need to develop extraordinary capabilities to ensure digital records can be kept".[28] However, opening archival materials up to the public, or providing controlled access to closed records, comes with numerous challenges. Aside from complying with data protection laws and FOI requests, providing access to large-scale digital collections requires user-focussed solutions, collaboration, and additional ethical considerations.

According to Mark Bell (Senior Digital Researcher at TNA), the transition for TNA from paper to digital since the 1990s has seen a "phenomenal increase" in the number of records.[29] It is estimated that 1.7MB of data was created every second in 2020, approximately 2.5 quintillion bytes a day.[30] Collecting and preserving an accurate record of our recent history is a challenging task, and that is without considering how archives can sort and present this information for researchers in a useful way. As archives "need to be used in order to be useful", developing and using AI to support archival preservation and accessibility is crucial, but the development of these technologies is often siloed.[31] Many archives and other sectors develop systems in-house, meaning the possibility of transference to another system, or integrating records developed with different models, will be increasingly problematic as our digital cultural assets grow and alter with new technologies.

However, first and foremost, archives need to know what to archive. The judgment of what should be online and accessible, what needs reviewing, and what requires limited or case-by-case access, is currently still employing the same methodologies as paper archiving. TNA's digital strategy explains that, as a first-generation digital archive, digital records are currently "appraised and selected like physical records".[32] But these processes were never

[27] TNA (2019) *Areas of Research Interest* [online], available at: https://www.nationalarchives.gov.uk/documents/areas-of-research-interest.pdf (Accessed 27 August 2021).
[28] TNA (2017) *Digital Strategy* [online], p. 3, available at: https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf (Accessed 8 September 2021).
[29] Bell, M., TNA (2018) *Machine Leaning in the Archive* [online], available at: https://blog.nationalarchives.gov.uk/machine-learning-archives/#note-39468-1 (Accessed 28 August 2021).
[30] DOMO (2018) *Data Never Sleeps 6.0* [online], available at: https://www.domo.com/assets/downloads/18_domo_data-never-sleeps-6+verticals.pdf (Accessed 25 August 2021).
[31] TNA (2017) *Digital Strategy* [online], p. 3.
[32] Id., p. 5.

designed to deal with the sheer volume and multiple formats of born-digital records, and the same can be said of TNA's online catalogue, Discovery. Primarily designed for users to search descriptions of physical records and services, it is not an adequate system to present born-digital records.[33] In other words, practices for paper-archiving cannot deal with the unprecedented number of born-digital records that archival institutions now hold, or present records in an accessible way. As TNA have observed, the preservation of and access to digital records "requires nothing less than a revolution".[34] But on the brink of revolution, while we must remember that born-digital records are historical records – not everything can or should be kept.

1) Releasing Records Responsibly

Users may find it strange to think that archives want to dispose of documents, but this is a part of the curation process, and is done through a rigorous criterion to avoid the accidental disposal of important documents. But even for those documents that are preserved, archivists, as well as researchers, must accept that not everything can or should be released to the public. While archives may hold and preserve relevant records of our collective history, they also have a responsibility to present those records, not only for user access, but also with the consideration of legal and ethical factors. In this way, digital technologies have transformed how archives are used by the public, as online catalogues and their search boxes give users instant results for millions of digital and digitised records.[35] For TNA, their catalogue search results indicate whether a record is available online, must be viewed onsite, or if the record is closed access. However, while making more born-digital materials accessible to users may be the goal, offering access to, or even keeping all digital materials, is unrealistic.

To ensure data protection laws are met for archival records, many archives set a high, overly cautious bar on sensitivity review. As John Sheridan points out, it is not a case of "transparency above everything", as

> archives are not Wikileaks, and we're not in the Wikileak 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 need to build the techniques to provide access responsibly.[36]

That said, with continuing advances in AI, the potential to offer more, albeit limited, access to born-digital records is possible. Discussing the balance between risk management and providing access to potentially sensitive materials, Sheridan notes that publishing materials

[33] Id., pp. 3-5.
[34] Id., p. 1.
[35] Id., p. 3.
[36] Sheridan, J. (2021), Interview.

online “is fundamentally a very different act from providing reasonable facilities for someone to inspect a record”.[37] So how can archive services design and present online access systems to meet the needs and expectations of their users, while also managing risk levels?

2) Developing Tools for Access

Discussing the development of tools to access digital records, TNA’s research priorities emphasise the necessity of understanding their researchers – and preparing for how research needs and skills will develop in the future. Providing access to users must not only accommodate records in multiple formats, but also the researcher’s and the archive’s capabilities. Acknowledging these elements to design new approaches to delivery and research with aggregated data, TNA are investing in new tools for quantitative analysis and the manipulation of data at scale.[38] The following sections discuss current issues and solutions to increasing usability and providing safe access to closed materials.

i) Discovery and Access

Accessing physical materials has inevitably been made harder by the COVID-19 pandemic, but archives have also had to face the reality of digital accessibility sooner than they might have expected. With a growing demand for remote access to records, archives have focussed on numerous ways to provide digital content to users. Providing an online catalogue is a vital part of an archive’s usability in today’s modern world, and while not everything can be listed, the catalogue is often the first step for the user.[39] With no or limited access onsite, TNA offered free downloads to registered users from April 2020, allowing access to almost nine million of their digital records from the Discovery online catalogue. But while TNA have digitised over 80 million records, there are just over 24 million records available to search via Discovery, because some of the catalogue entries are closed.[40]

However, at the other end of the scale, the UK Government Web Archive, curated by TNA, has over 500 million digital records, dating from 1996 to 2021. However, with only four filters available to users (keyword, website, file type, year), the searchability functions are insufficient for such a huge number of records; there are over 11 million results for “COVID-19” alone. With the sheer volume of born-digital records in various

[37] Id.
[38] TNA. *Openness, Access and Use* [online], available at: https://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/our-research-priorities/priority-research-themes/openness-access-and-use/ (Accessed: 25 August 2021)
[39] See Dunley, R. and Pugh, J. (2021) ‘Do Archive Catalogues Make History?: Exploring Interactions between Historians and Archives’, *Twentieth Century British History*. Available at: https://doi.org/10.1093/tcbh/hwab021 (Accessed: 7 September 2021)
[40] TNA. (2020) *Digitisation and Digital Archives*, available at: https://www.nationalarchives.gov.uk/about/our-role/transparency/digitisation-and-digital-archives/ (Accessed 8 September 2021)

formats passing on to TNA, as well as other cultural institutions, the development of new tools and methodologies need to ensure digital records are not only made accessible (or at the very least discoverable), but that their searchability is adequate enough to allow users to sort through records and find what they need.

