



Digital archives and recombinant historical geographies

Progress in Human Geography
2022, Vol. 46(6) 1298–1310
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/03091325221103603
journals.sagepub.com/home/phg



Jake Hodder 

School of Geography, University of Nottingham, Nottingham, UK

David Beckingham

School of Geography, University of Nottingham, Nottingham, UK

Abstract

This article considers how digitisation is reshaping archival research in geography. Digitisation is more than a technical convenience, something that simply speeds up existing ways of working. Through novel practices of recombination, digital archive platforms enable researchers to extract and recombine fragments of historical information, drawn across multiple periods, places, collections and contexts. This represents a fundamental change in how we research the past. In this paper, we conceptualise recombination as an uneven geographical phenomenon, we situate it within the shifting political and economic infrastructures of archives, and pose a series of ethical questions for geographers to consider.

Keywords

digital archives, ethics, historical geography, platforms, recombination

‘From enabling researchers to mine vast amounts of data ... to facilitating the most laser-focused search and discovery on its award-winning platforms, ProQuest is committed to research excellence’ (ProQuest, n.d.a).

1 Introduction

Historical research in geography is changing. Over the past thirty years, digital technologies have significantly altered the way that historical sources are stored and retrieved. The term ‘digital archives’ captures both this move of sources online, as well as the wider ecosystem of digital tools and platforms that scholars use to find and analyse archive materials (see [Owens and Padilla, 2021](#)). The ability, within seconds, to locate exact words or phrases across inconceivably vast databases of text is now a routine part of our daily practice.

The creation of digital archives raises distinct issues of how best to preserve and disseminate historical material, how to manage notoriously unstable file formats, and how to approach intellectual property rights. These are important questions, but they are also predominantly technical ones. It is easy to mistake digital archives as tools of convenience: that they simply speed up the kind of work *we were already doing* ([Ramsay, 2010](#)). This may explain the lack of formal, methodological reflection from historical geographers ([Bressey, 2020](#)). But, as the quote from ProQuest above intimates, digitisation is more

Corresponding author:

Jake Hodder, School of Geography, University of Nottingham, Sir Clive Granger Building, Nottingham NG7 2RD, UK.

Email: jake.hodder@nottingham.ac.uk

than a technical convenience. It brings with it ‘new ways of reading, viewing, and structuring archives, new forms of value and their extraction, and new infrastructures of control’ (Thylstrup, 2018: 4).

In this paper, we argue that digital archives not only bring new ways of doing historical geography; they engender a deep and philosophical shift in our relationship to the past itself. Following the work of the cultural theorist Steve Anderson, we argue that digital archives are giving rise to new recombinant historical geographies. Traditional research – once highly ordered by place and archival arrangement – is being supplanted by the rise of digital, text-searchable databases that offer a world of ‘infinitely retrievable fragments’ (Anderson, 2014: 101). These fragments, alongside the platforms that host them, and the algorithms we use to find them, are reorienting our research around practices of *recombination*. Digital technologies enable the researcher to extract, recombine, and even reformat, snippets of historical information, drawn across multiple periods, places, collections and contexts.

This is not the same work *we were already doing*. Geographers have written extensively of the ways in which the location, arrangement and display of archives structure our engagement with them (e.g. Craggs, 2008, 2016; Ferretti, 2020; Gagen et al., 2007; Hammond, 2020; Hodder et al., 2021). The archive is both a collection of materials and a form of ordering and control rooted in the entanglement of document and place (Mbembe, 2002). Digitisation disrupts these relations. As Lara Putnam (2016a: 379) argues, ‘increasing reach and speed by multiple orders of magnitude is transformative. It makes new realms of connection visible, new kinds of questions answerable’. But as Putnam goes on to show, it also creates new blind spots and forms of erasure that demand our critical attention. Digital archives dissolve the former place-based economy of historical research which incentivised in-depth, *in situ* inquiry, whilst making scholarship that drew across multiple archives and locations prohibitively expensive (Putnam, 2016a, 2016b).

The geographical promise of digitisation is to unbind archives from place, giving us the ability to view sources from anywhere. But digital space is not synonymous with anywhere. Instead, a different set

of geographical configurations emerge over what gets digitised, where it is viewed, and who has access. Online, archives often exist in a very particular space: that of the proprietary research platform. These platforms are presented to us as benign digital replications of the physical world in a way that profoundly underplays their determinative role in our research. Although they appear to overcome the hierarchy of the archive, in practice they supplant one form of ordering for another. In so doing, digitisation makes the historical and geographical claims of archives harder to see, even as it makes information more widely accessible.

By bringing work from history and archive studies into dialogue with geographical scholarship, this paper develops its argument in three parts. In the first part, we conceptualise recombination. We show how recombination differs from the principles of original order and *respect des fonds* that have traditionally governed archive use and arrangement. In the second part of the paper, we turn to the role played by platforms. We interrogate how the digitisation of archives has taken place in tandem with a process referred to as ‘platformisation’ (Poell et al., 2019), in which recombination and aggregation are central modes of value capture. In the final part, we explore the ethical implications of recombination. We pose that recombination be read as a geographical phenomenon. Remote access is changing from where and how we view archives, while uneven processes of digitisation prompt new questions of what and whose records get left behind.

II Conceptualising recombination

Until recently, virtually all historical work in geography was done with physical sources in a so-called ‘analogue’ archive. Since the mid-nineteenth century, archives have been organised by a basic principle: records which come from different creators or origins *should not be combined together*. Materials should not be reorganised, for example, based on their subject matter, date or geography as one might reorganise a library collection. Instead, a group of records should be maintained in the same organisational system as they were placed by the record’s creator – a principle known as ‘original order’.

