

Decolonizing the Digital Manuscript: Local Islamic Archives, Codicological Materiality, and the Remapping of Malay-Islamic Intellectual Worlds

Dzul Iskandar^{1✉}, Ridhwan Sabri¹

¹Department of Malay Literature, Academy of Malay Studies, Universiti Malaya, Kuala Lumpur, Malaysia

²Department of Islamic History and Civilization, Academy of Islamic Studies, Universiti Malaya, Kuala Lumpur, Malaysia

✉ dzuliskandardzul@gmail.com

📄 [10.65840/jllcd.v2i2.57](https://doi.org/10.65840/jllcd.v2i2.57)

Article Info

Submitted:

06/22/2025

Revised:

07/12/2025

Accepted:

08/15/2025

Available Online:

08/21/2025

Abstract. This article examines how digitisation changes the intellectual visibility of Islamic and Malay manuscript traditions in Southeast Asia. It analyses a curated dataset of 99 records covering manuscript collections, digitisation projects, literary works, oral traditions, inscriptions, and institutional archives from Brunei Darussalam, Indonesia, and selected transregional sites. Instead of treating digitisation as a neutral technical act, the study reads the dataset as an archival ecology shaped by custody, material form, language, institutional power, and scholarly access. The findings show that the corpus is dominated by Brunei state-cultural and literary collections (55 records) and Indonesian local/community digitisation clusters (35 records), with smaller but analytically important entries from Vietnam, Myanmar, Thailand, Cambodia, Laos, and transregional Malay collections. Four evidence fields structure the dataset: literary/oral worlds (41 records), Islamic learning (28), archive and polity (18), and language/knowledge infrastructure (12). The article argues that local digitisation remaps Islamic Southeast Asian writing traditions by shifting attention from colonial collection bias toward surau, pesantren, private, royal, and community-held materials. Its contribution is a model of archival visibility that links codicological materiality, local custody, digital access, and decolonial manuscript interpretation.

Keywords: Islamic manuscripts; digital humanities; Malay manuscripts; Southeast Asia; codicology; archival visibility; decolonial philology



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Manuscript digitisation has become one of the most consequential transformations in the study of Southeast Asian intellectual history. It does more than preserve fragile objects or provide online images for distant readers. It changes the archive itself by making certain custodians, scripts, materials, religious traditions, and local knowledge systems visible in new ways. For a long time, the study of Malay and Islamic writing traditions was mediated through collections held in colonial libraries, national archives, and elite courtly institutions. Those collections remain indispensable, but they do not exhaust the region's written culture. Recent digitisation initiatives have brought forward another archive: manuscripts kept in surau, pesantren, private houses, royal families, local mosques, ritual communities, and regional institutions whose materials are often less canonical but intellectually

decisive. This article takes that shift as its central problem. It asks how a dataset of manuscript collections and projects from Southeast Asia can reveal not only what survives, but also what digitisation makes newly legible as Islamic, Malay, literary, codicological, and regional evidence (Gallop 2020; Gallop and Fathurahman 2022; van der Meij and van der Putten 2023; DREAMSEA 2026; Endangered Archives Programme 2025).

The problem is especially significant for Islamic Southeast Asia because colonial collection histories often privileged what collectors found legible, valuable, or administratively useful. Court chronicles, legal codes, letters, genealogies, and literary romances entered metropolitan collections because they served philological, diplomatic, antiquarian, or imperial interests. By contrast, local Islamic texts such as khutbah, mawlid, fiqh manuals, Sufi treatises, Qur'anic excerpts, pesantren teaching manuscripts, village inventories, and talismanic or ritual objects were more likely to remain embedded in community custodianship. Digitisation projects such as the Endangered Archives Programme and DREAMSEA have not simply added new images to the global archive; they have altered the evidentiary base from which scholars infer the intellectual geography of the region. Gallop's argument that newly accessible community-held collections can remap the writing traditions of Islamic Southeast Asia is therefore methodological as much as historical: it requires scholars to treat digital access, local custody, and codicological form as part of interpretation rather than as background infrastructure (Gallop 2020; Peacock 2022; van der Meij 2022; Peacock 2024; Mursyid, Aizan, and Faisal 2026).

The dataset analysed here is particularly useful because it does not resemble a clean catalogue of one archive. It is a mixed workbook of 99 records covering Brunei literary and institutional materials, Indonesian EAP and DREAMSEA projects, British Library Malay manuscripts, Southeast Asian Buddhist and epigraphic entries, and contemporary language and periodical infrastructures. At first sight, such heterogeneity may appear methodologically inconvenient. This article treats it differently: the unevenness is the evidence. The dataset reveals a layered ecology in which manuscript traditions travel across state institutions, community collections, foreign archives, academic projects, online repositories, print publications, and oral documentation. Brunei Darussalam dominates the corpus numerically, but Indonesia supplies the densest cluster of locally digitised Islamic materials. The result is not a representative sample of all Southeast Asian manuscripts, and it is not presented as such. It is an analytically productive snapshot of how digital and institutional visibility is currently being assembled around Malay-Islamic written culture.

The article contributes to three debates. First, it contributes to digital humanities by insisting that digitisation should be studied as a cultural and institutional process, not merely as preservation technology. Digital surrogates create access, but they also produce metadata regimes, viewing conditions, research hierarchies, and new asymmetries between photographed objects and the communities that historically sustained them (Pierazzo 2015; Terras 2016; Drucker 2021; Poleg 2022; Spina 2023; Risam 2018). Second, it contributes to manuscript studies by bringing codicological materiality into conversation with platform

visibility. Paper, palm leaf, horn, bamboo, bark, scrolls, Jawi script, Arabic interlinear translation, lithography, and oral recording each carry different evidentiary affordances and risks (Déroche 2006; Gacek 2009; Blair 2006; Gallop 2020). Third, it contributes to Southeast Asian Islamic studies by arguing that the intellectual landscape of the region is polycentric. It cannot be reduced to metropolitan libraries, national canons, or single-language textual histories. Malay, Arabic, Jawi, Wolio, Acehnese, Javanese, Cham, Bugis, Makassarese, Ulu, Pali, and other script-language constellations coexist in a region where Islamization, court culture, trade, pedagogy, and local ritual authority intersect (Johns 1993; Ricci 2011; Laffan 2011; Fathurahman 2012; Peacock 2024).

