





COMMENT

## Material and Digital Archives: The Case of Wills

Harry Smith  and Emily Vine 

Department of Archaeology and History, University of Exeter, Exeter, UK

**Corresponding author:** Harry Smith; Email: [h.j.smith2@exeter.ac.uk](mailto:h.j.smith2@exeter.ac.uk)

(Received 23 July 2024; revised 24 July 2024; accepted 24 July 2024;  
first published online 26 September 2024)

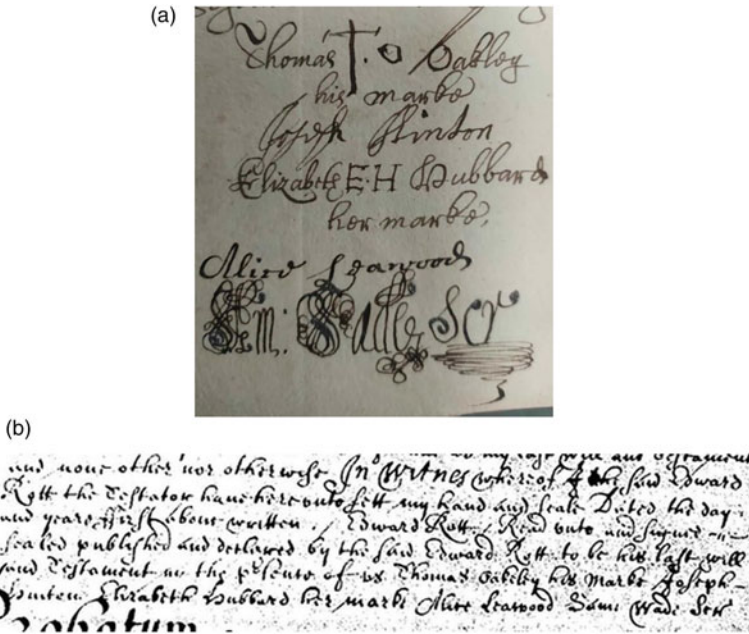
### Abstract

The range of digital sources available to historians has expanded at an enormous rate over the last fifty years; this has enabled all kinds of innovative scholarship to flourish. However, this process has also shaped recent historical work in ways that have not been fully discussed or documented. This article considers how we might reconcile the digitisation of archival sources with their materiality, with a particular focus on the probate records of the Prerogative Court of Canterbury (PCC). The article first considers the variety of digital sources available to historians of the United Kingdom, highlighting the particular influence of genealogical companies in shaping what material is available, how it has been digitised and how those sources are accessed. Secondly, we examine the PCC wills' digitisation, what was gained and what was lost in that process, notably important material aspects of the wills. This article does not seek to champion archival research in opposition to digitally based scholarship; instead, we remind historians of the many ways in which the creation of sources shape their potential use, and call on historians to push for improvements in the United Kingdom's digital infrastructure to avoid these problems in future.

**Keywords:** Archives; Digitisation; Wills; England; Wales

This article explores one of the key questions faced by historical researchers in the twenty-first century – how to reconcile the digitisation of archives with their materiality – through a consideration of the digitisation of one of the largest and most-used archives of English and Welsh historical records: the probate records of the Prerogative Court of Canterbury (PCC). These records, specifically the registered copies of wills made between 1384

© The Author(s), 2024. Published by Cambridge University Press on behalf of The Royal Historical Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.



**Figure 1.** Signatures as they appear in the original will of Edward Rott PROB 10/980 (top) vs in the registered copy PROB 11/317/402 (bottom). (Sources: TNA, PROB 10/980; PROB 11/317/402. Photo © Emily Vine.)

and 1858, are used by thousands of historical researchers and genealogists every month, but the overwhelming majority of users access these records though their digital surrogates: the manuscripts are rarely viewed. Indeed, these digital surrogates are several steps removed from the original will drawn up by the scribe or scrivener, signed and sealed by witnesses and the testator, and from the sickbed in which the deceased had made their final wishes known. These digital surrogates reproduce not the original will containing the writing or marks of the scribe, testator or witnesses, but the microfilms of the registered copies subsequently made by the church court clerk.

Some of the issues we will explore in this article become apparent when illustrated by a visual example. Figure 1 compares excerpts from two versions of the will of Edward Rott, a blacksmith of the City of London who died in 1665. It shows the signatures and marks of the witnesses and scribe as they appear in the original will (PROB 10/980, top) and the digital surrogate of the registered copy (PROB 11/317/402, bottom).<sup>1</sup> The original copy contains the signatures and marks of several witnesses who were present when the will was made, including the elaborate autograph of the scrivener, Samuel Wade. The

<sup>1</sup> Original will of Edward Rott, The National Archives (TNA), PROB 10/980 and registered copy, TNA, PROB 11/317/402, d.1665.

registered copy was then made when the will was proved after Edward Rott's death. It was written out by a church court clerk, and of course does not replicate the signatures or marks of the witnesses, but neither does it replicate their formatting (the signatures are listed at the bottom of the original will, but appear as part of main text of the copy, separated by commas). The registered copy replicates the content but not the materiality of the original will, and the microfilm or digital surrogate replicates the content but not the materiality of the registered copy. In a climate where researchers increasingly rely on digital surrogates as a proxy for a physical document, as a means of catching a glimpse of the circumstances in which that document was drawn up, we need to be constantly mindful of the complex and centuries-old archival and reproduction histories that have led us to the electronic image on our screen. In the case of Edward Rott, this archival history means we lose any information about literacy and authorship we might have discerned from the individuals' handwriting, as well as evidence pertaining to the stylistic features of early modern wills and the work of the scrivener.

The issues surrounding access and use of wills are faced by all historians who make use of digital sources, so this article begins by surveying existing digitised resources pertaining to British and Irish history. It considers the accessibility of these resources and demonstrates how their temporal coverage largely maps onto the time periods most used by genealogists. It then considers the digitisation and materiality of the PCC's archive of registered will copies and shows how the decision to microfilm the registered copies, but not the original wills, has shaped the types of research that can be conducted. In doing so this article touches upon some important methodological issues facing historical research (broadly defined) including the benefits and limitations of digitisation, the influence of keyword searchable interfaces and the difficulties of replicating the materiality of these sources and their archives. One of its key arguments pertains to the significance of the context of the archive and the problems that arise when digital surrogates remove sources from that context. In concluding it offers some suggestions for best practice for researchers and for the future of digitisation projects.

### The current state of digitised material

Over the last fifty years an ever-expanding amount of historical material has been digitised. Others have traced the history of this process, exploring how different institutional, commercial and intellectual factors have shaped what is available and in what format; here, we are concerned with what is available to historians today and how the characteristics of these online sources enable and limit the kind of historical writing that can be produced using them.<sup>2</sup> Our

---

<sup>2</sup> For useful accounts of the British case, see Tim Hitchcock, 'Digitising British History since 1980', *Making History*, Institute of Historical Research, [https://archives.history.ac.uk/makinghistory/resources/articles/digitisation\\_of\\_history.html](https://archives.history.ac.uk/makinghistory/resources/articles/digitisation_of_history.html) (accessed 7 May 2024); Adam Crymble, *Technology and the Historian: Transformations in the Digital Age* (Urbana, IL, 2021), 46–78.

fundamental argument is that, in most cases, users of these sources lack even the most basic information on how the resource was created. The move from the material to the digital is not simply a case of photographing records, or even transcribing them; rather digitisation is the whole process from the selection process, through imaging, storage of images, processing of images whether this involves transcription or not, formatting of resulting data, the creation of metadata and other documentation, and the distribution of the resulting material. Every stage of this process affects the resulting source and impacts how a user can utilise the material. All too often users do not know how the material in a digital collection was selected, what Optical Character Recognition (OCR) software was used, or how the output was checked. These factors and more can have profound effects on the nature of the resource and need to be fully considered by users.<sup>3</sup> Following on from this, we argue that access is not just about having permission to use a given resource; instead, as with non-digital sources, not knowing how and why a digital resource was created will lead the user astray.

Several previous papers have demonstrated how digitisation methods and storage practices can affect what kinds of historical work are possible. For example, Tim Hitchcock noted the problems with poor-quality OCR and obscure search algorithms have created substantial problems in using many digitised texts.<sup>4</sup> Michael Moss and Tim Gollins have drawn attention to the impact of often poorly understood algorithmic and machine processes on what is available digitally, especially when it comes to born-digital content, and urged archivists and historians to avoid a preoccupation with digitising as much as possible and to return to older archival processes of appraisal and reviewing what should and can be kept.<sup>5</sup> Several authors have noted the political aspects of digitisation, highlighting how choices in what material is selected and how it is digitised can create new inequalities and heighten pre-existing ones between the global north and south, and between different kinds of institutions and users.<sup>6</sup> In the British case, much discussion has revolved around the use of digitised newspapers, with many highlighting the opportunities they present, but recent work has shown just how difficult they are to use given the biases baked into them by the digitisation methods used and the

<sup>3</sup> OCR is the process by which images of text are converted into machine-readable text.

<sup>4</sup> Tim Hitchcock, 'Confronting the Digital: Or How Academic History Writing Lost the Plot', *Cultural and Social History*, 10 (2013), 9–23.

<sup>5</sup> Michael Moss and Tim J. Gollins, 'Our Digital Legacy: An Archival Perspective', *The Journal of Contemporary Archival Studies*, 4 (2017), article 3. Michael Moss drew particular attention to these issues in the context of state records, see Michael Moss, 'The Hutton Inquiry, the President of Nigeria and What the Butler Hoped to See', *English Historical Review*, 120 (2005), 577–92.

<sup>6</sup> Gerben Zaagsma, 'Digital History and the Politics of Digitization', *Digital Scholarship in the Humanities*, 38 (2023), 830–51; Joseph Nockels, Paul Gooding and Melissa Terras, 'Are Digital Humanities Platforms Facilitating Sufficient Diversity in Research? A Study of the Transkribus Scholarship Programme', *Digital Scholarship in the Humanities* (advance access, 2024); Ian Milligan, *The Transformation of Historical Research in the Digital Age* (Cambridge, 2022), 18–19; Alexandra Ortolja-Baird and Julianne Nyhan, 'Encoding the Haunting of an Object Catalogue: on the Potential of Digital Technologies to Perpetuate or Subvert the Silence and Bias of the Early Modern Archive', *Digital Scholarship in the Humanities*, 37 (2022), 844–67.

particular titles selected.<sup>7</sup> Other sources have recently started to receive similar consideration, notably newsreels, court papers, eighteenth-century and early modern published works.<sup>8</sup> Despite the varied subjects, this literature, in common with our paper, stresses the importance of understanding why and how digital materials are created.