ii) Providing Physical Access

TNA's Discovery catalogue has just under 340,000 of its closed documents listed. These sensitive materials, much like those held by other archives, must be requested through a FOI request. Traditionally, if a request is granted, these types of records need to be viewed within the specific archive. However, as in-person research is slowly returning, TNA has signed up to SafePod. Developed by Prof. Chris Dibben (University of Edinburgh) and Darren Lightfoot (University of St Andrews), the SafePod Network (SPN) provides access to sensitive datasets through a series of secure pods located throughout the UK – with an estimated 25 locations by the end of 2022.[41] According to Mark Bell, the TNA's SafePod will be mainly for "sensitive administrative data" and will allow a researcher to remotely access different data sets, without being able to take anything away with them.[42] The Data Centres that can currently be accessed from a SafePod include the Secure Anonymised Information Linkage (SAIL) Databank, UK Data Service and Office for National Statistics. While a physical space that allows multiple de-identified or anonymised datasets to be examined securely in the same location is one answer to providing greater accessibility, there are still practical issues.

SafePods will be primarily based at universities and aimed at researchers, but they require users to be onsite, and this could potentially cause issues for public users. They also need users to register, and complete a short training questionnaire, to book.[43] Additionally, there is a capacity issue. While a single SafePod may be adequate for a university library, is one SafePod enough for TNA? Until the demand is recorded, it is difficult to predict users' needs long-term. But the number of records a researcher might need to consult adds time restraints, potentially meaning multiple visits to a SafePod. Looking ahead, the secure technology behind SafePod, providing users with remote access to its partnered Data Centres, could be adapted to remove the necessity of the 'pod'. Registered users could be provided with temporary access to the required resources through a remote desktop, with the session recorded via webcam and screen capture to prevent issues of photography, copying or misuse. Such steps could allow

[41] Lightfoot, D. (2021) *The SafePod Network (SPN)* [online], available at: https://safepodnetwork.ac.uk (Accessed: 26 August 2021)
[42] Bell, M. (2021) Interview.
[43] https://safepodnetwork.ac.uk

the technology to be used by far more archives, libraries, and universities across UK, and beyond. But for now, SafePods are offering a timely and necessary solution – a first step on a long road to making closed records more accessible.

iii) Providing Remote Access

While providing physical access can placate user-demand, it is not a practical solution in the long term. Much of the content contained in digital records could be made remotely accessible if the sensitive information was efficiently redacted. Personal emails, for example, can contain large amounts of sensitive information, scattered across a potentially huge number of records. While they can provide evidence of prominent individual's lives and information that could interest researchers and the public alike, email preservation and review can be a laborious process.[44]

AI tools can make the process of sensitivity review more efficient and less time consuming. Stanford University Library's email archive, a system developed through their open-source software programme ePADD, was designed to address this mammoth task. Started in 2010, ePADD uses machine learning and NLP to meet the multiple challenges of email archiving.[45] The programme screens emails for confidential and legally protected information, offering a lexicon-based search for sensitive topics and image browsing. These tools allow ePADD's users to prepare records for preservation, while making them accessible and discoverable for researchers.

While providing access to preserved emails is a pressing issue, it must also be addressed with future users in mind. How will these sources be engaged with once issues of access have been navigated?[46] While content may be preserved, processes of sensitivity review increase a risk of de-contextualisation, presenting challenges for historical researchers and general users. To this end, the TNA has been involved in the project eConDist, a context-based search tool developed using NLP and deep learning. This advanced search tool helps to incorporate human intuition into user queries. It has been developed as a part of the AHRC-NEH funded project 'Contextualisation of Email Archives', where TNA partnered with University of Bristol (UK), De Montfort University (UK) and University of Maryland (US).[47]

---

[44] Schneider, J. et al., (2019) 'Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives', *Archives and Manuscripts*, 47(3), pp. 305-326. Available at: https://doi.org/10.1080/01576895.2019.1622138 (Accessed: 25 August 2021)

[45] Stanford Libraries Projects. (2021) *About ePADD* [online], available at: https://epadd.stanford.edu/epadd/about (Accessed: 25 August 2021)

[46] See Nix, A. et al. (2021) 'Finding Light in Dark Archives: Using AI to Connect, Context and Content in Email', presentation at AURA: Artificial Intelligence and Archives: What comes next? Online conference, available at: https://www.aura-network.net/wp-content/uploads/2021/04/Adam-Nix-slides.pdf

[47] Decker, S., Kirsch, D. Venkata, S., and Nix, A. (2021) 'Finding Light in Dark Archives: Using AI to Connect, Context and Content in Email', accepted for publication in the *Journal of Knowledge, Culture and Communication*, *AI & Society*.

Through these examples, it is clear that AI and other technologies are being applied effectively to address issues of preservation and access. However, the employment of such technologies in archives requires skilled training, and, without a dedicated digital department or external assistance, archival staff would be required to learn digital skillsets on top of their existing archival expertise. In this way, and at a policy level, the infrastructure for digital archiving is severely lacking, and changes need to be implemented across the sector.