The research questions are therefore as follows. First, what does the dataset reveal about the distribution of custody, language, materiality, chronology, and thematic domains in Southeast Asian manuscript collections and digitisation projects? Second, how does local digitisation challenge inherited collection biases by making Islamic learning and community-held materials more visible? Third, what model of archival visibility can be developed from the interaction between codicological evidence, custodial regimes, and digital access? The article answers these questions through a mixed qualitative design: corpus audit, category normalization, evidence stratigraphy, close reading of selected records, and interpretive modelling. The aim is not to produce another inventory of manuscripts. The aim is to show how a structured reading of a relatively modest dataset can generate a stronger theory of digitised manuscript culture in maritime Southeast Asia.

2. Method

This study uses a qualitative-dominant documentary corpus design. The primary dataset is a workbook titled *Daftar Koleksi dan Proyek Manuskrip Islam di Asia Tenggara*, consisting of 99 records and eight fields: collection or project title, geographical location, manuscript content or genre, language used, owner or custodian, estimated period, digitisation status, and source index. A second sheet contains 22 source references. The unit of analysis is not the manuscript tradition in total, but the record as a trace of archival visibility. Each row is treated as evidence that a particular object, collection, project, or institutional pathway has become visible through description, digitisation, publication, cataloguing, or scholarly reporting. This design follows recent calls in manuscript and digital humanities research to treat metadata and collection description as evidence structures rather than transparent containers (Krippendorff 2018; Saldaña 2021; Drucker 2021; Poleg 2022; Mursyid, Aizan, and Faisal 2026).

The first analytical procedure was corpus auditing. Duplicate or repeated titles were not removed automatically because repetition itself indicates different circuits of visibility. *Syair Rakis*, *Syair Awang Semaun*, Brunei royal letters, British Library Malay collections, and EAP-related projects appear through more than one source layer, and those repetitions are analytically meaningful: they show that certain texts circulate simultaneously as manuscripts, printed classics, national heritage, digital surrogates, and scholarly objects. Instead of deduplicating the

corpus into a single title list, the study retained all records while distinguishing between collection, text, material form, and institutional pathway. This is important because manuscript cultures are rarely reducible to title identity. A text may survive as a handwritten witness, a lithographic print, a colonial copy, a national reprint, a digitised image, or an oral performance record.

The second procedure was category normalization. Geographical labels were grouped into Brunei Darussalam, Indonesia, transregional Malay world, and smaller country entries such as Myanmar, Vietnam, Thailand, Cambodia, and Laos. Custodians were coded into five regimes: state or court heritage institution; community, private, or religious custody; foreign or global archive; academic, NGO, or digital initiative; and unspecified or other. The content field was grouped into four interpretive evidence fields used in the results: literary/oral worlds; Islamic learning; archive and polity; and language or knowledge infrastructure. These fields were not intended to replace philological classification. They function as analytical lenses that link content to institutional and material visibility. The use of broad domains is appropriate because the dataset itself combines detailed manuscript entries, project records, literary collections, and institutional descriptions.

The third procedure was case-oriented interpretation. After the dataset was normalized, selected records were read closely as micro-cases: EAP229 and EAP329 in Aceh, EAP144 and EAP205 in Sumatra, EAP061 in East Java, EAP117 in Kerinci, EAP212 and DS10 in Buton, EAP276 in Ambon, Brunei History Centre and DBP collections, British Library Malay manuscripts, and the Cek Midi collection in Aceh. These records were chosen not because they are statistically dominant, but because they expose different dimensions of the same problem: local Islamic pedagogy, fragile private custody, codicological materiality, royal historiography, foreign archival displacement, and the politics of digital access. This combination of normalized corpus mapping and close record analysis helps avoid two common weaknesses: overgeneralizing from isolated examples and flattening manuscripts into counts (Hsieh and Shannon 2005; Schreier 2012; Nowell et al. 2017; O'Connor and Joffe 2020).

The study adopts a cautious interpretive posture. The dataset is not a probability sample, and its geographical imbalance is part of its meaning. Brunei Darussalam and Indonesia dominate because the source sheets emphasize Brunei literary heritage and Indonesian digitisation projects. Rather than concealing this imbalance, the article uses it to ask what kind of archive is being produced. Quantification is therefore descriptive and diagnostic, not inferential. The article does not claim that 99 records represent Southeast Asian manuscript culture in general. It claims that the dataset makes visible a specific archival ecology: one in which local Islamic materials, state heritage institutions, global repositories, and digital platforms intersect to reshape scholarly perception.

3. Results and Discussion

3.1. The corpus as archival stratigraphy

The corpus contains 99 records distributed unevenly across geography, custody, and content. Brunei Darussalam accounts for 55 records, while Indonesia

accounts for 35. The remaining records include transregional Malay, Malay world/Singapore-Malaysia, Myanmar, Vietnam, Thailand, Cambodia, and Laos entries. This imbalance is analytically important. It means the corpus is not a neutral map of Southeast Asia; it is a layered evidence structure generated by particular collecting and digitisation pathways. Brunei appears primarily through literary, state, language, and historical institutions, while Indonesia appears strongly through local and community-held manuscripts digitised or documented by EAP, DREAMSEA, Manassa-related, and regional preservation efforts. The dataset therefore presents two different forms of visibility: Brunei as a state-cultural and literary archive, and Indonesia as a field of local manuscript rescue and Islamic knowledge preservation.

Table 1 summarizes the corpus audit and normalization logic. The strongest methodological lesson is that this corpus must be read through evidence layers rather than through nation-by-nation comparison alone. A simple country count would imply that Brunei is the main intellectual centre in the data. A custody-sensitive reading gives a more accurate interpretation: Brunei is highly visible because institutional and literary records are well represented, whereas Indonesia is visible through rescue-oriented digitisation of dispersed local collections. The difference is not merely quantitative; it reflects different archival regimes.