While not disagreeing with these existing arguments, this article seeks to move on from debates over the benefits and drawbacks of digitisation. Instead, in this section, we survey existing digitised primary sources with a view to expanding our understanding of accessibility. Access is not simply a matter of whether or not a source is open to all or subscription based, it is also related to what can be done with that source. If material can be freely accessed by searching a source through a web application but cannot be downloaded in total then only certain kinds of research are possible. Similarly, if a source can only be downloaded in a particular file format, then the data are at least partially closed to those without the resources to process such material.

To examine this issue, we have attempted to identify all currently existing and available digital sources for the history of the United Kingdom and Ireland.<sup>9</sup> This is no simple task given the amount of digitisation which has happened over the last fifty years and it is likely that we have missed some resources and datasets. However, we believe we have covered most key collections and, at least, cover all types of digital resource available to scholars, even if not every example of a given type of source has been identified. Additionally, all sources discussed start their temporal coverage no later than the 1980s; partly to assist us in keeping a handle on the volume of material discussed, but also to ensure that the historical material examined is not overwhelmed by the sheer volume of data and texts produced and available online from the last thirty years. Our focus was on resources which provide historians with access to primary sources, so bibliographies and biographical sources,

---

<sup>7</sup> Adrian Bingham, 'The Digitization of Newspaper Archives: Opportunities and Challenges for Historians', *Twentieth Century British History*, 21 (2010), 225–31; Gioria Tolfo, Olivia Vane, Kaspar Beelen, Kasra Hosseini, Jon Lawrence, David Beavan and Katherine McDonough, 'Hunting for Treasure: Living with Machines and the British Library Newspaper Collection', in *Digitised Newspapers: A New Eldorado for Historians?*, ed. Estelle Bunout, Maud Ehrmann and Frédéric Clavert (Oldenburg, 2023), 25–46.

<sup>8</sup> Sam Rutherford, 'Researching and Teaching with British Newsreels', *Twentieth Century British History*, 32 (2021), 441–61; Stephen H. Gregg, *Old Books and Digital Publishing: Eighteenth-Century Collections Online* (Cambridge, 2020); Jonathan Blaney and Judith Siefing, 'A Culture of Non-citation: Assessing the Digital Impact of British History Online and the Early English Book Online Text Creation Partnership', *Digital Humanities Quarterly*, 11 (2017), <https://digitalhumanities.org/dhq/vol/11/1/000282/000282.html>; Sharon Howard, 'Bloody Code: Reflecting on a Decade of the Old Bailey Online and the Digital Futures of Our Criminal Past', *Law, Crime and History*, 5 (2015), 12–24.

<sup>9</sup> There are considerable numbers of sources which should theoretically be available but cannot be accessed because the website hosting them is now defunct. When searching for oral history collections this seems to be a particular problem. This survey was carried out in March and April 2024; since then, more material has become available. For example, between 22 April 2024 and 5 June 2024 Ancestry.co.uk has added 11 new datasets comprising 187,715,831 individual records related to the United Kingdom and Ireland.

such as the *Oxford Dictionary of National Biography* or the *Bibliography of British and Irish History*, were not included.

The resources discussed below were identified through several avenues. Firstly, university library lists of databases were consulted. No university library subscribes to all historical databases, so, secondly, all major commercial providers of historical data were surveyed.<sup>10</sup> This produced a list of 247 historical databases covering a range of aspects of British history. Information on the volume of pages digitised was not available for all these databases, but the 165 cases where we do know the volume of digital material amounted to just over 152 million pages. These sources are broad in topic and type of material available. Many of these resources provide access to digitised newspapers and periodicals. Some are personal archives, although mainly related to important politics figures, such as the papers of the Chamberlain family, or in some cases these are not personal archives but instead state papers organised by person rather than function, such as the Cecil Archives.<sup>11</sup> Others are thematic, collecting various archival and published material around a theme; thus, Adam Matthew has collections such as ‘Gender: Identity and Social Change’, or ‘The Grand Tour’. British Online Archives offers many similar collections, such as ‘The Industrial Revolution: Technological Innovation in the Textile Industry, 1672–1929’ or ‘Slavery, Exploitation and Trade in the West Indies, 1759–1832’. Such collections are somewhat different in nature from the archival or newspaper collections in that they offer a curated selection of material on a given topic rather than attempting to reproduce entire runs of a given publication or the entirety of a personal archive. They are often rather narrower in scope than their titles imply: the grandly titled ‘Science and Marxism’, for example, actually contains the papers of William Wainwright, the British communist activist and theorist of scientific socialism, an interesting resource but perhaps a smaller topic than its title implies.

Thirdly, there are an increasing number of open access historical resources, some deriving from projects funded by the Research Councils and other bodies, some comprising digitised versions of long-standing analogue primary sources and some arising from university or other libraries digitising their own material, such as the LSE’s various digital collections.<sup>12</sup> We have identified 126 of these open resources, along with two others which are partially open: one (Queen Victoria’s Journals) is free to UK residents and the other (JSTOR’s Ireland Archive Collection) is free to further and higher education institutions.<sup>13</sup> These were identified through several avenues. Some were simply

---

<sup>10</sup> The following vendors were checked, not all of which provide databases fitting the criteria of starting pre-1980 and covering the United Kingdom: ALCS, Adam Matthew, Alexandra Street Press, Bloomsbury, Brepolis, Brill, British Online Archives, Cambridge University Press, Coherent Digital, Eastview, Ebsco, Elsevier, Gale, Heinonline, Irish Newspaper Archives, Iter, JISC, John’s Hopkins Press, JSTOR, Liverpool University Press, Macmillan, North Waterloo Academic Press, Oxford University Press, Proquest, Readex, Sabinet, UK Press Online, Wiley, Yale University Press.

<sup>11</sup> Provided by Gale and Proquest respectively.

<sup>12</sup> <https://lse-atom.arkivum.net/> (accessed 7 May 2024).

<sup>13</sup> <http://www.queenvictoriasjournals.org/home.do> and <https://subscriptionsmanager.jisc.ac.uk/catalogue/1101> (both accessed 7 May 2024).

resources known to the authors.<sup>14</sup> Others were found using lists of historical resources available online, such as those provided by the Institute of Historical Research or the Bodleian Library.<sup>15</sup> Again, information on the number of pages available was not provided for all these resources, but the seventy-two where these data was available provided over 441 million pages. Nearly 400 million of these pages contained the data provided by three sources: FreeCen, FreeBMD and FreeReg, the volunteer-based, open access versions of the UK censuses, civil registration records and parish registers.<sup>16</sup> These open access resources are extremely varied in topic. Some are produced by the state and typically cover political topics, such as Historic Hansard. Others are the outputs of academic research projects and so cover all manner of areas from the Legacies of British Slavery, to the Pulter Project, to the Survey of Scottish Witchcraft.<sup>17</sup> Finally, others are the outputs of various other organisations such as the membership records of the Inns of Temple.

There is a considerable number of other resources that are available for free but that require registration to access them. The United Kingdom Data Service (UKDS) stores and provides access to a wide range of data on the history of the UK to researchers. There are three types of resources in the UKDS; first, databases in the ReShare repository, which are freely available to anyone. Secondly, safeguarded datasets that can be accessed by anyone but require the user to register with the UKDS. Thirdly, secure access datasets for which the user has to complete a specific request and, in some cases, undertake safe researcher training. In total, the UKDS holds 698 datasets relevant to British history as defined in this article. Of these, 136 are open access, 556 require you to register with the UKDS before downloading and six require more stringent special access licences. We have also included two other datasets which are held elsewhere but similarly require registration in order to access them: the Calum Maclean Project and Around 1968: Activism, Networks, Trajectories.<sup>18</sup> The total size of these databases is hard to judge, but will run to the hundreds of millions of records given the presence of large datasets covering census, civil registration and parish records.<sup>19</sup> Once again the range of historical topics covered by these records is substantial; however, these do tend to be more data heavy. Many are databases derived from historical resources, such as the British Business Census of Entrepreneurs, or the Digest of Welsh Historical Statistics; others are deposits of interviews and

<sup>14</sup> Indeed, the authors themselves have contributed to the creation of four of them.

<sup>15</sup> <https://www.history.ac.uk/library-digital/collections/online-resources> (accessed 25 April 2024); <https://www.diiigo.com/profile/hfloxford> (accessed 25 April 2024).

<sup>16</sup> <https://www.freeukgenealogy.org.uk/> (accessed 3 May 2024).

<sup>17</sup> <https://www.ucl.ac.uk/lbs/> (accessed 19 April 2024); <https://pulterproject.northwestern.edu/> (accessed 19 April 2024); <https://witches.hca.ed.ac.uk/home/> (accessed 25 April 2024).

<sup>18</sup> <https://www.calum-maclean-project.celtscot.ed.ac.uk/database-access/register-for-access/> and <https://around1968.history.ox.ac.uk/> (both accessed 7 May 2024).

<sup>19</sup> The Integrated Census Microdata dataset, for example, is just one resource available from the UKDA but includes 183,470,912 individual records, see Kevin Schürer and Edward Higgs, *Integrated Census Microdata (I-CeM), 1851-1911* [data collection], UK Data Service (2023), SN: 7481, DOI: <http://doi.org/10.5255/UKDA-SN-7481-2>.

surveys which are now being increasingly used for historical research.<sup>20</sup> These resources, therefore, are similar to the open access ones in their breadth of coverage, but much narrower in the kind of material available, tending to take the form of machine-readable files ready to be processed by statistical software.

The remaining historical sources can be found through various genealogical websites. The expansion of these websites in the last few decades has been one of the main driving forces behind the digitisation of historical material. Of these websites, FamilySearch is mostly free, although you have to register an account to use it.<sup>21</sup> FamilySearch provides access to 231 different databases related to the United Kingdom and Ireland, totalling 876,835,233 records. Other genealogical websites require the user to pay subscription fees, but provide access to vast numbers of historical records. Ancestry has 1,921 datasets covering the United Kingdom and Ireland, containing a remarkable 3,398,576,084 records.<sup>22</sup> FindMyPast has fewer databases, but still provides 1,253 datasets to subscribers, totalling 1,919,888,010 records.<sup>23</sup> Finally, the Genealogist offers a smaller array of records, but has a number of notable differences, notably a set of historical mapping tools, as well as alternative transcriptions of the British censuses. They offer 1,195 datasets. Although it is not easy to discover how many records this dataset covers, it is likely to be in the hundreds of millions. These datasets are all focused on allowing people to trace their ancestors and so tend to have a different focus from many of the sources discussed above, most of which were designed either to provide a complete run of a particular source, whether *The Times* or Newton's personal papers, or were collections created to investigate a given topic, whether that is demographic transition, popular protest in medieval England or the history of the 1641 Irish Rebellion.<sup>24</sup> In contrast the collections on genealogical websites seek to gather as many names as possible; this purpose affects the kind of sources they digitise, the method of digitisation and the format in which the data are stored. Thus, they have a large preponderance of religious and state records, sources related to probate and land ownership, publications which list individuals, such as trade directories and organisation member lists, but relatively few personal archival records, little in the way of political sources and, where they have newspapers or periodicals, the point is not to provide an accurate transcription of such publications, but instead one that is good enough for names to be searched for in their contents. However, given their

<sup>20</sup> See 'Roundtable: Historians' Uses of Archived Material from Sociological Research', *Twentieth Century British History*, 33 (2022), 392–459.