Addressing this issue, a recent multinational project has paved the way for a more collaborative approach to digital strategy. The European Archival Records and Knowledge Preservation (E-ARK) project has focussed on ensuring digital archives and technologies remain usable and consistent over time, and internationally.[48] Running from 2014 to 2017, the project brought together national archives across Europe, Chile and the US, to research consistency in digital archiving with support from the University of Brighton (UK) and the Digital Preservation Coalition (DPC). The collaborative project shared pioneering digital tools and expertise, which in turn improved skills and lowered costs for archives – drawing the attention of TNA.[49]

With a similar intention, TNA are taking the lead in the UK digital archival sector with the projects *Archives for Everyone* in 2015, *Archives Unlocked* in 2017 and most recently, *Plugged In, Powered Up* in 2020.[50] Through these initiatives, they have set up training for AHRC Collaborative Doctoral students, delivered seminars, and launched *Bridging the Digital Gap*, a National Lottery Heritage Fund training programme for 24 technical apprentices in UK archives.[51] They have also worked with the DPC on the online learning pathway *Novice to Know-How*.[52] TNA, like many across the sector, have recognised the importance of collaborative action to secure future access to and use of digital records. Through collaborative partnerships, built on feedback, justification, and the exchange of knowledge, those with fewer resources can benefit from those with more.[53] Although AI methods and technologies may

[48] www.eark-project.com
[49] TNA. (2017–2020) *Strategic Vision for Archives* [online], p. 5. Available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/strategic-vision-for-archives/ (Accessed: 8 September 2021).
[50] TNA. (2015–2021) *Archives for Everyone* [online], available at: https://www.nationalarchives.gov.uk/about/our-role/plans-policies-performance-and-projects/our-plans/archives-for-everyone/ *Archives Unlocked* [online], available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/strategic-vision-for-archives/ *Plugged In, Powered Up* [online] available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/plugged-in-powered-up/ (Accessed: 8 September 2021).
[51] TNA. (2021) *Bridging the Digital Gap* [online], available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/bridging-digital-gap-technical-traineeships-archives/ (Accessed: 8 September 2021).
[52] TNA. (2021) *Novice to Know-How* [online], available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/plugged-in-powered-up/novice-to-know-how/ (Accessed: 8 September 2021).
[53] Gurciullo, S. (2017) 'Keeping born-digital literary and artistic archives in an imperfect world: theory, best practice and good enoughs', *Comma*, 1, pp. 49–65 (p. 65). Available at: https://doi.org/10.3828/comma.2017.4 (Accessed: 27 August 2021).

continue to develop in-house to meet the specific needs of individual archives, the experience and advice gained through the employment of these technologies can not only help to address issues of access, but inform ethical considerations and highlight potential dangers, across the sector.

3) Managing Risk

While AI continues to offer solutions for providing user access to digital records, there is also a danger of making records more vulnerable to abuse, misuse or corruption. The dangers of digital corruption and the potential solutions are more formally addressed in the next section, but here we discuss how archives must address these issues through a risk management approach. TNA's current risk assessment for digital continuity addresses several considerations, from types of risks and timely reaction, to learning from past issues. For digital records, they complete a risk assessment at least every two years, or when there is a significant change within the technical environment.[54]

However, according to Sheridan, the hardest element in anticipating risk is knowing how people will use digital collections. Although there are interventions within the data, are these enough to prevent a user piecing together potentially sensitive information?[55] TNA have experience using named entity recognition and statistical models to decipher sensitive information (names, addresses, contact information, account numbers, etc), but unfortunately an adequate digital risk model has not yet been built. As Sheridan has discussed, TNA's current approach, like many other archives, is to manage risk "through expert knowledge" rather than systems.[56]

Until systematic risk models are an inherent part of archives' digital practice, adaptive approaches can be employed; one of those is gradation. A notion that "maximises use but manages the risks of publication", gradating access can allow for a more flexible system of publishing potentially sensitive records (but not in breach of data protection laws), by identifying varying levels of risk, and determining necessary exemptions.[57] Publishing born-digital materials with an exemption from search engine indexing is one way to make the record less discoverable, but still accessible to those who want to use it. For example, the court jurisdiction database, British and Irish Legal Information Institute (BAILII), prohibits the "external indexing of documents" or the publishing of any materials on external websites, as a

[54] TNA. (2017) *Risk Assessment Handbook* [online], available at: https://www.nationalarchives.gov.uk/documents/information-management/Risk-Assessment-Handbook.pdf (Accessed: 25 August 2021)
[55] Sheridan, J. (2021), Interview.
[56] Id.
[57] Id.

form of risk management.[58] But while imposing publishing and use exemptions may work for certain categories of documents, working at scale requires varying and intuitive approaches.

Even when legal requirements for sensitive data are met, determining the level of risk from an ethical point of view is still at the archives' discretion. In a recent interview, a director of Special Collections at a prestigious American library discussed the ethical nature of the archive.[59] They recalled an occasion with the collection of Susan Sontag's emails held by UCLA, where a student published an essay on a personal relationship detailed in the emails. The legally protected information had been redacted, but some personal information had remained, revealing the additional ethical concerns archives must consider when presenting an individual's intimate correspondence.

AI technology can help to bring archival material to light, but it cannot replicate (at least not yet) the human processing that goes into ethical decision making or anticipate the connections users may potentially make from redacted materials. While placing prohibitions for external indexing and unauthorised use on sensitive documents may help to moderate risk, it is not sufficient to prevent harmful use. In this sense, finding a balance between usability and limiting the potential misuse of archive materials must be addressed by adequate risk management. As the number of born-digital records continues to increase, and the capacity of expert assessment becomes overstretched, there are AI technologies that are enabling risks to be managed and reduced.

[58] British and Irish Legal Information Institute (BAILII) [online], available at: www.bailii.org (Accessed: 8 September 2021). Sheridan, J. (2021) Interview.
[59] Anon. (2021), Interview with L. Jaillant.

(III) Risk, Uncertainty and Trust

The scale and growth rate of the digital world has strong implications on digital recordkeepers. Technology provides an opportunity to empower archivists with new capabilities of processing and inference from digital collections otherwise lost in the deluge of information. But in this process, two questions need to be answered: 1) To what extent do we try and use AI to help us? and 2) How can we be mindful of the harmful uses of material that we are trying to prevent through the application and use of AI?[60] To answer these, TNA is investigating technologies and tools to manage risk and uncertainty while reconciling with trust.