Table 1. Corpus audit and normalization logic

Analytical layer	Corpus evidence	Interpretive implication
Records	99 rows across 22 source groups	The dataset is large enough for structured qualitative mapping but not representative sampling.
Major geographies	Brunei Darussalam = 55; Indonesia = 35; other/transregional = 9	The corpus is dominated by two complementary regimes: Brunei state-literary visibility and Indonesian local digitisation.
Evidence fields	Literary/oral = 41; Islamic learning = 28; archive and polity = 18; language/knowledge infrastructure = 12	Content is not reducible to religion or literature alone; manuscript visibility is multidomain.
Custody regimes	State/court = 39; community/private/religious = 27; foreign/global = 12; academic/NGO/digital = 6; unspecified = 15	Custody determines how manuscripts become visible, interpretable, and accessible.
Digital condition	55 records are digitally accessible or digitised; 44 are print/offline, in process, restricted, unspecified, or mixed	Digitisation is substantial but incomplete; access remains uneven and institutionally mediated.
Chronology	Pre-17th = 16; 17th-18th = 20; 19th = 30; 20th = 14; 21st/mixed = 19	The corpus foregrounds late manuscript cultures and modern documentation more than classical origins.

The numerical profile matters because it changes the kind of claim the article can responsibly make. The dataset does not prove that Brunei or Indonesia contains more manuscripts than other Southeast Asian regions. It shows which collections and projects have been foregrounded by the workbook. This is precisely why the article uses the term archival visibility. Visibility is not the same as existence. Manuscripts may exist without being catalogued, digitised, or cited; conversely, a highly visible record may reflect institutional documentation rather than regional abundance. The corpus therefore teaches a decolonial lesson: manuscript history must ask not only where texts were produced, but also how they become countable.

Figure 1 translates this point into a stratigraphic diagram rather than a conventional bar chart. It treats custody as a layer of evidence. State and court heritage institutions constitute the largest layer, followed by community/private/religious custody, unspecified records, foreign/global archives, and academic or digital initiatives. The figure is intentionally designed as a stratigraphy because the article's argument is not about which category wins numerically. It is about how manuscript visibility is built from overlapping layers of custody, authority, and access.

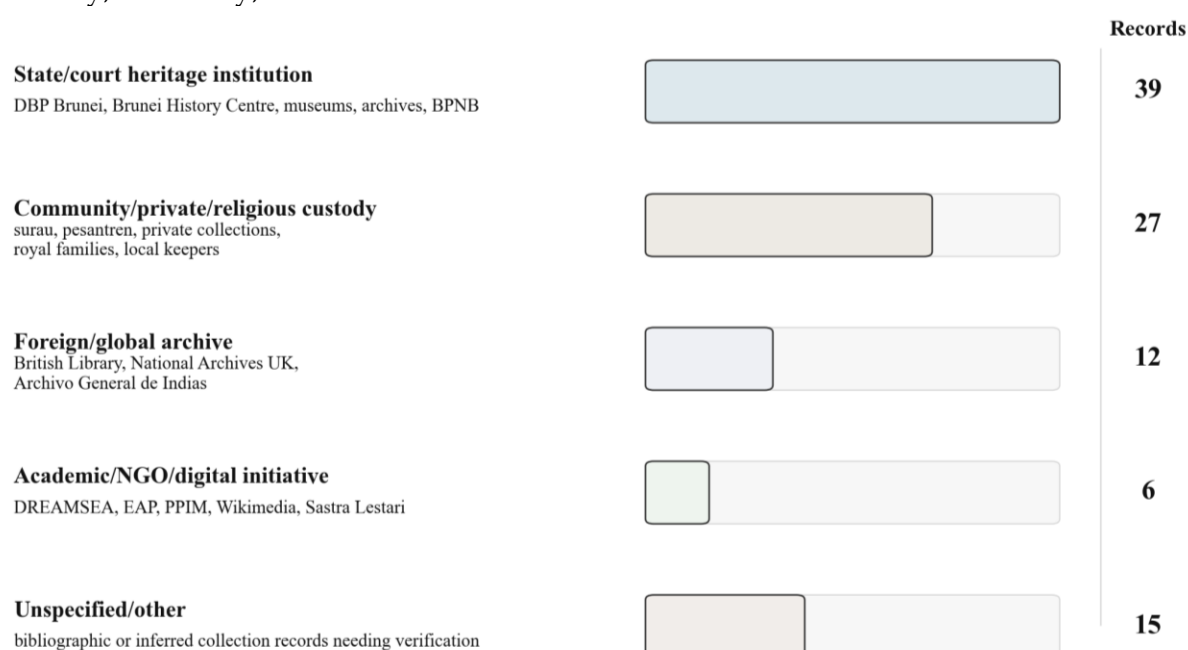


Figure 1. Archival stratigraphy of the manuscript dataset.

Figure 1 clarifies that digitisation does not eliminate institutional hierarchy. State and court institutions remain highly visible, especially in Brunei, where DBP, the Brunei History Centre, museums, archives, and language institutions shape the documentary field. Yet the second layer, community/private/religious custody, is crucial for remapping Islamic intellectual history. It includes surau, pesantren, Sufi orders, private families, royal households, local custodians, and manuscript owners whose materials often escape classical library-centred narratives. The smaller academic/NGO/digital layer is not unimportant; its function is catalytic. Projects such as DREAMSEA and EAP may represent fewer records as custodians, but they restructure the archive by bringing local custodians into global digital visibility.

3.2. Local digitisation and the remapping of Islamic knowledge

The second finding concerns collection bias. The dataset confirms a contrast already raised in recent scholarship: Western and colonial collections have long shaped the study of Malay and Islamic writing traditions, but locally held digitised collections disclose different intellectual priorities. Gallop argues that newly accessible community-held collections digitised through EAP and DREAMSEA are primarily concerned with Islamic topics and often written in Arabic, thereby enabling a remapping of maritime Southeast Asian writing traditions (Gallop 2020). The present dataset supports that argument, but it also nuances it. Across the full mixed corpus, Islamic learning accounts for 28 of 99 records, while literary/oral worlds account for 41. However, within Indonesian local digitisation clusters, Islamic learning is far more prominent, especially in Aceh, West Sumatra, East Java, Buton, Ambon, and private Acehnese collections. The apparent difference between a 'literary' corpus and an 'Islamic' corpus is therefore produced by the mixture of Brunei state-literary records and Indonesian local digitisation records.

This distinction matters. A conventional catalogue organised around famous titles may foreground *Syair Rakis*, *Syair Awang Semaun*, *Hikayat Raja Pasai*, *Taj al-Salatin*, royal correspondence, and DBP literary publications. A community digitisation lens foregrounds different objects: *khutbah* scrolls, Sufi manuals, *fiqh* texts, *pesantren* teaching manuscripts, *mawlid* texts, Qur'anic excerpts, Arabic-Jawi works, and village or family collections. These are not marginal supplements. They are the everyday infrastructure through which Islamic knowledge circulated in local contexts. The dataset therefore demonstrates a shift from literary monumentality to pedagogical and devotional ecology.