The sanctity of this principle is enshrined in the concept of *respect des fonds*. *Respect des fonds*

emerged in the wake of the French Revolution to enable consistency as new records and those surviving from the *ancien régime* were brought together. Over the course of the later 19th and early 20th century, *respect des fonds* was revised and codified to become the most important principle of archival management (Bartlett, 1991; Duchein, 1983). Maintaining original order is so significant because it is what gives archives their evidentiary quality. By making creatorship the organising principle, archive documents offer insights far beyond the text-on-the-page. Where a record was made, by whom, and for what purpose is central to understanding the work that record did in the world and frames, through the arrangement of the archive, how we retrieve and analyse it (Craggs, 2016; Roche, 2021). This is something we instinctively know. Historians and geographers have written extensively of the challenge of working with colonial archives, for example, precisely because doing so risks reproducing the racial power structures that brought those records into existence (Clayton, 2021; Jazeel and Legg, 2019; Stoler, 2010).

Yet in the digital world, creatorship is no longer the primary access point or organising principle. Digital archives exist as databases of information and, as Jefferson Bailey (2013: np) argues, in 'a database, objects are related but not ordered' – certainly not in the ways of traditional archives. Bailey argues that 'nonlinear retrieval supplants the narrative logic of respect des fonds with a broader notion of context and discoverability' (Bailey, 2013: np). Navigation is not predetermined by a creator, an archivist, or a finding aid, but arrangement is dynamic and dependent upon our search terms. Materials are open to new connections and *recombination* (Bailey, 2013). Original order is not technically lost – metadata accompanying results often give details of provenance – but neither does it hold a monopoly on how we access and understand the archive (Zhang, 2012).

If you have used a digital archive – or platforms like Google Books or JSTOR – you will be familiar with the search box that invites us to enter our 'key terms'. It has the look and feel of a finding aid that we might use to identify a call number in a physical archive. But, as Ted Underwood (2014) argues, the underlying technology and philosophical principles are vastly different. The search bar does not help us

navigate the arrangement of the archive; it allows us to circumvent it. Digital platforms privilege searching over browsing, and that searching has more in common with data mining than document retrieval. The more precise the phrasing, the more efficiently digital search can personalise our results. This is part of the nature of computer search or 'information retrieval' – it is very effective at identifying exact terms and can do so across millions of data points in seconds.

When we receive our search results, fragments of historical information are recombined from multiple collections and places with little fidelity to original order or provenance. As Sassoon (cited in Sternfeld, 2011: 565) notes, digital archives return results as 'a databank of orphans which have been removed from their transactional origins and evidence of authorial intent'. Digital archives, then, offer us an unprecedented means to quickly find exactly *what we were already looking for*. But rarely do we know what we are looking for, even when we think we do. And even more rarely do we know the precise, historical wording that would find it. So, we search by trial and error. We enter a term as a proxy for a broader theme. We refine it. We search again. If we are lucky, we might be able to tie our research question to a distinct and historically stable watchword that reveals dozens of new sources.

The ability to extract and recombine historical information based on its proximity to our search terms brings a clear confirmatory bias. Digital search is not akin to a finding aid, but to an experiment – and, as Underwood (2014: 65) reminds us, 'there's something a bit dubious about experiments that get repeated until they produce a desired result'. And how representative are those results? How well can we understand them when they are decontextualised from their historical site of meaning? Results sorted by relevance filter out historical ideas that might contradict the assumptions underpinning our search terms. Take the example of newspapers. They constitute the largest bulk of digitised material, with multiple national and local titles and therefore thousands of potential data points for digital search to 'hit' (Gooding, 2016). Against that volume of material, platforms show the information you need and little more. Articles appear separate from information

that offered credibility and context to historical readers of the newspaper: mastheads, page layout, even other articles in the same issue. Searches also return multiple versions of syndicated stories, comparatively small in their own day but now given artificial prominence. This is the version of research ‘efficiency’ and ‘empowerment’ that digital platforms promise, and it is these to which we now turn.

III Platforms and recombination

Digital archives have emerged in the context of an increasingly competitive marketplace of digital platforms and software providers. These companies, in partnership with leading public and private institutions, offer high-quality research platforms alongside support for preserving and cataloguing archive collections. But we should be in no doubt, they are commercial enterprises. They include *Preservica*, which counts the EU, Associated Press, UK National Archives and 18 US State Archives as customers. It promises the ability to ‘dynamically rearrange and enrich your archive to stay relevant’ (*Preservica*, n.d.). Clarivate’s *ProQuest* claims to be used in over 26,000 libraries across 150 countries (*ProQuest*, n.d.b). And Cengage’s *Gale*, often regarded as the most scholar-orientated platform, offers 600 years of aggregated primary-source content combined with ‘advanced humanities computing tools’ (*Gale*, 2020).

Aggregation and recombination are central to the business model of these platforms and their mode of value capture. As the opening quote from *ProQuest* attests, in a crowded market companies are required to mark themselves out through either advanced, proprietary search and visualisation tools which allow for best-in-class research ‘efficiency’ (read: recombination). Or, by offering access to aggregations of the most extensive and unique archive materials. In that sense, the digitisation (and concurrent platformisation) of archives is driven by the same ‘future-facing processes of valuation and capitalisation’ that *Langley and Leyshon* (2017: 14) have identified in other areas of the digital economy. In terms of digital research tools, this is tellingly reflected in the semantic shift from ‘portals’ to ‘platforms’ (*Sherratt*, 2013), whereas a portal took you

somewhere else, a platform denotes a demarcated, (pay)walled digital space; a foundation on which to build immersive and proprietary tools (*Poell et al.*, 2019; *Srnicek*, 2017).

