

From Shelfmark to Platform: Provenance, Description, and Digital Access to Jawi Manuscripts in European Repositories

Allif Syahputra Bania

allifbania@unsam.ac.id

Samudra University

Berlin Sibarani

State University of Medan

Isli Iriani Indiah Pane

State University of Medan

Research Article

Keywords: Jawi manuscripts, archival netnography, digital archives, provenance, archival description, decolonial access

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11179851/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Jawi manuscripts produced across the Malay-Indonesian world now circulate through European archival infrastructures as physical holdings, catalogue records, digitised surrogates, exhibitions, and curatorial narratives. This article examines how those infrastructures mediate provenance, description, and access. Using a focused archival netnography, it analyses a purposive corpus of publicly accessible digital artefacts from the British Library, Cambridge University Library, Leiden University Libraries, and the Bibliothèque nationale de France, supplemented by verified scholarship on Malay manuscript collections and digital archives. The analysis identifies four recurring patterns: provenance is often institutionally legible but locally thin; Jawi is described through uneven combinations of language, script, genre, and colonial collecting history; digitisation expands discoverability without eliminating technical and interpretive barriers; and curatorial essays can partially restore historical relationships that catalogue fields compress or omit. The article argues that digitised Jawi manuscripts should be treated not only as textual surrogates but as archival interfaces in which custodial histories and descriptive choices are continuously remade. It proposes a reciprocal-description model that preserves legacy metadata while adding source-community terminology, acquisition histories, transliteration support, linked provenance, and multilingual contextual notes. The findings contribute to archival debates on digital access, provenance, descriptive positionality, and decolonial archival praxis while offering practical recommendations for repositories stewarding Southeast Asian manuscript heritage outside its regions of origin.

Introduction

Jawi is an Arabic-derived writing system historically used for Malay and related languages across maritime Southeast Asia. Manuscripts and documents written in Jawi encompass court correspondence, law, chronicles, Islamic scholarship, devotional works, literary texts, grammars, and everyday administrative records. A significant portion of this documentary heritage entered European libraries and archives through diplomacy, trade, scholarship, missionary activity, colonial administration, collecting, purchase, and the transfer of private libraries. Today, researchers increasingly encounter these materials first through digital catalogues and digitised surrogates rather than through the physical reading room. The digital interface has therefore become part of the archival encounter itself. (Gallop et al. 2015; Ricklefs et al. 2014).

This shift matters for archival science because digitisation does not simply reproduce a manuscript. It creates a new representational layer. A digital catalogue decides which fields are searchable; a finding aid stabilises a particular account of provenance; a digitisation portal determines what can be seen, downloaded, or cited; and a curatorial blog may provide contextual information absent from the catalogue. As Conway (2011) argues, digital surrogates acquire informational characteristics and user expectations of their own. Jeurgens and Karabinos (2020) similarly show that digitising colonial archival legacies produces new infrastructures rather than neutral copies of older ones. Those infrastructures

can reproduce inherited asymmetries, but they can also create opportunities for multivocality and recontextualisation. (Agostinho 2019; Hawkins 2022).

European repositories offer an especially productive setting for examining this problem. The British Library has publicly documented a major programme to digitise its Malay manuscripts and describes a collection of about 100 manuscript volumes and more than 300 documents in that project. Cambridge University Library reports 93 Malay manuscripts within its South and Southeast Asian holdings and preserves the Erpenius collection, which includes some of the oldest surviving Malay manuscripts in Europe. The Bibliothèque nationale de France describes a Malay manuscript collection of 131 volumes, including seventeenth-, eighteenth-, and nineteenth-century material. Leiden University Libraries holds extensive Malay and Minangkabau manuscript collections, documented through catalogues, research publications, and institutional curatorial writing. These repositories do more than preserve texts: they shape contemporary knowledge about what the manuscripts are, where they came from, and how they can be found. (British Library n.d.; Cambridge University Library n.d.; Bibliothèque nationale de France / Catalogue collectif de France 2009, updated 2022; Ricklefs et al. 2014).

The central problem is therefore not whether European repositories contain Jawi manuscripts; that is well established. The more analytically useful question is how digital archival systems make those manuscripts intelligible. This study asks: (1) How are provenance and custodial histories represented in public-facing digital descriptions of Jawi manuscripts in selected European repositories? (2) How do catalogue, digitisation, and curatorial interfaces describe language, script, genre, and regional origin? (3) What forms of access and friction emerge when researchers move from search result to digital surrogate? and (4) What descriptive practices could strengthen reciprocal and decolonial access without erasing the evidentiary value of legacy archival description? (King 2024; Jaillant 2022; Hawkins 2022).

The article contributes to Archival Science in three ways. First, it adapts netnography to an archival setting in which the digital field site consists primarily of institutional interfaces, catalogue records, finding aids, digitised objects, and curator-authored contextual texts rather than social-media interaction. Second, it brings Southeast Asian manuscript heritage into debates on provenance, archival description, and colonial digital infrastructures. Third, it proposes a practical model of reciprocal description designed for dispersed manuscript heritage: legacy metadata is retained as evidence of institutional history, while additional descriptive layers connect manuscripts to source-community names, historical geographies, script practices, and related collections. (Hine 2015; Kozinets 2020; Ghaddar and Caswell 2019; King 2024).

Literature and conceptual framework

Archival representation is never a purely technical operation. Finding aids and metadata are interpretive systems that condense judgments about creators, contexts, hierarchy, and relevance. King (2024) argues for archival meta-metadata that makes the revision history and positionality of finding aids more visible. This is directly relevant to Jawi collections because descriptions often have long histories: a manuscript

may have been classified by an early European collector, redescribed by a twentieth-century cataloguer, digitised by a twenty-first-century project, and then reintroduced to users through a modern discovery layer. Each stage can add knowledge while also sedimenting older terminology. (King 2024; Namhila 2016).

Provenance is similarly more dynamic than a single chain of custody. Recent archival scholarship has emphasised the social and organisational contexts that archival systems can obscure when they retrospectively describe records. Hurley et al. (2024) argue for a broader understanding of provenance within records continuum thinking, while Anderson (2024) demonstrates the centrality of contextual relationships to provenance across archives and archaeology. For displaced or diasporic manuscript heritage, provenance may need to include both custodial movement and the manuscript's prior social life: place of copying, scribal networks, courtly or religious use, circulation, acquisition, and later scholarly intervention. (Anderson 2024; Hurley et al. 2024).