<sup>21</sup> <https://www.familysearch.org/en/uk-ireland/> (accessed 3 May 2024). Some of their documents can only be accessed through FamilySearch centres, see <https://www.familysearch.org/en/centers/about> (accessed 10 July 2024).

<sup>22</sup> <https://www.ancestry.co.uk/> (accessed 22 April 2024).

<sup>23</sup> <https://www.findmypast.co.uk/home> (accessed 23 April 2024).

<sup>24</sup> See, respectively, Alice Reid, *Demographic and Socio-economic Data for Registration Sub-Districts of England and Wales, 1851–1911* [data collection] (2020), UK Data Service, SN: 853547, DOI: <http://doi.org/10.5255/UKDA-SN-853547>; Samuel Cohn, *Popular Protest in Late Medieval English Towns, 1196–1452* [data collection] (2012), UK Data Service, SN: 6979, DOI: <http://doi.org/10.5255/UKDA-SN-6979-1>; <https://1641.tcd.ie/> (accessed 26 April 2024).



resources it is clear that they are happy to digitise all manner of sources and so Ancestry and FindMyPast have several intriguing datasets, which are not primarily of genealogical interest but instead add historical colour or serve to entice users, such as FindMyPast's 'Views of Ireland' database.<sup>25</sup> These points will be discussed below in relation to wills, a key resource for genealogists and academic historians, but those two constituencies have substantially different interests and desires in terms of access, searching, format and, most fundamentally, what is digitised – the whole will or just names and dates.

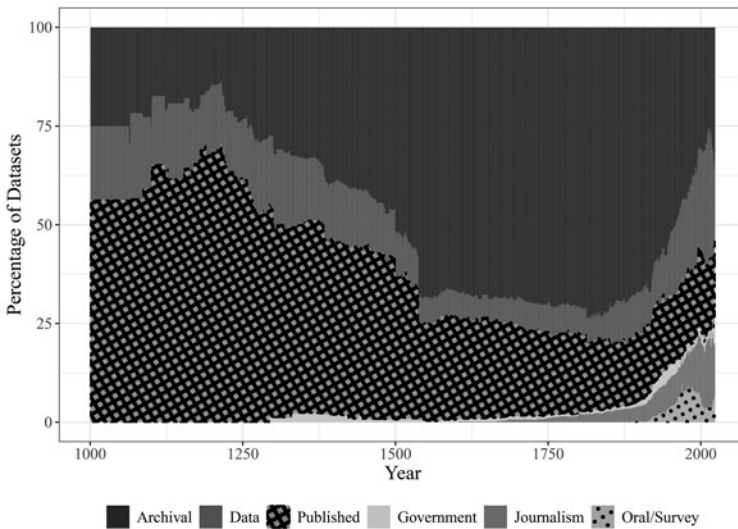
Taken together these subscription, registration and free historical resources provide scholars with access to billions of historical records. We have identified 5,675 historical databases and other resources that can be accessed online. They include a wide range of different kinds of material: untranscribed images of manuscript material, OCR'd newspapers, partially transcribed indexes to civil registration records, hand-transcribed personal correspondence and databases of millions of census records are just some of the different kinds of material available. In some cases, the original images can be consulted, in others they cannot, but in all cases at least part of the information from the original source is now reproduced in a digital format. We do not seek to ignore the differences between these resources; they have been created in diverse ways, are available in different formats, come with various access conditions and are used in different ways by varied audiences: academics, genealogists, archivists and other members of the public. Figure 2 gives some sense of the variety of material involved in these datasets and how the balance between different types changed over time.<sup>26</sup> However, we group them together here because despite these differences, there are important issues related to all digital resources which affect how any user utilises them. Issues around access, format, metadata and documentation fundamentally link the digital version to the materiality of the original sources and affect how academics and others can reliably use, and interpret others' use of, these digital resources.

Of the resources surveyed, 264 are entirely open source; 795 are free but require the user to register with the service providing them; and the remaining 4,616 are subscription based. These subscription ones are divided between the genealogical websites (4,369 datasets) where the usual method of access is through individual subscriptions and the remaining 247 where access is usually managed through institutional subscriptions.<sup>27</sup> The sheer size of the genealogical resources is such that they dominate the online historical environment

<sup>25</sup> Ancestry and FindMyPast's digitisation policies are interestingly discussed in Adam Kriesberg, 'The Future of Access to Public Records? Public-Private Partnerships in US State and Territorial Archives', *Archival Science*, 17 (2017), 5–25; it is not clear to what extent the practices discussed in this article apply to the United Kingdom or whether they are still current.

<sup>26</sup> Only resources with known date coverage have been included; before 1000 there are too few resources to be worth graphing.

<sup>27</sup> Ancestry offers educational institutions access to some of their resources through AncestryClassroom, <https://ancestryclassroom.co.uk/k12/k12home> (accessed 8 May 2024) and other institutions can subscribe to a program called AncestryLibrary for access to some of the Ancestry holdings. Similarly, FindMyPast offers a library subscription for institutions <https://www.findmypast.co.uk/help/articles/360009035958-does-findmypast-offer-group->



**Figure 2.** Databases surveyed by category of record, 1000–2024. (Sources: see text.)

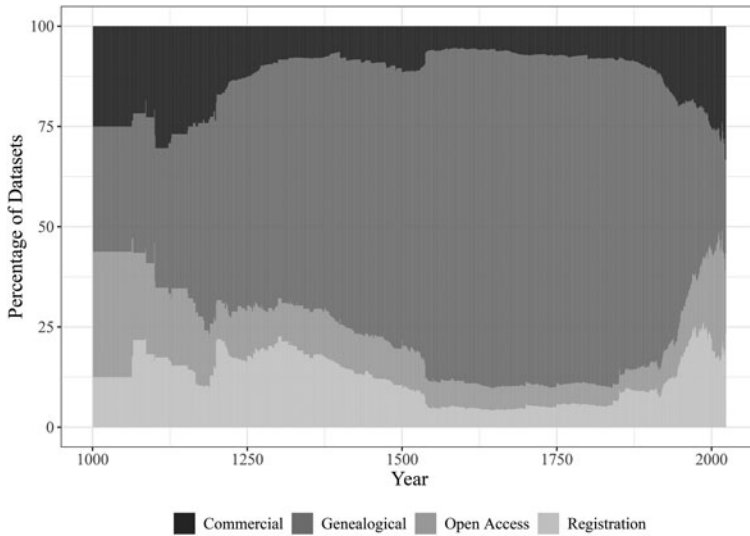
*Note:* The categories necessarily cover rather disparate resources. Archival covers all non-published records not found in the other categories; Data contains all resources which provide tabulated or other data derived from sources; Published are all materials which have been published in one way or another including artworks and film, apart from newspapers and periodicals which are contained in Journalism; Oral/Survey includes all oral history archives and all outputs of surveys (whether conducted in person or not). In each year the total number of active databases is counted and the percentage available from each category is calculated. For example, for the year AD 1000 there are 17 active databases: 5 archival, 3 data and 9 published.

to an astonishing extent, providing access to vast amounts of material, but shaping access and coverage in particular ways. This point will be returned to below, but at the most simple level, therefore, access to many digitised historical resources is mediated by an individual's ability to afford personal subscriptions and whether or not they are a member of a body which offers institutional subscriptions. Furthermore, many of the free to access resources that require registration are easier to access with a university affiliation, particularly for the UKDS. Subscriptions are within reach of some, but far from all, and university membership is even more restricted.

We can take this point further. The three categories not only differ in how access is obtained, but also in terms of the material available. Much has been made of how digitised sources are not representative of the totality of the historical record, but instead are selected; particular kinds of sources and particular topics tend to be digitised and others are not.<sup>28</sup> Within digitised material, however, there are patterns in the type of document and time frame covered

[subscriptions](#) (accessed 8 May 2024). These are generally used by schools and libraries rather than universities.

<sup>28</sup> See references in notes 2–5 above.



**Figure 3.** Types of database available by year, 1000–2024. (Sources: see text.)

Note: Genealogical covers sources accessible through the four family history websites discussed in the text: Ancestry, FindMyPast, the Genealogist and FamilySearch; Commercial refers to any source provided by a commercial body; Open Access and Registration are all products of academics, government, charities or private individuals.

that vary by access. As noted above, open access resources cover a more varied set of material and topics than those available through subscriptions. This is unsurprising given that the open sources tend to be the outputs of academic projects coming from all areas of historical scholarship, whereas subscription datasets are driven by commercial imperatives, hence focusing on genealogical sources or material which aligns closely with historical curriculums, and by how easy a source is to access and digitise. This also fundamentally affects the time period covered by different kinds of resource and drives much of the variation in the utility or otherwise of digital material to historians of different periods, as shown in Figure 3, a point of great importance that we will return to in the next section. This shows that genealogical databases dominate the available resources for the period 1500 to the late twentieth century, when between half and three quarters of active databases in any given year are accessible through the genealogical websites. This is the key period for genealogy in the United Kingdom and Ireland given the nature of the sources most often used for such work: parish registers, civil registration material and the census. Before and after that period the other kinds of resource become proportionately more important. Partly this reflects simply not wanting to compete with the genealogical websites, which have far greater resources available for digitisation, and so other companies and academics focus elsewhere on materials such as personal letters and archives, newspapers, other manuscript sources, political papers and so on, which are of less import for genealogists and so of lower priority for Ancestry, FamilySearch, FindMyPast and the

Genealogist. However, it is also driven by the lack of funds available for digitisation; it is difficult for academics to obtain funding solely for digitising resources, as opposed to digitisation as part of a wider research project, and this means that most large-scale digitisation is left to commercial companies.<sup>29</sup> This is only to consider the origins of these resources, however, not their use. As we will discuss below, sources provided by genealogical companies are often used for scholarship other than family history.