The term digital archive is a misnomer, then. Research platforms are rarely archives in any actual sense. Rather, they act as intermediaries that monetise the digital reproduction and exchange of materials whose original versions continue to exist in a pre-existing location, copyright status and arrangement. The platform is an on-screen interface that allows researchers to view archive materials but, as Lizzie *Richardson* (2020: 460) notes, it is also a ‘flexible spatial arrangement’. It reorganises archives through ‘novel technologies of coordination’ that rearrange sources already in existence elsewhere. Value (or investment return) is generated by creating increasingly advanced coordinating activities and elaborate recombinative effects. Disparate collections are brought together; search is enhanced to read further and deeper; sources are repeatedly recombined into new thematic collections based on evolving trends.

For commercial publishers, recombination is crucial to securing licensing agreements and justifying charges to access materials that are ultimately owned by someone else or are outside copyright. For example, *Gale* has spent years ‘developing and refining’ its ‘unique search technology’. As such, it offers, ‘more than simple text-searchable scans of original documents: our digitized archives ... [help researchers] take advantage of efficiencies in their research process that they do not get with other digitized archives’ (*Gale*, n.d.). Its cross-search platform means that scans can be made across all *Gale Primary Sources* collections at once ‘to reveal unseen connections’ (*Gale*, 2020). Likewise, *Wiley Digital Archives* claims to be the only programme in the library market to offer Automated Text Recognition (ATR) of handwritten sources. *Preservica* speaks for them all, in its description of ‘Information sitting in a dark archive or backup system [as] value waiting to be unlocked’ (*Preservica*, n.d.). In short, the development of advanced search and recombination tools is not tangential to the other benefits or design features of digital archives; it is their *raison d’être*.

As elsewhere in the digital economy, commercial logics are not unchallenged. For public providers, recombination matters because it offers the opportunity to unlock collections, widen access and increase usage. In this respect California, the birthplace of the platform, offers a compelling example. The California Digital Library, part of the University of California, was founded in 1997 and today it constitutes one of the world's largest digital research archives – accessible through its purpose-built platform, *Calisphere*. The platform offers the ability to search over two million items, across 2,000 collections, from more than 300 Californian cultural heritage organisations. *Calisphere's* origins can be seen in relation to concerns about the model of commercial platforms. As Richard Lucier, the founding director of the California Digital Library commented at its launch, 'We can't trust commercial entities to do archiving. There may be knowledge we want to preserve that just does not supply sufficient economic return for commercial publishers... Institutions like libraries have a traditional responsibility' (Quint, 1998: 49). Digitisation, therefore, does not automatically equate to the privatisation of knowledge, and the hegemony of the commercial platform is far from complete. Instead, given that many libraries, archives, museums or historical societies lack the resources or scale to create their own digital platforms, a more hybrid landscape is emerging that includes public institutions that own materials and private companies that offer solutions for scanning, hosting and searching them. Irrespective of the precise relations between these parties, the same forces of recombination and aggregation are locked into the logic of digital platforms – commercial or otherwise.

Power in the digital archive appears to move from the archivist to the researcher. After all, online we arrive at materials not through the arrangement of the archive, but through tools that prioritise *our* search terms. This is the 'efficiency' and 'empowerment' promised by all digital platforms: the ability to surgically extract only those sources that speak directly to our research questions. But it is also the illusion of digital archives: it is not the researcher who is creating these new connections. The language

of end-user empowerment masks the fact that it is the platform that delivers recombination – not us. The shift of power in digital archives is not from archivists to researchers, but from archivists to algorithms.

We need to situate digital 'discovery tools' within the context of critical historical geographies of archival research that interrogate practices of selection and ordering. For example, scholars rarely know (or can know) the relevance metrics that an algorithm is using to organise and display their results (Underwood, 2014). If archiving not only represented the world but participated in it – today algorithms increasingly shape our understandings of the past. This is more widely appreciated. Kitchen and Dodge (2014: 44), for example, note that 'software needs to be understood as an actant in the world; it augments, supplements, mediates and regulates our lives... Software transforms and reconfigures the world in relation to its own systems of thought'. This presents a clear research challenge. As Louise Amore (2020: 20) writes: 'To attend to algorithms as generating active, partial ways of organizing worlds is to substantially challenge notions of their neutral, impartial objectivity'.

Although digital platforms often feel neutral and comprehensive, they are no less selective and deterministic of our research findings. As Sternfeld (2011: 557) notes, in its representational form, the platform *is* the archive: 'From traditional finding aids and indexing schema to sophisticated digital design features, an archival interface governs use'. Material documents needed to be ordered and consulted in physical space which went on to shape the linear-narrative historical geographies we wrote from them. Digital technology represents a fundamental change here. For Anderson (2014: 112), the capacity for recombination based on search terms facilitates 'increasingly volatile visions of the past'. How many examples does one need to prove a claim? It pays to be mindful that in a 'database of millions of sentences you can find twenty examples of practically anything' (Underwood, 2014: 66). Let us turn to explore how the recombinative power of platforms is tied to their ability to draw across ever-growing collections of source material.

IV Platforms and aggregation

Digitisation promotes forms of aggregation. Source material which existed in separate archives and places is bundled together on digital platforms. [ProQuest \(n.d.a\)](#) notes how its ‘vast content sets’ span centuries of newspapers and primary sources. *Gale Primary Sources* offers dozens of aggregated thematic series, such as Women’s Studies Archive, Slavery and Anti-Slavery: A Transnational Archive, or Public Health Archives. One of the major draws of *Calisphere* is its approach to topical groupings and ‘themed collections’ which are routinely curated and reassembled from across its 300 contributing organisations. And the *Wiley Digital Archives* platform allows users simultaneously to search across collections such as the Royal Geographical Society (with IBG) (RGS-IBG), British Association for the Advancement of Science and New York Academy of Sciences. By combining sources onto a digital platform, the vagaries of an individual archive’s collections play a diminishing role in structuring our research outcomes. This is especially true as recent advances in scanning technology have facilitated a shift from early digitisation efforts, that focussed primarily on published materials in libraries – notably books and newspapers – towards unpublished, manuscript collections in archives, of which *Wiley Digital Archives* is an exemplar ([Hahn, 2006](#)).