This expanded view of provenance is reinforced by Hurley et al. (2024), who argue that provenance in a records-continuum perspective encompasses multiple social and organisational relationships rather than a single retrospective custodial account. For manuscript collections assembled across centuries, such an approach is valuable because one codex may simultaneously carry evidence of textual production, scribal intervention, ownership, scholarly annotation, colonial collecting, institutional accession, conservation, cataloguing, digitisation, and contemporary reuse. A catalogue record cannot narrate all of these contexts at equal depth, but it can signal that provenance is layered and can link outward to evidence held elsewhere. In this article, 'layered provenance' therefore refers to a descriptive practice that distinguishes, where evidence permits, between the manuscript's production context, subsequent circulation and use, acquisition by a collector or intermediary, and eventual institutional custody. The term is used as an analytical heuristic rather than as a claim that complete provenance can always be reconstructed.

The colonial history of collecting further complicates the apparent neutrality of custody. Ghaddar and Caswell (2019) argue that decolonial archival praxis requires moving beyond the inherited assumptions of Western archival theory. Jeurgens and Karabinos (2020) show that digitisation can reproduce coloniality when new infrastructures simply restate inherited descriptive systems, but can support decolonising work when they allow multivocality, multiple agency, and multiple provenance. Their argument is particularly salient for Malay-Indonesian manuscripts in Europe: digital return is not equivalent to physical repatriation, yet well-designed access can reconnect dispersed materials with communities of language, scholarship, and heritage. (Agostinho 2019; Jeurgens and Karabinos 2020).

The problem can also be framed through the literature on displaced archives. Lowry (2019a) defines displaced archives as records removed from the context of their creation where ownership is disputed, while also arguing that archival displacement needs to be analysed beyond narrow inter-state claims. Jawi manuscripts in European libraries do not all meet that definition in a strict legal sense: many entered collections through purchases, gifts, scholarly exchange, diplomatic activity, or the collecting

practices of colonial officials, and the present study does not make restitution claims about particular items. Nevertheless, the displaced-archives literature is analytically useful because it foregrounds the consequences of geographic separation between records, the communities and knowledge environments from which they emerged, and the institutions that subsequently control description and access. Lowry's (2019b) related discussion of radical empathy and postcolonial records is especially relevant to digital description because it redirects attention from custody alone to relationships, affects, claims, and responsibilities that persist after records have moved.

Recent work in Archival Science further demonstrates that digital access can support forms of reconnection without erasing questions of authority. Jones and Taylor (2024) argue that access guides can be redesigned for changing archival landscapes and note that physical custody is not necessarily a prerequisite for intellectual management or community connection. Rayburn et al. (2024), studying a community-driven knowledge infrastructure for Indigenous heritage, show that aggregation and digital sharing generate collaborative, data, and systemic frictions rather than a frictionless technical solution. These studies concern different communities and historical circumstances from Malay-Indonesian manuscript heritage, and their concepts should not be transferred uncritically. Their relevance lies in a common archival problem: digitised cultural materials can be globally visible while the descriptive systems governing them remain institutionally local. This distinction helps position Jawi manuscripts not simply as digitisation successes, but as cases through which to examine how archival infrastructures distribute descriptive authority. (Jones and Taylor 2024; Rayburn et al. 2024).

Access is not resolved by scanning. Jaillant (2022) documents legal, technical, and organisational barriers that continue to shape access to digitised and born-digital archives. Conway's (2011) work on digital surrogate quality likewise warns against treating digitised content as self-evidently adequate for research. In manuscript studies, high-resolution imaging may make palaeographic analysis possible, but only if the user can discover the item, understand the metadata, navigate the image viewer, and identify how the digital surrogate relates to the physical object. (Jaillant 2022; Conway 2011; Hawkins 2022).

The ethical dimension of description also matters. Caswell and Cifor (2016) propose radical empathy as a framework that positions record creators, subjects, users, and communities within relationships of mutual responsibility. Flinn, Stevens and Shepherd (2009) and Flinn (2010) demonstrate the importance of community agency in archival representation. Applied to Jawi heritage, these perspectives suggest that repositories should not treat Malay-speaking and manuscript-source communities merely as downstream users. Their terminology, expertise, transliteration practices, and historical knowledge can form part of description itself. (Caswell and Cifor 2016; Flinn et al. 2009; Flinn 2010).

This article uses netnography as a method for examining those digital relations. Kozinets (2020) defines netnography as an adaptation of ethnographic inquiry to online cultural environments, while Hine (2015) emphasises that internet field sites are embedded, connected, and institutionally situated rather than separate virtual worlds. The present study is therefore not an ethnography of manuscript-reading

communities. It is an archival netnography of the institutional online environments through which manuscripts become discoverable, interpretable, and actionable for users. (Hine 2015; Kozinets 2020).

Method

Research design. The study employed a focused, non-participant archival netnography conducted through publicly accessible institutional websites and digital catalogues. The field site was defined as the online archival environment in which Jawi manuscript heritage is represented by European repositories. The approach follows the connective logic of internet ethnography (Hine 2015) and the purposive, context-sensitive principles of netnography (Kozinets 2020), but shifts the primary unit of observation from social interaction to archival interface and descriptive practice.

The unit of observation was the public-facing digital representation rather than the manuscript as a material object. This distinction is important. The study did not attempt codicological verification, palaeographic dating, or textual criticism from digital images, tasks that require specialist manuscript analysis and, in many cases, examination of the original object. Instead, it treated catalogue records, collection guides, project pages, repository datasets, and curatorial essays as archival acts: institutional statements that select and organise information about what an object is, where it came from, how it can be found, and how users are invited to understand it. The approach therefore sits between digital ethnography and document analysis. Netnographic sensibilities were used to follow pathways across interfaces and linked resources, while archival concepts supplied the categories through which those pathways were interpreted. (Hine 2015; Kozinets 2020; King 2024).