These points about access have been made by others before, in terms of the impact both on users and on archives.<sup>30</sup> However, the question of what history can be written with these digitised historical resources goes beyond a question of open access vs subscription schemes.<sup>31</sup> Not least because the amount of data and other sources available open access is so extensive, as shown above. Furthermore, it is frequently possible to negotiate access to material digitised by companies such as FindMyPast or Ancestry, even if such data often come with certain conditions and perhaps in a format which is designed primarily for genealogical users.<sup>32</sup> Some commercial companies allow users to access the underlying data from their resources through proprietary interfaces. For example, Proquest offer TDM Studio, a platform in which users can undertake various kinds of textual analysis on their digitised collections.<sup>33</sup> Where access to an entire dataset is not available it is also feasible to scrape data from genealogical websites for use in historical study, although there is some uncertainty over whether this is or is not allowed given the terms and conditions of such websites.<sup>34</sup> Alternatively, use is increasingly being made of datasets created

<sup>29</sup> Ruth Ahnert, Emma Griffin, Mia Ridge and Giorgia Tolfo, *Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project* (Cambridge, 2023), 23–5.

<sup>30</sup> Barry Godfrey, 'Future Perspectives on Crime History as "Connected History"', *Crime, histoire et sociétés*, 21 (2017), 45–6; David Thomas and Michael Moss, 'The Commercialisation of Archives: The Impact of Online Family History Sites in the UK', in *Do Archives Have Value?*, ed. Michael Moss and David Thomas (2019), 141–66.

<sup>31</sup> The open vs closed issue is surveyed in the context of nineteenth-century sources by Ahnert *et al.*, *Collaborative Historical Research*, 25–32.

<sup>32</sup> For some examples of work where access to such records was granted see Neil Cummins, Morgan Kelly and Cormac Ó Gráda, 'Living Standards and Plague in London, 1560–1665', *Economic History Review*, 69 (2016), 3–34 (Ancestry's parish records); Richard Ward, 'State Authority and Convict Agency in the Paper Panopticon: The Recording of Convict Ages in Nineteenth-Century England and Australia', *Australian Historical Studies*, 52 (2021), 509–32 (Ancestry's Hulks registers and FindMyPast's Prison registers, see also [https://www.digitalpanopticon.org/About\\_The\\_Project](https://www.digitalpanopticon.org/About_The_Project) (accessed 9 May 2024)); Fabon Dzogant, Thomas Lansdall-Welfare, FindMyPast Newspaper Team and Nello Cristianini, 'Discovering Periodic Patterns in Historical News', *PLoS ONE* 11/11 (2016), <https://doi.org/10.1371/journal.pone.0165736> (FindMyPast newspapers); Carry van Lieshout, Joe Day, Piero Montebruno and Robert J. Bennett, 'Extraction of data on Entrepreneurs from the 1871 Census to Supplement I-CeM', Working Paper 12, ESRC project ES/M010953, 'Drivers of Entrepreneurship and Small Businesses', <https://doi.org/10.17863/CAM.27488> (The Genealogist census data).

<sup>33</sup> <https://proquest.libguides.com/tdmstudio/home> (accessed 20 March 2024).

<sup>34</sup> Jorge L Contreras, Kyle Schultz, Craig C. Teerlink, Tim Maness, Laurence J Meyer and Lisa A Cannon-Albright, 'Legal Terms of Use and Public Genealogy Websites', *Journal of Law and the Biosciences*, 7 (2020), 1–24; for an example of historical work that uses scraped data which is definitely available for academic use, see Neil Cummins, 'Where is the Middle Class? Evidence from 60

from the outputs of genealogists' work, such as crowd-sourced genealogies which underpin the Familinx database.<sup>35</sup> Such databases combine large numbers of family trees created by genealogists from digital or other sources to produce datasets with millions of linked individuals which are then used for historical studies of migration, fertility, mortality and so on. It is possible, therefore, for historians and others to get access to sources even when they are not explicitly open access; however, this does not settle the issue raised in our introduction concerning the impact of digitisation and access on historical scholarship, because access is not simply a matter of obtaining the source in question. Access is also about the process of digitising a source, the interface used to work with it, and the metadata and documentation provided. All these factors shape what can and cannot be done with a given historical resource. These factors have often not been given sufficient consideration and have had a profound impact on historical scholarship in our digital age. If you have access to a source but have no real understanding of how or why it was created, then research based on that source will probably be flawed, just as it would be if a non-digital source was approached with a similarly uncritical mindset. This section finishes with a brief general consideration of this problem, before returning to the issue in the next section, where we will consider this in more detail in the context of the PCC wills.

The most serious issue with many of these historical resources is the lack of documentation. The sources made available by commercial companies and those from the genealogical providers give little indication of how they were digitised. For example, the sources available on Ancestry that have been transcribed may have been transcribed in a number of different ways. Some are already transcribed datasets which Ancestry has licensed from other organisations, such as many of the UK parish registers; others have been keyed by volunteers through the now-discontinued Ancestry World Archives project. In other cases, there is no information provided on how a source was transcribed.<sup>36</sup> In both cases details on conventions, checking and other aspects of the transcription process are not readily available, meaning that even if you have the entirety of a given dataset you cannot be sure exactly how it was created. This uncertainty limits what can be done; for example, if you wished to study the distribution of surnames in a Welsh parish register you would need to know how Welsh language names were dealt with by the

---

Million English Death and Probate Records, 1892–1992', *The Journal of Economic History*, 81 (2021), 359–404.

<sup>35</sup> Joanna Kaplanis, Assaf Gordon, Tal Shore, Omer Weissbrod, Dan Geiger, Mary Wahl, Michael Gershovits, Barak Markus, Mona Sheikh, Melissa Gymrek, Gaurav Bhatia, Daniel G. Macarthur, Alkes L. Price and Yaniv Erlich, 'Quantitative Analysis of Population-Scale Family Trees with Millions of Relatives', *Science*, 360 (2018), 171–5.

<sup>36</sup> The article noting the discontinuation of the Ancestry World Archive project alludes to the use of 'new technologies' for transcription, suggesting they are increasingly using automated transcription. Whether these are Optical Character Recognition or Handwritten Text Recognition algorithms is unclear; the sections about each source also do not record this information, [https://support.ancestry.com/s/article/Discontinuing-the-Ancestry-World-Archives-Project?language=en\\_US](https://support.ancestry.com/s/article/Discontinuing-the-Ancestry-World-Archives-Project?language=en_US) (accessed 9 May 2024).

original scribes and by the transcribers; this information is not available on Ancestry for their Welsh parish registers. If you wish to carry out corpus linguistic analysis on the text present in these commercially produced sources, such as the court records, town directories and histories held on ancestry and FindMyPast, it is essential to know whether spellings have been changed, whether to 'correct' them, or to update them to modern spellings; again this information is difficult to obtain for many of these datasets. Finally, genealogical websites often allow users to provide corrections to transcription errors they find while using the website. In many cases these improve the transcription, but because there is no style guide, they sometimes change a name from what is actually written in the source to what the user 'knows' the name to be. For example, PROB 11/319/128 is John Vaughan's will, probated on 18 January 1665. However, this will cannot be found on Ancestry by searching for John Vaughan, or even just Vaughan; instead to find it you need to search for Vaughn, because it has been transcribed as Johes Vaughn by Ancestry. This entry has a user-suggested correction, John Vaughn, but this still has the incorrect surname, despite it clearly being spelled 'Vaughan' in the will itself.<sup>37</sup> Such problems are only increased in cases where the image is unclear. Once again, there is considerable uncertainty about how sources were digitised and, consequently, how scholars should use them and how confident they can be in doing so.

These issues are multiplied when using datasets like Familinx, which ultimately derive from family history sources, many found through these genealogical websites, but which give little to no information about where and why links were created. Such derived databases are, therefore, based on an unknown set of records, which vary, to an unknown extent, in quality and coverage. In terms of quality, both the quality of the original record and the quality of the digitisation can vary. In terms of coverage, this changes by date covered by the link and by the date on which the genealogist created the link, given the expansion and change in records available and methods of searching. In many cases, the creation of these genealogical sources is a black box that fundamentally changes how historians and others can use such material.

This lack of clarity about the methods of digitisation and transcription is not limited to genealogical websites, however. Many commercial historical databases include little or no information on how a given collection was created. Proquest, for example, provides no details on how texts and other sources were digitised. Early English Books Online (EEBO), for instance, has a brief note about transcription being undertaken through the Text Creation Partnership between Proquest and the Universities of Michigan and Oxford, but offers no details on the conventions used, severely impacting how that data can be used. This information can be found on a separate Text Creation Partnership website, but given the absence of such information from EEBO itself, and the fact that text mining can be carried out on the EEBO corpus through TDM Studio, the potential for misleading analysis is considerable. Other providers give more information, but it is still usually less fulsome than scholars would want. Gale and Adam Matthew, for example, both note where OCR

<sup>37</sup> Will of John Vaughan, Gentleman, Montgomeryshire, 18 Jan. 1666, TNA, PROB 11/319/128.

and HTR models have been used to generate transcriptions, but provide no details on how the models were trained, or how the results were checked. Such models are undoubtedly of better quality than they were when Tim Hitchcock warned us about the perils of low-quality transcriptions a decade ago, but the quality of information given to users about the models used has not, leading to a considerable amount of uncertainty even when just searching such datasets, let alone using them to undertake quantitative and statistical analysis.<sup>38</sup> Of the 247 datasets from commercial providers surveyed for this article, 59 gave no information on how the records included were digitised. Of the remaining 188, the only collection which provides a good description of the method used to transcribe the material are the 272 volumes of primary material provided by British History Online, which notes that double rekeying is the usual method of transcription, with closely checked OCR'd material also being added recently. Even they, however, do not provide details on the conventions used, however.<sup>39</sup>

In general, the resources available from the UKDS provide better documentation about how the datasets were created, but even here this is not perfect given that some of the resources have been deposited long after their initial creation as originally analogue databases. For example, see the documentation for a set of transcribed nineteenth- and twentieth-century vaccination registers, which notes just that 'the layout of the spreadsheets is not consistent, and the information contained in them varies. Standard field names have not been used. Some columns are not named. Information is not always placed in the same fields. Standard codes have not been used consistently in all spreadsheets.' This is an artefact of how these records were created; it can be overcome but again it presents a barrier to use and limits what kind of analysis can be reliably undertaken.<sup>40</sup>

Finally, the other open access resources present a mixed picture in terms of details about digitisation. Of the 126 datasets surveyed here, 56 gave no detail on how they were created; the remaining 70 did provide at least some information, but it was often vague or simply notes the kind of transcription performed, whether manual or by OCR or HTR model. Some, however, such as the Prosopography of Anglo-Saxon England, provide excellent documentation on the methods and conventions used in the creation of that particular resource.<sup>41</sup>

As we noted at the start of this section, users of digital historical resources often lack basic information about the material. This is a major problem for using these sources in historical research and for people reading or trying

<sup>38</sup> Hitchcock, 'Confronting the Digital'.