The drive to aggregation reveals how digitisation is simultaneously a business model, a political ideology and a distinct perspective on understanding archives. As with other areas of the platform economy, like Google for search, the goal of totality and market dominance is inherent to the commercial strategy of digital platforms, driven by the stark imbalance between high start-up cost and low marginal cost. This economic logic also shapes public providers who have their own need to demonstrate value for money. In the case of digital archives, platforms frequently pitch that the virtually limitless capacity and global reach of the internet age offers the possibility that we might search across the world’s knowledge, near perfectly preserved. But this is to present a political point merely as a technical one. The idea of the ‘total archive’, which has been

reactivated in our digital present, has a longer, discernibly analogue history and geography.

Here, several accounts turn to the historical example of the Mundaneum. Planned by Belgian internationalists Paul Otlet and Henri La Fontaine in the late-nineteenth century, the Mundaneum was to gather together the world’s knowledge, and organise it under a standard system of decimal classification in the Brussels-based Palais Mondial. Otlet’s vision may seem a historical eccentricity, if it was not for the fact that in 2012 Google, responding to regulatory and cultural concerns about its dominance of online search, announced a partnership with the Mundaneum Archive Center, now in Mons. Its innovative *Google Books*, notes [Thylstrup \(2018: 14\)](#), ‘situated Google as a utopian, even ethical, idealist project’. But mass digitisation has ‘highly contingent spatio-temporal configurations’. When French officials reacted to the dominance of *Google Books* by debating an alternative – which would eventually materialise as the European Commission’s *Europeana* – mass digitisation was confirmed as a ‘process that not only neutrally scanned and represented books but could also produce a new mode of world-making, actively structuring archives as well as their users’ ([Thylstrup, 2018: 15](#)).

That same tension continues to play out on digital platforms. Supporters and the companies themselves are keen to pitch digital archives as placeless, neutral tools for scholarship. It is Google’s self-proclaimed mission ‘to organise the world’s information and make it universally accessible’ ([Google, n.d.](#)). However, critics see those same platforms as the enclosure and privatisation of knowledge. In short, digital archives cannot be separated from wider political struggles over the control and ownership of cultural memory. There is a dynamic set of tensions between international capital and sovereignty in this digital landscape, most vividly expressed in the ambition of aggregation. National claims to archives collide with commercial interests that ostensibly drive towards internationalist agendas, even while having to defend themselves in specific legal jurisdictions. For example, when challenged by American authors over its digitisation of books, Google appealed to US fair use copyright law ([Liptak and Alter, 2016](#)).

Thylstrup's important account lays bare how digital assemblages are remaking political, cultural, economic and historical geographies of knowledge in profound and far-reaching ways. New technical and economic infrastructures are 'governed less by the hierarchical world of curators, historians, and politicians, and more by feedback networks of tech companies, users, and algorithms' (Thylstrup, 2018: 51). And there is not simply an economics to digital archives, there is a geopolitics too: 'The *scale* of Big Data, it transpires, is much less impressive than the ways in which scale is mobilised within a rhetoric of completeness, and the circulation of digital information comes to define the limits of inquiry' (Jardine and Drage, 2019: 4). We know to be suspicious of the perceived neutrality of archives and this is no less true for understanding the often implicit political and philosophical assumptions that underpin digitisation. These assumptions raise new ethical injunctions for historical geographers (Moore, 2010). In the remainder of this paper, we explore two particularly salient ethical issues in turn: first, the politics of remote access and, second, the issue of absence and erasure.

V Ethics and recombination

I Remote access

As we have argued, recombination is an inherently geographical process; the act of digitisation is also always an act of dislocation. And if we consider the location and display of records to be central to the task of understanding the work that they historically performed, how do geographers remain attentive to the impress of setting when accessing materials remotely (see Griffiths and Baker, 2020)? We can see that the locational geography of archives matter in the case of disputed collections (Lowry, 2017; Shepard, 2015) or those themes, such as internationalism (Hodder et al., 2021) or race (Hyacinth, 2019), marginalised by state-centric recording practices. The promise of digitisation is precisely its ability to work against the locational geography of archives by pulling together disparate collections and thereby removing the time and cost involved in research travel. Remote access is therefore central to

the version of 'efficiency' that platforms promise, and to the task of opening up collections to wider audiences. These benefits are clearly significant, but there is nothing inherently egalitarian about digital platforms that make it possible for *some* people, in *some* places to do historical work remotely (Putnam, 2016a).

Putting to one side subscription fees and technology costs, remote access raises a larger ethical issue. How might we manage our vastly increased opportunities to write about people and places we never have to visit? Putnam argues that with these 'research efficiencies' we also risk losing the unintended, but important, experiential learning of fieldwork. How much do we really know about the fragments that surface in our search results? Digital platforms require 'almost no prior contextual knowledge: that's what happens when you piggyback on commercial technology honed to connect people to purchases as easily as possible' (Putnam, 2016a: 399).

