



Digitised Archives and the Uneven Visibility of Marginal Voices in Historical Research

Aibek Parthasarathy^{1*}, Ewa Pham², Beatriz Ngo³

¹Department of Social Sciences, Vistula Graduate Institute of Social Sciences, Chile

²Department of Social Sciences, Transylvanian Center for Social Sciences, Colombia

³Department of Social Sciences, Tagus University Centre for Social Sciences, Nepal

Article Info

Volume 1, Issue 1, 2026

doi: [10.62715/ijhcs.2026.v1.i1.pp19-23](https://doi.org/10.62715/ijhcs.2026.v1.i1.pp19-23)

Received: 03 February 2026

Accepted: 06 April 2026

Published: 24 May 2026

Abstract

Mass digitisation is often presented as a democratising force in historical research: sources once confined to reading rooms can be searched from anywhere, and a keyword can surface a servant, a migrant or an enslaved person in a newspaper column that no index ever listed. This Perspective argues that the gain is real but unevenly shared, and that digitisation adds new filters to an archive that was already selective. What is scanned depends on funding, copyright, institutional priorities and the technical ease of converting a page to text, and these factors favour printed, Latin-script, metropolitan and official material. Search then rewards what has been converted well and described fully. The result is that some marginal voices become far easier to find while others, held in manuscript, in minority languages or in poorly resourced repositories, fall further from view because research attention flows to what is online. The article sets out this argument, considers the objections that digitisation has recovered many hidden lives and that community-led digital archives redress the balance, and draws implications for how historians search, cite and teach, and for how institutions choose and describe what they digitise.

Keywords: Digitisation, Archives, Marginalised groups, Keyword search, Digital history, Archival silence

© 2026 Aibek Parthasarathy et al. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

Within a generation the working conditions of historians have been transformed. Newspapers, parliamentary papers, court proceedings, parish registers and colonial correspondence that once required travel and weeks of page-turning can now be queried in seconds. For those who study people who left few records of their own, the change has seemed especially promising. A name typed into a database of digitised newspapers may return a labourer's court appearance, a runaway notice or a letter to an editor that no finding aid would have led to.

The promise deserves a careful second look. Historians have long recognised that archives are made by the powerful and that the poor, the colonised and the enslaved appear in them largely as objects of

* Corresponding author: Aibek Parthasarathy, Department of Social Sciences, Vistula Graduate Institute of Social Sciences, Chile. E-mail: aibek.parthasarathy@vistula.edu

© 2026 Aibek Parthasarathy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

administration (Trouillot, 1995). Digitisation does not undo that history. It reproduces a portion of the archive in a new medium, and the choice of portion, the manner of reproduction and the tools of retrieval each have consequences for who can be seen.

This Perspective contends that digitised archives have made the visibility of marginal voices more uneven, raising some sharply and lowering others in relative terms. It then weighs the main objections to that view and suggests what researchers and institutions might do in response.

2. An Archive Already Filtered

Trouillot (1995) distinguished four moments at which silences enter the making of history: when sources are created, when they are assembled into archives, when narratives are drawn from them, and when significance is assigned in retrospect. Archivists have made a parallel case from within their own profession, insisting that archives are instruments of power that shape what a society can remember, and are far from neutral storehouses (Schwartz and Cook, 2002). Scholars of slavery and empire have shown what this means on the page. Spivak (1985) found an Indian queen present in East India Company records only where she served the Company's purposes, and Fuentes (2016) reconstructed the lives of enslaved women in Barbados from documents that register them through violence and sale.

These writers responded with methods of reading: along the grain of official reasoning to learn its habits (Stoler, 2009), against it to catch what it let slip, and in some cases beyond it, by openly imaginative reconstruction where the record gives out (Hartman, 2008). All of them assume a reader who knows what kind of archive is in hand. That assumption is the first thing that mass digitisation weakens.

3. What Digitisation Adds

Digitisation places further layers between a past life and a present researcher. Figure 1 sets them out in order, from the creation of a record to its discovery on a screen, and the subsections below take up the three that are new.

3.1. Selection

Only a small fraction of archival holdings has been scanned, and the fraction is not a sample. Projects are funded where demand is proven, where rights are clear and where material is uniform enough to process cheaply. National newspapers, printed government papers and records useful to family historians come first. Material in the languages and scripts of colonised peoples, manuscript collections in small repositories and records still in private or community hands come later or not at all. The politics of these choices, including the tendency of former imperial powers to digitise colonial records held in the metropole on their own terms, has become a subject of study in its own right (Zaagsma, 2023).