Fieldwork proceeded iteratively rather than through a single keyword search. Initial entry points included repository collection guides and searches for 'Malay', 'Jawi', known titles, and shelfmarks identified in published scholarship. From these entry points, the study followed institutional links to catalogue records, digitisation pages, research datasets, exhibitions, and curatorial commentary. Shelfmarks were used as anchors whenever possible because titles and romanisations vary across catalogues and publications. This strategy also reduced the risk of treating search-engine prominence as equivalent to archival significance. When a record was encountered through a general web search, its descriptive claims were checked against the repository's own catalogue or an identifiable scholarly publication before being incorporated into the analysis. (Hine 2015; Kozinets 2020; Gallop et al. 2015; Ricklefs et al. 2014).

Case selection. Four repositories were selected because they hold significant Malay manuscript materials, provide public digital discovery infrastructure, and represent different national and institutional histories: the British Library (United Kingdom), Cambridge University Library (United Kingdom), Leiden University Libraries (the Netherlands), and the Bibliothèque nationale de France (France). Selection was not intended to produce a census of all Jawi material in Europe. Rather, the cases form an information-rich comparative sample of institutions with established collections and visible digital traces. (British

Library n.d.; Cambridge University Library n.d.; Bibliothèque nationale de France / Catalogue collectif de France 2009, updated 2022; Ricklefs et al. 2014).

Corpus. The analysed corpus comprised collection-level descriptions, catalogue records, digitisation project pages, institutional curatorial essays, and selected scholarly publications that directly identify shelfmarks or collection histories. Core examples included British Library catalogue records such as Add MS 4828 and Add MS 12378; British Library documentation of the Malay manuscript digitisation programme; Cambridge collection guides and institutional material on the Erpenius manuscripts, including the early Malay Qur'anic commentary; Leiden Special Collections material concerning Malay manuscripts such as Or. 3286 and scholarship on Or. 7056; and the BnF/CCFr finding aid for the 131-volume Malay manuscript collection. Appendix A lists the principal digital artefacts used in the analysis. Access and factual claims were checked against the institutional or publisher pages available on 27 September 2026. (British Library 2026; British Library Research Repository 2025; Cambridge University Library n.d.; University of Cambridge 2024; Leiden University Libraries 2013; Harun 2013; Gallop et al. 2015).

Corpus construction prioritised evidence that could be independently revisited. A digital artefact was included when it met at least one of four criteria: it (a) represented a collection or item held by one of the four focal repositories; (b) documented the history of acquisition, cataloguing, or digitisation of those holdings; (c) provided structured metadata capable of being compared with another repository; or (d) offered scholarly re-identification or contextualisation tied to a stable shelfmark. General cultural commentary without a traceable relationship to a holding was excluded. Social-media posts were also excluded because the study's purpose was to analyse institutional archival representation rather than public reception. Appendix A lists the principal artefacts that structured the analysis; additional linked pages were used for contextual checking but were not treated as separate cases when they merely duplicated the same institutional description. (Hine 2015; Kozinets 2020; King 2024).

The corpus is deliberately purposive rather than statistically representative. The four repositories differ substantially in catalogue architecture, language, digitisation history, collection scale, and the degree to which item-level records are exposed to general web indexing. Those differences are part of the phenomenon under study, but they also make simple frequency comparison misleading. Accordingly, the analysis does not report percentages such as the proportion of records containing provenance fields across an entire repository. It instead identifies recurring descriptive patterns and contrasts them through traceable examples. The resulting claims are therefore analytical generalisations about the observed digital infrastructures, not an audit of every Jawi manuscript held in Europe. (Hine 2015; Kozinets 2020; Jaillant 2022).

Analytical procedure. Each artefact was read for six dimensions: (1) naming and discoverability, including language, script, title, transliteration, and keywords; (2) provenance and acquisition history; (3) geographic and community context; (4) digital access status and interface conditions; (5) relationships to other records, catalogues, or collections; and (6) evidence of descriptive revision or curator-mediated

recontextualisation. Notes were coded iteratively and compared across repositories. Rather than treating frequency as proof of significance, the analysis focused on recurring representational patterns and consequential contrasts between catalogue-level compression and contextual narrative. (King 2024; Hurley et al. 2024; Jaillant 2022; Hawkins 2022).

A two-stage coding procedure was used. In the first stage, descriptive observations were recorded under the six predefined dimensions, together with the repository, URL or DOI, shelfmark where available, and notes on the route by which the artefact was discovered. In the second stage, observations were compared across repositories and regrouped into higher-order themes. Four themes were prominent in the initial coding—acquisition-horizon provenance, uneven script terminology, the non-equivalence of digitisation and access, and the contextual role of curatorial narrative—while a fifth theme, cross-repository dispersal, emerged from repeatedly following the same text, collector, or descriptive problem across more than one institutional environment. The analysis therefore combined a concept-led framework with inductive refinement. (Hine 2015; Kozinets 2020).

Reflexivity was necessary because the researcher encounters these collections through the same search systems being analysed. Search ranking, interface language, broken links, institutional redesign, and prior knowledge of well-known manuscripts can all influence what becomes visible. The study addressed this limitation by keeping shelfmarks and persistent identifiers at the centre of source checking and by distinguishing three levels of claim: what a repository explicitly states; what published scholarship establishes about a manuscript or collection; and what the present comparative analysis infers from those descriptions. Where an institutional page and scholarship differed in emphasis, the difference was treated as data about representation rather than resolved by silently choosing one narrative. (Hine 2015; Kozinets 2020; King 2024).

Ethics and scope. No private accounts, closed groups, personal social-media data, or human-subject interactions were used. The study analysed public institutional descriptions and published scholarship. It did not evaluate the physical condition of manuscripts beyond what repositories reported, and it does not claim that online records fully represent internal cataloguing systems or curatorial knowledge. Because catalogue interfaces change, the study treats every online description as a versioned public representation rather than a final statement of archival truth. (Hine 2015; Kozinets 2020).

AI-assisted preparation. Generative AI was used during manuscript preparation to assist with organising source notes, language formulation, and document preparation. It was not treated as an author or as an evidentiary source. All archival claims, interpretations, and bibliographic entries were checked by the authors against the cited repository records, books, journal articles, and publisher pages, and the authors retain responsibility for the final scholarly content. This disclosure follows Springer Nature's current risk-based guidance on transparent, human-accountable AI use in manuscript preparation (Springer Nature 2026).