Without travel, how might we ethically stress-test our archival research? As geographers have highlighted (e.g. Haines, 2019), the archive is a space in which documents are read, but so too is the researcher. It is the space where we are forced to confront our positionality and our relationship to the material. By contrast, digital archives invite us to confront pasts – sometimes difficult and uncomfortable ones – from the safety of our world. We risk becoming insulated from the people and places we claim to know and write about. In their study of the digitisation of community archives in Scotland's Western Isles, Beel et al. (2015: 203) note that traditional archives existed 'like "silos" of local knowledge whereby you have to be in-place to add to them or view them'. In that sense, geography was an obstacle. Now the idea of the geographer being *emplaced* is being supplanted as researchers become *enmeshed* in a different infrastructure and geography of knowledge. There is still something ethically at stake, as Putnam notes, if the wider world becomes at once more present in our writing and less present in our working lives.

Digital archives have largely been welcomed with a benign sense of enthusiasm, and we fully recognise their convenience and their capacity to broaden our research. The novelty of this new digital

infrastructure, however, does raise a series of ethical questions. How can we maintain ethical standards of anonymity in a world of full-text search or guarantee the right to be forgotten (Allen, 2017; Crossen-White, 2015; Mkadmi, 2021)? What are the effects of making wholesale historical collections about marginalised groups available online? What kind of ethic of care is there to help users navigate sensitive materials? What does it mean to enable records to be broken apart, decontextualised and recombined? And what happens when we do not have to confront the racial and colonial power structures that fixed records in their current locations? These questions invite no clear-cut answers, and we do not wish to defend a narrow conception of expertise, but by considering them we can certainly better fit our training to meet the new ethical challenges and opportunities presented by digital technologies.

2 Absence and erasure

If one has any training in archive methods, it has likely highlighted how acts of erasure or silencing have shaped the production and management of records (see Mills, 2013; McGeachan, 2018). One of the great hopes of digitisation is that recombination might herald new, emancipatory historical geographies. Digital discovery tools offer new ways of navigating the older infrastructures and biases of collections previously separated by place, time and context. The sub-text of this is that aggregation (combined with powerful search tools) offers the possibility of scaling-up absence into presence; digital archives can be used to resurrect ‘forgotten’ individuals. However, Caroline Bressey repeats the warning of feminist historians that digitisation ‘has not transformed the nature of the sources we are searching’ (Hunter, 2017: 210, cited in Bressey, 2020). Here alternative methods might merit attention, such as Saidiya Hartman’s (2008: 11) notion of ‘critical fabulation’. Hartman’s creative response to absences in the archive on trans-Atlantic slavery was to blend archive fragments with fictional narratives. Such work marks one response to the limitations of using archival sources as evidence. For our purposes, mass digitisation might appear to represent another; but by itself it cannot overcome the silencing intrinsic

to historical recording practices – it simply scales it up.

Importantly, digitisation can compound forms of absence and erasure (Hodder, 2017). Platforms, and the commercial logics that power them, generate results, however tangential. The effect, as Brian Maidment (2012: 112) notes, is that ‘Any sense of what might be absent recedes under the press of what is so obviously and overwhelmingly present’. As more material becomes available online, the records left behind seem to become more hidden, less important. Leary (2005: 82) argues that soon analogue materials may ‘simply cease to exist’ to anyone but the most dedicated of specialists. He coins the term the ‘offline penumbra’ to refer to ‘that increasingly remote and unvisited shadowland into which even quite important texts fall if they cannot yet be explored [online]’ (Leary, 2005: 82). For Leary, the offline penumbra is one half of a new ‘digital divide’ that is fundamentally transforming historical work.

Digitisation has facilitated a wider shift in how we understand archives, then. In a world of disparate analogue archives, knowledge had tended to be imagined as scarce and therefore fiercely guarded, and the labour to retrieve it arduous. Conversely, digital platforms present a world of limitless information in which knowledge is being relentlessly expanded. This shift has undoubtably made archive work more appealing to those who would not traditionally have used them, reaching beyond the sub-discipline of historical geography. This is to be welcomed. However, to develop the ideas of Anderson (2014), the illusion of comprehensiveness obscures how digitisation is a spatially uneven and unrepresentative process. The rationale for choosing which collections are digitised differs between commercial and public providers. But in both cases, it usually includes some combination of ease of digitisation, perceived popularity and copyright issues. Beyond those points, however, we know that whole institutions and collections remain offline.

In practice, digitisation has a geography. Its origins in elite, well-funded institutions in the Global North still indelibly mark the boundaries of our scholarship. For some areas of enquiry, this has enabled a transformative expansion. It is not insignificant that Putnam (2016a) notes the enabling

effects of digitisation on the development of transnational history. Different regional platform configurations will continue to determine the geographical parameters of our work. As is hinted at by its very name, *Calisphere* might lead to a scalar focus on particular states if we lean into source collections because they can most conveniently be accessed. Here, the partiality of the archive is not necessarily in the sources, but rather emerges from the geography of the platforms. The implications are profound, if the future scope of historical geography is to be determined by the vagaries of platform building.

VI Conclusion

In 2002, when the RGS-IBG embarked on the first step of ‘unlocking its collections’ by digitising its archive catalogue, Charles Withers (2002: 309) wrote that ‘It is too soon to say whether the WWW will act simply as a means of recall from a global archive, or if it marks the beginning of “a new inventive relationship to knowledge, a relationship that is dissolving the hierarchy associated with the archive”’ (Caygill, 1999), cited in Withers, 2002: 309). Twenty years on, as hundreds of thousands of items from the RGS-IBG archive have been digitised, we can see more conclusively that digitisation is fundamentally changing historical geographical research. *Wiley Digital Archives* – the platform that hosts the RGS-IBG collection – promises tools that are built for research ‘efficiency’ and recombination. By using the latest scanning technology, it notes that archive content is ‘transformed into clear, crisp, searchable documents and made accessible on our platform where researchers can find, group, translate, download, manipulate and share historical materials with ease’ (Wiley Digital Archives, n.d.). The digitisation of one of the discipline’s preeminent collections offers an important moment to reflect on the broader challenges posed for geographical scholarship by the shift from finding aid to search algorithm, and from physical building to digital platform.