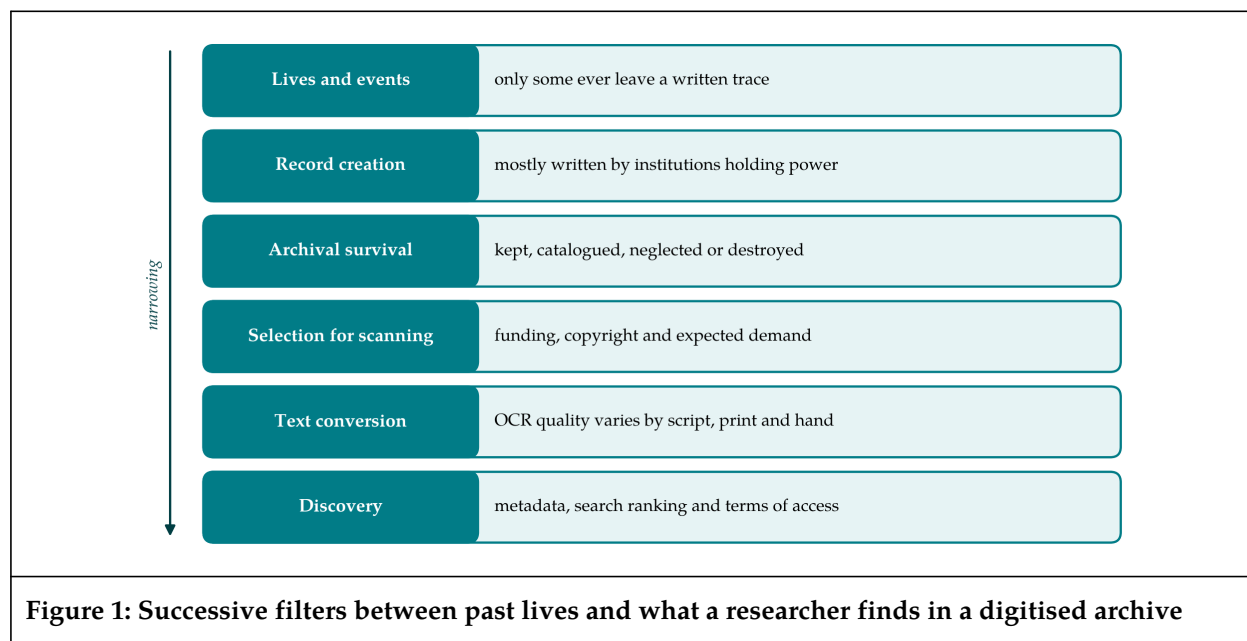
3.2. Conversion

A scanned image becomes searchable only when its text is extracted, usually by optical character recognition. Recognition works best on clean modern print in major languages and degrades on worn type, crowded columns, older letterforms and non-Latin scripts; handwriting requires separate and costlier methods. Cordell (2017) has argued that the error-strewn text behind a newspaper database is best understood as a new edition with its own bibliographic character. The practical consequence is that a failed search does not show absence. It may show only that the cheap paper and small type of a radical or minority-language weekly defeated the software.

3.3. Discovery

What can be found also depends on description and ranking. Catalogue records often carry the vocabulary of the period that produced them, so that people are retrievable under the terms their rulers used and seldom under the names they gave themselves. Full-text search bypasses the catalogue but brings its own distortions. Underwood (2014) pointed out that a search engine is a tool whose workings scholars rarely examine and which tends to confirm the hypothesis that framed the query. Putnam (2016) showed how text-searchable sources allow historians to follow people and topics across borders without

the contextual knowledge that long immersion in one place used to supply, and warned that those who lived beyond the reach of print are pushed further into shadow as a result.



4. Uneven Visibility in Practice

The effect on research habits is documented. [Milligan \(2013\)](#) traced citations in Canadian history dissertations and found that references to the major dailies rose markedly once those titles could be searched online. [Hitchcock \(2013\)](#) observed that historians routinely consult the digital copy and cite the paper original, which conceals the role of search in shaping what they found. Attention is finite, and when it moves to what is online it moves away from what is not.

For marginal groups the outcome is split. People who came into contact with institutions that printed, and whose records were in a favoured language, have become dramatically easier to trace: defendants in a metropolitan court, passengers on a manifest, subjects of a newspaper report. Even then they are met through the eyes of the court, the shipping clerk and the reporter, and [Johnson \(2018\)](#) has cautioned that databases of the enslaved can repeat the logic of the ledger by turning persons into rows. People whose traces lie in oral tradition, in manuscript, in vernacular print or in under-funded archives in the global South gain nothing comparable. Table 1 summarises how each stage skews visibility and what a researcher can do about it.

5. Counter-Arguments

Two objections carry weight. The first is that digitisation has plainly recovered hidden lives. Full-text search finds individuals whom no index recorded, and work on servants, sailors, convicts and the enslaved has been enriched by it. This is true and is not in dispute here. The claim is about distribution: recovery has been concentrated where sources were printed and scanned, so that the gap between the findable and the unfindable among the marginal has widened even as the total of findable lives has grown.

The second objection is that communities now build their own digital archives. Independent community archives have a substantial history ([Flinn et al., 2009](#)), and [Caswell \(2014\)](#) has described how seeing one's own group represented in a collection counters the symbolic annihilation produced by its absence from mainstream ones. Such projects matter a great deal. They are also small, dependent on volunteers and short grants, and exposed to the fragility of digital materials, which need continual maintenance to survive at all ([Rosenzweig, 2003](#)). They complement large-scale digitisation without yet balancing it.