One of these is the use of blockchains, or Distributed Ledger Technology (DLT). A blockchain is a series of blocks of digital data, stored in a digital ledger (like a database) that multiple organisations can maintain, check, share, and add to, but, most importantly, the data in the blockchain cannot be altered. The claim that blockchain is immutable is supported by the decentralised nature of blocks of data. If an attempt is made to change any of the data, this needs to be verified by the other blocks in the chain, making editing nearly impossible. The ARCHANGEL project combined the use of blockchain and AI to guarantee the authenticity of digital records and foster trust while accounting for some of the most important weaknesses of digital archives – including their dependency on ephemeral file formats.[61] Statistical risk management allows risk models to incorporate the uncertainty surrounding the digital collections in terms of the probability of known events and unknown variables. To this purpose, TNA has explored a Bayesian network. A Bayesian network is a graphical model that represents a set of variables and conditions, which are often used for probability analysis. These networks can either be specified by an expert, or, for larger models, trained using data, such as that stored in blockchains. In conjunction with the Applied Statistics and Risk Unit at the University of Warwick, TNA have developed the Digital Archiving Graphical Risk Assessment Model (DiAGRAM), to provide a decision support system capable of quantifying risks and benefits of possible interventions and help prioritise investments.[62]

The adoption of AI in many sensitive domains has raised awareness about the implication of this technology on decision-making processes, highlighting the requirements for fair, accountable, and transparent tools. A tool that has attempted to answer this need is explainable AI (XAI). Discussed in more detail in section III:3, XAI allows its decision-making

[60] Sheridan, J. (2021) Interview
[61] https://www.archangel.ac.uk/
[62] Barons, M., Bhatia, S., Double, J., Fonseca, T., Green, A., Krol, S., Merwood, H., Mulinder, A., Ranade, S., Smith, J.Q. and Thornhill, T. (2021) 'Safeguarding the nation's digital memory: towards a Bayesian model of digital preservation risk', *Archives and Records*, *42*(1), pp. 58–78.

processes to be understood by humans, unlike AI, and this enables users to ensure that the XAI is making good decisions. However, the focus of the research community has been mostly directed towards the technology, forgetting the role of humans and their environment when engaging with AI.

1) Blockchain to Establish the Authenticity of Records

It is natural, talking about archive preservation, to think about the physical artefacts and how to protect them from damages and natural deterioration. This concept, although less intuitive, also extends to digital archives. There are many challenges posed by the preservation of digital and born-digital archives, but one crucial aspect is related to integrity and trust. Indeed, while the possibility of copying digital content allows for escaping the natural expiration date of storage supports, it paves the way for digital corruption, tampering and modifications. Some of these are wanted. Redacting sensitive records or removing personal information are essential to open archives to the public. Others conceal malevolent intentions, spanning from rewriting history, generating fake or counterfeit artefacts, or simple faults and corruption of the supporting devices. How do we guarantee that unauthorised manipulations do not take place while still allowing authorised modification to happen? How can this process be carried out in a way that engenders trust towards archival and memory institutions?

Technology may provide the answer through the combination of DLTs and AI. As pointed out by Lemieux, "the discussion about trusted records or systems boils down to two interlinking concepts: reliability and authenticity, and closely related concepts such as identity, integrity and provenance".[63] Blockchain, the technology at the backbone of Bitcoin, can be exploited for ensuring trustful digital records:

> In medieval time, pages of court records were stitched together into a patchwork, an obvious hole would be left if anyone removed a page. Today, blockchain uses a similar idea to stitch together blocks of data to detect tampering.[64]

In a blockchain, trust is achieved by using a decentralised database to keep a record of transactions, usually packaged in *blocks* along with a hash code pointer (a unique reference code), used to check the integrity of the block. The focus of trust moves from the individual parties to the network of members (called nodes) which are now required to reach a consensus before a block is added to the chain.[65] Generally speaking, a consensus of 51% is

[63] Lemieux, V.L. (2016) 'Trusting records: is Blockchain technology the answer?', *Records Management Journal*, 26(2), pp. 110–139.
[64] ARCHANGEL Project. (2019) *Trusted Digital Archives* [Online video] available at: https://www.youtube.com/watch?v=xKCdKo6rQXw (Accessed: 7 September 2021).
[65] The ODI. (2018) *How can smart contracts be useful for business?* [online], available at https://theodi.org/wp-content/uploads/2018/05/378720579-How-can-smart-contracts-be-useful-for-businesses.pdf (Accessed: 7 September 2021).

required to modify data in the blockchain, making the addition of malicious blocks incredibly difficult. Additionally, to remove the risk of malevolent parties taking control over such data, permissioned access can be implemented.

ARCHANGEL, which brought together TNA, the Open Data Institute (ODI), and the University of Surrey, used permissioned access to provide reader access to the general public. In using a permissioned ledger, it is possible to reach a balance between control and transparency. From one side, the general public can access and view the records and openly verify their integrity. On the other, only authorised individuals are allowed to add to the ledger. The consensus to add a block to the chain is achieved through two practices based on a process called proof-of-work (PoW), i.e. proof that the work of a participant 'node' qualifies them to add to the blockchain (this is usually gauged through the completion of a complex computational puzzle). One implementation of the practice allows only private nodes, maintained by multiple Archives and Memory Institutions (AMIs), to generate such PoWs. Another form, allows access from the public, but controls write access by using a smart contract (a computer programme that acts like a third party) with a user key to verify the identity of the user.[66]

In a DLT, 'fingerprints' (i.e. hash codes) are used to uniquely identify a digital object. These fingerprints exist within the content metadata of a file, so even if the system metadata (such as its name or extension) is altered, the hash identifier remains the same. They are deposited in a DLT-based system in order to a) ensure that there were no unauthorised modifications since the deposit of the fingerprint and b) if there were authorised modifications, these leave a transparent auditable trail. In this way, it is possible to ensure the identity, integrity and provenance of digital objects. The verification of authenticity is based on the match between the deposited fingerprint and the one generated by the object.

However, to address integrity and authenticity in archival records, emphasis needs to be placed on new media, like audio-visual streams, as these forms of records are becoming a predominant way of documenting and capturing our society. The wealth of publicly available video, photo, and audio, combined with the unscrupulous use of AI to generate new content, is at the basis of phenomena like fake videos ("deepfakes") or simply video/photo editing. To ensure archival integrity and limit the risk of manipulation, the US National Archives incorporated hash codes within the metadata of the John F. Kennedy assassination archive.[67]

[66] Collomosse, J., Bui, T., Brown, A., Sheridan, J., Green, A., Bell, M., Fawcett, J., Higgins, J. and Thereaux, O. (2018) 'ARCHANGEL: Trusted Archives of Digital Public Documents' in: *Proceedings of the ACM Symposium on Document Engineering*, August 2018, Halifax, NS, Canada, ACM DocEng, pp. 1-4. See also, Porat, A. Pratap, A., Shah, P. and Adkar, V. (2017) 'Blockchain Consensus: An analysis of Proof-of-Work and its Applications' [online], available at: https://www.scs.stanford.edu/17au-cs244b/labs/projects/porat_pratap_shah_adkar.pdf (Accessed 22 October 2021).