Table 2. Evidence fields and representative records

Evidence field	Records	Representative records	Analytical significance
Literary/oral worlds	41	<i>Syair Rakis</i> ; <i>Syair Awang Semaun</i> ; <i>Hikayat alf layla wailayla</i> ; <i>Diangdangan</i> ; Brunei novels and oral traditions	Shows that manuscript and literary heritage cannot be separated from print, oral performance, and national cultural institutions.
Islamic learning	28	EAP061 <i>Langitan</i> ; EAP144 <i>Minangkabau surau</i> ; EAP205 Sufi orders; EAP229 Aceh; EAP276 Ambon; Cek Midi collection	Reveals the density of local Islamic pedagogy, devotional practice, and Arabic-Jawi textual circulation.
Archive and polity	18	Brunei royal letters; <i>Hukum Kanun Brunei</i> ; <i>Amanat 1885</i> ; <i>Selesilah Raja-raja Brunei</i> ; <i>Taj al-Salatin</i>	Connects manuscript culture to statecraft, diplomacy, legal authority, genealogy, and political ethics.
Language/knowledge infrastructure	12	DBP dictionaries and journals; <i>Unit Jawi</i> ; Wikisource <i>Loves Manuscripts</i> ; inscriptions; composite corpora	Shows that preservation also depends on language planning, digital infrastructure, periodical culture, and metadata systems.

Table 2 shows why a single genre label is inadequate. A manuscript ecology is not simply a set of religious, literary, or historical texts. It is a layered system in which pedagogical Islam, courtly memory, language planning, public literature, oral performance, and foreign archival displacement interact. The strongest evidence for remapping lies in the Islamic learning cluster: its records are not concentrated in one metropolitan repository but are distributed across local custodians and regional projects. This changes the archive's centre of gravity. Instead of asking how Southeast Asian Islam appears in colonial libraries, the dataset invites scholars to ask how local institutions generated, preserved, and transmitted Islamic knowledge before digitisation made them visible to external research.

The difference can be seen through Aceh, Minangkabau, East Java, Buton, and Ambon. EAP229 and EAP329 document Acehnese religious and historical manuscripts in Pidie and Aceh Besar. EAP144 and EAP205 foreground surau and Sufi networks in West Sumatra. EAP061 highlights pesantren collections in East Java. EAP212 and DS10 show the Buton sultanate as a site of Arabic, Wolio, and Malay textual production. EAP276 identifies Ambonese local manuscripts, including khutbah scrolls and mawlid texts. In each case, digitisation does not simply preserve isolated manuscripts. It reveals an institutional sociology of Islamic knowledge: teachers, surau, pesantren, families, ritual specialists, sultanates, and village custodians function as archives before they become data.

3.3. Codicological materiality and multiscriptal worlds

The third finding concerns materiality. The dataset challenges any assumption that manuscript digitisation is only about paper codices. It contains paper manuscripts, Jawi letters, palm-leaf and lontar traditions, horn and Ulu-script objects, khutbah scrolls, lithographic books, inscriptions, printed journals, oral recordings, and composite digital repositories. This codicological diversity is not decorative. Material form shapes interpretation. A palm-leaf manuscript implies a different production ecology from a Jawi diplomatic letter; a horn text from Kerinci requires different handling and description from a Brunei court genealogy; a lithographic Hikayat alf layla wa-layla travels through print capitalism differently from a handwritten pesantren commentary; an oral Diangdangan recording expands manuscript studies toward performance documentation.

Language ecology reinforces the same point. Malay appears in 81 records, making it the corpus's dominant language frame. Arabic appears in 23 records and is especially tied to Islamic learning. Jawi appears repeatedly as both script and cultural technology. Javanese, Wolio, Acehnese, Cham, Pali, Burmese, Thai, Khmer, Sanskrit, Bugis, Makassarrese, Sundanese, Ulu script, Lampung, Dusun, Tutong, Kedayan, and other labels appear in smaller numbers but with major interpretive consequences. The corpus is therefore not merely multilingual in the modern sociolinguistic sense. It is multiscriptal and multimaterial. Textual identity emerges from the interplay between language, script, support, custody, and use.

Table 3. Micro-case matrix of codicological and interpretive evidence

Micro-case	Material/codicological signal	Knowledge function	What the case changes analytically
EAP117 Kerinci	Pusaka objects, including horn with Ulu script	Local law, memory, sacred custody	Moves the archive beyond paper and beyond library custody.
EAP276 Ambon	Khutbah scrolls and mawlid materials	Ritual pedagogy and public Islamic speech	Shows that sermon and devotional texts are central to manuscript ecologies.
EAP061 East Java	Pesantren collections, including tafsir, hadis, fiqh, nahwu, tasawuf, and Cerito Ambiya	Scholastic transmission in pesantren networks	Links manuscript preservation to Islamic pedagogy rather than only heritage.
EAP212 / DS10 Buton	Arabic, Wolio, and Malay manuscripts in royal/local collections	Sultanate learning, literature, and political memory	Shows Arabic literary culture at the edge of Malay-Indonesian Islam.
British Library Malay manuscripts	Malay and Arabic manuscripts, early maps, Raffles-related papers	Colonial collection and global access	Represents high visibility but also inherited selection bias.
DBP Brunei collections	Manuscripts, language journals, dictionaries, Jawi services, literary publications	State language planning and literary canon formation	Shows that preservation includes language policy and periodical infrastructure.
Cek Midi Aceh	Private collection of Qur'an, tasawuf, astronomy, medicine, tauhid, fiqh, and ilmu falak	Household archive of wide-ranging Islamic knowledge	Shows private custody as a serious intellectual infrastructure.

The micro-cases in Table 3 demonstrate why codicology must remain central to digital manuscript analysis. Online images can create the illusion that all manuscripts become equivalent once photographed. The dataset says otherwise. A manuscript's support, script, storage, institutional pathway, and performance context continue to matter after digitisation. Kerinci's horn and Ulu-script materials, Ambon's khutbah scrolls, Buton's Arabic-Wolio-Malay collections, and the Cek Midi household archive each demand different questions. What is the ritual status of the object? Who has the authority to handle or read it? Does digitisation preserve form, or does it flatten material differences into pixels? What metadata is needed to make the object intelligible without removing it from its local world?