This paper has considered the impacts of digitisation, drawing on discussions underway in cognate disciplines such as history and archive studies. We

have argued that digital archives and their associated technologies are fundamentally changing our relationship to the geographical past – from one heavily structured by place and archival arrangement, toward one shaped by processes of aggregation, fragmentation and recombination. In doing so, our aim is not to reify the pre-digital age as the benchmark of ideal practice. Digitisation has many positive aspects, including widening access, preserving collections and promoting novel, interdisciplinary scholarship. Instead, digitisation invites us critically to consider how the material qualities of paper documents have an increasingly undue influence on how we conceptualise historical research in geography. Our methods teaching must go further to interrogate not simply the experience of working in archives, but the changing political and economic infrastructures of digitisation. As we have argued above, digital platforms make the user think they are in control, that it is their effort that returns new connections. But it is the platform that sorts, reorders, tabulates, extracts and recombines results in ways that are profoundly deterministic of our research outcomes.

As with analogue archives, working with digital platforms demands a consciously antagonistic approach and a clear theorisation of underlying technology (Beckingham and Hodder, 2023). At present, we are using digital platforms as tools of convenience without critically examining their role in determining the conclusions we draw. As Ted Underwood (2014: 69) argues, ‘Researchers can never afford to treat algorithms as black boxes that generate mysterious authority. If we’re going to use algorithms in our research, we have to crack them open and find out how they work’. Geographers do not need to be coders, but we do need to ask what relevance metrics an algorithm is using to organise our results? What are the basic assumptions that underpin them? What kind of information is likely to be included, prioritised or lost through that process? And how can we better report those parameters in our writing? The rise of recombinant historical geographies is not inherently worrisome, but it does demand us to take a different set of critical questions into the archive.

Digitisation prompts us to consider how we cite material. Advocates for greater source transparency – influenced by the open science movement – have

called for researchers to publish accompanying datasets. As we come to rely on digital archives, which are potentially accessible to our readers, could such demands be made of historical work? For example, Cope (2018) raises the prospect of providing hyperlinks to each source in full from our publications (also see Elman and Kapiszewski, 2017). But digitisation invites a far broader discussion about what it means to have ‘used’ a source (Leary, 2005). We can no longer assume that an archival footnote is evidence that the contextual education of fieldwork has been gained. Nor can we assume that a source is representative of the wider collection or context in which it sits. In a world of recombination, full transparency would require us to explain how we arrive at sources, not merely link to what they say.

We probably also need a better way to fully account for the hidden benefits of offline research – or, the ‘unsheddable contextualization that makes work with analog sources so inefficient’ (Putnam, 2016a: 393). We need to value those practices that embed our scholarship, as much as those that speed it up. With hindsight, we can see that the need to publish more, faster, laid the groundwork for the expansion and adoption of digital archives. Platforms cater to a demand, identified by Lorimer (2010: 254) over a decade ago, that researchers be ‘directed along the shortest, quickest and easiest search routes likely to lead to the desired archival object, or anticipated “find”’. In this way, platforms are a response to the demands we have made as researchers, as a discipline and as a scholarly industry.

By opening this discussion, we suggest that historical geographers not only have much to reflect on with respect to their own practice but can also contribute to disciplinary discussions on new digital geographies (e.g. Ash et al., 2018a, 2018b; Kinsley, 2014; Offer, 2013; Pickrell, 2018) and histories (Dougherty and Nawrotzki, 2013; Owens, 2018; Weller, 2012). The technologies mobilised in digital archives are part of a set of applications and platforms whose everyday use has transformed geographical relations, ‘altering space, time, memory, and collective knowledge’, shaping what pasts become available, when and to whom (Elwood and Mitchell, 2015: 147; Elwood, 2021).

Ultimately, whether we access sources in person or online, we need a new model of historical research in geography. A model that rewards geographers for taking the time to learn about the richness of what was going in particular periods and places, rather than one that disproportionately rewards us for recombining fragments that surface in search results. That is the intellectual challenge ahead of us but, as we have argued in this paper, it is an ethical challenge too.

Acknowledgements

We presented this work in the Cultural and Historical Geography Seminar Series at the University of Nottingham and are grateful to the audience for their feedback. We would also like to thank Andrew Leyshon, Charles Watkins, our editor Don Mitchell, and the journal’s anonymous reviewers for their generous comments on earlier versions of this paper.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the British Academy and Wolfson Foundation (WF21\210306), Arts and Humanities Research Council (AH/M008142/1) and a Royal Geographical Society (with IBG) Small Research Grant (20th International Geographical Congress Fund).

ORCID iD

Jake Hodder  <https://orcid.org/0000-0003-1898-1479>

References

- Allen R (2017) Lost and now found: the search for the hidden and forgotten. *M/C Journal* 20(5): 1290.
- Amoore L (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.
- Anderson S (2014) Past indiscretions: digital archives and recombinant history. In: Kinder M and McPherson T