[67] Bhatia, S., Douglas, E.K. and Most, M. (2020) 'Blockchain and Records Management: Disruptive Force or New Approach?, *Records Management Journal*, 30(3), pp. 277–286.

In addition to being highly subject to distortion and manipulation, video media forms are characterised by their ephemeral nature, often relying on formats that are quickly becoming obsolete. This can present issues, even with the use of hash codes, because although the code will only alter if the content of the file is altered – opening files in different format applications (like creating a PDF from a Word document) changes the embedded content, and therefore the hash. The changing of file format in this way is known as transcoding. The need to create a copy due to transcoding can easily result in errors and corrupted files, hence requiring methods capable of detecting accidental or malicious alteration of content while being invariant to format. It is important then to decouple the object from its format, which may change over time. While this can be easily done for textual data, it is more complex for formats like videos.

To solve this issue, the ARCHANGEL project created digital 'fingerprints' that were "sensitive to tampering, but invariant to the format".[68] Using blockchain technology, the content-based hash, the file identifier and a unique identifier of the process used to extract the hash, were stored with other metadata to ensure the file and format's integrity as technologies change.[69] In addition to an integrity check, smart contracts, which are essentially programmable contracts that sit on the blockchain and are run when predetermined conditions are met, could be used to access the metadata associated with the object fingerprint, providing support to implement indexing and search capabilities over these digital objects.

The need to provide mechanisms to check for authenticity and integrity by detecting attempts of tampering and forgery is also the motivation behind a Research and Development project conducted at the National Archives of Korea,[70] which led to two case studies to inform the adoption of this technology using hyperledger fabric (an open-source blockchain platform designed for use in enterprise).[71] The first study was inspired by ARCHANGEL and aimed at using blockchain to verify the authenticity of audio-visual content and provide an audit trail of transactions. The second focused on datasets generated by government agencies and aimed at ensuring the integrity of datasets from tampering and forgery when they are self-managed and stored in multiple institutions.

[68] Bui, T., Cooper, D., Collomosse, J., Bell, M., Green, A., Sheridan, J., Higgins, J., Das, A., Keller, J., Thereaux, O. and Brown, A. (2019) 'Archangel: Tamper-proofing Video Archives using Temporal Content Hashes on the Blockchain' in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, IEEE. Available at: https://arxiv.org/abs/1904.12059 (Accessed: 7 September 2021).

[69] Collomosse, J., et al. (2018) 'ARCHANGEL: Trusted Archives of Digital Public Documents'.

[70] Wang, H. and Yang, D., 2021. Research and Development of Blockchain Recordkeeping at the National Archives of Korea. *Computers*, 10(8), p. 90.

[71] Hyperledger Fabric (2020) [online] 'Introduction', available at: https://hyperledger-fabric.readthedocs.io/en/release-2.2/whatis.html (Accessed: 6 October 2021).

2) Balancing Risk and Access

Two main concerns of digital archives are inevitably connected with the risks associated with preservation, from faulty devices and file corruption to missing metadata and loss of integrity, and access. Risk management's "ultimate goal is to define prevention and control mechanisms to address the risk attached to specific activities and valuable assets, where risk is defined as the combination of the probability of an event and its consequence".[72] It is natural to think of pairing archive preservation and access with risk management processes, since these help to identify the limitations of the context, to assess the risks, and plan for treatments. J. Barateiro et al. propose an approach to risk management specific for digital preservation consisting of three steps: 1) identification of requirements; 2) classification of threats and vulnerabilities; and 3) treatment of the risks deriving.[73] However, one limitation with these types of approaches comes from the impossibility to quantify outcomes and probability, hence resulting in qualitative evaluation rather than quantitative indication for decision makers. In an environment often constrained by high volume of data and low available resources, how to select what to preserve and prioritise accordingly?

This is the concept behind "Safeguarding the Nation's Digital Memory" project: to approach preservation risk by employing statistical methods for decision support based on data and evidence.[74] The outcome, DiAGRAM, combines the knowledge of domain experts and statisticians into a Bayesian network used to infer the risks associated with four key areas:

I. **Preservation** – caused by the fluidity and fragility of digital artefacts;
II. **Context and provenance** – due to the facility with which these records can be moved around, lost or hidden;
III. **Transparency, trust and inclusion** – imputable to the greater complexity faced by digital archivists when creating a digital story; and
IV. **Policy** – when idealistic benchmarks, processes, standards and models collide with the lack of resources for local and small archives, and their necessity to prioritise.

The risk here is that standards and processes hinder the archival preservation process. Hence, the project aims to provide practical decision support tools that guide archivists through the quantitative assessment of risks and threats. Through these quantifications, decision-makers can examine different risks and benefits associated with threats and make informed decisions and plans. Additionally, the use of statistical risk models provides the flexibility to

[72] Barateiro, J., Antunes, G., Freitas, F. and Borbinha, J. (2010) Designing digital preservation solutions: A risk management-based approach [online] available at: https://doi.org/10.2218/ijdc.v5i1.140 (Accessed: 7 September 2021).
[73] Barateiro, J., Antunes, G., Borbinha, J. and Lisboa, P. (2009) June. Addressing digital preservation: Proposals for new perspectives. In *Proceedings of InDP-09, 1st International Workshop on Innovation in Digital Preservation. Austin, TX, USA*.
[74] Merwood, H. (2020) Risk Alert: Insufficient Technical Metadata. [Blog] *Digital Preservation Coalition*. Available at: https://www.dpconline.org/blog/risk-alert-insufficient-technical-metadata (Accessed: 8 September 2021)

adapt through time and navigate risks when it is hard to predict all possible outcomes and uses of a digital collection.[75]

The power of the model resides in its capability to tackle uncertainty, i.e., to provide estimates even in presence of limited or imperfect data. This is a characteristic of Bayesian networks, which provide a framework to model expert knowledge necessary to compensate for the lack of information and provide a robust tool for reasoning under uncertainty. However, even when risk management practices are employed, there are challenges due to the lack of shared details around the experience of system failures, as pointed out by Dearborn and Meister.[76] To this end, the authors discuss the past experience of failures within the MetaArchive Cooperative as a way to plan for success in the future. Openness and transparency can thus be interpreted in terms of shared experience and practices.