This material perspective also helps avoid a narrow understanding of Islamic manuscript culture. Islamic texts in the dataset are not limited to Qur'anic manuscripts. They include tafsir, hadis, fiqh, tasawuf, nahwu, mawlid, khutbah, tauhid, ilmu falak, medicine, tajul muluk, and moral narratives. These genres reveal Islamic intellectual life as practical, devotional, pedagogical, cosmological, and juridical. The same archive can contain teaching texts, ritual speech, astronomical

calculation, healing knowledge, and political memory. Digitisation becomes academically powerful only when it preserves that internal diversity rather than using 'Islamic manuscript' as a vague umbrella.

3.4. Custody, access, and the politics of digital legibility

The fourth finding concerns access. Digitisation status in the dataset is not binary. Fifty-five records are digitised or accessible online, but the remaining records are print-only, in process, restricted, unspecified, offline, or mixed. This unevenness matters because access is one of the main mechanisms through which scholarly canons are produced. A text that can be viewed online is more likely to be cited, taught, translated, and incorporated into comparative research. A manuscript kept in a private cupboard or an agency inventory may be culturally central yet academically peripheral. Conversely, a colonial or national repository can give a text high global visibility even if its historical custody is displaced from its local community.

The digitisation problem is therefore not simply a technical backlog. It is a politics of legibility. The Brunei records show a dense state-cultural infrastructure: DBP, the Brunei History Centre, museums, national archives, language departments, publications, and journals. Indonesian EAP and DREAMSEA records show dispersed local collections entering digital view through external or collaborative initiatives. British Library records show global accessibility created through a major metropolitan institution. These pathways differ in authority, metadata, access rights, and interpretive framing. A digital image hosted by a global repository does not have the same cultural meaning as a locally controlled repository, and a catalogue description produced by a foreign archive is not equivalent to one co-produced with the manuscript's custodial community.

Table 4 summarizes these access regimes. The table does not rank them morally. Instead, it identifies the epistemic risk and scholarly opportunity attached to each regime. The central issue is that every route to visibility has a cost. State institutions can stabilize preservation but may nationalize or standardize local variation. Community custody preserves living meaning but may lack conservation and access infrastructure. Foreign archives offer global visibility but often embody older extraction histories. Digital initiatives can widen access but may separate images from ritual and social context. Unspecified records remind researchers that metadata gaps are not minor technical defects; they shape what can be responsibly inferred.

Table 4. Access regimes and interpretive risks

Access regime	Dataset signal	Scholarly opportunity	Risk caution	requiring
State/court institution	heritage 39 especially institutions	records, Brunei cataloguing, publication, language planning, and national heritage	May canonical, courtly, or state-approved narratives.	privilege

Access regime	Dataset signal	Scholarly opportunity	Risk requiring caution
Community/private/religious custody	27 records, including surau, pesantren, private families, and local religious collections	Reveals local Islamic pedagogy and non-metropolitan knowledge ecologies	May remain fragile, under-described, or ethically sensitive.
Foreign/global archive	12 records, including British Library, UK National Archives, and AGI	Provides high-resolution access, stable catalogues, and international discoverability	May reproduce colonial collection bias and detach materials from local custodianship.
Academic/NGO/digital initiative	6 records, including DREAMSEA, EAP, Wikimedia, and Sastra Lestari	Creates new bridges between endangered collections and open research	Depends on metadata quality, community consent, and long-term repository sustainability.
Unspecified/other	15 records requiring verification	Marks the limits of available evidence	Can lead to overconfident claims if uncertainty is not reported.

The table supports a major interpretive claim: digital access is not the opposite of archival power; it is one form of archival power. Digitisation can democratize access, but it can also create new dependencies on platform design, institutional priorities, metadata categories, copyright policies, and funding cycles. The strongest digitisation projects are therefore not those that simply accumulate images. They are those that document provenance, local names, custodial authority, language, script, material support, genre, usage context, and restrictions. This is why manuscript digitisation must be evaluated through cultural accountability as well as technical quality.

3.5. Toward a model of archival visibility

The final finding is theoretical. The dataset suggests a model of archival visibility in which four elements must be read together: content, material form, custody, and access. Content tells us whether a record concerns Islamic learning, literature, genealogy, diplomacy, law, language infrastructure, or technical knowledge. Material form tells us whether the record is a paper manuscript, Jawi letter, palm leaf, horn text, scroll, lithograph, inscription, oral recording, or digital publication. Custody tells us who keeps the object or controls its description. Access tells us how the object enters scholarly circulation. When these elements are separated, interpretation becomes thin. When they are combined, the dataset becomes a map of how Islamic and Malay intellectual worlds become visible.

Figure 2 presents this model as evidence cards rather than a statistical chart. The purpose is to prevent the results from becoming a descriptive frequency exercise. The cards show four interpretive fields: Islamic learning, literary/oral worlds, archive and polity, and language/knowledge infrastructure. These fields overlap in practice, but separating them analytically shows how different materials

require different scholarly questions. Islamic learning asks about pedagogy, ritual, jurisprudence, and devotional transmission. Literary/oral worlds ask about performance, genre, canon formation, and print transition. Archive and polity ask about sovereignty, diplomacy, law, genealogy, and political memory. Language infrastructure asks about standardization, script maintenance, journals, dictionaries, and digital repositories.

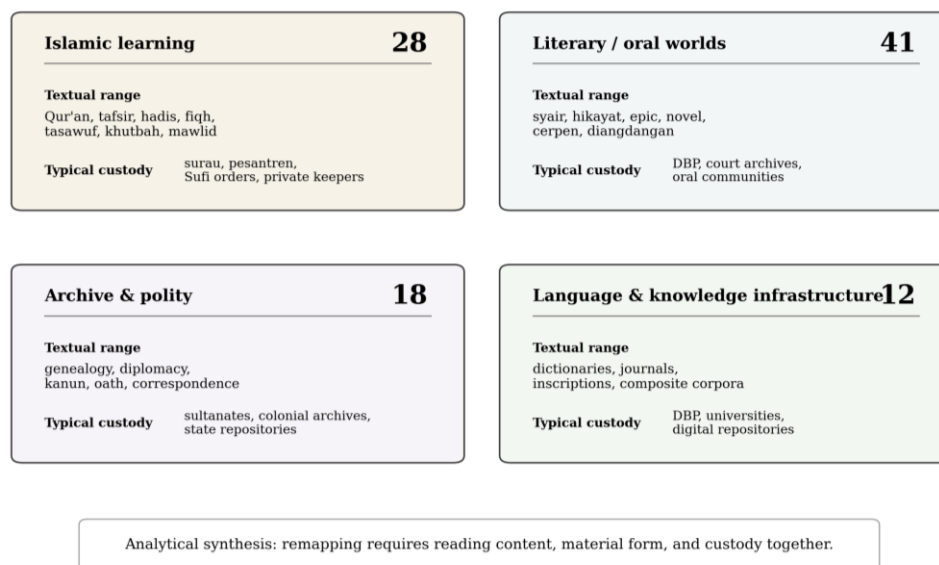


Figure 2. Evidence cards for post-digitisation visibility.