Findings

1. Provenance is visible, but often stops at the institution's acquisition horizon

The strongest records in the corpus demonstrate that European repositories can document custodial provenance with considerable specificity. The British Library record for Add MS 12378, *Tāj al-salāṭīn*, identifies the manuscript as Malay in Jawi script, dates it to the early nineteenth century, describes paper and binding features, and records that it was purchased from John Crawford in 1842. Such information is archival evidence in its own right: it makes acquisition history legible and allows the researcher to connect the object to a named collector and a wider colonial-era manuscript assemblage. (British Library 2026; Ricklefs et al. 2014).

Yet the acquisition horizon can also function as a descriptive boundary. A repository may describe how a manuscript entered a European collection more fully than how it circulated before that transfer. This asymmetry is understandable because institutional custody generates records, while earlier local circulation is often poorly documented. Nevertheless, when the acquisition event dominates the description, provenance can appear to begin with European possession. For Jawi manuscripts, a more relational description would distinguish at least three layers: production provenance (where, when, by whom, and for what textual community the manuscript was made), circulation provenance (owners, readers, courts, pesantren or religious networks, copyists, and intermediaries), and custodial provenance (collectors, purchases, donations, transfers, and institutional custody). (Hurley et al. 2024; Anderson 2024; Namhila 2016).

The archival consequence is that provenance appears asymmetrical: the later European phase is often more granular because it is supported by accession registers, collector papers, institutional catalogues, and shelfmark histories, whereas earlier Southeast Asian circulation may survive only in colophons, seals, marginalia, textual references, or scholarship external to the catalogue. This asymmetry should not automatically be read as negligence. It is partly an effect of the evidence available to the repository and partly an inheritance of descriptive traditions that privileged acquisition and custody. The problem arises when the better-documented institutional phase is allowed to stand for the whole provenance narrative. A user may then encounter a manuscript primarily as an item 'from' a named European collector even when the object had a much longer social life before collection. Layered provenance offers a way to mark that difference without claiming evidence that does not exist. (Hurley et al. 2024; Anderson 2024; King 2024).

Cambridge's Erpenius materials illustrate why such layering matters. Institutional histories explain that Erpenius's collection reached Cambridge in the seventeenth century after earlier movement through Leiden and English aristocratic ownership. Modern scholarship further links specific Malay manuscripts to Aceh and to early European engagement with Malay textual culture. The meaningful archival story is therefore neither solely 'Cambridge provenance' nor solely 'Acehnese origin', but a transregional chain of production, encounter, collecting, scholarly reuse, and institutional custody. (Cambridge University Library n.d.; University of Cambridge 2024; Anderson 2024; Hurley et al. 2024).

2. Jawi becomes legible through uneven metadata vocabularies

Across the corpus, Jawi is not represented through one stable descriptive formula. Some records name the language as Malay and the script as Arabic; others use the phrase 'in Malay in Jawi script'; older scholarship may use 'Arabic-Malay' or historically specific labels; and collection-level descriptions sometimes foreground literary genre rather than script. These differences matter for discovery. A user searching 'Jawi' may miss records indexed only under 'Malay' plus 'Arabic' script, while a user searching 'Arabic manuscript' may retrieve items whose language is Malay rather than Arabic. (Gallop et al. 2015; Hawkins 2022).

The British Library's catalogue is comparatively explicit in many records. Add MS 4828, for example, is described as a Malay letter in Jawi script and is linked to identifiable political actors and Bengkulu. The BnF collection-level finding aid, by contrast, offers a valuable overview of 131 Malay manuscript volumes and genre composition but functions at a coarser level. Leiden scholarship shows another challenge: key Jawi manuscripts may be most discoverable through shelfmarks and specialist literature, requiring users to know collection histories or established catalogue conventions. (British Library 2026; Bibliothèque nationale de France / Catalogue collectif de France 2009, updated 2022; Harun 2013; Gallop et al. 2015).

The problem is not that one vocabulary is inherently correct and others are wrong. Archival description serves different functions at different levels. The issue is interoperability between vocabularies. For dispersed manuscript heritage, discovery improves when repositories expose multiple controlled access points: language (Malay), script (Jawi/Arabic-derived), alternative script term (Arabic script), geographic association (Aceh, Riau, Malay Peninsula, Sumatra, etc., where supported), genre, personal names, and standard shelfmarks. Transliteration fields should sit alongside original-script titles where known, rather than replacing them. (Gallop et al. 2015; Hawkins 2022; Jaillant 2022).

This unevenness has practical consequences for discovery. A user who searches 'Jawi' may retrieve only records in which the term has been explicitly added, while a user who searches 'Malay manuscript' may retrieve a broader set that includes Jawi materials but gives no direct indication of script. Conversely, searching an Arabic-script title may fail where a catalogue exposes only a romanised form. The same manuscript can therefore occupy several partially overlapping discovery environments: language, script, title, genre, personal name, place, and shelfmark. For scholars already familiar with a collection, the shelfmark can bridge those environments. For students, community researchers, or first-time users in Southeast Asia, however, the absence of cross-indexed terminology can become a significant access barrier. Description here is not only about accuracy; it determines which conceptual doorways into the collection are available. (Gallop et al. 2015; Hawkins 2022; Jaillant 2022).

3. Digitisation produces access, but not equivalence

The corpus strongly supports the distinction between availability and usability. The British Library's Malay digitisation programme demonstrates the scale of institutional commitment to online access, and its 2025 research dataset provides structured metadata for hundreds of digitised Indonesian and Malay-world manuscripts. Such initiatives dramatically reduce geographic barriers and permit researchers in

Southeast Asia to identify and, when platforms are available, inspect materials without travelling to Europe. (British Library Research Repository 2025; Conway 2011; Jaillant 2022).

However, digital access is contingent. The British Library catalogue currently notes that some legacy digitised-manuscript images are unavailable, a consequence of the institution's continuing recovery from its cyber-attack. This is a reminder that digitised access is infrastructural, not permanent by default. A catalogue may survive while an image service is temporarily unavailable; a persistent shelfmark may remain stable while a viewer URL changes. Jaillant's (2022) broader argument that digital archives remain constrained by technical and organisational barriers is therefore directly visible in the Jawi case. (British Library 2026; Jaillant 2022).