- (eds) *Transmedia Frictions*. Oakland, CA: University of California Press, 100–114.
- Ash J, Kitchin R and Leszczynski A (eds) (2018a) *Digital Geographies*. London: Sage.
- Ash J, Kitchin R and Leszczynski A (2018b) Digital turn, digital geographies? *Progress in Human Geography* 42(1): 25–43.
- Bailey J (2013) Disrespect des fonds: Rethinking arrangement and description in born-digital archives. *Archive Journal*. Available at: <https://www.archivejournal.net/essays/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/> (accessed 21 January 2022).
- Bartlett N (1991) Respect des fonds: the origins of the modern archival principle of provenance. In: McCrank LJ (ed) *Bibliographical Foundations of French Historical Studies*. New York, NY: Howarth Press, 107–115.
- Beckingham D and Hodder J (2023) [forthcoming] Historical geographies: geographical antagonism and archives. In: Lovell SA, Coen SE and Rosenberg MW (eds) *The Routledge Handbook of Methodologies in Human Geography*. London: Routledge.
- Beel D, Wallace C, Webster G, et al. (2015) The geographies of community history digital archives in rural Scotland. *Scottish Geographical Journal* 131(3–4): 201–211.
- Bressey C (2020) Surfacing black and brown bodies in the digital archive: domestic workers in late nineteenth-century Australia. *Journal of Historical Geography* 70: 1–11.
- Caygill H (1999) Meno and the internet: between memory and the archive. *History of the Human Sciences* 12: 1–12.
- Clayton D (2021) Historical geography I: doom, danger, disregard—towards political historical geographies. *Progress in Human Geography* 45(6): 1692–1708.
- Cope M (2018) Qualitative methods and geohumanities. In: Ash J, Kitchin R and Leszczynski A (eds) *Digital Geographies*. London: Sage, 96–105.
- Craggs R (2008) Situating the imperial archive: the Royal Empire Society Library, 1868–1945. *Journal of Historical Geography* 34(1): 48–67.
- Craggs R (2016) Historical and archival research. In: Clifford N, Cope M, Gillespie T and French S (eds) *Key Methods in Geography*. 3rd edition. London: Sage, 111–128.
- Crossen-White HL (2015) Using digital archives in historical research: what are the ethical concerns for a “forgotten” individual? *Research Ethics* 11(2): 108–119.
- Dougherty J and Nawrotzki K (2013) *Writing History in the Digital Age*. Ann Arbor, MI: University of Michigan Press.
- Duchemin M (1983) Theoretical principles and practical problems of respect des fonds in archival science. *Archivaria* 16: 64–82.
- Elman C and Kapiszewski D (2017) Annotating qualitative social science. Parameters: knowledge under digital conditions (SSRC). Available at: <https://items.ssrc.org/parameters/annotating-qualitative-social-science/> (accessed 28 April 2022).
- Elwood S (2021) Digital geographies, feminist relationality, black and queer code studies: thriving otherwise. *Progress in Human Geography* 45(2): 209–228.
- Elwood S and Mitchell K (2015) Technology, memory, and collective knowing. *Cultural Geographies* 22(1): 147–154.
- Ferretti F (2020) History and philosophy of geography I: decolonising the discipline, diversifying archives and historicising radicalism. *Progress in Human Geography* 44(6): 1161–1171.
- Gagen E, Lorimer H and Vasudevan A (2007) *Practising the Archive: Reflections on Method and Practice in Historical Geography*. Historical Geography Research Series 40. London: Historical Geography Research Group.
- Gale (2020) Gale primary sources 2020–2021: collections and platform features. Available at: https://www.gale.com/binaries/content/assets/gale-us-en/intl-assets/intl-assets-uk-europe/gale_brochure_gps2021_web.pdf (accessed 21 January 2022).
- Gale (n.d.) Tools for discovery. Available at: <https://www.gale.com/intl/research-tools/tools-for-discovery> (accessed 21 January 2022).
- Gooding P (2016) *Historic Newspapers in the Digital Age: Search All About it!* Abingdon: Routledge.
- Google (n.d.) About Google. Available at: <https://about.google> (accessed 18 May 2022).
- Griffiths M and Baker K (2020) Decolonising the spaces of geographical knowledge production: the RGS-IBG at Kensington Gore. *Area* 52(2): 455–458.
- Hahn TB (2006) Impacts of mass digitization projects on libraries and information policy. *Bulletin of the American Society for Information Science and Technology* 33(1): 20–24.