<sup>39</sup> [https://www.british-history-ac-uk.uolibrary.idm.oclc.org/about](https://www.british-history.ac.uk/uolibrary.idm.oclc.org/about) (accessed 18 April 2024).

<sup>40</sup> L. James, C. Fellows, P. Birch, J. Walsh, J. Robinson, S. Green, J. Rider, J. Hack, H. Coleman, N. Cattell, M. Drake, W. Baird, M. Razzell, A. Dix, A. Clark, S. Smith, P. Buckingham, R. Proctor, L. Davies, E. Hall, G. Culshaw, V. Dodgson, T. James and S. Richens, *Decline of Infant Mortality in England and Wales, 1871-1948: A Medical Conundrum; Vaccination Registers, 1871-1913* [data collection] (2001), UK Data Service, SN: 4127, DOI: <http://doi.org/10.5255/UKDA-SN-4127-1>.

<sup>41</sup> <https://pase.ac.uk/about/research-methodology/> (accessed 19 April 2024). Others with notably good documentation include *Old Bailey Online*, *London Lives*, *The Digital Panopticon*, *Addressing Health*, *The British Business Census of Entrepreneurs*, *Hearth Tax Digital*, *Social Bodies*; apologies to others not explicitly mentioned.

to replicate research based on such data. If you do not know how a transcription was produced, how can you rely on any searches made within it? Such searches might claim to have identified all cases of a word or phrase in a given corpus, but if the transcription is faulty or unknown conventions have affected how given words were transcribed, the researcher and reader may only be considering a sample of that population, leading to errors of analysis and interpretation. It is unlikely that such issues will be resolved satisfactorily for existing datasets. Indeed, given the lack of documentation or metadata in some cases, it may be impossible to retrieve any information on how those records were digitised. However, it is possible, going forward, to ensure that such information is supplied with newly digitised resources and, indeed, as the next section will demonstrate, the provision of good documentation and metadata can expand the kind of work that can be accomplished with digital records beyond historians' previous focus on textual and statistical analysis, into the realm of material culture and beyond. Furthermore, we will see how consideration of the material aspects of documents has to be included when digitising documents as these aspects will directly affect the quality of the transcription provided.

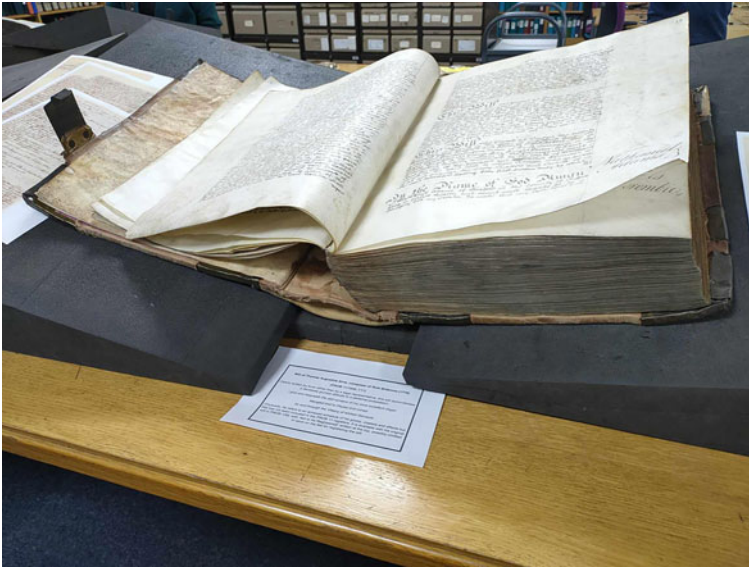
## **The materiality of historical sources**

### *The microfilming and digitisation of PROB 11*

Faced with a sea of searchable digitised historical data, it can be easy to forget about the physical manuscripts, printed books and other sources that underpin these databases. This article so far has considered the influence of digitised genealogical records (such as censuses and baptism records) within the landscape of digitised historical resources. It now turns to consider the specific context of digitised wills, and in particular, the registered copies of wills proved before the Prerogative Court of Canterbury (PCC). The digital surrogates of these registered copies of wills are consulted by thousands of researchers each month, for a broad range of historical and genealogical purposes. These registered copies of wills are held at The National Archives (TNA) at Kew, in a series known as PROB 11. Registered wills were copied by church court clerks onto quires of eight leaves, which were bound into large volumes with leather straps or brass bindings. It is important to consider the physical volumes of PROB 11, and the arrangement and referencing system of the wills in each, in order to understand the microfilm and digital surrogates that most researchers make use of. 2263 of these volumes exist: they date from 1384 to 1858 and contain hundreds of thousands of registered copies of wills. Where the original version of the will survives, it appears in a separate series known as PROB 10 – these have largely not been digitised. The decisions made about which series have been digitised (and how they were digitised and through which interfaces they are accessed) have a direct bearing on the type of historical research that can be produced.

The archival history of PROB 11 ensures that there are some inconsistencies with the arrangement and numbering of wills within each volume. The indexes





**Figure 4.** A photograph of one of the PROB 11 volumes, in this case PROB 11/1040, open at Thomas Arne's will. (Source: PROB 11/1040/181. Photo © Emily Vine.)

to these volumes were often compiled before the files were transferred to the Public Record Office, and therefore use a different reference system. These include contemporary manuscript indexes or calendars. Names were recorded as the grant of probate was made, ensuring that each entry could be easily consulted. These records were bound into a volume of the calendar for that year, and now form the series PROB 12. The indexes in PROB 12 use PROB 11 quire references, which mark out each quire of eight leaves. In the PROB 11 volumes themselves, a folio number appears on the recto of each leaf, and quire numbers appear on the first leaf of each quire. These individual volumes, which each contain a few dozen quires, then make up a register for that year which is assigned a name – either the first name in that register, or another ‘notable’ name that appears in it. For example, the register for 1778 is entitled ‘Hay’, and is made up of eleven volumes: PROB 11/1038–1048. The first volume, PROB 11/1038, contains Hay Quire Numbers 1–46, the final volume, PROB 11/1048, contains Hay Quire Numbers 474–519. PROB 11/1049, the next volume, is the first volume of the register for 1779, ‘Warburton’, and contains Warburton Quire Numbers 1–47. The introduction to the PCC will registers on the TNA’s website informs researchers: ‘It is not possible to tell from the catalogue reference alone which particular register an individual volume belongs to.’<sup>42</sup>

Figure 4 gives an indication of the size and materiality of the PROB 11 volumes. Each volume weighs over 10 kg, and measurements taken of a sample volume give a spine depth of 14 cm, a width of 36 cm and a height of 46 cm.

<sup>42</sup> <https://discovery.nationalarchives.gov.uk/details/r/C12122> (accessed 22 July 2024).

This example is an eighteenth-century volume, PROB 11/1040, one of the 11 volumes to contain the registered copies of wills proved in 1778. As discussed above, PROB 11/1040 can also be referred to in terms of its quire numbers, in this case 'Hay, Quire Numbers 93–140'. In this photograph, the volume is open on the will of Thomas Arne, proved 16 March 1778, which begins on folio 11. But the catalogue reference of this will is PROB 11/1040/181. As of 2013, the reference numbers of the wills were renumbered in chronological order. Accordingly, the will of Benjamin Jennings, proved on 5 February 1778, appears on folio 186 of PROB 11/1040, but carries the reference PROB 11/1040/1. The wills that have been assigned the numbers between PROB 11/1040/1 and PROB 11/1040/181 appear in a different order in the physical volume PROB 11/1040, but were all proved after Jennings's on 5 February, and before Arne's on 16 March. As with many archives with centuries of administrative history, differing indexing and numbering systems mean that the reference systems on the online catalogues do not represent the order of wills in the physical volumes themselves.

Alongside the physicality and arrangement of the PROB 11 volumes, we also need to consider the different circumstances in which they were microfilmed, and the reasons why. In the 1950s, while the wills were held by the Principal Probate Registry at Somerset House, PROB 11 was microfilmed by the Church of the Latter-Day Saints. These microfilm images contain no folio numbers, as the quire numbers were at this date the only form of internal marking in the volumes. When they were initially microfilmed, these images would only have been available at centres run by the Latter-Day Saints (now 'Family Search' centres). PROB 11 was later microfilmed again once it had been accessioned by the Public Record Office. At this point microfilming was carried out in order to preserve the manuscripts, rather than to make them widely accessible. The decision to initially microfilm PROB 11 (the registered copies of wills) rather than PROB 10 (the original wills) is one that was in itself rooted in the materiality of the manuscripts. The pages of PROB 11 are generally clean, unfolded and fairly uniform, appearing as they do in the standard volumes of bound parchment. The PROB 10 wills are more often folded and appear in bundles, rather than being arranged in bound volumes, making them more difficult to microfilm. The microfilms of PROB 11 were digitised and made available in stages on TNA's website c. 2001–4. In 2013, Ancestry.com also digitised the PROB 11 microfilm.<sup>43</sup> The decision to microfilm and subsequently digitise PROB 11 but not PROB 10 (and to digitise the microfilms, rather than the original manuscripts) continues to have ramifications for research conducted today.