Digitisation is also not equivalent to descriptive repair. A high-resolution image can reproduce the manuscript page while leaving historical labels, sparse provenance, or monolingual metadata untouched. Conversely, a richly contextual catalogue record without images may offer strong intellectual access but limited palaeographic access. The most useful digital archival environment is therefore composite: stable identifiers, descriptive context, high-quality images, rights information, download or citation guidance, and links among related manuscripts and scholarship. (Conway 2011; Jeurgens and Karabinos 2020; Hawkins 2022).

The distinction becomes clearer if digital access is separated into at least four layers: image availability, metadata availability, technical retrievability, and interpretive usability. A manuscript may score highly on one layer and poorly on another. Images may exist but be temporarily unavailable; metadata may be searchable but too sparse to support contextual interpretation; a record may be technically public but difficult to retrieve without an exact shelfmark; or an exhibition may provide rich interpretation while omitting stable links to the full surrogate. This layered view avoids an all-or-nothing language of 'digitised' versus 'not digitised'. It also clarifies why long-term digital stewardship must preserve relationships among image files, identifiers, descriptions, and contextual resources, not merely the image sequence. (Conway 2011; Jaillant 2022; Hawkins 2022; Bak 2024).

4. Curatorial narrative can reopen compressed archival contexts

Institutional blogs, exhibitions, and collection essays perform a function that conventional catalogue fields often cannot. Leiden's Special Collections discussion of notes and scribbles in Malay manuscripts uses seemingly minor traces to reconstruct relationships between the linguist Herman Neubronner van der Tuuk and a Malay copyist, Abdul Wahid. Cambridge's recent public writing on the Erpenius collection similarly revisits manuscripts that were misidentified or poorly understood in early European cataloguing and explains how new linguistic and codicological research changes their interpretation. (Leiden University Libraries 2013; University of Cambridge 2024; King 2024).

These narratives demonstrate that description is revisable knowledge. A legacy catalogue term is not merely an error to be deleted; it may also be evidence of how a manuscript was perceived by earlier custodians. The archival challenge is to preserve that evidentiary history while preventing outdated or

misleading labels from controlling present-day discovery. King (2024) calls attention to precisely this issue through the concept of archival meta-metadata: users benefit from knowing that finding aids have histories and that description changes over time.

Curatorial narrative also exposes a productive tension between catalogue brevity and interpretive depth. Catalogue records must remain structured enough for discovery and maintenance, yet manuscripts often require explanations that exceed a single note field. Essays and exhibitions can discuss scribes, collectors, annotations, historical actors, or new identifications in prose, but those insights are easily detached from the catalogue if reciprocal links are weak. The most useful model is therefore not to replace the catalogue with narrative, but to create an evidential loop: the catalogue provides stable identity and core metadata; the curatorial essay provides interpretation; and the essay's findings, when sufficiently established, feed back into revised or enriched description. In this way, public scholarship becomes part of archival maintenance rather than a parallel layer that users must rediscover independently. (King 2024; Hawkins 2022; Jones and Taylor 2024).

For Jawi materials, a layered interface could therefore show the current preferred description, previous descriptions or classifications where historically significant, the date and agent of revision, and a short rationale. This approach avoids two unsatisfactory extremes: freezing inherited colonial description as neutral fact, or silently overwriting it and thereby erasing evidence of the institution's own descriptive history. (King 2024; Jeurgens and Karabinos 2020).

5. Dispersal turns discoverability into a cross-repository problem

The corpus also shows that the archival identity of Jawi manuscript heritage is distributed across institutions. A single research question may require movement between an item record in one repository, a historical catalogue published elsewhere, a scholarly article that identifies an alternative title, and a digitisation or exhibition page that supplies contextual information. This distribution is not simply a consequence of digitisation; it reflects the longer history through which Malay-Indonesian manuscripts entered multiple European collections and were catalogued under different national, linguistic, and disciplinary traditions. Digital platforms make those dispersed traces easier to reach, but they do not automatically connect them. (Lowry 2019a; Ricklefs et al. 2014; Gallop et al. 2015).

Shelfmarks provide a partial solution because they function as durable local identifiers, yet local identifiers are not themselves a trans-institutional discovery system. The same text can survive in several copies under unrelated shelfmarks, while an individual codex may be discussed under variant titles or older catalogue labels. Cross-repository scholarship such as Gallop's Jawi sourcebook and the revised catalogue of Indonesian manuscripts in Great Britain demonstrates the value of scholarly work that normalises or cross-references such variation. The archival challenge is to translate that scholarly connective work into machine-actionable and user-visible links without collapsing meaningful differences between copies, descriptions, and institutional histories. (Gallop et al. 2015; Ricklefs et al. 2014; Hawkins 2022).

For this reason, dispersed heritage requires what may be called relational discoverability. The goal is not a single universal catalogue that absorbs every repository, but a network in which persistent identifiers, authority data, related-resource fields, and cited scholarship allow users to traverse institutional boundaries. Relational discoverability would be especially valuable for Jawi materials because a user may approach them through Aceh, Riau, Bengkulu, Islamic law, court correspondence, a personal name, a manuscript title, or the script itself. When these access points remain isolated inside repository-specific metadata, the geographic dispersal of the physical collection is reproduced as informational dispersal online. Digital archival design can mitigate that fragmentation even where physical custody remains unchanged. (Hawkins 2022; Jones and Taylor 2024; Rayburn et al. 2024).

Discussion

The findings show that digitised Jawi manuscripts in European repositories should be understood as archival interfaces rather than only as manuscript images. The interface is where provenance is assembled, scripts are named, geographic relationships are indexed, user pathways are enabled or blocked, and earlier descriptions are either reproduced or revised. In this sense, the digital surrogate participates in recordkeeping: it does not alter the historical manuscript, but it alters the contemporary conditions under which the manuscript can be discovered, cited, interpreted, and connected to other records. (Conway 2011; Agostinho 2019; Jeurgens and Karabinos 2020; Bak 2024).

The analysis therefore extends the concept of access from opening a digital object to constructing a pathway into archival knowledge. This is consistent with Jones and Taylor's (2024) argument that archival guides can mediate changing landscapes of custody, sovereignty, and knowledge. For Jawi manuscripts, the pathway is unusually dependent on translation between descriptive systems: Arabic-derived script and Romanised search; local and European place names; manuscript titles and supplied catalogue titles; older philological terminology and contemporary archival language. A platform that offers an image but does not support these translations can widen nominal access while leaving expert knowledge as the practical prerequisite for discovery. The implication is not that every catalogue must become a full research edition, but that digital access should be evaluated by the relationships it enables users to follow. (Jones and Taylor 2024; Hawkins 2022; Gallop et al. 2015).