Figure 2 reinforces the article's central argument: digitisation changes the manuscript landscape by changing the relationship between evidence fields. A digital archive that only preserves literary classics may strengthen national canon formation but leave Islamic pedagogy under-described. A repository that photographs local religious manuscripts but provides minimal metadata may preserve images while weakening interpretive context. A language institution that publishes dictionaries and journals may not appear to be a manuscript archive, yet it sustains the linguistic infrastructure that allows manuscript culture to remain readable. In this sense, manuscript digitisation is never merely about manuscripts. It is about the wider ecology of reading, authority, language, memory, and access.

The model also explains why the dataset's Brunei-Indonesia imbalance is useful rather than fatal. Brunei records foreground state-cultural, literary, language, and courtly infrastructures. Indonesian records foreground community-held, religious, and endangered manuscript collections. Read separately, each might appear partial. Read together, they reveal the two sides of a regional problem: how do manuscript traditions become culturally authoritative, and how do they become digitally accessible without losing local meaning? The answer requires decolonial philology, not as a slogan but as a method. It means reading manuscripts with attention to who preserved them, who digitised them, who describes them, who gains access to them, and whose intellectual history becomes visible as a result.

This methodological position also changes how digital humanities can be integrated into Southeast Asian manuscript studies. The next step is not simply to

create larger digital repositories. Larger repositories will be useful only if they allow researchers to compare material supports, scripts, genres, custodians, and access conditions across collections. A regional dataset should be able to ask, for example, whether khutbah scrolls cluster in particular islands, whether Jawi diplomatic letters are disproportionately located in foreign archives, whether Sufi texts remain more strongly tied to local religious custodians, or whether DBP-style language institutions produce a different form of visibility from EAP-style preservation missions. Such questions require structured metadata and philological interpretation together. The present article offers a modest demonstration: even 99 records, when read as structured evidence, can generate stronger claims than a larger unexamined inventory.

The methodological value of the dataset is also visible in its negative evidence. Fifteen records were coded as unspecified or other in terms of custody because the owner, keeper, or institutional pathway could not be clearly normalized. Rather than discarding these records, the article treats them as evidence of metadata vulnerability. In manuscript studies, missing information is not empty space; it is a historical and methodological signal. It may indicate undocumented transfer, weak cataloguing, reliance on secondary reporting, or a mismatch between local custodial categories and institutional metadata forms. Future research should therefore report uncertainty explicitly. A transparent record that says 'custody unclear' is more academically responsible than a polished claim that hides the limits of the source.

The corpus therefore supports a staged model of digital philology. The first stage is preservation: photographing, scanning, stabilizing, and storing fragile material. The second stage is discoverability: building searchable catalogues, persistent identifiers, and multilingual metadata. The third stage is interpretability: linking images to codicological description, provenance, ownership, genre, language, and historical use. The fourth stage is relational analysis: comparing collections across regions, custodians, and knowledge domains without detaching them from local social contexts. Most digitisation projects succeed at the first two stages, while the third and fourth stages remain uneven. This pattern is visible in the dataset's contrast between major repositories, which supply discoverability, and community-held collections, which supply interpretive richness but may depend on external projects for digital exposure.

A further implication of the results is methodological. The dataset shows that digitisation should not be treated as a final research outcome. A record marked as digitally accessible does not automatically become analytically usable unless the description also identifies content, language, script, custodian, estimated period, material condition, and access pathway. In this corpus, 55 records can be classified as digitised or online-accessible, but several of them still require additional verification because the status field contains inferred information, mixed descriptions, or reliance on secondary references. This distinction is important for Q1-level manuscript research: digital presence is not the same as data quality. A digitised image without reliable metadata may preserve visibility while weakening interpretive precision; conversely, an offline record with strong codicological detail may be analytically richer than an online record with minimal description.

3.6. Methodological implications: from digital access to structured evidence

The findings have implications for the study of language, literature, and cultural dynamics. First, they show that the category of literature is too narrow when used alone. In the dataset, literary and oral records are numerous, especially in Brunei, but they coexist with Islamic learning, legal-political documents, letters, language planning materials, and technical knowledge. This means that Southeast Asian manuscript cultures should not be divided too quickly into literature, religion, history, and language. These domains were historically entangled. A *syair* can be political advice; a genealogy can be state theory; a Sufi manuscript can be pedagogy, devotion, and metaphysics; a dictionary can be a tool of cultural preservation; a royal letter can be both diplomatic document and aesthetic artefact. The digital archive should be designed to preserve these overlaps rather than forcing each record into a single modern category.

Second, the findings support a critique of colonial collection bias, but they also complicate it. It would be too simple to oppose colonial archives to local archives as if one were false and the other authentic. Foreign repositories often provide stable preservation and high-quality access, and they contain materials that would otherwise be difficult to consult. The problem is not their existence; the problem is their historical dominance as interpretive gateways. Local digitisation corrects that dominance by bringing forward manuscripts whose value is pedagogical, ritual, communal, and religious rather than merely diplomatic or literary. The strongest future research will therefore compare rather than replace archives: it will read British Library Aceh materials alongside Acehnese private collections, Brunei court manuscripts alongside DBP language institutions, and Buton royal collections alongside Arabic literary networks across the Indian Ocean.