Having sketched out the history of the microfilming and digitisation of PROB 11, it is necessary to reflect on the ways in which researchers access these different versions, and how this shapes the type of research that is possible, or indeed the type of research that users are directed towards. Many

<sup>43</sup> Many thanks to Ruth Selman, Early Modern Records Specialist at TNA, for providing a scan of the PROB 11 Introductory Note and for providing further information about its recent archival history.

researchers access the digitised microfilms of PROB 11 through the TNA's 'Discovery' catalogue. They are directed to search for an individual's will by name. Downloading a pdf image of a single will costs, at the time of writing, £3.50 per will. Alternatively, users can create a free account and download up to 100 items (e.g. individual wills) in a 30-day period. While the requirements of many researchers would fall within these restrictions, this still limits large-scale studies. By clicking on the volume reference that the individual will appears in, it is possible to see the references of other wills in the same volume, but it is not possible to browse freely through the digitised scan of an entire volume. Indeed, as has already been established, the numbering system in the catalogue arranges the wills in each volume in chronological order and does not correspond with the order of the wills as they appear in the physical volume. Through this interface individual wills need to be identified, selected and downloaded. Ancestry is also geared around viewing individual wills which are searched for by name. It does however provide the option to browse through whole volumes of the PROB 11 wills. A subscription that allows access to this function costs, at the time of writing, a minimum of £13.99 a month. These records are also available on TheGenealogist.co.uk. At the time of writing, a subscription to The Genealogist which provides access to 'Wills, Probates, and Testaments' costs between £98.95 and £139.95 per year. Both Ancestry and The Genealogist have produced their own indexes to the digitised PROB 11 records. As Jerome de Groot reminds us, sites such as Ancestry are ultimately 'profit-based' businesses; this has a significant bearing on how information is presented and how its websites influence the 'historical imaginary'.<sup>44</sup>

Access to the digital surrogates of PROB 11 is therefore divided and mediated in an uneven manner. PROB 11 is accessible only behind a paywall, or it requires registering for an account and limitations on the number of individual wills that can be downloaded per month. The way in which these interfaces permit the researcher to access these digital surrogates is also uneven: the access granted does not always provide the option to browse through the whole volumes. These interfaces are predicated on the assumption that PROB 11 would be used for certain purposes, primarily genealogical research, or studies which involve searching for named individuals. In the TNA's search interface for PROB 11, users can search by 'First name', 'Last name', 'Occupation', 'Place', 'Date range', or 'Other keywords' – these fields are the metadata that appear in the title of each will. Ancestry's user interface is similar and is orientated around the date and location of an individual's death or another event in their life. The interfaces and presentation of the digital surrogates therefore tend to hamper the type of research that necessitates browsing through volumes of wills or sampling at scale (e.g. studies of long-term patterns of bequests or economic change), that is interested in questions of wills as a source (e.g. the changing nature of preambles, religious language, or notarial practice over time) or ways of accessing the volumes without searching for an individual's name. As Richard Dunley and Jo Pugh have

<sup>44</sup> Jerome de Groot, 'Ancestry.com and the Evolving Nature of Historical Information Companies', *The Public Historian*, 42 (2020), 10, 26.

shown, archive catalogues such as TNA Discovery have been increasingly geared towards enabling genealogical research in recent years. Resources have accordingly been directed towards expanding item-level description that enables searching for named individuals.<sup>45</sup> At the same time, the interface has not developed to permit forms of research that look beyond the individual wills of named people, and which would be facilitated by making entire volumes browsable or downloadable.

As noted above, more than ten years ago Tim Hitchcock raised the problems of working from digital surrogates that were 'inaccurate representations of text, hidden behind a poor quality image'.<sup>46</sup> The digital surrogates of PROB 11 that appear on TNA, Ancestry and The Genealogist are digitised versions of the microfilm, and not of the original wills themselves. The versions that researchers have access to are therefore already several steps removed from the original manuscript. Katie Lanning, in a discussion of the digitisation of the Burney Newspaper Collection, which was scanned not from the original papers or the unused master microfilm, but from a used microfilm, asks 'should a popular archive prioritise preservation or access?'<sup>47</sup> For a collection like PROB 11 that contains hundreds of thousands of images, there is the possibility that each stage of microfilming or digitisation introduces human error, such as pages that are overlooked, mislabelled, or that appear in the wrong order. L. W. C. Van Lit cautions us on the use of 'low resolution digital scans of microfilms', which are 'surrogates of surrogates', and notes that a reliance on bad digital images can result in 'bad analyses'.<sup>48</sup> Microfilm copies are often of poor quality and their greyscale or black and white rendering fails to capture some details of the original manuscript, ensuring that low-contrast regions are harder to read. For these reasons, those who frequently transcribe PROB 11 wills have pointed to the difficulty of reading marginalia and interlined text on the digitised microfilms.<sup>49</sup> There are also examples of microfilmed PROB 11 wills which are partially unreadable due to damp or damage to the page, and it is not possible either to read the text or to determine the underlying reason for its unreadability until the original is viewed. 'Bleed through' text from the other side of the parchment also occasionally renders a microfilm unreadable. The digital surrogates would be more readable had they been produced using more recent, high-quality colour digital photography techniques. Text is also frequently obscured by folds or creases on the page which cannot be rectified other than by viewing the original volume. Even without better-quality scans, however, if the record came with metadata on the type and quality of the scan, it would render them easier to use as it would provide some explanation for the problems users experience while

<sup>45</sup> Richard Dunley and Jo Pugh, 'Do Archive Catalogues Make History? Exploring Interactions between Historians and Archives', *Twentieth Century British History*, 32 (2021), 591.

<sup>46</sup> Hitchcock, 'Confronting the Digital', 14.

<sup>47</sup> Katie Lanning, 'Scanner Darkly: Unpopularization in the Burney Newspaper Collection' *Archives and Records*, 41 (2020), 222.

<sup>48</sup> L. W. C. Van Lit, *Among Digitized Manuscripts. Philology, Codicology, Paleography in a Digital World* (Leiden, 2019), 69–70.

<sup>49</sup> Many thanks to Teresa Goatham for this point.

viewing them and give guidance on where future digitisation efforts should be focused.

There are also factors which render the use of PROB 11 (the registered wills, copied out by a Church court clerk in a uniform legal hand) distinct from PROB 10 (the original wills, written in the hand of a scrivener or perhaps the testator themselves). The introductory note to PROB 11 claims ‘there is usually no advantage to be gained from examining the original wills’, and indeed the TNA website suggests that ‘for most research purposes the registers are easier to use’.<sup>50</sup> This of course precludes a range of avenues of future research, which could for example analyse the original handwriting (as opposed to that of the Church court clerk), which could examine the signatures or marks of witnesses, or which could compare originals with their copies, to cross-check for edits or omissions.

A brief comparison between a sample of original wills in PROB 10 and their counterparts in PROB 11 reveals what can be lost in a focus on the registered copies only. The copies made by Church court clerks were written in one hand (the clerk’s), in a uniform secretary or legal hand, without retaining the original formatting and often without retaining the original spelling. This means we can no longer access the distinctions between the hand of the scrivener, or another tasked with writing up a will, and the varied signatures or ‘marks’ made by the testator or witnesses. This is exemplified by the comparison between the signatures in the original and registered copies of the will of Edward Rott which opened this article.

This also means that we cannot always identify from the registered copies alone wills like that of Benjamin Rogers, who ostensibly wrote out his own will. The writing in this will is hurried and distinct from the formal, considered, style of the scrivener and Rogers’s own signature appears to match the hand that wrote the main body of the text. We might attribute this to the rushed and dangerous circumstances in which Rogers found himself – he died in London at the height of the plague, making a will on 30 August 1665 that would be proved three weeks later.<sup>51</sup> The registered copies cannot provide an insight into either the presence of multiple hands or one: all nuance is subsumed by the uniformity of the clerk’s hand. Other original wills written during the plague of 1665 provide material insights into the health of the testator.<sup>52</sup> Aspects of this can of course be inferred from information that appears in the registered copies: the short length of time between when a will was written and when it was proved, or the presence of phrases such as ‘weake of Bodye’ or ‘beinge sick in her bed of the sickness whereof she dyed’.<sup>53</sup> But occasionally such phrases are coupled with the testator’s markedly

<sup>50</sup> Introductory note to PROB 11, with thanks to Ruth Selman; <https://discovery.nationalarchives.gov.uk/details/r/C12121> (accessed 10 July 2024).

<sup>51</sup> Will of Benjamin Rogers, made 30 Aug. 1665, proved 26 Sept. 1665, TNA, PROB 10/980.

<sup>52</sup> Many thanks to Judy Lester of Kerrywood Research for sharing her experience of finding ‘shaky’ handwriting in wills.

<sup>53</sup> Will of William Newarke or Newark, Factor of Saint Michael Bassishaw, City of London, 30 Aug. 1665, TNA, PROB 11/317/382; Will of Jane Rokeby, Widow of Saint Giles without Cripplegate, Middlesex, 8 Sept. 1665, TNA, PROB 11/317/460.

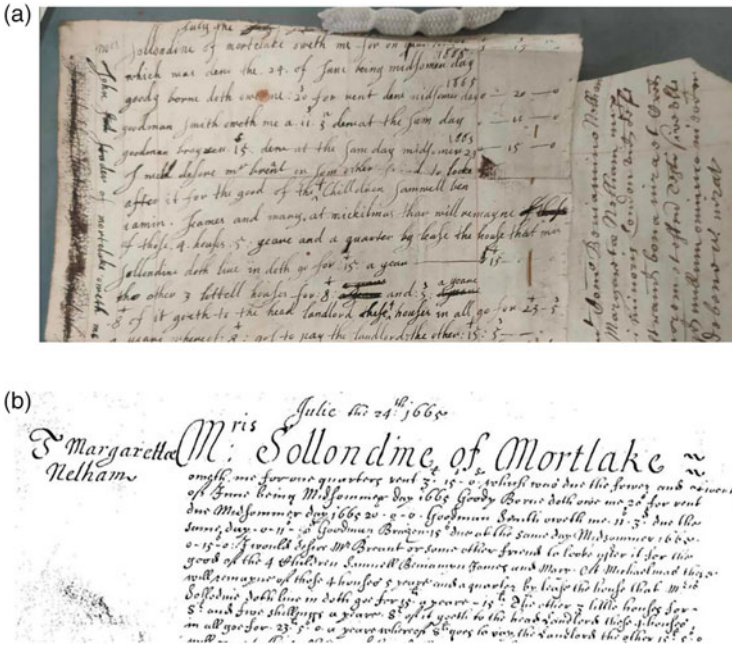
shaky or weak signature, such as the unusually short will of Thomas Roe, who was described as ‘being very weake in body’. Roe managed to scrawl his shaky mark when his will was written on 18 September: he died almost immediately, as it was proved two days later.<sup>54</sup> We do not know, of course, precisely how indicative shaky or weak signatures are of the testator’s health – they could be comparable with their usual writing. But nonetheless, that avenue of research is precluded if only the registered copies are consulted. There are many reasons why researchers would want to see the original signatures and handwriting of those involved in a will’s production, including for studies of literacy, to determine whether a testator or witness had written the rest of a will, or to infer information about the health of the writer.<sup>55</sup> The digitisation of the registered copies, but not the originals, pushes users towards historical research based around people, dates, places and the content of wills, but at the same time ensures researchers are less likely to pursue other questions that wills as a source provoke.