This has implications for provenance. Traditional custodial description remains essential, especially because acquisition records can reveal collectors, dates, purchases, and institutional transfers. But a manuscript's archival context exceeds its accession story. The records-continuum emphasis on multiple contexts and the decolonial critique of singular provenance both suggest a more expansive model. For a Jawi manuscript copied in Aceh, acquired by a European intermediary, transferred through a private library, and later catalogued in London, Cambridge, Leiden, or Paris, each stage is evidentially significant. None should erase the others. (Hurley et al. 2024; Anderson 2024; Ghaddar and Caswell 2019).

The study also suggests that 'digital return' should be treated cautiously. Online access can restore practical access to communities and scholars in the regions where manuscripts originated, but it does

not by itself rebalance authority. If search terms remain monolingual, if local place names are absent, if script terminology is inconsistent, or if acquisition histories are detached from colonial contexts, the user may receive the image while still encountering an epistemic structure centred elsewhere. Jeurgens and Karabinos (2020) are therefore correct to frame digitisation as a new archival infrastructure whose design choices matter. (Jeurgens and Karabinos 2020; Agostinho 2019).

The displaced-archives literature sharpens this point while also setting a limit on the argument. Lowry (2019a) emphasises that displacement includes questions of contested ownership and unresolved archival claims. The present study does not establish that the manuscripts in its corpus are legally disputed archives, nor does it infer present-day community positions from historical provenance. Its claim is narrower: geographic displacement and colonial collecting histories create descriptive responsibilities even when ownership is not formally contested. Digital access can support reconnection, but the terms of that reconnection are shaped by repositories through metadata, interface design, rights statements, language choice, and decisions about whose scholarship is linked. (Lowry 2019a, 2019b).

This distinction also guards against treating digitisation as a substitute for decolonial archival work. Jeurgens and Karabinos (2020) describe a paradox in projects that seek to decolonise colonial archives through digitisation while retaining archival systems formed through colonial logics. A similar caution applies here. Adding a scan to a platform can increase visibility without changing the categories through which an object is described. At the same time, deleting every inherited term would erase evidence of earlier archival practice. The more productive strategy is transparent revision: retain significant legacy description as history, identify the preferred contemporary description, document when and why changes were made, and connect users to the scholarship or community knowledge that informed the revision. King's (2024) argument for archival meta-metadata offers a concrete conceptual basis for this approach. (Jeurgens and Karabinos 2020; King 2024).

A reciprocal-description model offers a pragmatic path. 'Reciprocal' does not mean that repositories surrender professional responsibility or that every description becomes an open wiki. It means that descriptive systems deliberately create space for knowledge that flows back toward the repository from source-language scholarship, communities of expertise, and related collections. The model proposed here has six components: preserve legacy description as versioned evidence; add contemporary preferred terminology; record layered provenance; support original-script and transliterated access points; link dispersed related materials; and make contribution/revision mechanisms transparent. (King 2024; Hurley et al. 2024; Rayburn et al. 2024).

Reciprocity should also be understood infrastructurally, not only ethically. Rayburn et al. (2024) show that community-driven knowledge infrastructures encounter friction at collaborative, data, and systemic levels. For a Jawi manuscript network, comparable frictions would include disagreement over preferred romanisation, incompatible metadata schemas, uneven technical capacity, uncertainty over contributor authority, and the labour required to moderate additions over time. A credible reciprocal model therefore

needs governance: defined contribution routes, attribution, evidence requirements, versioning, and institutional responsibility for final publication. Without these mechanisms, invitations to 'community contribution' risk becoming symbolic gestures or short-lived project features rather than durable archival practice. (Rayburn et al. 2024).

Such a model is particularly appropriate for Jawi because the script itself is a discovery problem. Search infrastructures built around Romanised metadata may separate users from materials whose titles, names, and textual identities circulate in Arabic-derived script. Providing original-script fields, transliteration variants, and mapped authority terms would improve both scholarly retrieval and public accessibility. These changes need not wait for complete re-cataloguing. They can be implemented incrementally at collection, series, and item level. (Gallop et al. 2015; Hawkins 2022).

There are also implications for collaboration. European institutions already collaborate internationally on digitisation and scholarship. Future projects could extend that collaboration into description by involving Southeast Asian manuscript specialists, librarians, philologists, historians, and community-based knowledge holders in targeted metadata enrichment. This aligns with the participatory archival arguments developed by Flinn and colleagues and with Caswell and Cifor's emphasis on relational responsibility. The objective is not to replace archival standards with unstructured commentary but to create accountable pathways through which additional context can be attached to stable identifiers. (Flinn et al. 2009; Flinn 2010; Caswell and Cifor 2016; Rayburn et al. 2024).

Such collaboration could begin modestly. Repositories do not need to redesign entire catalogues before improving a Jawi collection. A pilot could select a bounded set of manuscripts with strong existing scholarship and add original-script title fields, variant romanisations, geographic authority terms, links to related copies, acquisition-history notes, and citations to Southeast Asian scholarship. Changes could be documented as a revision layer and evaluated through user testing with researchers and students in Indonesia and neighbouring Malay-reading communities. The outcome would provide evidence about which descriptive interventions most improve discovery and interpretation. This incremental model is more feasible than treating decolonial access as a one-time mass-recataloguing exercise. (King 2024; Jones and Taylor 2024; Rayburn et al. 2024; Gallop et al. 2015).

Finally, archival netnography proved useful for analysing public-facing archival infrastructures. It makes visible how researchers actually encounter institutional knowledge online: through search interfaces, snippets, collection guides, digitisation portals, blogs, and publisher pages. At the same time, the method has limits. It cannot reveal internal cataloguing decisions, staff workflows, unpublished metadata, or user analytics unless those are separately studied. Interviews with archivists and users would therefore be a valuable next phase, especially to test whether the descriptive frictions identified here correspond to repository constraints and user experience. (Hine 2015; Kozinets 2020; Jaillant 2022).