Third, the article proposes that digital humanities in Southeast Asian manuscript studies should move beyond access rhetoric. Access is necessary, but access without metadata, codicological description, community context, and interpretive method can reproduce shallow scholarship. Mursyid, Aizan, and Faisal's recent mapping of Qur'an manuscript studies observes that digitisation is often mentioned but rarely integrated as a methodology. The present article responds by treating the dataset itself as a digital humanities object: a structured corpus whose fields reveal the interaction between content, materiality, custody, and access. The methodological shift is small but important. Instead of using digitised manuscripts only as image sources, scholars can use digitisation metadata to study archival formation, regional visibility, and the politics of preservation.

Fourth, the article highlights the importance of language and script. Malay and Arabic dominate the dataset, but they do not operate as isolated languages. They appear with Jawi, Wolio, Acehnese, Javanese, Cham, Pali, Burmese, Thai, Khmer, Sanskrit, Bugis, Makassarese, Sundanese, Ulu script, and other local forms. This multiscriptal ecology matters for cultural dynamics. Translation, transcription, OCR, handwritten text recognition, metadata search, and digital cataloguing often privilege Latin-script and standardized languages. Manuscripts in Jawi, Pegon, Ulu, Lontara, old Lao, or palm-leaf formats require different technological and philological infrastructures. Digital equity therefore depends not only on scanning

manuscripts but also on building tools, fonts, authority files, transliteration standards, and community expertise for low-resource scripts and languages.

Finally, the dataset invites a more ethical understanding of manuscript preservation. The language of rescue can unintentionally imply that value begins when external projects digitise local materials. The evidence here suggests the opposite. Local custodians already preserved manuscripts before digitisation; digitisation changes access but does not create cultural value from nothing. An ethical model must therefore recognize manuscript owners, religious communities, families, local institutions, and regional scholars as knowledge agents. The archive is not a passive storehouse waiting to be activated by technology. It is a living network of trust, memory, authority, and practice. The future of Islamic Southeast Asian manuscript studies will depend on whether digital projects can make this network visible without extracting it from its local moral worlds.

4. Conclusion

This article has analysed a 99-record dataset of Southeast Asian manuscript collections and digitisation projects to show how digital access remaps Islamic and Malay intellectual worlds. The corpus is dominated by Brunei Darussalam and Indonesia, but this imbalance is not merely a limitation; it is an interpretive clue. Brunei records reveal the strength of state-cultural, literary, language, and heritage institutions, while Indonesian records foreground community-held and locally digitised Islamic manuscript collections. Together, they show that manuscript visibility is produced through custody, digitisation, institutional description, and scholarly circulation rather than by textual survival alone.

The main contribution of the article is the concept of archival visibility. This concept allows manuscript researchers to ask how materials become countable, citable, searchable, and comparable. The findings show that four evidence fields structure the dataset: literary/oral worlds, Islamic learning, archive and polity, and language/knowledge infrastructure. The most significant analytical insight is that digitisation remaps the archive only when it is joined to codicological materiality and custodial context. Future research should expand the corpus, verify titles through primary catalogues, include community interviews, compare metadata across repositories, and develop digital tools for Jawi, Pegon, Ulu, Wolio, and other low-resource manuscript scripts. Such work would move Southeast Asian manuscript studies from digital access toward digitally informed, decolonial, and community-accountable philology.

Acknowledgement

The authors acknowledge the value of publicly accessible manuscript repositories, local custodians, digitisation initiatives, and regional cultural institutions that continue to preserve Southeast Asian manuscript heritage. The authors also acknowledge the dataset supplied for this study as the empirical basis for the corpus audit and interpretive analysis.

Author Contributions

First Author: conceptualization, methodology, data auditing, formal analysis, writing-original draft, and final approval. Second Author: manuscript studies review, codicological interpretation, reference validation, and writing-review and editing. Third Author: digital humanities framing, visualization design, data verification, and writing-review and editing. All authors read and approved the final manuscript.

Funding

This research received no external funding.

Data Availability Statement

The study is based on a curated workbook of Southeast Asian manuscript collections and digitisation projects supplied for the present analysis. The normalized coding sheet and visual summaries can be provided as supplementary material upon reasonable request.

Declaration on the Use of Artificial Intelligence

Artificial intelligence support was used only for language refinement, structural organization, and document preparation. The analytical categories, interpretation of results, and final scholarly responsibility remain with the authors.

References

- Abid, Abdelaziz. 2007. *Memory of the World: Preserving Documentary Heritage*. Paris: UNESCO.
- Aljunied, Khairudin. 2013. *Colonialism, Violence and Muslims in Southeast Asia: The Maria Hertogh Controversy and Its Aftermath*. London: Routledge.
- Aljunied, Khairudin. 2019. *Islam in Malaysia: An Entwined History*. Oxford: Oxford University Press.
- Baried, Siti Baroroh, Siti Chamamah Soeratno, Sawoe, Sulastin Sutrisno, and Mohamad Syakir. 1994. *Pengantar Teori Filologi*. Yogyakarta: Badan Penelitian dan Publikasi Fakultas Sastra UGM.
- Blair, Sheila S. 2006. *Islamic Calligraphy*. Edinburgh: Edinburgh University Press.
- Braginsky, Vladimir I. 1993. *The System of Classical Malay Literature*. Leiden: KITLV Press.
- Braginsky, Vladimir I. 2004. *The Heritage of Traditional Malay Literature: A Historical Survey of Genres, Writings and Literary Views*. Leiden: KITLV Press.
- British Library. n.d. *Successful Projects: Malay Manuscripts*. London: British Library.
- Chambert-Loir, Henri. 2009. *Sadur: Sejarah Terjemahan di Indonesia dan Malaysia*. Jakarta: Kepustakaan Populer Gramedia.