- Haines E (2019) Visions from behind a desk? Archival performance and the re-enactment of colonial bureaucracy. *Area* 51(1): 25–34.
- Hammond T (2020) Papering, arranging, and depositing: Learning from working with an Istanbul archive. *Area* 52(1): 204–212.
- Hartman S (2008) Venus in two acts. *Small Axe* 12(2): 1–14.
- Hodder J (2017) On absence and abundance: biography as method in archival research. *Area* 49: 452–459.
- Hodder J, Heffernan M and Legg S (2021) The archival geographies of twentieth-century internationalism: nation, empire, race. *Journal of Historical Geography* 71: 1–11.
- Hunter K (2017) Silence in noisy archives: reflections on Judith Allen's 'Evidence and silence – feminism and the limits of history' (1986) in the era of mass digitisation. *Australian Feminist Studies* 32(91–92): 202–212.
- Hyacinth N (2019) Black archives in the UK report: opportunities, challenges and moving forward. Race, culture and equality working group report, royal geographical society. Available at: <https://raceingographydotorg.files.wordpress.com/2019/08/race-group-report-july-2019-proofread.pdf> (accessed 16 January 2022).
- Jardine B and Drage M (2019) The total archive: data, subjectivity, universality. *History of the Human Sciences* 31(5): 3–22.
- Jazeel T and Legg S (eds) (2019) *Subaltern geographies*. Athens: University of Georgia Press.
- Kinsley S (2014) The matter of “virtual” geographies. *Progress in Human Geography* 38(3): 364–384.
- Kitchin R and Dodge M (2014) *Code/Space: Software and Everyday Life*. Cambridge: MIT Press.
- Langley P and Leyshon A (2017) Platform capitalism: the intermediation and capitalisation of digital economic circulation. *Finance and Society* 3(1): 11–31.
- Leary P (2005) Googling the Victorians. *Journal of Victorian Culture* 10(1): 72–86.
- Liptak A and Alter A (2016) Challenge to google books is declined by supreme court, The New York Times, 18 April. Available at: <https://www.nytimes.com/2016/04/19/technology/google-books-case.html> (accessed 28 April 2022).
- Lorimer H (2010) Caught in the nick of time: archives and fieldwork. In: DeLyser D, Herbert S, Aitken S, Crang M and McDowell L (eds) *The Sage Handbook of Qualitative Geography*. London: Sage, 248–272.
- Lowry J (ed) (2017) *Displaced Archives*. Abingdon: Taylor & Francis.
- Maidment B (2012) Writing history with the digital image: a cautious celebration. In: Weller T (ed) *History in the Digital Age*. Abingdon: Routledge, 123–138.
- Mbembe A (2002) The power of the archive and its limits. In: Hamilton C, Harris V and Taylor J (eds) *Refiguring the Archive*. Dordrecht: Springer, 19–27.
- McGeachan C (2018) Historical geography II: traces remain. *Progress in Human Geography* 42(1): 134–147.
- Mills S (2013) Cultural–historical geographies of the archive: fragments, objects and ghosts. *Geography Compass* 7(10): 701–713.
- Mkadmi A (2021) *Archives in the Digital Age: Preservation and the Right to Be Forgotten*. London: ISTE Ltd and John Wiley & Sons.
- Moore FP (2010) Tales from the archive: methodological and ethical issues in historical geography research. *Area* 42(3): 262–270.
- Offer K (2013) Historical geography II: digital imaginations. *Progress in Human Geography* 37(4): 564–577.
- Owens T (2018) *The Theory and Craft of Digital Preservation*. Baltimore, MD: Johns Hopkins University Press.
- Owens T and Padilla T (2021) Digital sources and digital archives: historical evidence in the digital age. *International Journal of Digital Humanities* 1: 325–341.
- Poell T, Nieborg D and Van Dijck J (2019) Platformisation. *Internet Policy Review* 8(4): 1–13.
- Pickrell G (2018) The global assemblage of digital flow: Critical data studies and the infrastructures of computing. *Progress in Human Geography* 42(2): 225–243.
- Preservica (n.d.) Welcome to digital preservation 101. Available at: https://preservica.com/digital-preservation-101?gclid=CjwKCAiA24SPBhB0EiwAjBgkhvO85N64dLwvNVpjK0-MYKvEs5Bo_rt2RuE6diNDKBJyVm7qxnW5xoCV60QAvD_BwE (accessed 21 January 2022).
- ProQuest (n.d.a) Libraries empower researchers to take on today's global challenges. Available at: https://go.proquest.com/GlobalChallenges/?_ga=2.10750071.2121940063.1642151810-849271170.1642151810 (accessed 21 January 2022).

- ProQuest (n.d.b) ProQuest home page. Available at: <https://about.proquest.com/en/> (accessed 21 January 2022).
- Putnam L (2016a) The transnational and the text-searchable: digitized sources and the shadows they cast. *The American Historical Review* 121(2): 377–402.
- Putnam L (2016b) Daily life and digital reach: place-based research and history's transnational turn. In: Puri S and Castillo DA (eds) *Theorizing Fieldwork in the Humanities*. New York, NY: Palgrave Macmillan, 167–181.
- Quint B (1998) The California digital library: an interview with Richard Lucier. *Searcher* 6(5): 48–49.
- Ramsay S (2010). The hermeneutics of screwing around; or what you do with a million books. In: Kee K (ed) *Pastplay: Teaching and Learning History With Technology*. Ann Arbor, MI: University of Michigan Press, 111–120.
- Richardson L (2020) Coordinating the city: platforms as flexible spatial arrangements. *Urban Geography* 41(3): 458–461.
- Roche M (2021) From dusty to digital: archival research. In: Hay I and Cope M (eds) *Qualitative Research Methods in Human Geography*. Don Mills, ON: Oxford University Press, 222–243.
- Smicek N (2017) *Platform Capitalism*. Cambridge: Polity Press.
- Shepard T (2015) Of sovereignty: disputed archives, wholly modern archives, and the post-decolonization French and Algerian Republics, 1962–2012. *The American Historical Review* 120(3): 869–883.
- Sherratt T (2013) *From Portals to Platforms—Building New Frameworks for User Engagement. Paper Presented at LIANZA 2013*. Hamilton.
- Sternfeld J (2011) Archival theory and digital historiography: Selection, search, and metadata as archival processes for assessing historical contextualization. *The American Archivist* 74(2): 544–575.
- Stoler AL (2010) *Along the Archival Grain*. Princeton, NJ: Princeton University Press.
- Thylstrup NB (2018) *The Politics of Mass Digitization*. Cambridge: The MIT Press.
- Underwood T (2014) Theorizing research practices we forgot to theorize twenty years ago. *Representations* 127(1): 64–72.
- Weller T (ed) (2012) *History in a Digital Age*. Abingdon: Routledge.
- Wiley Digital Archives (n.d.) <https://www.wileydigitalarchives.com> (accessed 21 January 2022).
- Withers CWJ (2002) Constructing “the geographical archive”. *Area* 34(3): 303–311.
- Zhang J (2012) Original order in digital archives. *Archivaria* 74: 167–193.

Author biographies

Jake Hodder is an Associate Professor in the School of Geography, University of Nottingham. His research interests focus on the historical geographies of internationalism, race and archives in the twentieth century.

David Beckingham is an Associate Professor in the School of Geography, University of Nottingham. His research examines the relationship between liberal government and moral regulation, with a particular focus on the regulation of drinking and drunkenness.