3) Explainable AI to Build Trust

As AI is creeping into many aspects of our life, it raises questions regarding the reliability of these systems, and consequently, the risks deriving from AI-based decision making. While the black-box model fuelled by big data and complex deep architectures has determined the popularity of AI solutions in many domains in the past decade, now we are faced with the requirements of interpretable and explainable models as the top prerequisites to establish fairness, accountability, and trust in this technology. Although often used interchangeably, explainability and interpretability refer to two separate aspects of AI algorithms. Interpretability looks at how an AI works, focussing on what the algorithm does. Explainability is concerned with how the AI behaves, and as such, it aims at creating a trust link between the AI and its users, producing insights into their decision making.[77]

But how can an AI explain itself? In 'Explaining Explanations: An Overview of Interpretability of Machine Learning',[78] Gilpin et al. survey current work on XAI, trying to identify challenges and foundational concepts used to create a taxonomy of XAI approaches. There are three categories for XAI:

i. **Processing**, i.e. methods that try to rebuild the internal decision process of the algorithm trying to identify connections between input and output. This category of XAI is mainly concerned with the impact of AI on users, and how to create transparency, and eventually trust, in their use.

[75] Sheridan, J. (2021) Interview
[76] Dearborn, C. and Meister, S. (2017) 'Failure as Process: Interrogating Disaster, Loss, and Recovery in Digital Preservation', *Alexandria*, 27(2), pp. 83–93.
[77] Bunn, J. (2020) 'Working in Contexts for which Transparency is Important: A Recordkeeping view of Explainable Artificial Intelligence (XAI)', *Records Management Journal*, 30(2), pp. 143–153, available at: http://dx.doi.org/10.1108/RMJ-08-2019-0038 (Accessed 8 September 2021).
[78] Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M. and Kagal, L. (2018) 'Explaining Explanations: An Overview of Interpretability of Machine Learning' in: *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, October 2018, IEEE, pp. 80-89.

ii. **Representation** looks at the internal model of the AI trying to understand how data are represented. This is particularly important to help understand the role of bias in data and how this propagates in the algorithm.
iii. The last category looks at ways of ***producing explanation***, i.e. the capability of AI to engage in a conversation around its own decisions.

It is interesting to note how concepts of transparency and accountability are shared between AI and digital archives. In this way, the next step is to direct efforts towards a shared definition that benefits people.

As Abdul et al. have pointed out, while the AI community is working toward explainable algorithms, "their focus is not on usable, practical and effective transparency that works for and benefits people".[79] Building towards a shared view, the workshop on "Human-centred Explainable AI" (HeXAI) organised by University College London (UCL) and TNA focussed on human-centered multidisciplinary exploration of and engagement with XAI. Working towards XAI there is a "need to understand a lot more about explanation as a contextual human behaviour with a role in cementing social cohesion and trust".[80] Building explainable AI is not just an algorithmic matter, but needs to consider the individuals, and the environment in which it will operate.

[79] Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y. and Kankanhalli, M. (2018) 'Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda' in: *Proceedings of the 2018 CHI conference on human factors in computing systems*, April 2018, CHI, pp. 1-18.
[80] UCL. (2019) 'From Black Box to Tip of the Iceberg: Creative Engagement with the Emergence of XAI (Explainable Artificial Intelligence)', available at: https://cpb-eu-w2.wpmucdn.com/blogs.ucl.ac.uk/dist/e/653/files/2019/10/HeXAI-leaflet.pdf (Accessed: 8 September 2021).

Conclusion – AI and Ethics

Over the past five years TNA has been at the forefront of the exploration of AI applied to archives and has contributed substantial thought leadership and pragmatic case studies for the wider LAM sector to reflect on and learn from. The work has been driven by the knowledge that how archives select, appraise, manage, preserve and provide access to their collections has changed, and continues to change dramatically as digital technologies develop. From early trials with eDiscovery and computation tools to support selection and appraisal processes, to collaborative projects with emerging technologies like blockchain, TNA has an established appetite for experimentation. However, whilst the adoption of AI has often been thought to solve problems related to preservation and access of digital archives, it is also raising concerns regarding bias and ethics.

Indeed, AI can increase the risk of amplifying data and algorithmic bias, reinforcing stereotypes and skewed perceptions of the world, reframing discourse around popular topics statistically more prominent while filtering out niche views, and inducing decisions based on uncertain assumptions, without enough consideration of their confidence. Explainability is seen as a panacea for trust, in turn relying on fairness, accountability and transparency (FACT). However, there is still a gap between the interpretation of trustworthiness and FACT within the AI and the archive community. While sharing a common vocabulary, they prioritise different aspects, algorithmic from one side, and humans and their environment from the other. Convergence among these points-of-view can inform both sides, and lead to a greater progress in their disciplines. E. S. Jo and T. Gebru[81] have already highlighted the exemplary role that archives can play in creating datasets for learning algorithms. Building on the concepts of consent, inclusivity, power, transparency, and ethics and privacy in archival and library science, the authors describe how these can be applied in machine learning to limit bias and ethical concerns. However, we also need to be pragmatic about the expectation we set forth and the desired outcomes. Using data 'in the wild' is doomed to raise issues similar to those recently found in J. Buolamwini and Gebru's study, which outlined the way "machine learning algorithms can discriminate based on classes like race and gender".[82] But would library staff have been aware of demographic bias in the data sets before such a study? The

[81] Jo, E.S. and Gebru, T. (2020) 'Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning' in: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020, pp. 306-316. Available at https://dl.acm.org/doi/10.1145/3351095.3372829 (Accessed: 8 September 2021).
[82] Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency,* January 2018, PMLR, 81, pp. 77–91.

answer, as pointed out by C. N. Coleman, is “likely not”.[83] This is why an honest discussion needs to take place, where the focus shifts from ways to *remove* bias to ways of *managing* it. In doing this, guiding principles of beneficence, nonmaleficence, autonomy, justice and explicability can “serve as the architecture within which laws, rules, technical standards, and best practices are developed for specific sectors, industries, and jurisdictions”, in which these principles can have either an enabling or a constraining role.[84]

[83] Coleman, C.N. (2020) ‘Managing Bias When Library Collections Become Data’, *International Journal of Librarianship*, *5*(1), pp. 8–19.
[84] Floridi, L. and Cowls, J., (2019) ‘A Unified Framework of Five Principles for AI in Society’, HDSR, 1(1), pp. 1–15. Available at: https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/7 (Accessed: 7 September 2021).