- DREAMSEA. 2026. Digital Repository of Endangered and Affected Manuscripts in Southeast Asia. Jakarta: PPIM UIN Jakarta.
- Drucker, Johanna. 2021. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*. London: Routledge.
- Déroche, François. 2006. *Islamic Codicology: An Introduction to the Study of Manuscripts in Arabic Script*. London: Al-Furqan Islamic Heritage Foundation.
- Endangered Archives Programme. 2025. *Endangered Archives Programme: Collections and Projects*. London: British Library.
- Fathurahman, Oman. 2012. *Ithaf al-Dhaki: Tafsir Wahdatul Wujud bagi Muslim Nusantara*. Jakarta: Mizan.
- Fathurahman, Oman. 2015. *Filologi Indonesia: Teori dan Metode*. Jakarta: Prenada Media.
- Fickers, Andreas, and Anne-Katrin Weber. 2020. Towards a New Digital Historicism? Doing History in the Age of Abundance. *Journal of European Television History and Culture* 9 (17): 1-9.
- Gacek, Adam. 2009. *Arabic Manuscripts: A Vademecum for Readers*. Leiden: Brill.
- Gallop, Annabel Teh, and Oman Fathurahman. 2022. Islamic Manuscripts from Aceh in the British Library. *Journal of Islamic Manuscripts* 13 (2): 115-166. <https://doi.org/10.1163/1878464X-01302002>.
- Gallop, Annabel Teh. 2020. Shifting Landscapes: Remapping the Writing Traditions of Islamic Southeast Asia through Digitisation. *Humaniora* 32 (1): 1-14.
- Hasbullah, Wan Mohd Dasuki Wan, Mohd Taufik Arridzo Mohd Balwi, and Eizah Mat Hussain. 2023. Pusat Sejarah Brunei: Beberapa Aspek Perbincangan tentang Koleksi Manuskrip Melayu-Islam. *Journal of Al-Tamaddun* 18 (1): 1-18.
- Hirschkind, Charles. 2006. *The Ethical Soundscape: Cassette Sermons and Islamic Counterpublics*. New York: Columbia University Press.
- Hodel, Tobias. 2023. Best Practices for Handwritten Text Recognition in Historical Documents. *Journal of Open Humanities Data* 9: 1-11.
- Hsieh, Hsiu-Fang, and Sarah E. Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15 (9): 1277-1288. <https://doi.org/10.1177/1049732305276687>.
- Johns, Anthony H. 1993. Islamization in Southeast Asia: Reflections and Reconsiderations with Special Reference to the Role of Sufism. *Southeast Asian Studies* 31 (1): 43-61.
- Jones, Russell. 1986. *Hikayat Raja Pasai*. Petaling Jaya: Fajar Bakti.
- Kirschenbaum, Matthew G. 2008. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: MIT Press.

- Knysh, Alexander. 2017. *Sufism: A New History of Islamic Mysticism*. Princeton: Princeton University Press.
- Krippendorff, Klaus. 2018. *Content Analysis: An Introduction to Its Methodology*. 4th ed. Thousand Oaks, CA: Sage.
- Laffan, Michael. 2011. *The Makings of Indonesian Islam: Orientalism and the Narration of a Sufi Past*. Princeton: Princeton University Press.
- Lombard, Denys. 1996. *Nusa Jawa: Silang Budaya*. Jakarta: Gramedia.
- McGann, Jerome. 2014. *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, MA: Harvard University Press.
- Mursyid, Aizan, and Faisal. 2026. Codicology, Islamization, and Qur'an Manuscripts in Southeast Asia: A Systematic Mapping. *Jurnal Studi Ilmu-Ilmu al-Qur'an dan Hadis* 27 (1): 181-216.
- Nowell, Lorelli S., Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods* 16 (1): 1-13.
- O'Connor, Cliodhna, and Helene Joffe. 2020. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods* 19: 1-13.
- Peacock, A. C. S. 2022. The Arabic Manuscripts of the La Ode Zaenu Collection, Buton. In *Malay-Indonesian Islamic Studies*, edited by A. C. S. Peacock, 45-80. Leiden: Brill.
- Peacock, A. C. S. 2024. *Arabic Literary Culture in Southeast Asia in the Seventeenth and Eighteenth Centuries*. Leiden: Brill.
- Pierazzo, Elena. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham: Ashgate.
- Poleg, Eyal. 2022. New Technologies, Training Initiatives and the Future of Manuscript Studies. *Manuscript Studies* 7 (2): 333-350.
- Reid, Anthony. 1988. *Southeast Asia in the Age of Commerce, 1450-1680*. Vol. 1. New Haven: Yale University Press.
- Ricci, Ronit. 2011. *Islam Translated: Literature, Conversion, and the Arabic Cosmopolis of South and Southeast Asia*. Chicago: University of Chicago Press.
- Risam, Roopika. 2018. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston: Northwestern University Press.
- Robson, Stuart O. 1994. *Prinsip-Prinsip Filologi Indonesia*. Jakarta: RUL.
- Sahle, Patrick. 2016. What Is a Scholarly Digital Edition? *Digital Scholarly Editing* 4: 19-39.
- Saldaña, Johnny. 2021. *The Coding Manual for Qualitative Researchers*. 4th ed. London: Sage.

- Schreier, Margrit. 2012. *Qualitative Content Analysis in Practice*. London: Sage.
- Spina, Salvatore. 2023. Artificial Intelligence in Archival and Historical Scholarship Workflow: HTS and ChatGPT. *Digital Scholarship in the Humanities* 38 (4): 1-15.
- Sweeney, Amin. 1987. *A Full Hearing: Orality and Literacy in the Malay World*. Berkeley: University of California Press.
- Teeuw, A. 1984. *A Critical Survey of Studies on Malay and Bahasa Indonesia*. Dordrecht: Foris.
- Terras, Melissa. 2016. Crowdsourcing in the Digital Humanities. In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 420-439. Chichester: Wiley Blackwell.
- UNESCO. 2015. *Recommendation concerning the Preservation of, and Access to, Documentary Heritage including in Digital Form*. Paris: UNESCO.
- van der Meij, Dick, and Jan van der Putten. 2023. DREAMSEA Digital Repository of Endangered and Affected Manuscripts in Southeast Asia: A Multiple Religious and Cultural Digitization Experience. In *DREAMSEA Digital Repository of Endangered and Affected Manuscripts in Southeast Asia*, 73-96. Berlin: De Gruyter. <https://doi.org/10.1515/9783110747607-004>.
- van der Meij, Dick. 2022. Islamic Palm-Leaf Manuscripts from Lombok. *Journal of Islamic Manuscripts* 13 (3): 265-290. <https://doi.org/10.1163/1878464X-01302003>.
- Witkam, Jan Just. 2007. *Inventory of the Oriental Manuscripts of the Library of the University of Leiden*. Leiden: Ter Lugt Press.
- Wormser, Paul. 2012. Manuscrits et imprimés malais: les collections de manuscrits en Europe. *Archipel* 84: 37-70.
- Yusuf, Muhammad. 2020. Naskah Kuno dan Transformasi Digital di Indonesia. *Manuskripta* 10 (1): 1-25.