Table 1: Stages of digitisation, how each skews the visibility of marginal voices, and possible responses

Stage	How visibility is skewed	What a researcher can do
Selection	Official, printed and metropolitan material is scanned first	Ask what was left out; consult collection-level documentation
Conversion	Poor OCR hides worn print, minority scripts and handwriting	Browse page images; test searches on known items
Description	Catalogues retain the terms of the record creators	Search under period terms and self-given names alike
Search	Queries return what was expected and rank the well converted	Record queries; vary spellings; read around each hit
Access	Subscription databases exclude poorer institutions	Prefer and cite open collections where they exist
Citation	Digital use is hidden behind references to originals	Cite the database and date of consultation

A further consideration cuts the other way. More visibility is not always what a community wants. [Christen \(2012\)](#) has shown that Indigenous knowledge systems commonly restrict who may see certain materials by kinship, gender or ritual status, and that an ethic of universal open access overrides those rules. The CARE principles for Indigenous data governance, which set collective benefit, authority to control, responsibility and ethics beside technical openness, give that position a formal statement ([Carroll et al., 2020](#)). Unevenness, then, is not cured simply by scanning more.

6. Implications

For researchers the first implication is that source criticism must extend to the interface. A historian should know why a collection was digitised, what it omits, how its text was produced and how its search works, just as one would ask who compiled a printed calendar of documents. Queries should be recorded and reported, and the digital resource cited as what was consulted. Searching should be paired with browsing, reading whole issues and files so that a hit is understood in its setting. And the physical archive retains its value, both for what has not been scanned and for the knowledge of a collection's shape that handling it gives.

For institutions and funders the implication concerns priorities. Selection criteria that weigh under-representation alongside demand, investment in text recognition for neglected scripts and for handwriting, catalogue revision that adds self-given names to inherited headings, and long-term support for community-held collections on terms those communities set would each reduce the skew. Teaching matters as well. Students who meet the past first through a search box need to be shown what the box cannot reach.

7. Conclusion

Digitised archives have widened access to the past and have brought many obscure lives into view. They have done so selectively, along lines laid down by old archival hierarchies and by new technical and commercial ones. Recognising this is no argument against digitisation. It is a reason to treat every database as a made thing with a history, to keep asking who is missing from it, and to direct effort towards the records and communities that the first wave passed over.

Conflict of interest

To be declared by the author(s) before publication.

Funding

To be declared by the author(s) before publication.

References

- Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020). *The CARE Principles for Indigenous Data Governance*. *Data Science Journal*, 19: 43. <https://doi.org/10.5334/dsj-2020-043>
- Caswell, M. (2014). *Seeing yourself in history: Community archives and the fight against symbolic annihilation*. *The Public Historian*, 36(4): 26-37.
- Christen, K. (2012). *Does information really want to be free? Indigenous knowledge systems and the question of openness*. *International Journal of Communication*, 6: 2870-2893.
- Cordell, R. (2017). *"Q i-jtb the Raven": Taking dirty OCR seriously*. *Book History*, 20: 188-225.
- Flinn, A., Stevens, M. and Shepherd, E. (2009). *Whose memories, whose archives? Independent community archives, autonomy and the mainstream*. *Archival Science*, 9: 71-86.
- Fuentes, M.J. (2016). *Dispossessed Lives: Enslaved Women, Violence, and the Archive*. University of Pennsylvania Press, Philadelphia.
- Hartman, S. (2008). *Venus in two acts*. *Small Axe*, 12(2): 1-14.
- Hitchcock, T. (2013). *Confronting the digital: Or how academic history writing lost the plot*. *Cultural and Social History*, 10(1): 9-23.
- Johnson, J.M. (2018). *Markup bodies: Black [life] studies and slavery [death] studies at the digital crossroads*. *Social Text*, 36(4): 57-79.
- Milligan, I. (2013). *Illusionary order: Online databases, optical character recognition, and Canadian history, 1997-2010*. *Canadian Historical Review*, 94(4): 540-569.
- Putnam, L. (2016). *The transnational and the text-searchable: Digitized sources and the shadows they cast*. *American Historical Review*, 121(2): 377-402. <https://doi.org/10.1093/ahr/121.2.377>
- Rosenzweig, R. (2003). *Scarcity or abundance? Preserving the past in a digital era*. *American Historical Review*, 108(3): 735-762.
- Schwartz, J.M. and Cook, T. (2002). *Archives, records, and power: The making of modern memory*. *Archival Science*, 2: 1-19.
- Spivak, G.C. (1985). *The Rani of Sirmur: An essay in reading the archives*. *History and Theory*, 24(3): 247-272.
- Stoler, A.L. (2009). *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton University Press, Princeton.
- Trouillot, M.-R. (1995). *Silencing the Past: Power and the Production of History*. Beacon Press, Boston.
- Underwood, T. (2014). *Theorizing research practices we forgot to theorize twenty years ago*. *Representations*, 127(1): 64-72.
- Zaagsma, G. (2023). *Digital history and the politics of digitization*. *Digital Scholarship in the Humanities*.