A reciprocal-description framework for Jawi manuscript heritage

The comparative findings can be translated into six repository practices. The framework below is intentionally pragmatic: it does not prescribe a new universal standard, but identifies interventions that can sit alongside existing cataloguing rules and collection-management systems. Its central premise is that dispersed manuscript heritage benefits when repositories make provenance more layered, terminology more interoperable, descriptive change more transparent, and external knowledge easier to connect to stable archival identities. (King 2024; Hurley et al. 2024; Jones and Taylor 2024; Rayburn et al. 2024).

Principle	Repository practice
Layer provenance	Separate production, circulation, acquisition, and institutional-custody provenance where evidence permits; do not let the accession event function as the manuscript's origin story.
Use multi-entry script and language access	Index Malay as language, Jawi as script tradition/term, Arabic script where required by standards, and known title/transliteration variants. Support original-script fields when reliable forms are available.
Expose descriptive history	Retain significant legacy terms in version history or notes while presenting an updated preferred description and the reason/date for revision.
Link dispersed collections	Use persistent shelfmarks and authority identifiers to connect manuscripts, copies, related correspondence, catalogues, digitised surrogates, and scholarship across institutions.
Design for Southeast Asian users	Provide multilingual collection introductions where feasible, stable citation instructions, image-access status, rights information, and low-friction routes from catalogue record to surrogate.
Create reciprocal enrichment workflows	Invite structured contributions from qualified external scholars and source-community partners, with attribution, moderation, and preservation of revision history.

Source basis for the framework: the six practices synthesise scholarship on descriptive revision and positionality, expanded provenance, digital/linked archival access, decolonial praxis, and community-driven knowledge infrastructures (King 2024; Hurley et al. 2024; Hawkins 2022; Ghaddar and Caswell 2019; Jones and Taylor 2024; Rayburn et al. 2024).

Conclusion

European repositories preserve indispensable parts of the Jawi documentary heritage of the Malay-Indonesian world. Their digital platforms have widened access, but they also demonstrate that access is an archival construction rather than a simple consequence of scanning. Catalogue fields, shelfmarks, script labels, acquisition histories, image services, and curatorial narratives collectively determine what users can know about a manuscript before they ever read its text. (Conway 2011; Jaillant 2022; Jeurgens and Karabinos 2020).

The article also demonstrates why manuscript digitisation should be studied as an archival process in its own right. Once a manuscript enters a digital platform, users encounter a constructed assemblage of image, metadata, identifier, interface, rights statement, contextual narrative, and external links. Each component influences what can be known and by whom. Evaluating digital manuscript heritage therefore requires attention not only to preservation quality but also to descriptive history, searchability, cross-repository relations, and the capacity for knowledge to be revised. (Bak 2024; Hawkins 2022; King 2024).

The archival netnography presented here identifies a recurring tension between strong institutional custody metadata and thinner representation of earlier social provenance; between discoverability through standardised fields and the linguistic complexity of Jawi; and between the promise of digitisation and the fragility or incompleteness of digital infrastructure. At the same time, institutional curatorial work shows that these descriptions are not fixed. Re-examination can recover misidentified languages, forgotten copyists, acquisition pathways, and regional relationships. (King 2024; Hurley et al. 2024; Gallop et al. 2015; Jeurgens and Karabinos 2020).

For Archival Science, the key implication is that the digital stewardship of displaced manuscript heritage should combine preservation with descriptive reciprocity. Retaining legacy metadata, documenting revisions, adding layered provenance, supporting Jawi-aware discovery, and enabling accountable contextual contributions can make European-held manuscripts more intelligible to the communities and scholarly traditions to which they remain historically connected. Digitisation then becomes more than remote viewing: it becomes an opportunity to redesign archival relationships without pretending that the histories of custody and displacement can be erased. (Ghaddar and Caswell 2019; Lowry 2019a, 2019b; Jones and Taylor 2024; King 2024).

Declarations

Author Contribution

ASB led the conceptualisation, research design, archival data collection, data analysis, interpretation of findings, and preparation of the original manuscript. BS contributed to conceptual development, methodological refinement, interpretation of findings, and critical revision of the manuscript. IIP contributed to methodological review, literature contextualisation, interpretation of findings, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript and agree to be accountable for the integrity of the work.

Acknowledgement

The authors gratefully acknowledge the Institute for Research and Community Service (LPPM), Universitas Samudra, for its institutional support. The authors also acknowledge the use of ChatGPT, a generative artificial intelligence tool developed by OpenAI, solely to assist with language refinement,

structural organisation, and readability during manuscript preparation. ChatGPT was not used as a source of scholarly evidence. All archival records, references, interpretations, analyses, citations, and conclusions were independently reviewed and verified by the authors, who take full responsibility for the content and integrity of the manuscript.

Data Availability

The research data used in this study consist of publicly accessible archival descriptions, digital manuscript records, catalogue metadata, repository interfaces, and published scholarly sources from institutional repositories and publisher platforms. The principal digital materials analysed are identified in Appendix A of the study, together with their repository or source information. No new restricted, confidential, or personally identifiable dataset was generated. All materials were publicly accessible through the cited institutional websites, catalogues, repositories, and publisher platforms at the time of data collection and analysis.

References

1. Agostinho D (2019) Archival encounters: rethinking access and care in digital colonial archives. *Arch Sci* 19:141–165. <https://doi.org/10.1007/s10502-019-09312-0>
2. Anderson BG (2024) Kindred contexts: archives, archaeology, and the concept of provenance. *Arch Sci* 24:761–781. <https://doi.org/10.1007/s10502-024-09459-5>
3. Bak G (2024) Digital provenance. *Arch Sci* 24:847–869. <https://doi.org/10.1007/s10502-024-09462-w>
4. Bibliothèque nationale de France / Catalogue collectif de France (2009) updated 2022) *Manuscripts malais*. Département des Manuscrits, Bibliothèque nationale de France, Paris. <https://ccfr.bnf.fr/portailccfr/ark:/16871/006FRFONDS-751041006-1013>. Accessed 27 Sep 2026
5. Borgerud C, Borglund E (2020) Open research data, an archival challenge? *Arch Sci* 20:279–302. <https://doi.org/10.1007/s10502-020-09330-3>
6. British Library (2026) Archives and Manuscripts Catalogue: Add MS 4828, Malay letter from Datuk Raja Kuasa to Richard Farmer of Bengkulu, 1718; and Add MS 12378, *Tāj al-salāṭīn*. <https://searcharchives.bl.uk/catalog/032-002110351>; <https://searcharchives.bl.uk/catalog/040-002042142>. Accessed 27 Sep 2026
7. British Library (n.d.) Successful projects: Malay manuscripts. British Library. <https://support.bl.uk/Page/Malay-manuscripts>. Accessed 27 Sep 2026
8. British Library Research Repository (2025) Datasets of digitised Malay and Indonesian manuscripts in the British Library. <https://doi.org/10.23636/ghz6-vt54>
9. Cambridge University Library (n.d) Asian & Near and Middle Eastern manuscripts. Cambridge University Library. <https://www.lib.cam.ac.uk/collections/departments/manuscripts-university-archives/subject-guides/asian-near-and-middle-eastern>. Accessed 27 Sep 2026