There are other, more material facets that we lose in focusing on the registered copies in PROB 11 rather than the originals in PROB 10. Original wills were written in varying formats: some on expansive pieces of parchment, some on smaller scraps of paper or indeed paper repurposed from other sources, including account books.<sup>56</sup> Some wills amount to a couple of lines that fill half a side, while others stretch to ten, twenty or more pages, with the testator’s signature and seal in the bottom right-hand corner of each page. The variety in the format of the original wills, and the fact that they have been tightly folded up in bundles for storage, is of course one reason why microfilming PROB 10 was too complex a task. In some cases, a later codicil on a separate scrap of paper has been stitched or stapled onto the original will. The codicil was often written by a different hand and signed by different witnesses. These varieties of format are lost in the uniform work of the Church court clerk, who in copying out the content of the will into the registered copy books, could not replicate features such as stapled addendums, or the scale or materiality of the paper or parchment used. Where the original wills feature seals in black and red wax, either stamped directly onto the page of the will or attached onto a fold-out of paper, the registered copies can only mark the place of the seal. This practice too is inconsistent, but eighteenth-century wills often represent the seal with ‘L.S’ or ‘Locus Sigilli’. Yet other features of the will, such as the formatting of the text, could have been replicated in the registered copies, but are generally not. We have already seen an example where signatures are incorporated into the main text of the will, rather than retaining their original list formatting. In the unusual will of Margaret Nelham, probably written by Nelham herself onto a repurposed account-book page, money owed to her is listed in a column of the right-hand side of the

<sup>54</sup> Will of Thomas Roe, made 18 Sept. 1665, proved 20 Sept. 1665, TNA, PROB 10/980.

<sup>55</sup> Mark Hailwood, ‘Rethinking Literacy in Rural England, 1550–1700’, *Past & Present*, 260 (2023), 38–70.

<sup>56</sup> One example of this is the will of Margareta Nelham, TNA PROB 10/979. Our thanks to Laura Sangha for pointing out that the will appears to have been written on account book paper.



**Figure 5.** Original will of Margarett Nelham (above, PROB 10/979) and its registered copy (below, PROB 11–317–321). (Sources: TNA, PROB 10/979; PROB 11/317/321, Photo © Emily Vine.)

page, arranged in the format of pounds-shillings-pence. In the registered copy version, Nelham’s itemised list is not retained: instead these amounts appear as run-on sentences, removing this insight into her numeracy. Figure 5 shows both the original copy of Nelham’s will (PROB 10/979), with an addendum and probate clause stitched onto the right hand-side of the page, and its registered copy (PROB 11–317–321). The registered copy of Nelham’s will, alongside other examples, also demonstrates that church court clerks had an inconsistent approach to retaining original spelling or capitalisation. Here ‘mortelake’ has been corrected to ‘Mortlake’, ‘gody borne’ has been corrected to ‘Goody Borne’, ‘Children’ to ‘Children’. At the end of the will, the number ‘4’, written numerically in the original, is rendered in the copy as ‘fower’. There is also no replication of the writer’s (possibly Nelham’s) own deletions, scribbles and additions. Very little of the character of the original writings, occasionally produced by the testator themselves, is retained in the registered copies. More concerning still is the possibility that this correction or failure to retain original formatting or spelling has introduced errors or omissions, or has rendered the intentions of the testator inaccurately.

There were practical reasons why PROB 11 and not PROB 10 was originally microfilmed and digitised, and for many research purposes pertaining to names, places, dates and the general content of the wills, the registered copies are sufficient. In other words, the registered copies are useful particularly for

forms of genealogical research, and for mass processing of their substantially more regular handwriting, which makes them particularly good candidates for automatic transcription using handwritten text recognition algorithms. But they are not sufficient for other forms of historical research that are predicated on the retention of original spelling, formatting or handwriting, or that are interested in questions of materiality. Digitised collections are of course over-represented in research in comparison to those which are harder to access. The microfilming and subsequent digitisation of PROB 11 has made a valuable archive of English and Welsh history widely accessible, but the decision to digitise it instead of PROB 10, accessible only through a visit to TNA in London itself, has had a direct influence on the forms of research that have been and continue to be produced.

### *The materiality of PROB 11 and its digital surrogates*

There are parallels in how PROB 11 replicates the content but not the materiality of PROB 10, and in how the digital surrogates of PROB 11 replicate the content but not the materiality of the volumes themselves. It is of course not only the readability of the text that can be compromised by digitisation, but also understandings of a manuscript's materiality and archival context. When isolating individual wills from their materiality (bound alongside quires of other wills in leather volumes and arranged into named registers), we isolate them from the complexity and scale of the archive and its history. The materiality of books and manuscripts has long been a point of discussion within wider questions of digitisation and accessibility. As Johanna Green has discussed, manuscript digitisation focuses on 'the page and text, rather than the 3-D codicological object' and is 'not a process of replication but transformation'.<sup>57</sup> Digitisation captures the largely two-dimensional visual features of a book or manuscript, yet even then important visual features can be lost or diminished. The vibrant colours of illustrated manuscripts, or the fainter marks of some marginalia, are not always well replicated in digitisation.<sup>58</sup> Equally, it is often necessary for scholars to get a sense of the size, materiality and weight of a book in order to understand how users would have interacted with it: could it be carried around, placed in a pocket or easily concealed? Some medievalists and codicologists have argued for the importance of being able to view, touch and hold a physical book or manuscript in order to understand fully its usage and reception. Taking this a step further, Ryan Szpiech argued in 2014: 'The manuscript cannot only be seen – it must be touched, smelled, read, received, interpreted in order to be appreciated and understood.'<sup>59</sup> This positioning has drawn more criticism in recent years,

<sup>57</sup> Johanna M.E. Green, 'Digital Manuscripts as Sites of Touch: Using Social Media for "Hands-On" Engagement with Medieval Manuscript Materiality', *Archive Journal*, 6 (2018), <https://www.archivejournal.net/essays/digital-manuscripts-as-sites-of-touch-using-social-media-for-hands-on-engagement-with-medieval-manuscript-materiality/> (accessed 10 July 2024).

<sup>58</sup> Van Lit, *Among Digitized Manuscripts*, 67.

<sup>59</sup> Ryan Szpiech, 'Cracking the Code: Reflections on Manuscripts in the Age of Digital Books', *Digital Philology: A Journal of Medieval Cultures*, 3 (2014), 90.



including from L. W. C. van Lit who questions how essential haptic or multi-sensual interactions with original books and manuscripts are for understanding them.<sup>60</sup> Equally both Green and Aengus Ward have shown that only the privileged few ever have the chance to touch or hold original books and manuscripts anyway: the general public can only ever interact with them from a distance, or behind glass. In Ward's words, 'if sensory access to the unique object is required for materiality truly to be appreciated ... it cannot be available to those scholars who are unable to access the artifact in person. In this counsel of despair, the previous hierarchy of privilege remains.'<sup>61</sup> This conclusion is drawn from quite a different context to PROB 11, a series that has been continuously consulted by members of the public since the wills were first registered: firstly at Doctors' Commons and subsequently at Somerset House. Yet since PROB 11 was digitised, it has not been possible for members of the public to order up the physical volumes.

This article has already considered the loss of material context that has resulted from the decision to microfilm and digitise PROB 11 rather than PROB 10. It has shown that the registered copy volumes replicate the content of the original wills in a standardised way, without reproducing original formatting, handwriting or materiality. For the purposes of this discussion of the registered copies in PROB 11, the concern is less with the inability to see, touch or smell the physical volumes, but with how digitisation removes the individual wills from the materiality of the volumes and the context of the archive. Unlike, for example, a small pocket bible, very few people would have actually held these volumes at the time in which they were produced, and their size or weight does not have a direct bearing on how they would have been used and consulted. But there are still important material aspects that are lost or complicated by digitisation, and which are worthy of consideration. In losing sense of the physicality of the large, leather-bound parchment volumes, we lose a sense of the scale and physicality of the archive, its expanse and its administrative history. While the digitised version on Ancestry provides the option of viewing pages of some of the volumes as an open book or double-paged spread, they elsewhere appear as isolated pages, cut off from the context of the will preceding or facing (even though wills generally end, and begin, halfway through a page). L. W. C. van Lit has pointed out that the 'cut' of a digitised image – the decision of what is included within the frame of the image and what is excluded, can have an important bearing on how the images are used and understood.<sup>62</sup> When downloading the will of a named individual on TNA Discovery, the final paragraphs of the preceding will often appear on the first page. These paragraphs often appear out of context and do not always contain the name of the preceding individual.

<sup>60</sup> Elaine Treharne, 'Fleshing out the Text: The Transcendent Manuscript in the Digital Age', *Postmedieval: A Journal of Medieval Cultural Studies*, 4 (2013), 274; Van Lit, *Among Digitized Manuscripts*, 61–2.

<sup>61</sup> Aengus Ward, 'Of Digital Surrogates and Immaterial Objects: The (Digital) Future of the Iberian Manuscript in Textual Editing', *Journal of Medieval Iberian Studies*, 14 (2022), 45.

<sup>62</sup> Van Lit, *Among Digitized Manuscripts*, 68.

Inconsistent numbering in the catalogue (where wills are categorised chronologically and not in the order they appear in the physical volume) means it is not always easy to locate or identify a preceding will. One of the key complications of the digitisation of PROB 11 is therefore the isolation of individual wills from their archival context (wills that have often been ‘pulled up’ through keyword-searching), and the ‘flattening’ of the archive more generally. Katie Lanning, in a study based around the Burney Newspaper Collection, has warned of how microfilming and digitisation decontextualises texts and shifts the ‘shape’ of the archive. Lanning notes that the British Library had continued to add additional newspapers to the original Burney Collection and that it is now no longer possible to determine where Charles Burney’s original collection begins and ends.<sup>63</sup> This has some implications for the way in which the digital surrogates of PROB 11 are accessed. On the TNA website, there is an interface where researchers are directed to search within the catalogue for PROB 11. Yet they can also search for wills, or names of testators, on TNA Discovery, the catalogue which comprises not only the TNA’s entire holdings but collections from but other British archives too. There is potential for researchers to be confused about what is and is not PROB 11, or indeed for the boundaries of PROB 11 to be blurred and discarded: for its shape and content to be subsumed within a broader meta-structure of British archives. On Ancestry this is perhaps even more acute: researchers can ‘call up’ digitised wills from several collections using the same search interface. Entering a name can pull up results from across a number of databases, not only ‘England & Wales, Prerogative Court of Canterbury Wills, 1384–1858’, but also ‘UK, Extracted Probate Records, 1269–1975’, ‘Irish Records Index, 1500–1920’ and ‘American Wills and Administrations’.<sup>64</sup> Furthermore, it is not clear, as we saw in the case of Johes/John Vaughn/Vaughan, where the names you are searching came from, or if your search has found every instance or variation. In being presented with isolated results drawn from different source types, databases and archives, aspects of the archival structure and context are lost.