Bibliography

Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y. and Kankanhalli, M. (2018) 'Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda' in: *Proceedings of the 2018 CHI conference on human factors in computing systems*, April 2018, CHI, pp. 1-18.

Allan, A. (2015) *Review of Government Digital Records.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/486418/Report_-_Digital_Records_Review.pdf

ARCHANGEL Project. (2019) *Trusted Digital Archives* [Online video] available at: https://www.youtube.com/watch?v=xKCdKo6rQXw

Barateiro, J., Antunes, G., Borbinha, J. and Lisboa, P. (2009) June. Addressing digital preservation: Proposals for new perspectives. In *Proceedings of InDP-09, 1st International Workshop on Innovation in Digital Preservation. Austin, TX, USA*.

Barateiro, J., Antunes, G., Freitas, F. and Borbinha, J. (2010) Designing digital preservation solutions: A risk management-based approach [online] available at: https://doi.org/10.2218/ijdc.v5i1.140

Baron, J.R. (2005) 'Toward a Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery', *Sedona Conference Journal* 6, pp. 237–239. Available at: https://thesedonaconference.org/sites/default/files/publications/237-246%20Baron_237-246%20Baron.qxd__0.pdf

Barons, M., Bhatia, S., Double, J., Fonseca, T., Green, A., Krol, S., Merwood, H., Mulinder, A., Ranade, S., Smith, J.Q. and Thornhill, T. (2021) 'Safeguarding the nation's digital memory: towards a Bayesian model of digital preservation risk', *Archives and Records*, *42*(1), pp. 58–78.

Bell, M., TNA (2018) *Machine Leaning in the Archive* [online] available at: https://blog.nationalarchives.gov.uk/machine-learning-archives/#note-39468-1

Bhatia, S., Douglas, E.K. and Most, M. (2020) 'Blockchain and Records Management: Disruptive Force or New Approach?, *Records Management Journal*, 30(3), pp. 277–286.

British and Irish Legal Information Institute (BAILII) [online], available at: www.bailii.org (Accessed: 8 September 2021).

Bui, T., Cooper, D., Collomosse, J., Bell, M., Green, A., Sheridan, J., Higgins, J., Das, A., Keller, J., Thereaux, O. and Brown, A. (2019) 'Archangel: Tamper-proofing Video Archives using Temporal Content Hashes on the Blockchain' in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, IEEE. Available at: https://arxiv.org/abs/1904.12059

Bunn, J. (2020) 'Working in Contexts for which Transparency is Important: A Recordkeeping view of Explainable Artificial Intelligence (XAI)', *Records Management Journal*, 30(2), pp. 143–153, available at: http://dx.doi.org/10.1108/RMJ-08-2019-0038

Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' in: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency,* January 2018, PMLR, 81, pp. 77–91.

Cabinet Office (UK). (2017) *Better Information for Better Government.* Available at: https://www.gov.uk/government/publications/better-information-for-better-government (Accessed: 29 August 2021).

Coleman, C.N. (2020) 'Managing Bias When Library Collections Become Data', *International Journal of Librarianship*, *5*(1), pp. 8–19.

Collomosse, J., Bui, T., Brown, A., Sheridan, J., Green, A., Bell, M., Fawcett, J., Higgins, J. and Thereaux, O. (2018) 'ARCHANGEL: Trusted Archives of Digital Public Documents' in: *Proceedings of the ACM Symposium on Document Engineering*, August 2018, Halifax, NS, Canada, ACM DocEng, pp. 1-4.

Collomosse, J., et al. (2018) 'ARCHANGEL: Trusted Archives of Digital Public Documents'.
Day, C. (2020). *Computing Cholera? 'Distant Reading' General Board of Health catalogue data.* Available at: https://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/our-research-projects/2020-annual-digital-lecture-staff-research-poster-exhibition/#computing-cholera

Day, C. (2021) *Cholera! Public health in mid-19th century Britain.* Available at: https://media.nationalarchives.gov.uk/index.php/cholera-public-health-in-mid-19th-century-britain/

Dearborn, C. and Meister, S. (2017) 'Failure as Process: Interrogating Disaster, Loss, and Recovery in Digital Preservation', *Alexandria*, 27(2), pp. 83–93.

Decker, S., Kirsch, D. Venkata, S., and Nix, A. (2021) 'Finding Light in Dark Archives: Using AI to Connect, Context and Content in Email', accepted for publication in the *Journal of Knowledge, Culture and Communication*, *AI & Society*.

DOMO (2018) *Data Never Sleeps 6.0* [online], available at: https://www.domo.com/assets/downloads/18_domo_data-never-sleeps-6+verticals.pdf

Dunley, R. and Pugh, J. (2021) 'Do Archive Catalogues Make History?: Exploring Interactions between Historians and Archives', *Twentieth Century British History*. Available at: https://doi.org/10.1093/tcbh/hwab021 (Accessed: 7 September 2021)

Fesharaki, Mehdi N., et al. (2020) 'A Conceptual Model for Socio-Pragmatic Web Based on Activity Theory', *Cogent Education*, 7(1). Available at: https://doi.org/10.1080/2331186X.2020.1797979 (Accessed 6 October 2021).

Floridi, L. and Cowls, J., (2019) 'A Unified Framework of Five Principles for AI in Society', HDSR, 1(1), pp. 1–15. Available at: https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/7

Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M. and Kagal, L. (2018) 'Explaining Explanations: An Overview of Interpretability of Machine Learning' in: *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, October 2018, IEEE, pp. 80-89.

Gurciullo, S. (2017) 'Keeping born-digital literary and artistic archives in an imperfect world: theory, best practice and good enoughs', *Comma*, 1, pp. 49–65. Available at: https://doi.org/10.3828/comma.2017.4

Hoffman, G. L. (2019) *Organizing Library Collections: Theory and Practice*, London, Rowman & Littlefield.