10. Cambridge University Library (n d.) The seventeenth century. Cambridge University Library historical sketch. <https://www.lib.cam.ac.uk/about-library/historical-sketch/seventeenth-century>. Accessed 27 Sep 2026
11. Caswell M, Cifor M (2016) From human rights to feminist ethics: radical empathy in the archives. *Archivaria* 81:23–43. <https://archivaria.ca/index.php/archivaria/article/view/13557>
12. Conway P (2011) Archival quality and long-term preservation: a research framework for validating the usefulness of digital surrogates. *Arch Sci* 11:293–309. <https://doi.org/10.1007/s10502-011-9155-0>
13. Flinn A (2010) Independent community archives and community-generated content: 'writing, saving and sharing our histories'. *Convergence* 16(1):39–51. <https://doi.org/10.1177/1354856509347707>
14. Flinn A, Stevens M, Shepherd E (2009) Whose memories, whose archives? Independent community archives, autonomy and the mainstream. *Arch Sci* 9:71–86. <https://doi.org/10.1007/s10502-009-9105-2>
15. Gallop AT (2021) Who reads digitised Malay manuscripts? British Library Asian and African Studies blog (archived in British Library Research Repository). <https://bl.iro.bl.uk/entities/publication/60c66480-94fc-4e2f-b783-05743fb7ce62>. Accessed 27 Sep 2026
16. Gallop AT, Mamat WAW, Akbar A, Braginsky V et al (2015) A Jawi sourcebook for the study of Malay palaeography and orthography. *Indonesia Malay World* 43(125):13–171. <https://doi.org/10.1080/13639811.2015.1008253>
17. Ghaddar JJ, Caswell M (2019) To go beyond': towards a decolonial archival praxis. *Arch Sci* 19:71–85. <https://doi.org/10.1007/s10502-019-09311-1>
18. Harun J (2013) Mir'at al-Tullab by Syekh Abdul Rauf Singkel: a preliminary study of manuscripts kept in the Special Collections, Leiden University Library. *Malay Lit* 26(2):119–138. [https://doi.org/10.37052/ml.26\(2\)no1](https://doi.org/10.37052/ml.26(2)no1)
19. Hawkins A (2022) Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. *Arch Sci* 22:319–344. <https://doi.org/10.1007/s10502-021-09381-0>
20. Hine C (2015) *Ethnography for the Internet: embedded, embodied and everyday*. Bloomsbury Academic, London
21. Hurley C, McKemmish S, Reed B et al (2024) The power of provenance in the records continuum. *Arch Sci* 24:825–845. <https://doi.org/10.1007/s10502-024-09463-9>
22. Jaillant L (2022) How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. *Arch Sci* 22:417–436. <https://doi.org/10.1007/s10502-022-09390-7>
23. Jeurgens C, Karabinos M (2020) Paradoxes of curating colonial memory. *Arch Sci* 20:199–220. <https://doi.org/10.1007/s10502-020-09334-z>
24. Jones M, Taylor R (2024) Beyond access: (re)designing archival guides for changing landscapes. *Arch Sci* 24:143–166. <https://doi.org/10.1007/s10502-024-09441-1>

25. King OC (2024) Archival meta-metadata: revision history and positionality of finding aids. Arch Sci 24:509–529. <https://doi.org/10.1007/s10502-024-09443-z>
26. Kozinets RV (2020) Netnography: the essential guide to qualitative social media research, 3rd edn. SAGE, London
27. Leiden University Libraries (2013) Short notes and scribbles in Malay manuscripts. Leiden Special Collections Blog. <https://www.leidenspecialcollectionsblog.nl/articles/short-notes-and-scribbles-in-malay-manuscripts>. Accessed 27 Sep 2026
28. Lowry J (2019a) Displaced archives': proposing a research agenda. Arch Sci 19:349–358. <https://doi.org/10.1007/s10502-019-09326-8>
29. Lowry J (2019b) Radical empathy, the imaginary and affect in (post)colonial records: how to break out of international stalemates on displaced archives. Arch Sci 19:185–203. <https://doi.org/10.1007/s10502-019-09305-z>
30. Namhila EN (2016) Content and use of colonial archives: an under-researched issue. Arch Sci 16:111–123. <https://doi.org/10.1007/s10502-014-9234-0>
31. Rayburn AJ, Punzalan RL, Thomer AK (2024) Persisting through friction: growing a community driven knowledge infrastructure. Arch Sci 24:61–82. <https://doi.org/10.1007/s10502-023-09427-5>
32. Ricklefs MC, Voorhoeve P, Gallop AT (2014) Indonesian manuscripts in Great Britain: a catalogue of manuscripts in Indonesian languages in British public collections. New edn with addenda et corrigenda. École française d'Extrême-Orient. Perpustakaan Nasional Republik Indonesia, Yayasan Pustaka Obor Indonesia, Jakarta
33. Springer Nature (2026) AI use in manuscript preparation. Springer Nature editorial policies. <https://www.springernature.com/gp/policies/editorial-policies/ai-manuscript-preparation>. Accessed 27 Sep 2026
34. University of Cambridge (2024) Endless Stories at Cambridge University Library: ancient manuscripts on display for the first time. <https://www.cam.ac.uk/stories/endless-stories-rare-manuscripts>. Accessed 27 Sep 2026

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixA.docx](#)