Tim Hitchcock, among others, has warned of the complications of keyword searching of a database, which produces results that are isolated or ‘deracinated’ from their archival context. Keyword searching, and the digitisation of seemingly whole archives or series, can also give a false impression of completeness or comprehensiveness. There is a danger that the digitisation of PROB 11, rendered on archival and genealogical websites as ‘England & Wales, Prerogative Court of Canterbury Wills, 1384–1858’, gives the impression of a complete archive of wills throughout the time period 1384–1858, when the probate process was disrupted for example during the English Civil Wars and Interregnum (there is no extant register for wills proved only at Oxford).<sup>65</sup>

<sup>63</sup> Lanning, ‘Scanner Darkly’, 218–19.

<sup>64</sup> [https://www.ancestry.co.uk/search/categories/clp\\_wills/](https://www.ancestry.co.uk/search/categories/clp_wills/) (accessed 10 July 2024).

<sup>65</sup> Introductory note to PROB 11, with thanks to Ruth Selman; Van Lit, *Among Digitized Manuscripts*, 54.

Knowledge of the archival context, and the historical context in which documents were produced, remains a prerequisite for using and understanding digital surrogates and searchable databases. Hitchcock cautions that keyword searching 'lets us escape this post-Enlightenment knowledge system, but it also removes the framework of source criticism and classification that we have come to rely upon'. He uses this to argue for more honesty when citing digital sources and for researchers to acknowledge when they have viewed the digital surrogate rather than the original book or manuscript. Researchers 'must be even more honest than is required by the form of a traditional footnote, about how we are searching evidence, and what it is we are searching'.<sup>66</sup> Van Lit has echoed this, arguing 'Being honest about this means that we should refer to the surrogate in our bibliography and we should include a description of the digital materiality of the photos'.<sup>67</sup> The wills in PROB 11 have been widely cited by a broad range of historians or genealogical researchers, the majority of whom will have accessed these records via the digital surrogates only and will not have acknowledged this. For understandable reasons of preservation, TNA discourages researchers from viewing original manuscripts where a surrogate exists 'in digital, microfilm or microfiche formats'. It provides the following exemptions: 'the surrogate is illegible or obscured' or 'viewing the original record provides information not available from the surrogate'.<sup>68</sup> At the time of writing it is not possible to order PROB 11 volumes through the usual document-ordering channels.

Historians in the last two decades have had to become more adept at navigating the complexities of the digital landscape. Jon Coburn's study of historians' digital practices suggests that, despite Tim Hitchcock's earlier warnings, many are cognisant of the limitations of archival databases and are mindful that the biases of keyword searching and algorithms can influence research in unintended ways. Those whom Coburn interviewed suggested that accessing digital surrogates should not be a substitute for viewing the original manuscript and indeed that sometimes a hybrid approach would be adopted: viewing some of a collection online could help a researcher decide whether to make a trip to the archive itself.<sup>69</sup> One of the key issues raised by Coburn's respondents about digital surrogates and keyword searching is that they remove the chance for 'serendipitous finds' – useful discoveries that are stumbled across only when one is browsing through a box or volume in an archive, or flicking through the papers adjacent to the ones that had actually been identified and requested through the catalogue. These are findings that cannot be captured by keyword searching and this in turn has an indirect effect upon the type of historical research that is produced. There is growing

<sup>66</sup> Hitchcock, 'Confronting the Digital', 14.

<sup>67</sup> Van Lit, *Among Digitized Manuscripts*, 71.

<sup>68</sup> <https://www.nationalarchives.gov.uk/about/visit-us/researching-here/how-to-order-view-and-copy-documents/> (accessed 10 July 2024).

<sup>69</sup> Jon Coburn, 'Defending the Digital: Awareness of Digital Selectivity in Historical Research Practice', *Journal of Librarianship and Information Science*, 53 (2021), 398–410.

awareness within the historical discipline, and beyond it, of the limitations of consulting digital surrogates in isolation, but also of the benefits of using digital surrogates alongside viewing original manuscripts in person. The microfilming and digitisation of PROB 11 has made these sources accessible to a global community of researchers and has had a profound impact upon the forms of historical and genealogical research that can be produced. But in using the digital surrogates, researchers have a duty to consider the loss of archival context and materiality and how the purely pragmatic decision to microfilm the registered copies, and not the originals, has shaped and will continue to shape the forms of research that can be carried out. So too do researchers need to be mindful of how they are directed to search these digital surrogates: through interfaces that prioritise genealogical research, keyword searching for individual people and metadata that captures names, dates and places, but not other aspects of wills as a source. The PCC wills are an example of a fairly well-documented digital collection. The potential pitfalls we have identified in their case are all substantially more severe in the wider landscape of digital historical sources where, as we have seen, information on methods of digitisation and cataloguing are often worse or entirely absent.

### Conclusions: towards the future of digitisation and archival materiality

Faced with these challenges, how can scholars and institutions reconcile digitisation efforts with the recognition of the materiality of their holdings? TNA itself has grappled with how best to digitally image and represent the materiality of its wax seal mould collection. After experimenting with different forms of colour photography and 3D scanning, it opted for greyscale scanning using a flat-bed scanner, alongside recording detailed metadata on colour and size.<sup>70</sup> Aengus Ward has pointed to how data encoding could be used ‘to represent or recreate the physical dimension of manuscripts’.<sup>71</sup> Bill Endres’s work has produced interactive 3D renderings of pages from the St Chad Gospels, allowing users to ‘rotate’ the book, and moving beyond the representation of pages as flat surfaces.<sup>72</sup> Other digital approaches to materiality have not only replicated the three-dimensional aspects of manuscript sources, but have been used to conduct further research on them. The ‘Letterlocking’ project has not only created three-dimensional scans of sealed seventeenth- and eighteenth-century letters, but has used X-ray

<sup>70</sup> Amy Sampson, ‘TNA Wax Seal Moulds – from Drawer to Discovery’, 18 Mar. 2019, <https://blog.nationalarchives.gov.uk/wax-seal-moulds-drawer-discovery/> (accessed 10 July 2024).

<sup>71</sup> Ward, ‘Of Digital Surrogates and Immaterial Objects’, 43.

<sup>72</sup> Bill Endres, *Digitizing Medieval Manuscripts: The St. Chad Gospels, Materiality, Recoveries, and Representation in 2D and 3D* (Leeds, 2019); and Bill Endres, ‘More than Meets the Eye: Going 3D with an Early Medieval Manuscript’, in *Proceedings of the Digital Humanities Congress 2012*, ed. Clare Mills, Michael Pidd and Esther Ward (Sheffield, 2012), available online at <https://www.dhi.ac.uk/books/dhc2012> (accessed 10 July 2024).

microtomography to ‘virtually unfold’ and read their contents for the first time (without opening the letters or breaking their seals).<sup>73</sup>

Limitations of cost and other practicalities mean it is unlikely that PROB 11, or indeed PROB 10, would be re-digitised in light of developments in 3D digital technology. It is also unlikely that the metadata of either series will be updated to account for material features. But it is possible for future digitisation projects to comprise documentation and metadata that comprehensively accounts for material features, even when those sources, like the registered will copy volumes of PROB 11, have traditionally been used for textual rather than material analysis. We cannot predict the ways in which future researchers will approach sources such as these. While most researchers have little influence over future digitisation projects, there are other modes of best practice that we can all follow. We can follow the call of Hitchcock, Van Lit and others that we be honest in our citation practices and honest when we have viewed a digital surrogate rather than an original manuscript. This is true even of sources drawn from PROB 11, and in cases where the interest is primarily in a will’s content, such as for genealogists, or social or economic historians, digital surrogates are entirely satisfactory. We also need to be more consistent in reporting methodologies, how material was identified, what biases or lacunae such methodologies may have introduced into the work and how we have sought to overcome them. And throughout our use of digital surrogates, we need to be ever mindful of their archival context, the reasons why these manuscripts were microfilmed or digitised (and other manuscripts overlooked) and the ways in which archival websites and other interfaces direct us towards certain forms of research and away from others.

Digital resources are of profound importance to the way that history is written today and that importance is likely only to increase in the future. Much of this article has been concerned with the limitations and problems arising from the way such sources are produced and accessed, yet we do not mean to give the impression that we are dubious of the value of digitisation, not least in terms of allowing access to material that for reasons of distance, cost or format is difficult to engage with and with regard to allowing other scholars to check others’ published accounts directly. Indeed, it would be foolish of us, given the nature of our careers, to make such an argument. However, we have seen in this article the difficulties faced by scholars using these digitised resources: subscription costs, library or university access, poor or uncertain transcription practices, historical artefacts shaping the format of the resulting digital material, dead links, substandard or entirely absent documentation and metadata. All of these issues reflect the particular nature of the data/digital infrastructure in the United Kingdom, and, just as individual scholars need to adopt and push for better practices in using and citing such material, the historical profession

<sup>73</sup> <https://letterlocking.org/> (accessed 10 July 2024); J. Dambrogio, A. Ghassaei, D. S. Smith *et al.*, ‘Unlocking History through Automated Virtual Unfolding of Sealed Documents imaged by X-ray Microtomography’, *Nature Communications*, 12 (2021), <https://doi.org/10.1038/s41467-021-21326-w>.

as a whole needs to push for improvements in the infrastructure which will ensure that future digitisation projects produce material that is easier to use with confidence and that can be employed for a wider range of historical enquiries.<sup>74</sup>

**Acknowledgements.** The authors would like to thank Mark Bell, Tim Hitchcock, Jan Michielsen, Laura Sangha, Ruth Selman, Jane Whittle and the readers at *TRHS* for their constructive feedback on this article.

**Financial support.** This research was conducted as part of Leverhulme Trust Research Project Grant, RPG-2023-07, 2023-27.

---

<sup>74</sup> The sorry state of data infrastructure in the United Kingdom is surveyed in Ahnert *et al.*, *Collaborative Historical Research*, 23–32. The country’s weakness in this area is epitomised by the much-lamented discontinuation of the Historical Texts service in July 2024, which brought together several important digital resources: EBO, ECCO, UK Historical Medical Library, British Library Nineteenth Century Collection; the impending retirement of this service makes it hard to cite, but see (hopefully) <https://web.archive.org/web/20240513075051/https://historicaltexts.jisc.ac.uk/news#2024-02-29> (accessed 5 June 2024).

**Cite this article:** Smith H, Vine E (2024). Material and Digital Archives: The Case of Wills. *Transactions of the Royal Historical Society* 2, 441–470. <https://doi.org/10.1017/S0080440124000094>