Hyperledger Fabric (2020) [online] 'Introduction', available at: https://hyperledger-fabric.readthedocs.io/en/release-2.2/whatis.html

Isaacson, S. (2013) *Computer Technology Review*, March.

Jo, E.S. and Gebru, T. (2020) 'Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning' in: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020, pp. 306-316. Available at https://dl.acm.org/doi/10.1145/3351095.3372829

Lemieux, V.L. (2016) 'Trusting records: is Blockchain technology the answer?', *Records Management Journal*, 26(2), pp. 110–139.

Lightfoot, D. (2021) *The SafePod Network (SPN)* [online] available at: https://safepodnetwork.ac.uk (Accessed: 26 August 2021)

Mas, S., Maurel,D., and Alberts, I. (2011) 'Applying Faceted Classification to the Personal Organization of Electronic Records: Insights into the User Experience', *Archivaria*, 72, pp. 29–59.

Mcdonald, G., Macdonald, C., and Ounis, I. (2020) 'How the Accuracy and Confidence of Sensitivity Classification Affects Digital Sensitivity Review' ACM Transactions on Information Systems, 39, 4, pp. 1-34. Available at: https://doi.org/10.1145/3417334

McDonald, G., Macdonald, C., Ounis, I. (2020) 'Active Learning Stopping Strategies for Technology-Assisted Sensitivity Review', in: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York (USA): Association for Computing Machinery, pp. 2053–2056.

Merwood, H. (2020) Risk Alert: Insufficient Technical Metadata. [Blog] *Digital Preservation Coalition*. Available at: https://www.dpconline.org/blog/risk-alert-insufficient-technical-metadata

Nix, A. et al. (2021) 'Finding Light in Dark Archives: Using AI to Connect, Context and Content in Email', presentation at AURA: Artificial Intelligence and Archives: What comes next? Online conference, available at: https://www.aura-network.net/wp-content/uploads/2021/04/Adam-Nix-slides.pdf

The ODI. (2018) *How can smart contracts be useful for business?* [online] available at https://theodi.org/wp-content/uploads/2018/05/378720579-How-can-smart-contracts-be-useful-for-businesses.pdf (Accessed: 7 September 2021).

Porat, A. Pratap, A., Shah, P. and Adkar, V. (2017) 'Blockchain Consensus: An analysis of Proof-of-Work and its Applications' [online], available at: https://www.scs.stanford.edu/17au-cs244b/labs/projects/porat_pratap_shah_adkar.pdf

Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., Stuart, K. (2019) 'More human than human? Artificial intelligence in the archive', *Archives and Manuscripts*, 47, pp. 179–203. Available at: https://doi.org/10.1080/01576895.2018.1502088

Schneider, J. et al., (2019) 'Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives', *Archives and Manuscripts*, 47(3), pp. 305-326. Available at: https://doi.org/10.1080/01576895.2019.1622138

Seles, A. (2021) Interview for the AURA project (Archives in the UK/ Republic of Ireland and AI), 28 May.

Sheridan, J. (2021) Interview for the "Unlocking our Digital Past" project, Loughborough University, 29 June.

Stanford Libraries Projects. (2021) *About ePADD* [online] available at: https://epadd.stanford.edu/epadd/about (Accessed: 25 August 2021)

The National Archives UK. (2016) *The application of technology-assisted review to born-digital records transfer, Inquiries and beyond* [online] available at: https://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf

TNA (2017) *Digital Strategy* [online], available at: https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf

TNA (2019) *Areas of Research Interest* [online] available at: https://www.nationalarchives.gov.uk/documents/areas-of-research-interest.pdf

TNA. (2016) *The Application of Technology-Assisted Review to Born-Digital Records Transfer, Inquiries and Beyond.*

TNA. (2017–2020) *Strategic Vision for Archives* [online], available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/strategic-vision-for-archives/

TNA. (2017) *Risk Assessment Handbook* [online], available at: https://www.nationalarchives.gov.uk/documents/information-management/Risk-Assessment-Handbook.pdf

TNA. (2020) *Digitisation and Digital Archives*, available at: https://www.nationalarchives.gov.uk/about/our-role/transparency/digitisation-and-digital-archives/

TNA. (2021) *Bridging the Digital Gap* [online], available at: https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/bridging-digital-gap-technical-traineeships-archives/

TNA. (2021) *Novice to Know-How* [online], available https://www.nationalarchives.gov.uk/archives-sector/projects-and-programmes/plugged-in-powered-up/novice-to-know-how/

TNA. *Openness, Access and Use* [online] available at: https://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/our-research-priorities/priority-research-themes/openness-access-and-use/

TNA. *Priority Research Themes.* Available at: https://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/our-research-priorities/priority-research-themes/
UCL. (2019) 'From Black Box to Tip of the Iceberg: Creative Engagement with the Emergence of XAI (Explainable Artificial Intelligence)', available at: https://cpb-eu-w2.wpmucdn.com/blogs.ucl.ac.uk/dist/e/653/files/2019/10/HeXAI-leaflet.pdf

Venkata, S. (2021) Interview for the AURA project (Archives in the UK/ Republic of Ireland and AI), 21 May.

Venkata, S., Young, P., Bell, M. and Green, A. (2021) 'Alexa, is this a Historical Record?', accepted for publication in the special edition *Computational Archival Science (CAS) of Journal on Computing and Cultural Heritage*.

Wang, H. and Yang, D., 2021. Research and Development of Blockchain Recordkeeping at the National Archives of Korea, *Computers*, 10(8).

Winters, J., and Prescott, A. (2019) 'Negotiating the born-digital: a problem of search', *Archives and Manuscripts*, 47, pp. 391–403. Available at: https://doi.org/10.1080/01576895.2019.1640753

Woods, K., and Lee, C.A. (2015) 'Redacting Private and Sensitive Information in Born-Digital Collections' in: *Archiving 2015 Final Program and Proceedings*, *May 2015*, Los Angeles: Society for Imaging Science and Technology, pp. 2–7.

Websites:

www.aeolian-network.net
https://www.archangel.ac.uk/
https://bitcurator.net/
https://www.dbpedia.org.
www.eark-project.com
https://safepodnetwork.ac.uk